Evaluating the Causal Role of Unobserved Variables

Christian C. Luhmann (christian.luhmann@vanderbilt.edu)
Department of Psychology, Vanderbilt University
301 Wilson Hall, Nashville, TN 37203 USA

Woo-kyoung Ahn (woo-kyoung.ahn@yale.edu)
Department of Psychology, Vanderbilt University
301 Wilson Hall, Nashville, TN 37203 USA

Abstract

Current psychological models of causal induction assume that causal relationships are inferred based on observations about whether the cause and effect are present or absent. The current study investigated how people infer the causal roles of unobserved events. In Experiment 1 we demonstrate that participants are indeed willing to evaluate the causal roles of unobserved events. We then suggest that the basis for these judgments may be situations in which effects occur in the absence of observed causes. Experiment 2 provides evidence that such information does influence participants' judgments about unobserved causes.

Introduction

People oftentimes infer causal relationships between two events based on information on how these two events co-vary. For instance, new parents seek out things that make their baby sleep through the night. A parent might hypothesize that giving their child a bath in the evening would facilitate sleeping. To evaluate this possibility, the parent gathers data on whether or not the baby takes a bath in the evening and whether or not the baby sleeps through that night. Further suppose that the parent discovers that an evening bath does not make the baby sleep through the night and instead hypothesizes that it is the amount of exposure to sunlight that facilitates nighttime sleeping. Yet, in the initial set of observations about evening baths, the parent did not keep track of the amount of sunlight the baby received each day. Will this initial set of data influence inferences about the causal efficacy of the unobserved quantity (e.g., sunlight)?

The current study examines what people do when inferring the causal efficacy of an event that is not observed in the data. Because no observations have been made about this variable, one highly plausible possibility is that people would be reluctant in making any inferences about the causal role of the unobserved variable. The current study, however, provides evidence that people are willing to make inferences about the causal efficacy of an unobserved variable. In this introduction, we will first describe an experimental design involving unobserved variables. We will then describe why this situation may present difficulties.

Our study employs a design similar to many causal learning experiments. Participants are told about a cause, in this case a colored button, and an effect, in this case a light. Participants are presented with a series of observations each of which portrays the causal candidate and the effect as either present or absent. This design results in four possible observations (see Figure 1; hereafter, a tilde indicates the absence of that variable). After viewing the series of observations participants are asked to evaluate the casual relationship between the cause and effect.

Our design is modified to explore how participants deal with unobserved causes. To do this, participants are told that there are two causal candidates for an effect but receive no presence/absence information about one of the causes.

Intuitively, evaluating the role of unobserved causes poses a problem. How is one to determine the causal role of an event that has never been observed? Even if there is only one unobserved cause and participants assume that the two buttons are the only possible causes of the light, the statistical relationship between the unobserved causal candidate and the effect is undefined.

If the two buttons are the only possible causes, trials on which the effect is present and the observed cause is absent

	E	~E
C	A	B
~C	C	D

Figure 1 – Table summarizing the contingency between a cause (C) and an effect (E). A tilde indicates absence.

(~CE) logically imply the presence of the unobserved cause. This certainty does not, however, imply an unambiguous answer about the overall relationship between the unobserved cause and effect. This is because on all other trials (i.e., CE and ~C~E) the state of the unobserved cause is indeterminate.

We will illustrate this more concretely using one common measure of contingency, ΔP, which is defined as the difference between the probability that an effect is present in the presence of a causal candidate and the probability that an effect is present in the absence of the same causal candidate (Jenkins & Ward, 1965; Cheng & Novick, 1992). According to ΔP, unobserved causes and the associated ambiguity leads to a range of possible relationships between the unobserved cause and the effect. Assume that a participant receives 40 observations, 10 from each cells of the contingency matrix (Figure 1) and that the two buttons are the only possible causes. This was the design used in our Experiment 1. In this situation the state of the unobserved cause is implied in 10 trials (~CE). If, on the remaining 30 trials, the unobserved cause is perfectly correlated with the effect, the resulting ΔP between the unobserved cause and the effect is 1.0 (e.g. the unobserved cause tends to lead to the effect). If, on the other hand, the unobserved cause is negatively correlated on the remaining trials, the resulting ΔP is –0.5 (the unobserved cause tends to prevent the effect). Thus, it can be proven that certainty about the unobserved cause on a subset of observations does not lead to certainty about the unobserved cause in general (see the conclusion section for more discussion on other models of causal learning).

