Confidence in Causal Inferences: The Case of Devaluation

Uwe Drewitz (uwe.drewitz@tu-berlin.de)
Stefan Brandenburg (stefan.brandenburg@tu-berlin.de)
Berlin Institute of Technology, Department of Cognitive Psychology and Cognitive Ergonomics
Franklinstraße 28/29, 10587 Berlin, Germany

Abstract

When people have to make predictions and diagnosis they make use of their causal knowledge. This knowledge refers to two constituting aspects of causality: sufficiency and necessity. In standard theories both aspects are considered as being independent from each other. The present research tests this assumption. In an experiment we examined how peoples confidence in one of both aspects is affected, if they receive negative evidence for the complementary aspect. The presented data show that peoples confidence related to the aspect that has not been challenged by negative evidence decreases under such conditions. This devaluation effect is not predicted by standard theories.

Keywords: causal models**;** causal learning; reasoning under uncertainty; induction

Introduction

When people make a causal statement like: *A causes B*, they attribute a causal relation. This attribution can be based on various cues to causality (Einhorn and Hogarth, 1986) like spatial and temporal contiguity. However, in many situations people need more information than these. Being repeatedly confronted with a phenomenon, people (can) look for regularities as well. Psychological theories claim, that under such circumstances causal attributions rely on contingency information. Contingency information describe how the occurrence or absence of one event (i.e. event C) goes together with the occurrence or absence of another event (i.e. event E). Based on this information people can determine how likely an effect of interest will occur, given the presence or absence of a putative cause. According to standard psychological theories (e.g. Waldmann & Holyoak, 1992; Waldmann & Hagmayer, 2001; Griffith & Tenenbaum, 2005) people integrate the information about the (co-)occurrence and (co-)absence to either infer a causal relation or estimate it' s strength, respectively. Therefore standard psychological theories claim that people base their judgments on all available data for contingency information. This is a reasonable assumption for situations where people do causal judgments. In contrast, in many real-world tasks people do not have to do such integrative judgments. They apply their knowledge to forecast events (i.e. E+ / E-) based on given data (i.e. C+ / C-). In probability calculus this is captured by conditional probabilities. The prediction of E+ for example can be made based on P(E+/C+) or P(E+/C-) (see Fig. 1) depending on whether C+ or C- is present. These conditional probabilities are independent of each other. Given these facts, standard theories do not predict effects of information integration over all contingency data. Hence, for a prediction of E given C, persons would not integrate over all the four possible pairings of the two events (i.e. *C* & *E*). However, we present experimental data that contradict this position.

Sufficiency and Necessity

Various so called rule-based models (see Allan, 1980) have been proposed in research literature (e.g. Jenkins & Ward, 1965; Cheng & Novick, 1992; Cheng, 1997; White, 2003). They assume that persons rely on frequencies of (co-) occurrence and (co-) absence of two events (i.e. *C* & *E*). The four cells in the contingency table in Figure 1 represent their four possible pairings. With respect to these two events, every observation can be assigned to one pairing and as such, to one cell of the contingency table. Every observation gives either positive or negative evidence to one of both aspects of causality: *sufficiency* and *necessity*. Positive evidence can be understood as strengthening an aspect (either sufficiency or necessity). Comparably, negative evidence weakens an aspect. Sufficiency and necessity are complementary building blocks of causality (e.g. Mill, 1869). An event C is recognized as *sufficient* to produce another event E, if the latter always follows the occurrence of the former. The same event C is considered as *necessary* to bring forth the event E, if its absence of C is always accompanied by the absence of E.

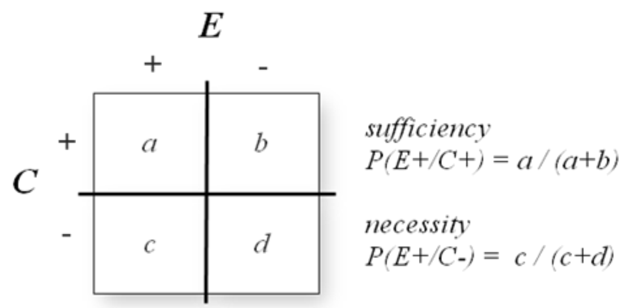


Figure 1. 2x2 contingency table (+ indicates presence, - indicates absence).

