Interpretation of Ambiguous Information in Causal Induction

Jessecae K. Marsh (Jessecae.Marsh@Yale.Edu)
Department of Psychology, Vanderbilt University; 111 21st Avenue South
Nashville, TN 37203 USA

Woo-kyoung Ahn (Woo-kyoung.Ahn@Yale.Edu)
Department of Psychology, Vanderbilt University; 111 21st Avenue South
Nashville, TN 37203 USA

Abstract

The current study investigates how people incorporate ambiguous information into judgments of causal relations. We presented participants with information that was not easily classified into the presence or absence of a candidate cause, breaking a traditional requirement of models of causal induction. We found that people were willing to incorporate this ambiguous information into their collected evidence, instead of ignoring the information as uninformative. Furthermore, people interpreted ambiguous stimuli as evidence most consistent with their prevailing causal hypothesis. These results give an idea of how people begin to determine what can function as a candidate cause in a causal induction problem.

Introduction

Determining the causal relationship between events is an essential skill in daily life. Humans are constantly presented with events for which we must determine probable causes. To be able to manipulate the occurrence of an event in our favor, we must determine what cause resulted in the event. Most of the current debates in the field of causal learning concern how information about presence or absence of one event followed by presence or absence of another should be combined in order to determine the causal relationship between the two events. The current study examines a more fundamental issue of how people first determine whether or not an event is present while they are learning a new causal relation.

Traditional studies in causal induction assume candidate causes exist in distinct, definable states. These states take the form of a candidate cause either being present or absent (e.g., fertilizer was dispensed or not, a button was pressed or not) or as polar opposites that vary along a given dimension (e.g., small or large) (Baker, Mercier, Vallée-Tourangeau, Frank, & Pan, 1993; Cheng, 1997; Shanks, Lopez, Darby, & Dickinson, 1996; Spellman, 1996). In all of these studies definitions of what constitutes presence or absence of a cause is clear. For instance, a button is either pressed or not, or a tank is camouflaged or not.

The assumption of well-defined candidates is also built into all existing models of causal induction (e.g., Cheng, 1997; Cheng & Novick, 1990; Pearle, 2000; Rescorla & Wagner, 1972). For instance, consider the well-known measure of contingency called ΔP. This measure is based on four types of evidence created by combining the presence and absence of the cause (X) with the presence and absence of the effect (E) as shown in Figure 1. The contingency model of causal induction hypothesizes that people's judgments of causal strengths approximate ΔP, which is the difference between the probability of the effect occurring in the presence of the candidate, P(effect | cause), and the probability of the effect occurring in the absence of the candidate, P(effect | no cause) (e.g., Cheng & Novick, 1990). More recently, the Power PC theory uses a ΔP measure weighted by the presence of the effect in the cause's absence (Cheng, 1997). Implicit in these different theories of how causal strength or covariation may be calculated is the assumption that an event would be unambiguously classified into one of the four possible types of information with little complication.

	Candidate present (X)	Candidate absent (~X)
Effect present (E)	XE	~XE
Effect absent (~E)	X~E	~X~E

Figure 1. The four possible types of information available in a well-defined causal induction task. Note: ~X notation is used both here to denote the absence of a candidate cause as well as in other situations to denote the presence of the opposite pole of the candidate.

It can easily be seen why clear-cut stimuli are desirable in an experimental setting or in modeling. However, this assumption is not warranted in many real-life situations. For instance, consider the following hypotheses that a layperson might entertain: stress causes insomnia, humid weather causes joint pain, and cheap shoes cause sore feet. In all of these examples, there is no definite boundary for what counts as presence or absence of stress, humid weather, or cheap shoes. How does one decide what counts as a potential cause?

