Causal Attribution As Mechanism-Based Story Construction: An Explanation Of The Conjunction Fallacy And The Discounting Principle

Woo-kyoung Ahn
Department of Psychology
University of Louisville
Louisville, Kentucky
wahn@cog.psych.lsa.umich.edu

Jeremy Bailenson
Brian Gordon
Department of Psychology
University of Michigan
Ann Arbor, Michigan 48104

Abstract

We propose that causal attribution involves constructing a coherent story using mechanism information (i.e., the processes underlying the relationship between the cause and the effect). This processing account can explain both the conjunction effect (i.e., conjunctive explanations being rated more probable than their components) and the discounting effect (i.e., the effect of one cause being discounted when another cause is already known to be true). In the current experiment, both effects occurred with mechanism-based explanations but not with covariation-based explanations in which the cause-effect relationship was phrased in terms of covariations without referring to mechanisms. We discuss why the current results pose difficulties for previous attribution models in Psychology and Artificial Intelligence.

Introduction

Judging causality when there are many possible explanations is a very common, yet poorly understood process. Our focus is on the reasoning processes involved in causal interactions, especially on what is known as the conjunction fallacy and the discounting principle.

In their description of the conjunction fallacy, Leddo, Abelson, and Gross (1984) showed that people commit the fallacy of rating the probability of conjunctive explanations as more likely than the probabilities of their constituents. For example, subjects received a story about John's decision to attend Dartmouth. They were asked to rate the likelihood of single explanations (e.g., "John wanted to attend a prestigious college," "Dartmouth offered a good course of study for John's major") and conjunctive explanations (e.g., "John wanted to attend a prestigious college and Dartmouth offered a good course of study for John's major"). Normatively speaking, the probability that two assertions are both true can never exceed the probability that either one alone is true. However, subjects rated two reasons more likely than one reason.

Another well-known phenomenon on causal interactions is the discounting principle proposed by Kelley (1972); people tend to discount the effect of one cause when another cause is already known to be true. Suppose Mary took John's radio. Perhaps this happened because Mary's radio was broken and / or Mary was angry with him. When John finds out that she took it because her own radio was broken, he would, according to the discounting principle, discount the possibility that it was because Mary was mad at him.

Several researchers have pointed out the paradox of these two effects (McClure, 1988; Zuckerman, Eghrari, & Lambrecht,1986). Whereas the discounting principle implies that when two causes are available, one is singly preferred, the conjunction effect implies that two causes are preferred to one. What causes these phenomena? One possibility is that they result from two different processes or strategies (one normative and the other non-normative), depending on the tasks or the experimental instructions. Another possibility is that the conjunction effect occurs with only certain types of materials whereas the discounting effect occurs with other types of materials. Using the same materials for the two tasks, however, Morris and Smith (in preparation) found that the two effects could simultaneously occur. In this paper we argue that the two phenomena are based on a single process and they can both occur with the same stimulus materials under appropriate conditions. Before presenting our own view, the following section briefly reviews previous causal attribution theories in order to clarify our approach to this issue.

Various Approaches To Causal Attribution

Covariation Approach

Traditionally, the principle of covariation has been considered the normative principle underlying causal attribution; "the effect is attributed to that condition which is present when the effect is present and which is absent when the effect is absent" (Kelley, 1967, p. 194). Suppose one wants to find out why Kim had a traffic accident last night. The covariation approach will start out with possible

candidate factors found in the event description, such as "Kim" and "last night." One would examine the covariation between these factors and the event. For instance, if other people did not have accidents at that time and Kim tends to have car accidents on other occasions, the event is attributed to be caused by something special about Kim rather than last night. Based on this principle, many models of causal attribution have been developed in Psychology (e.g., Cheng & Novick, 1990, 1992; Hewstone & Jaspars, 1987; Kelley, 1973).

Mechanism approach

Recently, Ahn, Kalish, Medin, and Gelman (submitted) demonstrated that people were not satisfied with explanations solely based on covariation information. Instead, it was found that when seeking the cause of an event, people attempted to discover the processes or mechanisms underlying the relationship between the cause and the effect. Suppose Kim had a traffic accident last night. The subjects in Ahn et. al's experiments were instructed to ask any questions they would like to know the answers to when attempting to explain the given event. Rather than asking such covariation questions like "Is Kim more likely to have traffic accidents than other people are?" they asked such questions like "Was Kim drunk?" or "Was the road icy?" referring to underlying mechanisms not stated in the event descriptions. In addition, the explanations spontaneously generated by the subjects went beyond the level of covariations; they involved a new set of theoretical entities, theories, or processes which were not present in the event description. These experiments showed that knowledge about underlying mechanisms was more important than covariation in making causal attributions. We call this the mechanism approach.