Moreover, sufficiency and necessity are statistically independent of each other. Whereas the *sufficiency* of a putative cause for an effect depends on the frequencies in the cells a and b, the *necessity* is determined by the frequencies in the cells c and d (see Fig.1). Two different, statistically independent conditional probabilities capture these facts (see Fig.1): the probability of the presence of E given the presence of C, P(E+/C+), and the probability of the presence of E given the absence of C, P(E+/C-). These probabilities are complementary in the sense of causality. People are willing to attribute a causal relation between two events if both aspects are met. This idea goes back to John Stuart Mill (1869) who claimed that causal knowledge does not arise from the repeated observation of the sequence of two events only. Instead people also acknowledge what happens if a putative cause *fails* to appear. From this perspective, causes can be characterized in terms of *sufficiency* and *necessity* and both of these aspects have to be satisfied. Every observation that belongs to the pairing of cell a gives positive evidence to the *sufficiency* of the putative cause C for E, the effect of interest. Just as all observations that belong to the pairing of cell d give evidence to the *necessity* of C for E.

How do the described facts fall into the scope of standard theories of causal learning (see introduction)? These theories describe how people come up with a judgment, when they are requested to rate the strength of a causal relation in a causal attribution task. In such tasks people base their judgments on both aspects (sufficiency and necessity), which means that they consider all four frequencies that can be presented in contingency information (see Figure 1). Of course, people do integrative judgments in real-world tasks. But very often they have to make predictions based on given data. In turn, as soon as people can rely for example on the presence or absence of C, i.e. on C+ or C-, their prediction is related to only one of both aspects. For example, given the presence or absence of C (C+ or C-), people act differently as if they were asked to rate the strength of the relation between C and E. Let us assume people have seen numerous pairings where one event C precedes another event E (frequency in cell a). Based on these observations people will predict E+ (presence of E) given C+ (presence of C). In such a case, there is no need to integrate the information about C- (absence of *C*), which is captured by the frequencies in cells c and d. On the other hand, if C- (absence of *C*) precedes E- (absence of *E*), which is represented by the frequency in cell d, people might use this information to predict that *E* will not occur given the absence of *C*. In that case information with respect to C+ (cells a and b) can be ignored. Consequently, given the independence of both aspects, neither positive nor negative evidence related to one of the aspects should affect inferences related to the complementary aspect. In contrast, we claim that such an effect exists. We tested this hypothesis based on the representation of causal knowledge, which is introduced in the next section.

Mental Causal Models

Several ways have been proposed to represent causal knowledge. For example Thüring and Jungermann (1992) suggest that people acquire mental models of causation in terms of conditional rules (e.g. If C+ then E+.). The conditional rules of a model reflect the characteristics of sufficiency and necessity of a causal relationship. This is in line with the conception of causes as sufficient and necessary conditions for their effects. As shown by Thüring, Drewitz and Urbas (2006) these conditional rules can be obtained by mere induction. In the case of the model of unique causation (see Table 1), which states that "C causes E", the event C is framed as a sufficient as well as *a* necessary condition for the event E. This is captured by the rules R1 and R2 in Table 1. When a situation calls for a causal inference, the available data (for instance C+ or C-) are matched with the rules (R1 and R2) and the required information is deduced.

Table 1: Model of unique causation.

Model statement: "C causes E"
R1: C+ → E+
R2: C- → E-

The importance of rules like R1 and R2 lies in the savings they provide. Rules save costs such as time, attention or memory capacity. However, to get all the benefit rules entail, they have to be linked into higher-order knowledge, like models. Let us have a look on both our rules R1 and R2. Neither R1 nor R2 tell us whether there is a causal relation between *C* and *E*, or not. Only when they are linked together one possesses this knowledge. We call the linking of rules the construction of a mental model. In this sense the statement "C causes E" is knowledge acquired by building the model, not by having the two rules R1 and R2 only. That also means that as soon as the rules are linked into a model, there is more than there was before. Or, in other words: The whole is more than the sum of its parts. Assuming that our considerations are right, we can ask the following question: If the whole - the mental model - is questioned because one of its parts fails, is there an effect on the other parts as well? In terms of the model of unique causation (see. Tab. 1): When people observe that one rule fails to predict the outcome, is there an effect on how they use the complementary other rule?

