Causal Models can be Used to Predict Base-Rate Neglect

Steven A. Sloman (Steven_Sloman@brown.edu)
Department of Cognitive & Linguistic Sciences, Box 1978
Brown University, Providence, RI 02912 USA

Abstract

Causal models are used to predict individuals' probability judgments on the taxicab problem of Tversky and Kahneman (1982). Predictions are based on the hypothesis that judgments take into account only those variables that are judged causally relevant. Two versions of the problem were tested, one with and one without causally-relevant base rates. The results showed that causal models were able to predict judgments reasonably well. However, the data failed to replicate Tversky and Kahneman's finding of a difference between the two conditions.

Introduction

"The fathers shall not be put to death for the children, neither shall the children be put to death for the fathers: every man shall be put to death for his own sin" (Deut. 24:16).

Judgments of probability can be contaminated by causal considerations. For instance, the probability of an effect given a cause is sometimes judged higher than the probability of the cause given the effect even in the absence of greater support for the former (Tversky & Kahneman, 1980). This paper pursues the hypothesis that judgments on a well-worn probability problem are mediated by beliefs about causal structure. In passing, we'll see that such behavior has a solid normative justification.

We can distinguish two types of evidence relevant to making a judgment of the probability of a unique event: class data and case data (see Tversky & Kahneman, 1982). Class data refers to evidence about the event emanating from its type. For example, if asked whether Juan spoke Spanish at the club, learning that Juan is Latino increases the probability that he did (perhaps only marginally) because the class of Latinos is more likely to speak Spanish at a club than the class of non-Latinos. In contrast, case data is evidence about the specific event itself, such as evidence that someone overheard Juan speaking Spanish at the club. Kahneman and Tversky (1973) argue that people sometimes neglect class data relative to case data when making judgments of probability. Evidence specific to the case at hand overwhelms the field, evidence about the general type of event can be treated as background and therefore neglected. One example Tversky and Kahneman (1982) used to make this point was the following cab problem:

Noncausal version

A cab was involved in a hit-and-run accident at night. Two cab companies, the Green and the Blue, operate in the city. Imagine you are given the following information.

85% of the cabs in the city are Green and 15% are Blue.

A witness identified the cab as a Blue cab. The court tested his ability to identify cabs under the appropriate visibility conditions. When presented with a sample of cabs (half of which were Blue and half of which were Green) the witness made correct identifications in 80% of the cases and erred in 20% of the cases.

What is the probability that the cab involved in this accident was Blue rather than Green?

The median response to this problem, as reported by Tversky and Kahneman, was .80. A Bayesian analysis, however, distinguishes 3 variables, W (the witness's report that the cab was blue), B (the event that cab involved in the accident was blue), and G (the event that the cab involved in the accident was green). The question asks for

$$P(B|W) = \frac{P(W|B) \bullet P(B)}{P(W|B) \bullet P(B) + P(W|G) \bullet P(G)} = .41$$

On this analysis, the cab is more likely to be green than blue despite the witness's report because of the high base rate of green cabs. Yet, the median response was closer to P(W|B). People are apparently using a procedure that neglects P(B) and P(G).

Since Kahneman and Tversky made their claim, a dispute has raged over the phenomenon of base-rate neglect (a summary of the debate can be found in Koehler, 1996, and ensuing commentaries. See also Sloman & Over, in press). Some have questioned the meaningfulness of the distinction between class and

case data. Some have disputed the normative value of the Bayesian analysis (e.g., Birnbaum & Mellers, 1983). Others have disputed whether people actually neglect class data. Whether or not base-rate neglect is, in general, normatively inappropriate, and whether or not the phenomenon generalizes to other problems, the data show clearly that the response to the cab problem above tends to be closer to P(W|B) than Bayes' rule prescribes.

However, a condition does exist that has elicited responses closer to the Bayesian prescription. Following work by Ajzen (1977), Tversky and Kahneman (1982) tested a version of the problem in which they replaced the statement of base rates (the second paragraph of the problem) with a statement that made the base rates appear to be causally relevant to the incident:

Causal version

> Although the two companies are roughly equal in size, 85% of the total accidents in the city involve Green cabs, and 15% involve Blue cabs.

Otherwise the problem was identical. The median response to this version was .60, closer but not identical to the Bayesian value. Participants were more likely to take into account base rates that were causally relevant to the event being judged.