In this paper, we present a more specific account of the mechanism-based model of judging causal strength. Given an event description (e.g., "Kim had a traffic accident"), we argue that people would first retrieve the most available and representative instance of mechanism information for the target event (e.g. drunk driver. Then, people construct a situational model by combining this mechanism information with the event description. In building the situational model, if there are slots that are not specified by the target event description (e.g., Kim's vision), then people would fill in the slots with default values (e.g., normal vision) as proposed by previous schema theories (Schank & Abelson, 1977). The judgment of causal strength is equivalent to judgment of the plausibility of this situational model. In judging the plausibility of the situational model, people run a simulation of the model and judge how likely it is that the given situation would lead to the effect. In the next section, we discuss how this model can account for the conjunction and the discounting effects given multiple candidate causes.

Mechanism-based Account Of The Conjunction And Discounting Effects

The Conjunction Effect

We propose that the causal attribution process is equivalent to using information about the mechanisms underlying the cause(s) and the effect to create a coherent explanation. Going back to our previous example of Kim's having a traffic accident, when asked to rate a single explanation (e.g., "Kim is near-sighted"), people would perform a mental simulation where Kim, who is near-sighted, had a traffic accident under normal driving conditions (default assumption), and make a judgment about the likelihood of this simulation (i.e., how likely this simulation is to actually occur).

Now, consider a conjunctive explanation, "Kim is near-sighted and there was a severe storm last night," which can be easily combined into a coherent model based on a single mechanism. Most people can easily imagine that when poor vision is combined with the severe storm condition, the traffic accident is more likely to occur. Compared to the single explanation, the conjunctive explanation would lead to a simulation which most people would believe is more likely to have led to a traffic accident in the real world. Therefore, they would rate the conjunctive explanation as more likely.

This account also specifies conditions in which the conjunction effect will not occur. We argue that conjunctive explanations tend to be judged more likely than a single explanation as long as the multiple causes can be combined into a single mechanism-based story. If people fail to simulate how the two causes together can lead to the event with respect to a single mechanism, the conjunction effect will not occur. Given the Kim's traffic accident event, suppose the conjunctive explanations are covariation-based without explicit reference to an underlying mechanism; "It was because Kim was much more likely to have traffic accidents than other people are and because it was much more likely to have traffic accidents last night than other nights." This type of covariation-based explanation does not allow people to come up with a single mechanism encompassing both explanations, resulting in no benefit of having two explanations. In sum, we predict that the conjunction effect will occur only with mechanism-based explanations and not with covariation-based explanations.

The Discounting Effect

As discussed earlier, in the discounting task, one cause is given to be true and the subject is asked to judge the probability that an additional cause also influenced the event. When receiving one cause as an established fact, people would construct a mechanism-based story. If this story conflicts or does not fit well with the additional cause to be judged, then the second cause would be judged as less likely (i.e., discounted).

Suppose that the explanation, "Kim is near-sighted," is given to be true for Kim's traffic accident. Then, people would construct a story in which the Kim's near-sightedness is severe enough to cause a traffic accident under normal driving condition. Again, people would assume that the road condition was normal based on their default information. When the additional cause, "There was a severe storm," is given to be judged, it would be extraneous in the framework of the initial simulation of the event where the driving condition was normal. Therefore, the additional cause would be discounted.

The discounting effect depend upon whether people initially have a mechanism-based explanation or not. Suppose both explanations were covariation-based (e.g., "Kim is much more likely to have traffic accidents than other people are"). Given covariation-based explanations without any mention of mechanisms, no story can be constructed based on mechanism information. Consequently the additional cause would not conflict with the first explanation and would not be discounted to the same degree, if at all. In brief, as in the conjunction effect, the discounting effect is predicted to occur only with mechanism-based explanations.

Experiment

These issues are tested in an experiment where the subjects received a series of event descriptions and performed either the conjunction task or the discounting task on two versions of explanations: mechanism or covariation-based explanations.