In the following experiment we examined how people interpret ambiguous information in a causal learning situation. We constructed a basic causal induction problem where participants were asked to track the relationship between one causal candidate and the presence versus absence of a given effect. Unlike the traditional causal reasoning experiment, we introduced variation in our causal candidate that did not make it immediately obvious what constituted presence or absence of the target value. Participants were given a causal candidate that varied along one dimension (e.g., height), producing three distinct forms of the candidate. Two candidate forms were extreme opposites of one another along the varying dimension (e.g., tall and short). The third candidate variation we created was a hybrid of the polar opposite forms, appearing as an absolute midpoint between the two poles along the varying dimension (e.g., medium height). This created a variation of the candidate that was not defined as belonging to either end of the dimension and was therefore ambiguous as to where it belongs in the four types of possible evidence depicted in Figure 1. The question we set out to answer is how such pieces of ambiguous information will be incorporated alongside clearly defined information to produce an estimation of causal status.

Possible Hypotheses

There are at least three plausible hypotheses concerning how people would deal with ambiguous candidate causes. First, people might completely ignore ambiguous information because it cannot be readily classified into one category or another. We call this the **Discrediting Hypothesis**. For instance, in testing the hypothesis of whether one's joint pain gets worse when it is humid, a day that is difficult to be classified as either dry or humid would not be figured into computation because it is an ambiguous case.

The second hypothesis is what we termed the **Coin Toss Hypothesis.** Instead of completely ignoring the ambiguous trials, people would randomly classify each piece of ambiguous information into either X or ~X as if they are tossing a coin in each ambiguous trial. For instance, those days that are difficult to be classified as either dry or humid would be classified as dry half the time, and humid half the time.

The third possibility is that people would assimilate ambiguous events in light of the most likely causal hypothesis they hold at that point in time (**Assimilation Hypothesis**). Consider the previous example of outside humidity as a cause of joint pain. Suppose a person is leaning towards believing that humid days cause her joint pain. One day when her joint pain was particularly bad, the weather was neither definitely humid nor definitely dry. Because she believed high humidity caused her joint pain, she would interpret the weather that day as being humid. Conversely, on a day with the same ambiguous humidity where she did not experience joint pain, she would interpret that day as dry, despite the exact same weather as in the previous case. To put it in more general terms, suppose the most plausible hypothesis at trial N is that X causes E. Further suppose that at trial N the reasoner is presented with an ambiguous value that can be thought of as either X or ~X. If E is obtained in this trial, this ambiguous value would be interpreted as X. If E is not obtained in this trial, this ambiguous value would be interpreted as ~X. Thus, each ambiguous trial is assimilated online to be consistent with the most dominant hypothesis held at that trial.

Methods

Overview

In each condition, participants received a series of 60 trials, each of which described which candidate cause was present and whether an effect was present or absent. The candidate cause took one of three values on a continuous dimension (e.g., short, medium, tall), where the intermediate value was selected through a pre-test to be a value that was judged to be equally similar to the two extreme values. After viewing all 60 trials, participants were asked to estimate the frequency of each of four cells in Figure 1, where X was one of the extreme values and ~X was the opposite extreme value (e.g., In how many cases were the bacteria TALL and white blood cells DID increase? In how many cases were the bacteria SHORT and white blood cells DID increase?). The main question is how the intermediate value was classified during causal learning. According to the Discrediting hypothesis, trials with the intermediate value would be ignored and therefore participants' frequency estimates will reflect only the trials with extreme values. According to the Coin Toss hypothesis, trials with the intermediate value will be sorted roughly in half. According to the Assimilation hypothesis, participants' hypothesis about what causes E will influence how ambiguous trials would be interpreted.

Materials and Design

Two sets of materials were developed (see Figure 2). In Set 1, a causal candidate, described as bacteria, varied in height (short, medium, and tall), and in Set 2, a causal candidate, described as spacing between rows of fruit trees, varied in spacing (near, intermediate, and far-spaced). A pre-test was conducted to verify that the medium value in each of the dimensions was perceived to be equally similar to extreme values on the same dimension.

