GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

EuroCogSci 2019
02.09.19 - 04.09.19

The Role of Effect and Sample Size in Causal Induction

Simon Stephan and Michael R. Waldmann, University of Göttingen, Germany

Summary

Background. Because causal relationships are not directly perceivable they must be induced based on observable statistical regularities. Generally, causes are assumed to raise the probability of their effects. An assessment of the covariation between a potential cause C and a potential effect E provides the epistemic tool allowing a reasoner to learn whether C and E are causally connected or not. Rational *Bayesian-inference models* of *causal-structure induction* (e.g., Griffiths & Tenenbaum, 2005; Meder, Mayrhofer, & Waldmann, 2014) rely on contingency information to estimate the probability of the existence of a causal link between a potential cause and effect factor (**Tab. 1**). These models derive their estimations by incorporating two types of information: the size of the empirical effect, operationalized as the *contingency* (Delta-P) between C and E, and the observed *sample size*. In accordance with the normative models, human causal induction has been found to be sensitive to contingency (**Tab. 2 A & B**). However, past studies suggest that reasoners struggle to recognize the role of sample size (**Tab. 2 B**). In the present research we aim at investigating why this is. One reason might be that the cover stories and instructions used in previous studies emphasized the role of contingency, relegating the importance of sample size to the background. Another possibility is that reasoners do not understand why sample size matters: it matters because it carries information about the outcome variability of hypothetical or actual repetitions of the causal process that generated the observed data, and thus on how compatible the observed data are with the non-existence of a causal relationship between C and E.

Experimental findings. Our research suggests that both explanations might be psychologically real. For instance, in a novel experimental paradigm (**Exp. 2**) in which participants could themselves determine the sample size in order to learn whether two factors are causally connected, we found that subjects who repeatedly observed weak effects indeed preferred to inspect larger samples than subjects who observed strong effects. However, we also found that subjects who observed weak effects (1) did not increase the observed sample sizes to an extent that would justify strong conclusions and (2) that they consequently refrained from doing so. In another experiment (**Exp. 3**), we found that this might be because many subjects fails to see the direct connection between sample size and measurement reliability. We presented subjects with information about the sampling variation of a fictitious experiment testing a particular effect with a particular sample size. We found that many subjects concluded that the observed sampling variation would remain constant if the sample size was increased in fictitious replications of the same experiment.

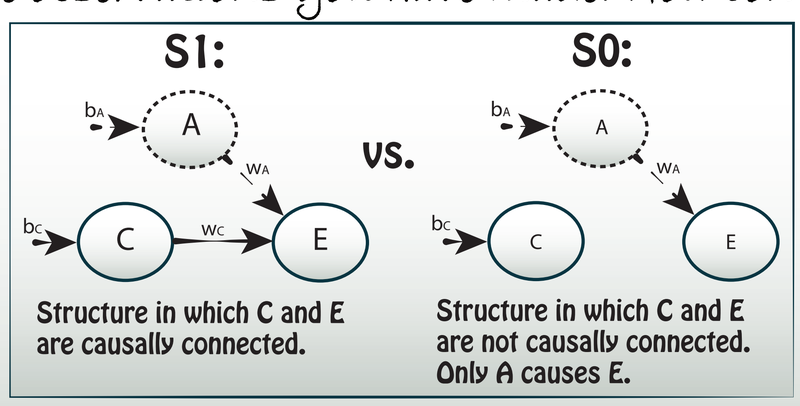
Tab. 1: The Framework of Bayesian Causal Structure Induction

„Based on the observations you have made: How confident are you that C causes E?"

	E	¬E
C	N(c,e) = 6	N(c,¬e) = 2
¬C	N(¬c,e) = 2	N(¬c,¬e) = 6

P(e|c) = 6/8 = .75
P(e|¬c) = 2/8 = .25
Delta-P = .75 - .25 = .50

Example of a contingency table listing observations



Alternative causal strucutres explaining observed data (C = target Cause, A = alternative causes)

$$P(S_i|D) = \frac{P(D|S_i)P(S_i)}{P(D)}$$

Posterior probability of each structure obtained by applying Bayes' Rule

with the Likelihood function for structure S1 being:

(illustrates incorporation of effect and sample size. effect size)

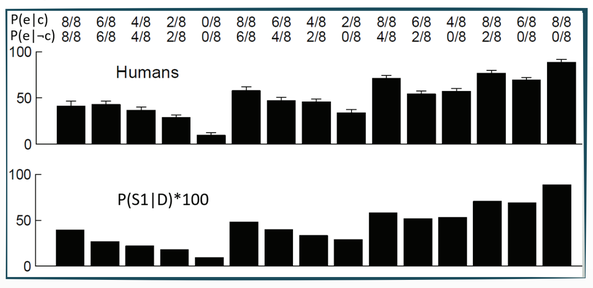
$$P(D|S_1) = [(1-b_C)(1-w_A)]^{N(\neg c,\neg e)} \cdot [(1-b_C)w_A]^{N(\neg c,e)} \cdot [b_C(1-w_C)(1-w_A)]^{N(c,\neg e)} \cdot [b_C(w_C+w_A-w_Cw_A]^{N(c,e)}$$

Computation of posterior probability for S1. Generally: P(S1|D) increases with incr. Delta-P and N.