Method

Procedure. Subjects received a series of problems about 6 events consisting of a target event description and a candidate explanation. For the conjunction task, the subjects were asked "to rate on a scale of 1 to 7 how probable it is that the explanation constitutes at least part of the actual explanation." For the discounting task, the subjects were asked "to rate the magnitude of the various possible causes for the event on a 7-point scale." For both tasks, 1 on the scale indicated "very low," 7 indicated "very high," and the intermediate numbers indicated the intermediate values. They were also instructed that throughout the experiment, the same event descriptions would be presented several times but with different possible explanations. They were asked to treat each problem separately (i.e., "an explanation given in one problem has nothing to do with an explanation given in another problem even when the event description is the same"). The order of the problems given to the subjects was randomized across all the subjects. The subjects performed the task at their own pace.

Design and materials. There were 6 event descriptions consisting of person, stimulus, and occasion factors. For each event, two factors were chosen as candidate factors (e.g., Kim (person) and last night (occasion) in the event, "Kim had a traffic accident last night"). For each factor in question, we developed a mechanism-based explanation (e.g., "Kim is near-sighted and tends not to wear her glasses while driving") and a covariation-based explanation (e.g., "Kim is much more likely to have traffic accidents than other people are").

For the conjunction task, three problems were developed for each event description: ratings for (a) single explanation for Factor A (P(A) henceforth), (b) single explanation for Factor B (P(B) henceforth), and (c) both explanations (P(A&B) henceforth). (See Table 1 for examples of each type of problem.) For each subject, all three problems for each event were either the mechanism type or the covariation type. Each subject received three events with two mechanism-based single explanations and one conjunctive explanation for a total of nine problems, and the other three events with two covariation-based single explanations and one conjunctive explanation for a total of nine problems.

For the discounting task, four problems were developed for each event: estimating (a) strength of factor A (P(A) henceforth), (b) strength of factor B (P(B) henceforth), (c) strength of factor A given factor B (P(A|B) henceforth) , and (d) strength of factor B given factor A (P(B|A) henceforth). (See Table 2 for examples of each type of problem.)

Subjects. There were 62 subjects who were undergraduate students at the University of Michigan, participating in partial fulfillment of course requirements for introductory psychology. Half of the subjects were randomly assigned for the conjunction task and the other half were assigned for the discounting task.

Results

The conjunction effect. As shown in Table 1, the conjunction effect was much stronger with mechanism-type explanation than with covariation-type explanation . There was a reliable interaction effect of the number and the type of explanations, F (1, 30) = 4.915, $p < .05$. In addition, the mechanism-type explanations were rated reliably higher than the covariation-type explanations, F (1, 30) = 10.28, $p < .001$.

The discounting effect. Similarly, the discounting effect occurred with the mechanism-based explanations but not with the covariation-based explanations. (See Table 2 for the mean ratings.) As in the conjunctive task, there was a significant interaction effect between type of explanation and the number of explanations, F (1, 35) = 12.93, $p < .001$. There was no reliable main effect of types of explanation, $p > .10$ but there was a reliable main effect of number of explanation, F (1, 35) = 7.98, $p < .01$.

TABLE 1 : Mean Ratings for the conjunctive task

	P(A), P(B)	P(A&B)
Covariation-based Explanat-ions	Kim is much more likely to have traffic accidents than other people are. 3.34	Kim is much more likely to have traffic accidents than other people are and traffic accidents were much more likely to occur last night than on other nights. 3.33
Mechanism-based Explanat-ions	Kim who is near-sighted tends not to wear her glasses while driving. 4.80	Kim who is near-sighted tends not to wear her glasses while driving and the road was very slick last night. 5.22

TABLE 2 :Mean ratings for the discounting

	P(A), P(B)	P(A|B),P(B|A)
Covariation-based Explanations	How much more likely is it that Kim has traffic accidents than other people? 4.08	Given that traffic accidents were much more likely to occur last night than on other nights, how much more likely is it that Kim has traffic accidents than other people? 4.17
Mechanism-based Explanations	How likely is it that Kim who is near-sighted tends not to wear her glasses while driving? 4.25	Given that the road was very slick last night, how likely is it that Kim who is near-sighted tends not to wear her glasses while driving ? 3.73

Discussion

Interpretation Of The Results

Unlike what some researchers have suggested, the conjunction effect and the discounting effect can occur with the same materials. In addition, both effects occurred only with the mechanism-based explanations. We have hypothesized that people have tendency to make a coherent story based on mechanisms when explaining events. With the conjunctive, mechanism-based explanations (e.g., Kim is near-sighted and tends not to wear her glasses, and there was a severe storm last night and the roads were very slick last night), we can easily imagine a coherent scenario about how each factor would cooperate for each other. Conjunctive explanations have an advantage over single explanations because the conjunctive explanations provide more evidence for the existence of a mechanism.