Causal Inferences under Uncertainty

Before we have a closer look on this question we want to make clear what known effects the failure of rule has. Depending on how successful the application of a rule was in the past, people will place more or less confidence into their predictions deduced from that rule. Let's think, for example, of a person that has rather limited causal knowledge as expressed by the model of unique causation. Whenever this person faces a situation where C+ is present, she will apply R1 and predict E+. Vice versa she will predict E- if C- is present, based on R2. As long as E+ goes always together with C+ all her predictions deduced from R1 are confirmed. Hence, her faith in R1 and therefore the confidence she places in her predictions should be high. The same holds for R2 and the related predictions as long as E- goes always together with C-. However, that will change as soon as it turns out that a rule is wrong. If people have build up incomplete or incorrect rules they will come up with wrong inferences in the course of events. Let's assume that in truth C+ in conjunction with another event X+ may be sufficient for E+ instead of C+ alone. In such a case people will observe E- subsequent to C+ whenever X is absent (X-). An observation like this will impair the *sufficiency* of *C* for *E.* Moreover, such an observation will *discredit* the rule R1. In general, every prediction that is not confirmed but contradicted by a subsequent observation *discredits* the rule it is derived from. Consequently, as long as a person cannot expand her model, she will loose confidence in the respective rule. All a person can do in such an uncertain situation is to reduce the confidence she places in her predictions based on that rule. Hence, the question we raised at the end of the last section remains unanswered. But now, we can reformulate and ask more specifically: Is the effect of the reduction of confidence always limited to the rule that was discredited?

Discrediting and Devaluating Causal Rules

To start with let us return to contingency information. Figure 1 shows a contingency table for the model of unique causation. A person's observations that fall into cell a provide evidence for the reliability of rule R1, while observations that fall into cell 'd' provide evidence for the reliability of R2 (see Table 1). On the other hand, all observations made in cell b discredit R1, while all observations made in cell c discredit R2. Therefore, the first row of the table provides information about the sufficiency of the cause and the second row about its necessity. As depicted in Figure 1 a cause is only *completely* sufficient, if observations are made in cell a, but not in cell b, and it is only *completely* necessary, if observations are made in cell d, but not in cell c. Only in these cases, the conditional probabilities are at their optimum with respect to a causal relation between *C* and *E*. From this point of view, the optimum of P(E+/C+) equals one and equals zero for P(E+/C-). Additionally, an increase or decrease of P(E+/C+) does not affect P(E+/C-) and vice versa. What does this mean from a psychological perspective? The first implication is consistent with the mechanism of *discrediting* a rule. When the sufficiency or necessity is weakened, the certainty of inferences based on the respective rule should decrease. For instance, if an observation of C+ together with E- (see cell b in Fig.1) is made, the aspect of sufficiency of C for E is weakened. Subsequently, inferences based on R1 go along with a reduced certainty. The second implication touches the central issue of this paper. It illustrates our assumption of devaluation. We assume that negative evidence for one aspect of causality will be reflected by increased uncertainty about the complementary aspect of causality. For instance, if R1 is discredited by negative evidence (observations that fall into cell b), confidence in R2 decreases as well. We call this the effect of devaluation. Table 2 shows which observation discredits and devaluates the rules of the model of unique causation.

So far, we have described the consequences of positive evidence that strengthens a rule and the consequences of negative evidence that weakens a rule in terms of discrediting and devaluation. This leads to three hypotheses:

1. **Strengthening:** Observations that fall into cell a and d of the contingency tables provide positive evidence for the respective rules of the models and should increase the confidence in inferences drawn from these rules.
2. **Discrediting:** Observations that fall into cell b and cell c provide negative evidence and *discredit* the rules as shown in table 2. In all these cases, the confidence in inferences drawn from the rules should get reduced.
3. **Devaluation:** Observations that fall into cell b and cell c should *devaluate* the complementary rules as shown in table 2. Again, the confidence in inferences from the affected rules should decrease.

The following experiment serves to test these hypotheses.

Table 2: Discrediting and devaluating causal rules.

Rule	Observation	Discrediting of	Devaluation of
R1: C+→E+	C+, E-	R1	R2
R2: C- → E-	C-, E+	R2	R1

Experiment

In our study, participants had to acquire causal knowledge about a simulated technical system based on inductive learning. Over the course of the experiment, positive as well as negative evidence was presented to investigate the consequences of discrediting and devaluation.