Despite the difficulty involving unobserved causes, people cannot, in general, avoid such situations in everyday life. For example, you may hear about a car accident, but not be told about the infinite number of potential causes (e.g. consumption of alcohol, road conditions, mechanical problems with the vehicle, mobile phone use, etc.). Furthermore, even if people were free to make any observations in a given situation, cognitive limitations would not allow them to keep track of information about all possible causes.

One way of getting around such limitations is to make the best use of all available information. We suggest that people evaluate every experience as possible evidence with respect to multiple causal hypotheses. Specifically, we argue that when people make causal inferences about unobserved causes, the basis for these inferences comes from situations in which an effect occurs in the absence an observed cause (i.e., ~CE or cell C in Figure 1).

Returning to our previous example, ~CE corresponds to situations in which a baby did not take a bath in the evening, but slept through the night. Henceforth we refer to these situations as "unexplained effects". Our suggestion is that when unexplained effects occur, people infer that an unobserved, but plausible cause is present. For instance, the parent, who suspects an influence of sunlight exposure, might infer that the baby was exposed to lots of sunlight on days when the baby did not take an evening bath but slept well.

The intuition behind this claim can be seen in another naturalistic example. When investigating possible causes of lung cancer, some have suggested that genetic factors could play a role. This hypothesis was generated, in part, because there were cases of lung cancer that occurred in the absence of any obvious cause (e.g. smoking). With no obvious cause, one possible conclusion is that an unobserved factor, such as genetics, is at least partially responsible for the cancer.

To summarize, unobserved causes create an ambiguity with regards to causal relations. Because of this, explicitly asking participants about events they have not witnessed may seem unnatural and raise objections. However, we argue that complete observability is unlikely to be found in many real-life situations. We further argue that there may be a basis for making causal inferences about unobserved events. Specifically, our proposal is that unexplained effects (i.e., ~CE) serve as basis for hypotheses about alternative, unobserved causes. If people indeed use such information, they may be willing to make causal inferences about an event that they did not observe at all. In Experiment 1, participants were explicitly provided with an option to say that they could not determine the causal potency of an unobserved cause in order to test this idea. Experiment 2 specifically examines whether inferences about unobserved causes are dependent on unexplained effects.

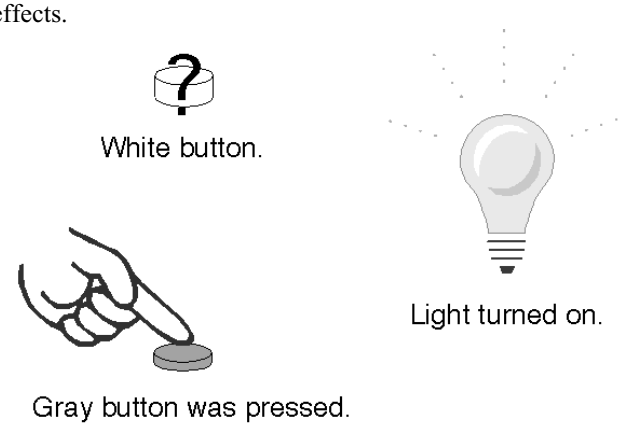

Figure 2 – Example stimuli from Experiment 1

Experiment 1

Method

Twenty participants were told about 3 electrical systems each consisting of a number of colored buttons (i.e. 2, 3, or 4 buttons) and a light (see Figure 2 for a system involving 2 buttons and a light). Participants were told that it was their

job to discover how each system worked. Each participant saw all three systems in a counterbalanced order. Participants were told that they would view a series of tests (i.e. trials) that had been run on the systems and that each test would contain which buttons had been pressed and whether or not the light had turned on.

Participants were then told that the data pertaining to some of the buttons had been lost and that because of this loss, information about only one of the buttons and the light would be presented for each system. The "lost" data were represented by a question mark over each appropriate button (see Figure 2).

	E	~E
C	10	0
~C	10	10

Figure 3 – Table summarizing the contingency between a cause (C) and an effect (E) Experiment 1.