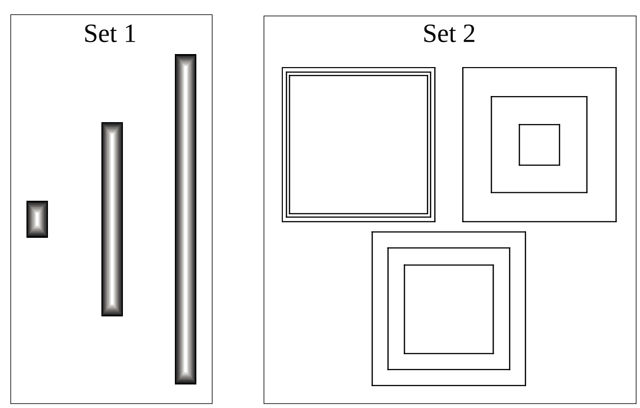

Figure 2. Set 1 and Set 2 targets and ambiguous items.

In the pretest, 17 undergraduate participants received 12 sets of stimuli, each containing two targets (Target A and B) and 15 test items. Within each set, the two targets varied greatly from one another along one dimension and the fifteen different test stimuli varied systematically between the targets. For instance, one set of materials consisted of a 0.1 inches long target and a 3.3 inches long target. The fifteen test items increased systematically in length by 0.2 inches between the two targets (i.e., 0.3, 0.5, …, 2.9, 3.1). All of the items in a set were identical except for the varying dimension.

The participants' task was to rate the similarity of each test item to Targets A and B on a nine-point scale. Participants were told that if they believed a test item was very similar to Target A to give it a number close to 1 and if they believed a test item was very similar to Target B to give it a number close to 9. Participants were also instructed that they could use any number between 1 and 9, keeping in mind that 5 was the midpoint of the scale. After reading the instructions for a given item set, a participant rated all 15 test items in a random order. For the two item sets that were chosen for the actual experiment, participants' mean similarity ratings for the intermediate value (5.0 and 4.76 for Set 1 and Set 2, respectively) were not significantly different from the scale's midpoint of 5 (all p's > .20). Thus, the pre-test verified that the intermediate values that were equally similar to either one of the two extreme values on the same dimension were perceived as such.

For each set of materials, 60 trials were developed for the learning phase. Specifically, 40 trials were developed using the two extreme values and 20 trials were developed using the ambiguous value. As shown in the contingency table in Figure 3, the 40 unambiguous trials were constructed to strongly suggest that X is a generative cause of E (with ΔP of 0.8). In the Ambiguous+ condition, E was present in all 20 trials involving the ambiguous value. In the Ambiguous– condition, E was absent in all 20 trials involving the ambiguous value.

	X	~X
E	18	2
~E	2	18

Figure 3. Number of trials for each type of evidence in the main experiment.

On each trial of Set 1, the value of a causal candidate was depicted on the left and the value of effect was depicted on the right. Figure 4 shows a sample trial. For Set 2 the candidate was shown at the top of the screen and the effect was shown on the bottom. The actual value of X or ~X was counterbalanced across subjects (e.g., In one version of Set 1, X was short bacteria, and in the other version, X was long bacteria).

Cover stories were built around each item set. For Set 1, participants were told that a newly discovered strain of bacteria was believed to be associated with white blood cell production in humans. Medical researchers were now trying to determine the relationship between the height of the bacteria and an increase in a person's white blood cell count. For Set 2, participants were told that the spacing between rows of fruit trees was believed to be associated with population numbers of a type of bird that lived in the trees. Agricultural researchers were trying to determine the relationship between the spacing of the trees and an increase in the bird's population.

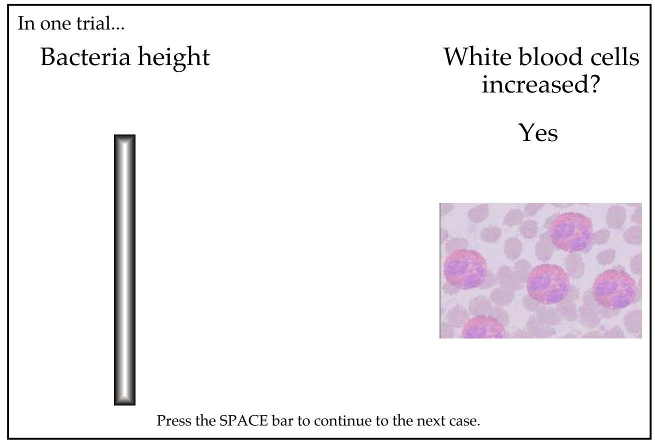

Figure 4. Sample trial for Set 1.