Method

Participants. Sixty graduate and undergraduate students at the Berlin Institute of Technology were recruited for the experiment. All of them were paid for their participation.

Material. Figure 2 shows the schematic screen layout of the simulated system that was presented to the participants. It was introduced as an electrical system of a power plant. The system was built up from four subsystems that were responsible for two output systems. Information about the state of these subsystems was displayed on four dials (for top boxes in Fig.2). Each dial represented the state of one subsystem, which was either DOWN (C+) or UP (C-), or unknown because its dial was switched off. Only one subsystem was causally relevant and served as cause C for the outcome of the relevant output system (either E+ or E-). The other three subsystems were irrelevant for the task. One of them was unused (the dial was switched off) while the other two were used as distractors to give the system a more diversified appearance. In the lower half of the screen, the displays for the output systems were shown. In some of the trials participants had to predict the outcome of only one of them and in the remaining trials they had to predict the outcome of both. If only the outcome of one system had to be predicted, the display of the other output system was not shown. Whereas one output system (*E*) was relevant for the experiment the other was used to make the task more realistic.

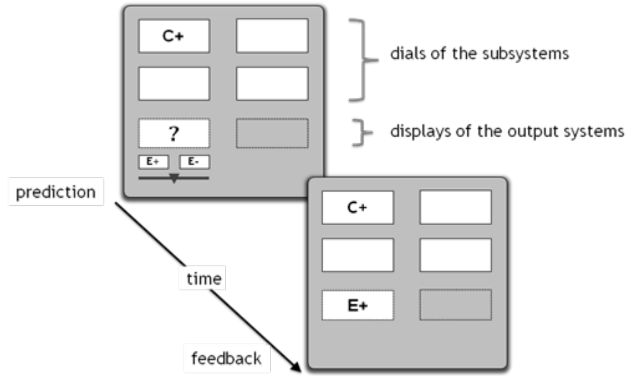


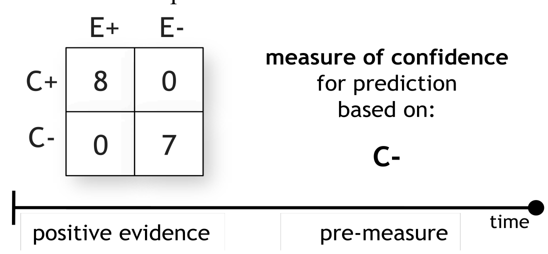
Figure 2. Screen layout (schematic) and sequence of one trial of the experiment.

Below the display of each output system two buttons were shown for the prediction of the outcome. One button served the prediction of MALFUNCTION (E+) and the other one the prediction of working operation (OK) (E-). Clicking on one of them was necessary to make the prediction. Finally, below these buttons a slider was presented that could be adjusted to rate the confidence of the judgments. The lowest confidence (0%) was set in the middle of the slider. Subjects were instructed to place the slider on the very right to indicate full confidence (100%) that E- will occur, and on the very left to mark full confidence (100%) for E+.

Procedure. The participants' task was to predict the outcomes (E+ or E-) of the output system(s). To solve this task, they had to understand the underlying causal relation between the subsystems and the output systems. In each trial, they were shown the layout of the device as presented in Figure 2. First, subjects had to check the operation of the subsystems. Then, based on the information, which was shown on the dials, they were requested to predict the state of the output system(s) by clicking on the respective buttons (OK or MALFUNCTION). Finally, they rated their confidence for each prediction by adjusting the respective slider(s). After participants finished their prediction and confidence rating, they had to click on a 'send' button and subsequently received feedback that showed the actual outcome(s). The experiment consisted of thirty-three trials. These trials were split up in a *reinforcement phase* and a *test phase*. Figure 3 depicts the experimental procedure schematically. Note that the frequencies in the cells of the contingency tables in Figure 3 (b) are summed up for both phases. In the reinforcement phase, which consisted of twenty-six trials, participants received information that enabled them to acquire a model of unique causation with two rules (R1 & R2, see Table 2). This was accomplished by providing positive evidence for R1 (eight trials, see Fig.3) and R2 (eight trials, see Fig.3). Additionally, there were two distractor trials in which information about an irrelevant subsystem was shown only. In the remaining eight trials participants had to predict only the outcome of the second output system that was irrelevant for the test of the hypothesis. After the twenty-six trials of the reinforcement phase the test phase started that consisted of seven trials. In four of these seven trials, negative evidence for one of the two rules (R1 or R2) was presented. The negative evidence always opposed the rule reinforced in the last trial of the reinforcement phase. In these trials people had to predict the outcome of the relevant output system (*E*) only. Another two trials were used as distractor trials presenting information about one of the irrelevant subsystems. In the seventh and last trial of the test phase the post-measure for the relevant test was recorded. Therefore data were presented that matched the same rule as in the last trial of the reinforcement phase (see Fig.3).