The statistical relationship between the observed button and the light was fixed at $\Delta P = 0.5$ by using identical frequencies for each of the systems. These frequencies are illustrated in Figure 3.

After viewing all the trials for each system, participants were asked to rate the causal role of each button. For example, for the unobserved cause in Figure 2 participants were asked to judge, "the extent to which pressing the white button caused the light to turn on." Participants responded on a scale from –100 (White button prevented the light from turning on) to 100 (White button caused the light to turn on), with zero labeled as, "White button had no influence on the light." To get an estimate of participants' willingness to respond, below the response scale of each question, participants were reminded that, "If you cannot make a judgment, please write 'N/A'."

Results

The question of interest was whether participants would be willing to give any judgments concerning unobserved causes. Figure 4 shows the percentage of judgments that received "N/A" responses, indicating that no judgment could be made. First and most surprisingly, all participants gave a causal strength judgment on both the observed and unobserved causes in the system with only one unobserved cause (e.g. Figure 2). Thus, people are willing to estimate the causal efficacy of a factor for which they have

	Number of Unobserved Causes		
Estimated Cause	1	2	3
Observed	0	0	10
Unobserved	0	45	65

Figure 4 - Percentage of N/A responses per judgment

ambiguous information. Despite the fact that, as demonstrated above, this relationship has no correct answer, participants apparently felt they had enough information to make a reasonable judgment.

This result, however, could be due to participants' overall unwillingness to use the "N/A" response due to unforeseen demand characteristics present in the experiment. This possibility prompts the second observation, that participants were willing to use the "N/A" response when evaluating systems with greater uncertainty (i.e. 2 or more unobserved buttons). Participants were significantly more likely to give an "N/A" response when there were 2 and 3 unobserved causes than when there was only one unobserved cause, $\chi^2(1, N=20)=9$, $p<.001$ and $\chi^2(1, N=20)=13$, $p<.001$, respectively using McNemer's test (McNemar, 1947).

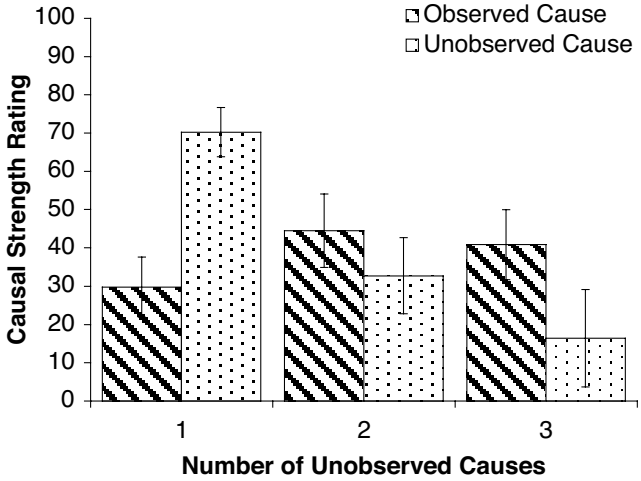

Figure 5 – Causal strength rating from Experiment 1

For responses other than "N/A", the causal estimates are summarized in Figure 5. Here too, the influence of increasing complexity can be seen. Participants indicated that an unobserved cause played less of a role when there were multiple unobserved causes than when there was only one unobserved cause. While participants' responses show a discernable pattern, the large number of "N/A" responses (and the resulting small number of numerical responses) in two of the conditions makes interpretation of this data difficult. Regardless, our main focus is on participants' willingness to respond at all. Our results show that participants are indeed willing to give causal estimates in at least some situations that include unobserved causes

	~CE present C~E absent	~CE present C~E present	~CE absent C~E absent	~CE absent C~E present
Contingency Structure	E ~E C 10 0 ~C **10** 10	E ~E C 10 **10** ~C **10** 10	E ~E C 10 0 ~C 0 10	E ~E C 10 **10** ~C 0 10
Predicted ΔP for Observed Cause	0.5	0	1.0	0.5

Figure 6 - The four conditions used in Experiment 2. Highlighted cells indicate critical differences between conditions.

indicating that missing data per se does not lead to a refusal to respond.