The Causal Modelling Hypothesis

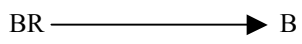
One interpretation of these data is that, when asked to judge the probability of a single event in a causal context like a car accident, people have a propensity to evaluate the likelihood of causal effects, rather than probabilities per se. That is, when trying to understand a situation involving causes and effects, people try to construct a causal model and reason from it. The implication is that evidence – even probabilistically relevant evidence – that is not part of a causal structure can be neglected. In particular, evidence that is taxonomically but not causally related to the object of judgment, like the number of cabs in a city, will be neglected.

Although such reasoning is not probabilistically sound, it is often legally and morally justifiable. Many moral and legal codes prohibit blame or the determination of guilt based on background or prior conduct, evidence that may increase the likelihood of guilt but does not support a specific causal chain leading from the accused individual's intentions and actions to the sin, crime, misdeed, or accident at hand. People should not be punished for a specific act due to the crimes of their fathers, their previous bad behavior, their nationality, or their race even if one or more of those facts indeed increases the probability of their guilt.

Evidence that this is consistent with modern American morés come from Wells (1992), who showed that both undergraduates and experienced courtroom judges deem certain probabilistically relevant base rates irrelevant when deciding guilt. Using a bus accident scenario related to the cab problem though without any form of case (witness) data, he found that participants considered a base rate – the proportion of buses passing the location of an accident that were owned by a certain company – irrelevant to a verdict against the company, even though participants were keenly aware of the relevance of the base rate data and used them to determine the judged probability of guilt. Indeed, Wells points out that legal cases resting on naked statistical evidence of this type are habitually thrown out of court. Even though the knowledge that one bus company is much more prevalent than another increases the probability that an accident was caused by the larger company, the consensus is that such knowledge is not a basis for assigning guilt in a particular case.

Beyond the normative justification for neglecting non-causal base rates, it is not surprising that people would depend on causal models to generate judgments and predictions in the context of a physical event like a car (or bus) accident. Much of the information we have about physical events is encapsulated in our causal models. Causal models capture the invariant aspects of events and they can be used to model, control, and predict both actual and counterfactual events. A formal framework for modeling probabilistic and deterministic causal systems using graphical models has recently been developed (Pearl, 2000; Spirtes, Glymour, & Scheines, 1993). A cornerstone of the approach is a means of representing an agent's actions or interventions on a causal system, a method that has passed an initial test of psychological plausibility (Sloman & Lagnado, 2002a; 2002b).

The hypothesis at hand is that base-rate neglect in the cab problem results from participants who do not perceive the base rate to be causally relevant to the judged outcome. To state this in causal modeling terms requires treating the judged event, a blue cab was involved in an accident (B), and the base-rate of the judged event (BR), as two separate variables. In the case of the noncausal scenario, this amounts to distinguishing the number of blue cabs in the city (BR) from the blue cab involved in the particular accident (B). In the causal scenario, the distinction is between the number of blue cabs involved in accidents in the city (BR) and B. The hypothesis is that participants whose causal models include a causal link from BR to B:



will be more likely to take account of base rates when asked for the probability of B than those participants whose causal models do not include a causal link from BR to B.

To test the hypothesis, we gave participants the cab problem as well as a series of questions intended to reveal their causal models of the problem. Their causal models were then used to postdict how they responded on the cab problem. We tested some participants on the causal and some on the noncausal version of the problem in order to increase the likelihood of getting participants both who had and who did not have causal relations from BR to B in their models.

Method

Participants

A total of 337 participants were obtained on the internet through advertising on various psychology websites; 161 were tested with the Causal cab problem and 176 with the Noncausal version. Each participant's name was entered in a raffle and two winners each received $80.00.

Design and procedure

Participants responded to a series of questions at their own pace. One question was their response to the cab problem above. They entered a numerical response between 0 and 100. A set of 6 questions was designed to determine their causal model of the problem scenario, one question for each ordered pair of the 3 variables, BR (the base rate of blue versus green cabs in the city), B (the event that the accident was caused by a blue rather than green cab), and W (whether the witness's identification was correct). The causal model questions asked the participant whether a functional relation obtained from one variable to another. For example, to assess the causal relation from BR to B, they were asked

> Would a change in the percentage of Blue versus Green cabs in the city that are involved in accidents cause a change in your belief that the cab involved in this accident was Blue or Green?

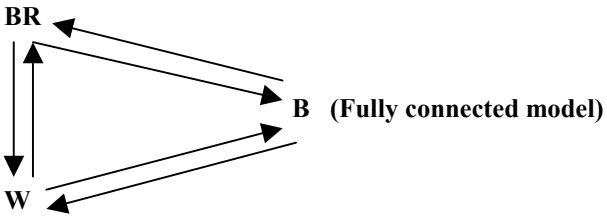
To assess the causal relation from B to W, they were asked

> Would a change in your belief that the cab involved in this accident was Blue or Green cause a change in the percentage of cabs of each color that were correctly identified by the witness in the court-ordered test?