Independent and dependent variables. Since the model of unique causation consisted of two rules, both were used to investigate the issues of discrediting and devaluation. For this purpose, the sample of sixty participants was split into two groups of thirty participants each. One group received negative evidence about R1, the other half about R2.

(a) Reinforcement phase.



(b) Test of Devaluation Effect phase.

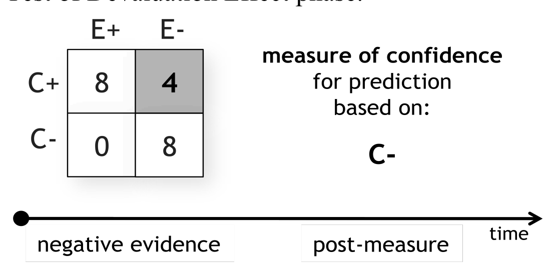


Figure 3. Experimental procedure (schematic). Reinforcement phase and test phase were presented in sequence. Values are exemplary for one of the two experimental groups. Contingency table in (a) displays frequencies for reinforcement phase. The trial of the pre-measure also was the eighth presentation of (C-, E-), which is updated in (b). The contingency table in (b) displays summed frequencies for reinforcement phase and test phase.

To investigate the strengthening of rules, the amount of *positive evidence* ranged from one to eight trials (see Fig. 3a, positive evidence) for each rule (R1 & R2). To test the impact of discrediting, the amount of *negative evidence* ranged from one trial to four trials (see Fig.3b, negative evidence) for each rule (R1 & R2). The factor measurement with the factor levels *pre* and *post* served the investigation of devaluation as described in the procedure (see Fig.3). Throughout the experiment, confidence ratings of inferences predicting the states of the relevant output system were used as dependent variable.

Results

For statistical analysis, we computed three ANOVAs with repeated measures, one for each effect. Additional to the significance of effects we report effect sizes after Cohen (1988). Cohen (1988) defines small effects from $0.10 < f < 0.25$, medium effects from $0.25 < f < 0.40$ and large effects from $f > 0.40$. The effect of *strengthening* was analyzed with a one-factorial ANOVA with repeated measures for each rule. We used the number of occurrences of positive evidence (1-8) as independent variable. Strengthening greatly affected subjects confidence ratings for R1 ($F(7,413)=57.12$, $p<0.01$, $f=0.98$) as well as R2, $F(7,413)=46.83$, $p<0.01$, $f=0.89$. Figure 4 shows the effects of strengthening on subjective confidence for both rules. As depicted, subjects' confidence in their prediction of the state of the output system strongly increases over time.

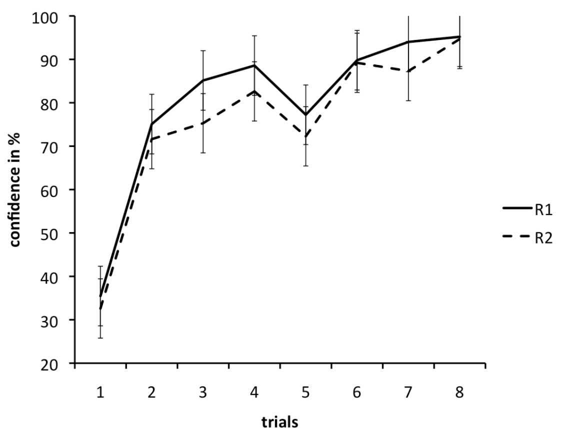


Figure 4. Effect of positive evidence on confidence ratings, depending on the number of trials. Error bars represent standard error.