Our predictions, however, go beyond this demonstration. As explained in the introduction, we argue that participants may base inferences to an unobserved, alternative cause on unexplained effects (~CE). Experiment 2 tests this prediction.

Experiment 2

Method

Procedure Twenty-four participants received instructions similar to Experiment 1 except for the following changes. All systems contained exactly two buttons and one light. To rule out any possible ambiguity in the method, participants were told that nothing other than the two buttons could activate the light and that on each trial any combination of buttons could be pressed (i.e. neither, one, or both).

After viewing the tests for each system, participants were asked to rate the causal role of each of the buttons. For example, for the observed cause in Figure 2 participants were asked, "Imagine running 100 new tests in which the gray button was pressed and the white button was not. On how many of these tests do you expect the light to turn on?" Participants responded with a number between 0 and 100.

Unlike Experiment 1, participants were not allowed "N/A" responses because that was not the main concern of Experiment 2. Although no participant in Experiment 1 gave N/A responses for systems involving one unobserved variable, we wished to maximize the number of numerical responses in Experiment 2 because we are primarily interested in comparing strength estimates across the conditions. In an attempt to disentangle participants' causal beliefs from confidence in those beliefs, participants were asked to provide a confidence rating on a 7-point scale ranging from 1 ("Not at all confident") to 7 ("Very confident").

Design and Materials Four conditions were used. The statistical relationships between the observed button (C) and the light (E) are summarized in Figure 6. Unexplained effects are represented by the ~CE cell of the contingency matrix.

As explained above, we predict that the occurrence of unexplained effects (~CE information) will influence participants' causal judgments of the unobserved cause. Situations that have unexplained effects should lead to beliefs about the causal role of unobserved causes. Thus, in the C~E present/~CE absent and C~E present/~CE present conditions, where unexplained effects occur, stronger causal attributions for the unobserved causes should be elicited than in the C~E absent/~CE absent and C~E absent/~CE present conditions where there are no unexplained effects. However, simply varying the frequency of ~CE also modulates the statistical relationship between the observed cause and the effect. To eliminate this confound we manipulated ~CE as well as C~E information.

Figure 6 shows the predictions of ΔP for the observed cause. Thus, the C~E present/~CE absent and C~E absent/~CE present conditions equate ΔP, but because the former has unexplained effects (~CE) and the latter does not, we predict that participants will believe that the unobserved cause has a greater causal role in the former condition than in the latter.

Results and Discussion

Figure 7 shows mean causal strength estimates for the unobserved cause in each condition. A 2 (C~E presented vs. C~E not presented) by 2 (~CE presented vs. ~CE not presented) ANOVA was conducted on causal judgments of unobserved causes. This analysis revealed a significant

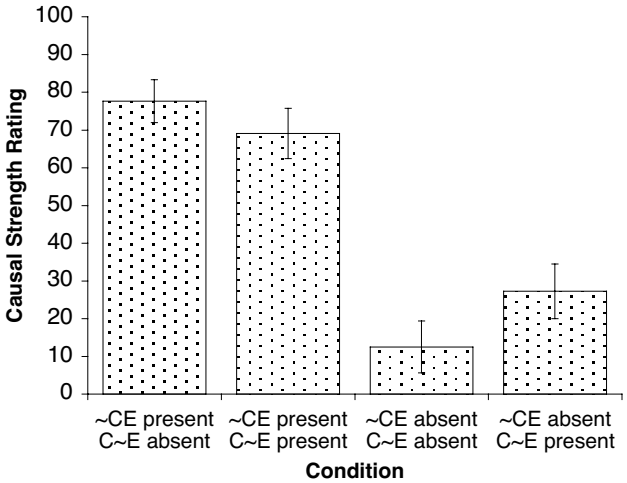

Figure 7 – Causal strength rating for the unobserved cause in Experiment 2

main effect of ~CE information, $F(1, 23) = 43.19$, $p<.0001$. No other main effects or interactions were significant.

The results show that the presentation of unexplained effects (~CE) led participants to believe that the unobserved cause was capable of exerting a causal influence on the light. However, it is possible that participants held these beliefs with little confidence.