However, with conjunctive, covariation-based explanations, a coherent story cannot be easily achieved because the mechanisms are unknown and have to be inferred from the information given to the subject. People can easily list a number of reasons why and how Kim would have had more traffic accidents than other people. Furthermore, when two covariation-based explanations are combined, the number of possible stories is multiplied. The uncertainty involved in coming up with a single story might have led the subjects to rate the probability of the conjunctive covariation-based explanations at the same level as for the single covariation-based explanations.

Similarly, the discounting effect did not occur with the covariation-based explanations presumably because there is no conflicting mechanisms with two abstract covariation-based explanations. In contrast, a given mechanism-based explanation might have led subjects to construct a story incoherent with the additional mechanism-based explanation, resulting in the discounting effect.

Comparison To Other Approaches

Covariation-based models. The covariation models, without additional assumptions, seem to have difficulty explaining the current data. As these models are normative, it would be difficult for them to account for the irrationality of the conjunction fallacy. It would never be logical, when assessing covariational probabilities, to rate the conjunction of two unrelated potential causes which are present while the effect is present as higher than one of its constituent parts.

The discounting principle was explained within the framework of Cheng and Novick's (1992) probabilistic contrast model. They argue that causal strength of a factor A increases as the probability of an event E given A (i.e., $P(E|A)$) increases and as the probability of E given the absence of A (i.e., $P(E|\bar{A})$) decreases. More formally, the causal strength of A equals $P(E|A) - P(E|\bar{A})$. According to Cheng and Novick (1992), if there is an additional cause for

E (say, B), then P(E|$\bar{A}$) would increase because E would occur due to B when A is absent. As a result, the causal strength for A is decreased and hence A is discounted in the presence of B. However, this account of the discounting principle is inconsistent with the conjunction effect for the following reason. According to the model, the strength of conjunctive causes, A and B, equals P(E| A&B) - P(E|$\bar{A}$&B) - P(E|A&$\bar{B}$) + P(E|$\bar{A}$&$\bar{B}$). In order to obtain the discounting effect, P(E| $\bar{A}$&B) has to be increased but this increase would decrease the strength of conjunctive causes. Therefore, these models are unable to explain the presence of the conjunction effect alone, much less the simultaneous occurrence of the conjunction effects along with the discounting principle.

Artificial Intelligence approach. Wellman and Henrion (1991) discussed when the discounting effect ("explaining away" in their terms) would or would not occur in terms of Bayesian theorem. Their account is based on probabilistic notions as follows. In order for the discounting effect to occur, our beliefs must satisfy the following relation;

$$\frac{P(E|\ A\&B)}{P(E|A\&\ \bar{B})} < \frac{P(E|\ \bar{A}\&B)}{P(E|\bar{A}\&\bar{B})}$$

That is, the discounting effect occurs when the proportional increase in probability of the event due to learning B is smaller given A than given $\bar{A}$. When the sign is reversed in the above equation, however, we believe that A and B must occur together.

Consequently, Wellman and Henrion's (1991) analysis implies that the discounting and conjunction effects are exclusive to each other. As a result, their analysis cannot explain why the two effects can occur with the same set of materials where the subjects' beliefs about the conditional probabilities should be the same across the two tasks.

Knowledge structure approach. In contrast to the covariation approach, some researchers sought to imbed causal attribution in the more general process of understanding, using knowledge structures such as schemas, scripts, plans, and goals (Schank & Abelson, 1977). According to the knowledge structure approach, causal attributions are made by matching an event with the appropriate script. The quality of a causal explanation is a function of how well the explanation matches the underlying script (Leddo et. al, 1984).