For *discrediting* rules 1 and 2 (R1 & R2), the four trials with *negative evidence* were run to weaken subjects' confidence in their predictions. Since rule one was discredited for half the subjects and rule two was discredited for the other half, rule became a factor in the analysis. Therefore a 2x2 ANOVA with repeated measurement was calculated in which the *rules* of the model (R1 and R2) served as between subjects factor and *negative evidence* (trials 1-4) was a within subjects factor. We found a significant large main effect of *negative evidence* ($F(3,174)=18.19$, $p<0.01$, $f=0.56$), but no effect of *rules* ($F(1,58)=0.03$, $p=0.95$, $f=0.00$) nor an interaction effect ($F(3,174)=1.96$, $p=0.12$, $f=0.18$). Figure 5 visualizes the results. To investigate the effect of *devaluating* a rule (Fig. 6), it seems necessary to highlight how we achieved the data for this computation. For all subjects rule 1 and rule 2 were strengthened. The last trial of the strengthening phase for each rule (trial 8) served as pre-measure. However, only for half of the subjects rule 1 was discredited. If these subjects' confidence for the prediction of rule two (post-measure) was lower after discrediting rule one, devaluation took place. Reversely, for the other half of the sample rule 2 was discredited. Hence, if subjects' confidence for rule 1 (post-measurement) also decreases, devaluation worked as well.

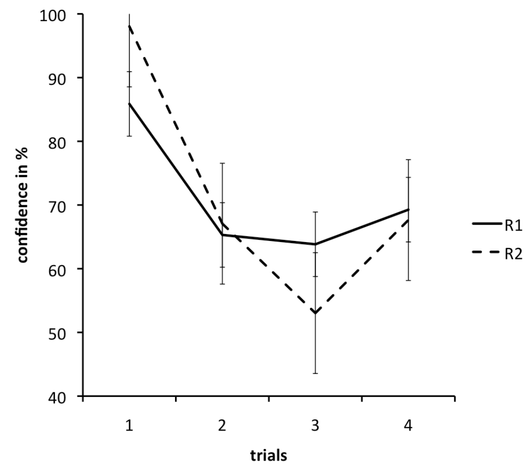


Figure 5: Effect of negative evidence on confidence ratings. Error bars represent standard error.

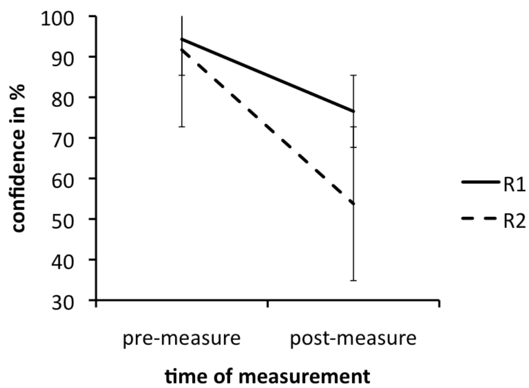


Figure 6: Effect of devaluation on confidence ratings for both rules. Error bars represent standard error.

A 2x2 ANOVA was calculated over the between subjects factor *rule* (either R1 or R2 was discredited) and the within subjects factor *measurement* (pre- and post-measure). This analysis revealed a medium main effect of *rule* ($F(1,58)=4.47$, $p=0.03$, $f=0.28$) and a large main effect of *measurement* ($F(1,58)=42.66$, $p<0.01$, $f=0.85$). Additionally, we observed a medium significant interaction, $F(1,58)=5.58$, $p=0.02$, $f=0.31$. Figure 6 visualizes these effects.