Turning to participants' confidence ratings (see Figure 8), the first result of note is that confidence ratings for unobserved causes were significantly greater than the

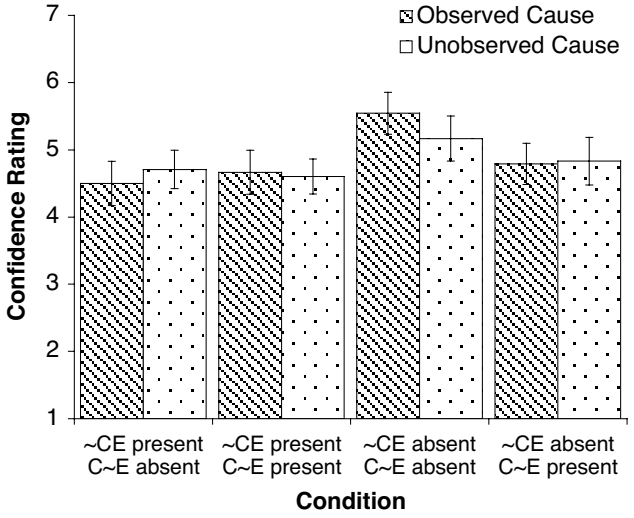

Figure 8 – Confidence ratings for Experiment 2

midpoint of the scale (i.e., 3.5), $t(95)=8.60$, $p<.0001$. Secondly, we compared participants' confidence ratings for observed causes with their confidence ratings for unobserved causes separately within each condition. There were no significant differences between these ratings in any of the conditions (all p's>.3). Thus, participants not only made causal judgments about unobserved causes according to our predictions, but they were just as confident in these judgments as they were in their judgments about observed causes.

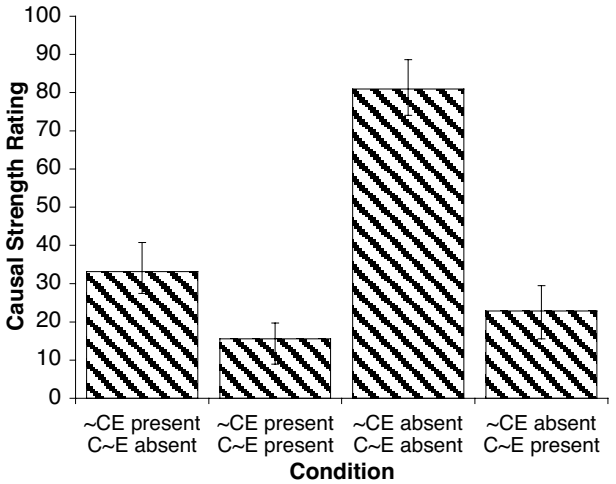

Figure 9 – Causal strength rating for the observed cause in Experiment 2

Conclusion

The two current experiments demonstrate that people are willing to estimate the strength of unobserved causes. Participants apparently believe that, despite the lack of direct observation, they had enough information to make causal judgments about unobserved causes.

In response to this finding, we have suggested that one possible source of information about the causal strength of unobserved causes is unexplained effects (~CE). These situations provide evidence for the influence of an unobserved cause. Our second study demonstrated that participants' estimates of unobserved causes are indeed influenced by the frequency of unexplained effects. In our original example, a parent who observes their child sleeping through the night without a bath might adopt a hypothesis about an alternative cause (e.g. sunlight). Of course our methodology constrained the possible alternative causes, but this constraint does not yield in a correct answer (as shown above) and ongoing studies demonstrate that the trend holds even when such constraints are omitted.

The situation set out to our participants is beyond the boundary conditions of current models of causal induction. As we will show, current theories rely on information about the presence and absence of causes and effects. Situations with unobservable causes result in ambiguity that prevents these models from making any predictions.

There are currently two main classes of models of causal induction. The first of these classes does not calculate causal relations per se, but rather calculates the associative strength of relationships between causes and effects. The Rescorla-Wagner model is the most well known of these models (Rescorla & Wagner, 1972). In calculating the association (V) between events, learners are theorized to update this association in proportion to the difference between the "expected" and actual outcomes. Formally, this updating is calculated according to Equation 1.

$$\Delta Vn = \alpha\beta\left(\lambda - \sum Vn-1\right) \qquad (1)$$

Equation 1 states that the association change experienced on the n^{th} trial is equal to the difference between the outcome (λ) and the associative strength present on the n^{th} trial (ΣV_{n-1}) weighted by two learning parameters (α and β). If on a given trial, an outcome is present (i.e. $\lambda = 1$) and the associative strengths predict otherwise (i.e. $\Sigma V_{n-1} < 1$), the left-hand side of equation 1 will be positive and thus increase the associative strength of the present cues for future trials.