Procedures

Each participant completed both the Ambiguous+ and the Ambiguous– conditions, instantiated in two different sets of stimulus materials. Each condition was presented to a participant as two experiments separated by a filler task. The order of the Ambiguous+ and Ambiguous– conditions as well as which value was chosen as X or ~X within a stimulus set was counterbalanced between subjects.

Participants began each condition by reading one of the previously described cover stories. They were also told that after seeing all of the cases for the given condition they would be asked to estimate the relationship between a target pole of the dimension and the effect event. (E.g., "You will be asked to estimate the relationship between a bacterium having a tall height and an increase in white blood cells.") Most critically, participants were not told what defines the target or given examples of the target; they were only provided with a target label (such as 'tall'). After finishing this introductory material, participants proceeded through the experiment by pressing the space bar. Each case was followed by a 500 millisecond black screen, so that participants could not easily compare the stimuli in adjacent trials. Regardless of conditions, the order of events was fixed as the order shown in Figure 5.[1]

After seeing all of the trials for a given condition, participants made a series of ratings for the trials they had just seen. The first rating asked participants to recall how many individual pieces of each evidence type they had seen. The following is an example for how participants were asked to respond for Set 1:

"In how many cases were the bacteria:
1) TALL and white blood cells DID increase?
2) TALL and white blood cells DID NOT increase?
3) SHORT and white blood cells DID increase?
4) SHORT and white blood cells DID NOT increase?"

[1] As will be further discussed in Discussion, the results we obtain appear to depend on when the ambiguous trials first appear in the sequence. In the current experiment, which did not investigate that issue, the order was fixed in order to minimize that effect.

XE	~X~E	Amb	~X~E	XE	Amb	~X~E	Amb	XE	Amb	~X~E	Amb
XE	Amb	~X~E	Amb	XE	~X~E	XE	Amb	~X~E	~XE	XE	Amb
~X~E	Amb	XE	Amb	~X~E	Amb	XE	~X~E	XE	Amb	XE	Amb
~X~E	X~E	XE	Amb	~X~E	~XE	XE	Amb	~X~E	Amb	XE	Amb
~X~E	Amb	XE	X~E	~X~E	XE	~X~E	Amb	XE	~X~E	XE	~X~E

Figure 5. Trial sequence for all experimental conditions. The sequence should be read from left to right, top to bottom. 'Amb' refers to ambiguous trials.

For each estimation, questions concerning the cued target dimension were listed first. Participants were told that they had seen 60 total trials, but were not explicitly told that their estimations must add to 60. Ambiguous stimuli were not directly asked about, and it would be under participants' discretion as to how to count ambiguous values.

In the second rating participants were asked to estimate to what extent X caused E (e.g., "To what extent does bacteria that are tall cause an increase in white blood cells?") Participants were asked to make their rating on a scale of 0 (not a cause) to 100 (strongly causes) and then asked to rate their confidence on a scale of 1 (not at all confident) to 7 (very confident).

In the last rating participants were asked to judge the similarity of the ambiguous stimuli to the target and alternative stimuli they had seen in the preceding block. The same type of rating system was used as in the pre-test. This final task was added to verify one more time that the same participants also agreed that the ambiguous values used in the main experiment are equally similar to their extreme values.

Thirty-four Vanderbilt University undergraduates participated in the experiment in partial fulfillment of a course requirement. Five additional subjects participated but were dropped from the study for not following the experimental directions (e.g., their total frequency estimates exceeding 60 trials). All experiments were conducted on iMac computers using SuperLab™.

Results

Figure 6 summarizes predictions of the three hypotheses discussed in the introduction, and the results from the experiment on frequency estimates. There was no effect of block order, materials, or which dimension served as X, so all results are collapsed across materials and order.