Tab. 2: Previous studies

(A) Data from Buehner and Cheng (1997) re-analyzed by Griffiths and Tenenbaum (2005)

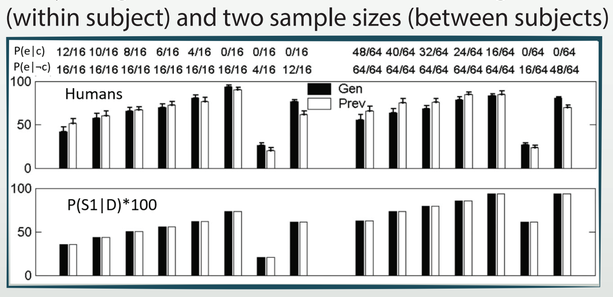
Study design: Manipulation of 15 contingencies (within subject)



Demonstration that reasoners' causal induction is sensitive to contingency (i.e., the observed effect size).
BUT: no manipulation of sample size.

(B) Exp. 3 in Lu, Yuille, Liljeholm, Cheng, and Holyoak (2008)

Study design: Manipulation of different contingencies (within subject) and two sample sizes (between subjects)



Again a demonstration that reasoners are sensitive to effect size. Results suggest, however, that reasoners don't incorporate sample size.

But: Sample size was manipulated between subjects, while contingency was manipulated within subject. The design emphasized contingency and may have led subjects to neglect sample size.

References

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology*, 51(4), 334-384.

Lu, H., Yuille, A. L., Liljeholm, M., Cheng, P. W., & Holyoak, K. J. (2008). Bayesian generic priors for causal learning. Psychological review, 115(4), 955.

Meder, B., Mayrhofer, R., & Waldmann, M. R. (2014). Structure induction in diagnostic causal reasoning. *Psychological Review*, 121(3), 277.

Contact: sstepha1@gwdg.de

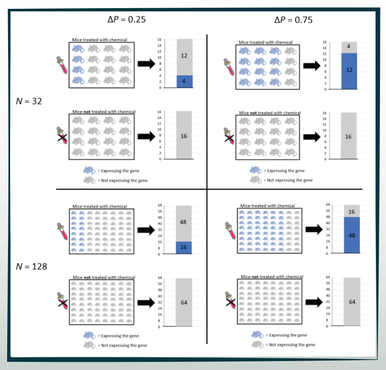
Experiment 1

Summary:

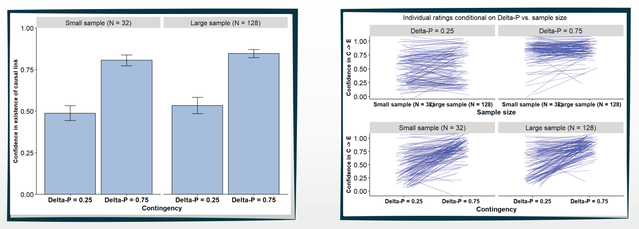
To test whether previous studies failed to detect an effect of sample size because sample size was manipulated between subjects and contingency within subject, we ran a study in which both factors were manipulated within subject.

We found only a small effect of sample size. Subjects were much more sensitive to contingency. The open question is: Why is this?

Design and Materials:



Results (N = 120):



Test questions asked for each data set

How confident are you that the tested chemical causes the expression of the tested gene?

Certain that it does not cause the expr. of the gene ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ Certain that it causes the expr. of the gene

50:50

Experiment 2

Summary:

Maybe Exp. 1 still emphasized effect-size information more than smaple-size information. Exp. 2 explored what happens when subjects are allowed to choose the size of the sample. We manipulated whether subjects rep. sampled from a small or large effect.

Subjects who sampled from a small effect tended to increase sample size, but only slightly.

Materials and Procedure:

Subjects tested the effects of five different chemicals. For each study, they could decide how many mice they wanted to test (see left panel above). The results of the fictitious study were displayed in a contingency table (see right panel). Subjects then answered a causal structure query as in Exp. 1.

Results (N = 200):

Selected sample size. Subjects sampling from a small effect tended to increase observed sample size. The effect was small, however.

Causal structure query. Subjects did not use sample size to increase confidence in the causal conclusions they were asked to draw.

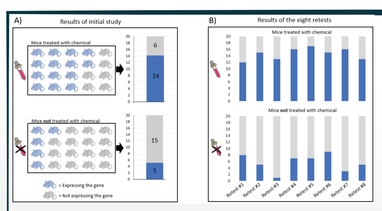
Experiment 3

Summary:

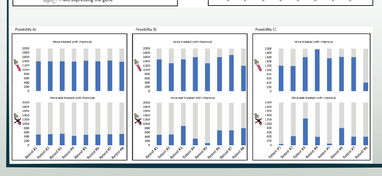
We tested whether subjects are less sensitive to sample size because they do not understand that larger samples imply smaller sampling variability. Subjects inspected eight replications of a fictitious study with sample size N = 40. They were then asked to indicate how the replications would have looked like if N had been 4000.

Most subjects indicated that the results would have looked the same.

Materials and Procedure:

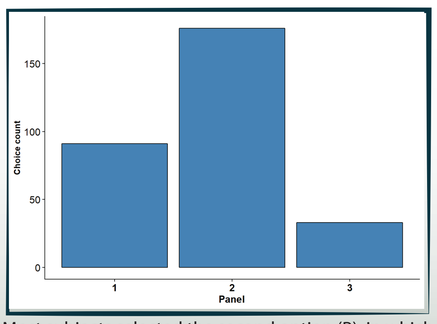


Results of initial study with N = 40 together with bar charts showing results of eight replications.



Three possibilities of how eight replications would look like with N = 4000. Subjects chose one.

Results (N = 300)



Most subjects selected the second option (B), in which the sampling variability for N = 4000 is the same as the one for N = 40.