The second class of models suggests that people's causal judgments are based on statistical information about the presence and absence of causes and effects. According to these models, people evaluate cause-effect relationships by

accumulating experience with a cause-effect pair and performing a calculation on the data represented in the covariation matrix (i.e. Figure 1).

One measure of covariation, as mentioned above, is ΔP. Because covariation does not necessarily imply causality, Cheng (1997) proposed the power PC model, which calculates causal power (i.e. the probability that a cause will lead to an effect in the absence of other causes). When the cause in question is independent of other causes, causal power can be calculated according to Equation 1. That is, causal power is ΔP, weighted by the presence of the effect in the absence of the cause.

$$Causal\ Power = \frac{P(E \mid C) - P(E \mid \sim C)}{1 - P(E \mid \sim C)} \quad (2)$$

It should be clear from this brief overview that both associative and statistical models assume that the relevant data are fully observable. Going back to our initial example of a new parent, these models may predict the associative strength or the causal power of an evening bath over sleeping through the night from the data about co-occurrence of these two events. However, if information is not available about the amount of sunlight the baby was exposed to each day, it is unclear whether the Rescorla-Wagner model should include sunlight as a cue in the summation used to predict the outcome[1]. Similarly, because we do not know whether the baby was exposed to sunlight each day, the causal power of sunlight cannot be calculated using the power PC model.

Thus, in their current forms, these models cannot (and were not designed to) deal with the ambiguity created by unobserved data. Our findings suggest the need for something akin to the Bayesian approach (Glymour, 2001) in which a wide variety of evidence can be evaluated with regards to multiple causal hypotheses.

Our findings suggest that observations can be used to evaluate hypotheses even about variables not included in those observations. Future research will be needed to determine exactly how people estimate the causal strength of unobserved causes. One possibility is that people's initial hypothesis about an unobserved cause, as formed by presence or absence of earlier ~CE trials, would guide their assumptions about presence or absence of the unobserved causes in other three types of trials. For example, if they believe that an unobserved variable is likely to be a cause, they should be more likely to believe that it is present when E is present (i.e., CE trials) and it is absent when E is absent (i.e., C~E and ~C~E trials). If people initially believe that an unobserved variable is unlikely to be a cause, different assumptions might be made about presence of an unobserved variable. Such dynamic updating of beliefs would be consistent with a constraint-based coherency account (e.g. Thagard, 1989; Hagmeyer & Waldmann, 2002).

[1] One might think that Recorla-Wagner could use the context cue to make predictions about unobserved causes. The context cue is present on every trial. We have since verified that participants, on the other hand, do not believe the unobserved cause to be constantly present. Conversely, one could think of the context cue as a composite cue that includes context and any unobserved causes. This remains problematic because it is still unclear whether any given unobserved cause is present or absent in CE, C~E, and ~C~E trials, as illustrated in the introduction, and because it is unclear how to partial out the context cue's predicted associative strength to each member of the composite.

Author Note

The current research is funded by NIMH grant (2 R01 MH57737-05A1) given to the second author.

References

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.

Glymour, C. (2001). T*he Mind's Arrows: Bayes Nets and Graphical Causal Models in Psychology*, Cambridge, MA: MIT Press.

Hagmayer, Y., & Waldmann, M. R. (2002). A constraint satisfaction model of causal learning and reasoning. In W. D. Gray & C. D. Schunn (Eds.), *Proceedings of the 24th Annual Conference of the Cognitive Science Society* (405-410). Mahwah, New Jersey: Lawrence Earlbaum Associates, Inc.

Ward, W. C., & Jenkins, H. M. (1965). The display of information and the judgment of contingency. *Canadian Journal of Psychology, 19,* 231-241.

McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika, 12*, 153-157.

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

Thagard, P. (1989). Explanatory coherence. *Behavioral and Brain Sciences, 12,* 435-467.