To test the Discrediting hypothesis, the data was first examined to see if participants were in fact not incorporating ambiguous trials into their estimates. If participants were following the Discrediting hypothesis, then their estimates should not exceed 40 (i.e., total number of trials not involving ambiguous values). Yet, 81% of the estimate sets reflected a total of 60 trials added across evidence types, and 9% reflected a total trial set of over 40. Therefore only 10% of the responses could possibly have been generated employing a discrediting strategy. In addition, the mean estimate of XE in the Ambiguous+ condition (28.71) was significantly greater than a hypothesized mean of 18 (the idealized estimation for these cells if no ambiguous information was included), $t(33) = 8.55$, $p < .0001$. The mean estimate of ~X~E in the Ambiguous– condition (28.24) was also significantly greater than 18, $t(33) = 7.57$, $p < .0001$).

If the Coin Toss hypothesis is supported, we should expect that in the Ambiguous+ condition, frequencies of both XE and ~XE would increase to the same extent compared to frequencies of extreme trials only (i.e., 18 and 2, respectively). Yet, in the Ambiguous+ condition, the degree to which participants' frequency estimate on ~XE (3.53) departed from 2 was much less than the degree to which their estimate on XE (28.71) departed from 18. More specifically, each participant's estimate on ~XE was subtracted from 2 and their estimate on XE was subtracted from 18, and a paired-sample t-test was carried out for these difference scores, finding a significant difference, $t(33)= 5.31$, $p < .0001$. Furthermore, participants' frequency estimate on ~XE did not differ from their estimate on X~E (3.15), $p > .12$, contrary to the expectations of the coin-toss hypothesis. Similarly, in the Ambiguous– condition, the Coin Toss hypothesis would predict that frequencies of both X~E and ~X~E would increase to the same extent.

		Hypotheses									Results	
		Discrediting			Coin Toss			Assimilation				
		X	~X		X	~X		X	~X		X	~X
Ambiguous+	E	18	2	E	18 +*10*	2 +*10*	E	18 +*20*	2	E	28.7	3.5
Condition	~E	2	18	~E	2	18	~E	2	18	~E	3.1	21.8
		X	~X		X	~X		X	~X		X	~X
Ambiguous–	E	18	2	E	18	2	E	18	2	E	21.2	3.5
Condition	~E	2	18	~E	2 +*10*	18 +*10*	~E	2	18 +*20*	~E	2.9	28.2

Figure 6. Division of trials according to each hypothesis and actual results. Highlighted cells note where ambiguous information should be parsed for that hypothesis. Italicized numbers reflect idealized number of trials parsed to that cell.

However, in the Ambiguous– condition, a paired-sample t-test, similar to the one conducted in the Ambiguous+ condition, found a significant difference for these difference scores, $t(33) = 6.78$, $p < .0001$. In addition, participants' estimate on X~E (2.85) did not differ from their estimate on ~XE (3.53), $p > .17$. Also, looking across Ambiguous+ and Ambiguous– conditions, there was no reliable difference between the estimates on the X~E in the Ambiguous+ (3.15) and Ambiguous– (2.85) conditions, $t(33) = .45$, $p = .65$, nor was there a difference between the estimates on the ~XE in the two conditions (both with a mean of 3.53).

The above results are most consistent with the Assimilation hypothesis. Recall that the non-ambiguous trials were constructed such that participants would be most likely to believe that X causes E. Thus, in the Ambiguous+ condition where ambiguous trials are accompanied with E, ambiguous trials would be more likely to be interpreted as XE rather than ~XE. Indeed, as reported earlier, this was the result that was obtained. Conversely, in the Ambiguous– condition where ambiguous trials are accompanied with ~E, they would be more likely to be interpreted as ~X~E rather than X~E (because participants presumably believe X causes E). Again, as reported earlier, this was exactly the pattern that was obtained.