Discussion

In the present paper we tested three. First, we assumed that positive evidence strengthens subjects' confidence for predictions they derived from a set of rules that was acquired in the course of an experiment. Empirical evidence supported that hypothesis. At the end of the strengthening phase peoples' confidence was close to 100%. Second, we expected a decrease in participants' confidence in their causal inferences if negative evidence discredited the respective rules. This hypothesis was empirically confirmed as well. Finally, hypothesis three claimed, that negative evidence for one aspect of causality results in decreased confidence in the complementary aspect as well. Empirical findings clearly supported this assumption. This result opposes a normative view that would require people to base their predictions solely on the given facts. For example, given C- subjects should predict E- with high confidence. In contrast, despite their correct prediction of E- (in case of C-) participants confidence decreased with respect to the critical test in the post-measure. This effect emphasizes the idea that humans do not consider sufficiency and necessity as independent of each other. Instead, once people have acquired causal knowledge, they take evidence for both aspects into account. They do so, even if the predictions they make are solely based on one of them. Hence, we conclude that people mentally construct causal models that relate sufficiency and necessity. These models can be seen as a whole. If one part or aspect of such a model proofs to be wrong, subjects loose their confidence for the complementary part as well. Existing models of causal learning and reasoning aim to explain integrative judgments. Hence people are required to integrate information over all four cells of the contingency table. Thus, they always have to consider both aspects of causality. Therefore these models do not fit to the conditions of the experimental task. Nevertheless, assuming that subjects frame the task in our experiment as to judge the strength of the relation of C and E, the Power PC model (Cheng, 1997) would predict a confidence level of 66% for the post-measure. This is within the range of our results. Hence, if subjects are asked to make predictions in a causal learning paradigm, they reframe the task to judge the strength of a causal relation of two events. According to Griffith and Tennenbaum (2005) parts of the experimental task can be described as causal structure learning. From this point of view presenting negative evidence for one aspect would favor a different causal structure (compound causation or alternative causation respectively). Hence, it might be that peoples' post-measure judgments reflect their preference for the new structure compared to the previous one. Alternatively the post-measure judgments might reflect participants' uncertainty regarding the new structure. Again, these alternative explanations require people to integrate over all contingency information. In contrast to these alternative explanations there are models of inductive causal learning that are based on cognitive architectures and that emphasize the role of declarative memory (Drewitz & Thüring, 2009; Drewitz & Brandenburg, 2012). These models account for peoples' judgments and their confidence ratings given positive as well as negative evidence. They provide a possible explanation for peoples' performance in inductive learning based on memory processes. Additionally they do not assume that people reframe the experimental task from prediction to integrative judgments. To discriminate between these explanations, future research should focus on the replication of the devaluation effect for more complex causal models and different dependent variables like reaction times and pupil dilation. If we can replicate the effect we also might be able to differentiate between possible alternative explanations.

References

Allan, L. G. (1980). A note on measurement of contingency between two binary variables in judgment tasks. *Bulletin of the Psychonomic Society, 15,* 147-149.

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

Cheng, P.W. & Novick, L.R. (1992). Covariation in natural causal induction. *Psychological Review*, *99*, 365-382.

Cohen J. (1988). Statistical power analysis for the behavioral sciences. Hillsdale, New Jersey.

Drewitz, U. & Brandenburg, S. (2012). Memory and Contextual Change in Causal Learning. In N. Rußwinkel, U. Drewitz & H. van Rijn (eds.), Proceedings of the 11th International Conference on Cognitive Modeling, Berlin: Universitätsverlag der TU Berlin.

Drewitz, U. & Thüring, M. (2009). Modeling the Confidence of Predictions: A Time Based Approach. In A. Howes, D. Peebles, R. Cooper (Eds.), 9th International Conference on Cognitive Modeling, Manchester, UK.

Einhorn, H. J. & Hogarth, R. M. (1986). Judging probable cause. *Psychological Bulletin*, *99*, 3-19.

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

Jenkins, H.M. & Ward, W.C. (1965). Judgment of contingency between responses and outcomes. *Psychological Monographs: General and Applied*, 79, 1, 1-17.

Mill, J. S. (1869). *A system of logic, ratiocinative and induc*tive. London: Longman, Roberts & Green.

Thüring, M. & Jungermann, H. (1992). Who will catch the Nagami Fever? Causal inferences and probability judgments in mental models of diseases. In D.A. Evans & V.L. Patel (Eds.), *Advanced models of cognition for medical training and practice* (307-325). Berlin: Springer.

Thüring, M., Drewitz, U., & Urbas, L. (2006). Inductive Learning, Uncertainty and the Acquisition of Causal Models. In R. Sun & N. Miyake (Eds.), Proceedings of the 28th Annual Cognitive Science Society. NJ: LEA.

Waldmann, M. R. & Hagmayer, Y. (2001). Estimating causal strength: The role of structural knowledge and processing effort. *Cognition, 82*, 27-58.

Waldmann, M. R. & Holyoak, K. J. (1992). Predictive and diagnostic learning within causal models: Asymmetries in cue competition. *Journal of Experimental Psychology: General, 121,* 222-236.

White, P. A. (2003). Making causal judgments from contingency information: the pCI rule. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 710-727.