Approximately half the participants answered the critical taxicab question first, followed by the six causal model questions. The other participants first answered the causal model questions and then the taxicab question. The six questions were always asked in the same random order and participants chose a response of "yes" or "no" on a pulldown menu to respond.

Causal model construction

Participants' answers to each question were used to derive a causal model for each participant. Each causal model consists of a (not necessarily proper) subset of the complete set of 6 causal relations. To illustrate, a participant who responded "yes" to each question would be assigned a completely connected causal model:



(Fully connected model)

A participant who responded "no" to each question would be assigned a completely disconnected causal model:

BR

B

W

(No connection model)

These particular models were relatively rare. There were $2^6 = 64$ possible models and most participants were assigned causal models of intermediate degree of connectivity.

Results

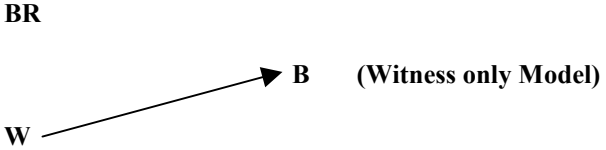
Responses to the cab problem

A lot of base-rate neglect was observed. Overall, 35% of participants gave the witness's credibility (80%) as their response. This was by far the modal response in both conditions (32% of participants in the Causal condition, 38% in the Noncausal; $z = 1.11$; n.s.). Six participants produced the precise Bayesian response of 41 (5 in the Causal and 1 in the Noncausal conditions). 5.6% of responses were within 2 standard errors of the Bayesian response in the Causal condition and 4.5% in the Noncausal condition. The overall differences between the Causal and Noncausal conditions were

small and failed to reach significance, thus not replicating Tversky and Kahneman (1982). The median responses were both 75. (Means were 60.2 and 57.5, respectively; $t < 1$.) Neither the medians nor the means differed substantially as a function of the order of question presentation.

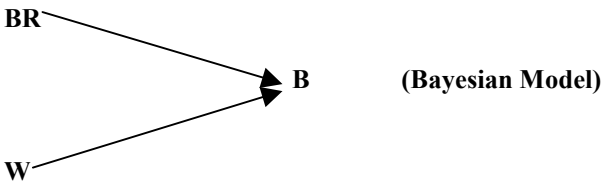
Causal Models Produced

In the Causal condition, 39 different models were generated. In the Noncausal condition, 37 of the 64 possible models were. In other words, the variability was enormous. But in both conditions, the model most frequently generated was



This is the simplest model consistent with Tversky and Kahneman's claim that people believe W (the witness's credibility) was relevant to their judgment but BR (the base rate) was not. It was generated by 22.4% of Causal and 26.1% of Noncausal participants ($z < 1$).

The next most frequent model in both conditions was the simplest one consistent with the Bayesian response; i.e., the simplest model stating the relevance to the judgment of both the base rate and the witness's testimony:



As predicted by greater sensitivity to the base rate in the causal condition of previous experiments, this model was more likely in the Causal (18%) than the Noncausal condition (12%; $z = 1.57$; $p = .06$). Thirty-eight participants constructed a model with no links (22 Noncausal and 16 Causal). No other model was chosen by more than 7 participants in either condition.

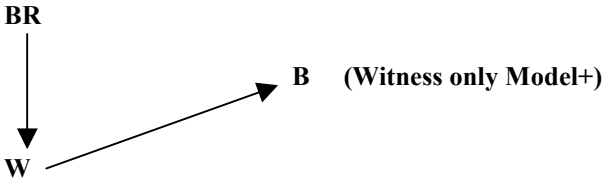
Aggregated Models

Using the causal modeling framework along with the distinction between BR and B, all 64 possible models can be divided into 4 groups depending on the prediction they make for judgments about B, the probability that the cab involved in the accident is blue.

Witness only Prediction. Models like the Witness only Model that include a link from W to B but not from BR to B predict that participants should consider only W in their assessment of B. That is, they predict that the judged probability of B should equal

$P(B|W) = .80$

as we saw above. Any other model that includes a W to B link but not a BR to B link will make the same prediction due to the Markov or "screening-off" property of graphical probability models (see, e.g., Pearl, 1990). For example, consider the model



Because BR has its effect on B only through W, BR and B are conditionally independent given W:

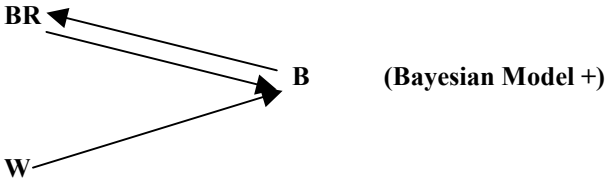
$P(B|W,BR) = P(B|W)$.