There was no difference for the overall causal strength estimate across conditions ($M = 79.30$ in the Ambiguous+ condition, $M = 79.62$ in the Ambiguous– condition, $p > .89$). There was also no difference in the confidence rating for the overall causal strength estimates across conditions ($p > .11$). Finally, participants in the main experiment also agreed that the ambiguous values were equally similar to the two extreme values, as shown by no significant difference for each rating from the midpoint of 5 on the similarity scale (all p's > .56).

Discussion

The current results show that an identical value is classified in different ways depending on a dominant causal hypothesis. Thus, if a person believes that tall bacteria increase white blood cell counts, then medium-height bacteria that also increased white blood cell counts are more likely to be classified as being tall. However, the bacteria that had the same medium-height, but did not increase white blood cell counts are more likely to be classified as being short. In this way, the power of the causal relationship gives the otherwise ambiguous stimuli an unambiguous nature.

Although the current experiment examined classification of causes only, we speculate that similar results would be obtained along the effect dimensions. Going back to our previous example, a person who believes that humid weather makes joint pain worse might be more likely to believe that joint pain in medium intensity is worse on a humid day than on a dry day.

The crux of the assimilation hypothesis is that interpretations of whether or not a cause is present or an effect is present are constrained by the on-going hypothesis that a reasoner holds. Furthermore, this assimilation would influence the kind of causal relation that a reasoner could induce. For instance, after interpreting medium-height bacteria that increase white blood cell count as "tall" bacteria, their hypothesis of what bacteria increase white blood cell counts would include "tall and medium values". Thus, causal induction involves a continuous interaction between top-down and bottom-up processes where the dominant hypothesis influences interpretations of incoming data, which in turn would influence causal induction (e.g., Wisniewski & Medin, 1994). The current models of causal induction are moot about this interaction in that they all assume that events come pre-classified as one of the four event types in Figure 1, and the induction process takes place over these pre-classified data. Yet, our results indicate that perceptually identical stimuli can be classified differently under the influence of a developing hypothesis of a causal relation.

The idea of changing one's classification of an object according to contextual information has also been demonstrated in categorization research. Research on similarity judgments has found that traits identified for an object can change according to the grouping of an object in a comparison task. For example Medin, Goldstone, and Gentner (1993) found that an object that has an ambiguous number of appendages would be described as having three appendages when compared to an object with three well-defined appendages but paradoxically would be described as having four appendages when compared to an object with four. It is impossible for a given object to physically vacillate between possessing three and four appendages. It is concluded that the context of the task is what drives the trait description. (See also Schyns, Goldstone & Thibaut, 1998).

Compared to this earlier study in categorization and similarity judgments, our results can be seen as a stronger demonstration of how the context of a situation can change the interpretation of ambiguous information. Whereas Medin et al.'s task directly asked participants to compare an ambiguous and a well-defined item, participants in our experiment were not explicitly forced to do such and could freely ignore ambiguous stimuli in the counting task. Despite this stronger manipulation, participants were still willing to evaluate and include ambiguous information into their estimates.

One possible objection to the current study is that our task forced participants to group ambiguous information in accordance with their existing hypothesis in a way that is not reflective of how such candidates would be treated in real life cases. That is, participants could normally treat ambiguous candidates as a third, intermediate value or as a sub-type of one of the defined variables. Most of the existing models of causal induction do not provide for such flexibility and are limited to situations involving candidates with binary values (e.g., presence or absence). Furthermoe, in every-day language people seem to explicitly verbalize causal rules in terms of a contingency between binary-valued variables (e.g., humidity causes joint pain) rather than a complex function that maps two continuous dimensions (e.g., varying levels of humidity and degrees of joint pain). Where does this tidy mapping come from if actual data present in the environment are not as neat as shown in the 2X2 contingency table of Figure 1? Our experiment gives the first idea of how this essential initial process might proceed when confronted with ambiguous information.