The knowledge structure approach has attempted to explain why people commit the conjunctive fallacy. Knowledge structures often have more than one goal. Thus, explanations with multiple reasons are rated as more representative of the underlying knowledge structure and are subsequently rated higher. Using similar logic, if a knowledge structure only had a single goal, then a single explanation would seem most representative, accounting for the discounting effect. Having filled in this available slot, the explanation is complete. This scenario, however, would predict that the conjunctive effect and the discounting effect could not occur for the same stimuli. Rather, the internal structure of the script would determine whether a given explanation would be more likely to demonstrate the conjunction or the discounting effects.

Follow-up Studies

Ahn, Bailenson, and Gordon (submitted) present follow-up studies of the current work. These experiments were conducted to test more directly when the conjunction fallocy or the discounting effect will occur. In their Experiment 2, subjects received apparently two independent explanations for each event. For example, given an event "Charles had to leave school," the subjects received explanations "It was because he was 21 year s old" and "The U.N. could not resolve conflict in middle East." Before the subjects judged the likelihood of single, conjunctive (P(A&B)), or conditional explanations (P(A|B), P(B|A)), different types of primes were briefly presented. These primes presumably determined which known mechanism information would be activated before the likelihood judgment. Given a prime that would allow a construction of coherent story over the two explanations (e.g., "DRAFT"), the conjunction effect was increased. That is, the conjunction effect occurred compared to the no prime situation, presumably because the context facilitated construction of a coherent story covering all available causes. But when the prime activated mechanism information supporting only one of the available explanation (e.g., "GRADUATE"), the opposite occurred. That is, the discounting effect was increased when people were primed with a mechanism supporting only one of the explanations. This study clearly shows the importance of mechanism information in causal reasoning processes by manipulationg the occurances of the conjunction or discounting effects.

Conclusion

The current result poses difficulties for most models of causal attributions in Psychology and Artificial Intelligence. People consistently rate the probability of conjunctive explanations as more likely than the probabilities of each constituent explanation. At the same time, there is a tendency to discount all other causes when there is support that a given cause is already responsible for a given event. As covariation theories assume that attribution is equivalent to the estimation of the strength of correlation between two factors, it is impossible for this normative paradigm to explain such contradictory processes. The experiment described in this paper shows that people do not rely solely on such covariation information. Instead, the attribution process relies on using information about the actual mechanisms underlying the causes and the effect to create a coherent story.

Acknowledgements

We would like to thank Charles Kalish, Frances Kuo, Douglas Medin, Brian Ross, Edward Smith, Michael Wellman for their helpful discussion of the study, Douglas

Medin, Susan Gelman, and Charles Kalish for their help on developing materials used for Experiment 1, and Kristin Sweitzer for collecting data for Experiments 1.

References

Ahn, W., Kalish, C. W., Medin, D. L., & Gelman, S. A. (Submitted). The role of covariation versus mechanism information in causal attribution.

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

Cheng, P. W., & Novick, L. R. (1990). A probabilistic contrast model of causal induction. *Journal of Personality and Social Psychology, 58*, 545-567.

Hewstone, M. R. C. (1989). *Causal attribution*, Cambridge, MA: Basil Blackwell.

Hewstone, M. R. C., & Jaspars, J. M. F. (1987). Covariation and causal attribution: A logical model of the intuitive analysis of variance. *Journal of Personality and Social Psychology, 53*, 663-672.

Kelley, H. H. (1972). Causal schemata and the attribution process. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins and B. Weiner, *Attribution: Perceiving the causes of behavior*. Morristown, NJ: General Learning Press.

Kelley, H. H. (1973). The processes of causal attribution. *American Psychologist, 28*, 107-128.

Leddo, J. , Abelson, R. P., & Gross, P. H.(1984). Conjunctive explanations: When two reasons are better than one. *Journal of Personality and Social Psychology, 47*, 933-943.

McClure, J. L.(1988). *The discounting principle in attribution theory*. Unpublished doctoral dissertation, University of Oxford, England.

Morris, M. W., & Smith, E. E. (in preparation). How much explanation does behavior require?: The paradox between discounting and the conjunction effects in attribution.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals and understanding: An inquiry into human knowledge structures*. Hillsdale, NJ: Lawrence Erlbaum.

Wellman, M. P., & Henrion, M. (1991). Qualitative intercausal relations, or explaining "explaining away." *Principles of knowledge representation and reasoning: Proceedings of the second international conference*. 535-546.

Zuckerman, M., Eghrari, H., & Lambrecht, M. R. (1986). Attributions as inferences and explanations: Conjunction effects. *Journal of Personality and Social Psychology, 51*. 1144-1153.