What are the boundary conditions for our demonstrations? Can any two items be grouped together given the appropriate dimension and causal hypothesis? We are currently exploring interesting boundary conditions, which will illuminate the process of online incorporation of ambiguous information. In the current study, we have demonstrated the assimilation phenomenon when within the first three trials of an experimental sequence our participants have seen an example of two extreme values and an ambiguous stimulus, as illustrated in Figure 5. In our pilot studies, however, when ambiguous stimuli begin appearing much later in a sequence, assimilation did not take place, presumably because participants had already developed a well-established notion of how they would define a cause (e.g., what counts as tall bacteria). Thus, this assimilation process appears to take place only when ambiguous stimuli are presented at an earlier stage of causal learning.

At the same time, the assimilation process probably would not take place too early in learning because at that point, people might not have developed a hypothesis about a causal relation. For instance, the first ambiguous stimulus they encountered (i.e., third trial in the current experiment) would not have had the benefit of a dominant hypothesis putting pressure on its interpretation. This first evidence might have been ignored or randomly classified as tall or short. Such an initial strategy would explain the discrepancy between the idealized prediction of the assimilation hypothesis and the current results shown in Figure 6 (e.g., the mean estimates on XE in the Ambiguous+ condition was 28.7 rather than the idealized 38). We are further investigating this avenue to discover how much information must be presented before assimilation will begin.

Allowing causal hypotheses to influence the interpretation of ambiguous evidence can speak to phenomena of human reasoning found in a variety of research fields. For instance in the illusory correlation literature, Chapman and Chapman (1969) found that in studying drawings supposedly produced by individuals with psychoses, participants reported higher than actual frequencies of features that matched their theories of the focus of the drawer's psychosis. We speculate that one possible mechanism underlying illusory correlations is that ambiguous or neutral features are assimilated differently under the light of the interpreter's theory of the mental disorder.

Indeed, the assimilation process appears to be ubiquitous. For instance in a now classic study, Asch (1946) found that the presence of one salient adjective could color the interpretation of an entire list of trait adjectives, in that adjectives with an ambiguous connotation were interpreted as positive or negative depending on the valence of a manipulated adjective. From illusory correlations to impression formation, ambiguous evidence may well be the mainstay, and not the exception, to the type of information available in everyday life. It is actually surprising that few previous studies have examined this issue in causal induction. This could be due to an attempt to unnecessarily separate out the top-down and bottom-up processes of causal induction. The idea forwarded by much of the traditional research (e.g., Cheng, 1997) that causal induction can operate separately from prior knowledge has become an even more tenuous claim in light of our findings.

Acknowledgments

The current research was supported by a National Institute of Mental Health Grant (2 R01 MH57737-05A1) given to the second author.

References

Asch, S. E. (1946). Forming impressions of personality. *Journal of Abnormal and Social Psychology, 41,* 258-290

Baker, A. G., Mercier, P., Vallée-Tourangeau, F., Frank, R., & Pan, M. (1993). Selective associations and causality judgments: The presence of a strong causal factor may reduce judgments of a weaker one. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 19*, 414-432.

Chapman, L. J., & Chapman, J. P. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology, 74,* 271-280.

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

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

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review, 100*, 254-278.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. New York, NY: Cambridge University Press.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. Prokasy (Eds.), *Classical conditioning II*. New York: Appleton-Century-Crofts.

Schyns, P. G., Goldstone, R. L., & Thibaut, J. (1998). The development of features in object concepts. *Behavioral and Brain Sciences, 21,* 1-54.

Shanks, D. R., Lopez, F. J., Darby, R. J., & Dickinson, A. (1996). Distinguishing associative and probabilistic contrast theories of human contingency judgment. In D. R. Shanks, D. L. Medin & K. J. Holyoak (Eds.), *Psychology of learning and motivation, Vol. 34: Causal learning* (pp. 265-312). San Diego, CA: Academic Press.

Spellman, B. A. (1996). Conditionalizing causality. In D. R. Shanks, D. L. Medin & K. J. Holyoak (Eds.), *Psychology of learning and motivation, Vol. 34: Causal learning* (pp. 167-207). San Diego, CA: Academic Press.

Wisniewski, E. J., & Medin, D. L. (1994). On the interaction of theory and data in concept learning. *Cognitive Science, 18*, 221-281.