The value of W is given in the problem and therefore BR cannot exert an effect on B. W screens off BR and B. So the prediction remains the same, .80.

Bayesian Prediction. Models like the Bayesian Model that include links from both BR and W to B predict that participants should consider both BR and W in their assessment of B. That is, they predict that the judged probability of B should equal

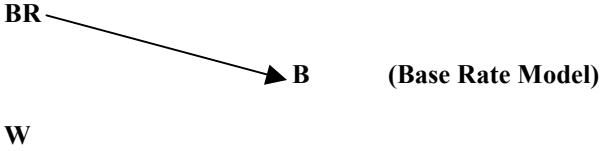
$P(B|BR,W) = .41$

as we saw above. Any other model that includes these two links will make the same prediction due to the screening-off property. For example, consider the model



In this case, we needn't worry about the link from B to BR because BR screens off B from itself.

Base Rate only Prediction. Models with links from BR to B and not from W to B, the simplest case being



predict that participants should consider only BR in their assessment of B. That is, they predict that the judged probability of B should equal

P(B|BR) = .15.

Again, due to the screening-off property, there exists a class of models that make this prediction. However, participants rarely produced models of this type and so not enough data exists to fairly evaluate the prediction.

No Prediction. No prediction can be made from the data given to participants in the problem for those models that do not include any link into B such as the No Connection Model.

Model Fits

Table 1 shows, for each model type in each condition (Causal and Noncausal crossed with whether the taxicab question or the causal model questions came first), the predicted likelihood judgment and the median judgment for all participants whose models corresponded to the type shown. Some of the cells have relatively few observations. The numbers of observations in the Causal conditions were 29, 34, and 5 when the taxicab question came first and 31, 28, and 5 when the base-rate question came second for the Bayesian, Witness, and Base Rate model types, respectively. In the Noncausal conditions, there were 32, 41, and 12 models when the critical question came first and 21, 36, and 5 when it came second for the 3 model types, respectively.

Table 1: Predicted and median observed likelihood judgments for the Causal and Noncausal conditions as a function of question order and derived model type. Observations based on 10 or fewer observations are in parentheses.

		Type of Base Rate	
Model Type	Prediction	**Causal**	**Noncausal**
		Taxicab Question First	
Bayesian	41	50	46
Witness	80	80	80
Base Rate	15	(15)	40
		Causal Model Questions First	
Bayesian	41	68	60
Witness	80	80	77.5
Base Rate	15	(85)	(70)

Table 1 shows a consistent correlation between predicted and observed judgments. In all 4 cases, the median judgments for Witness model participants was equal or close to the predicted value of 80. In all 4 cases, participants with a Bayesian model gave a substantially lower median judgment, as predicted. The Base Rate predictions were highly variable because very few participants had such models and therefore the sample sizes were small. The Base Rate prediction of 15 was actually the modal response in the Noncausal condition for participants with such a model.

Conclusions

This study shows that individual differences in base-rate neglect for the cab problem can be predicted by examining participants' causal models of the problem. More generally, causal models can be used to predict whether people are likely to make their judgment based on case data or both case and class data. Participants who reported that only the witness's testimony was causally relevant to the object of judgment – the likelihood of an accident – were more likely to neglect the base rate and those who reported that both the base rate and the witness's testimony were relevant were more likely to give a judgment closer to the Bayesian response that takes both into account. We did not replicate Tversky and Kahneman's (1982) finding of less base-rate neglect in likelihood judgment with causal base rates. However, participants were marginally more likely to report that the base rate was causally relevant when it was presented in causal terms.

This study suggests that causal modeling might play a central role in the process of judgment when the object of judgment can be construed as a causal effect. Such a construal is almost always appropriate in the legal domain as well as in scientific domains (modulo atomic physics on some accounts), indeed in any domain in which physical, social, or abstract events cause other events. Causal models may well be the primary determinant of what is considered relevant when reasoning, when making judgments and predictions, and when taking action within such domains.

A weakness of the current study is that no validity check ensures that participants who answered "yes" to the causal questions were actually endorsing causal relations rather than ignoring the specific question asked and instead endorsing some more general probabilistic relation. This can be assessed in the future by taking advantage of the fact that not all base rates can be construed in causal terms. For problems involving disease diagnosis (e.g., Casscells, Schoenberger, & Grayboys, 1978) for example, the causal role of base rates is, at best, indirect and probably not the critical element in learning to incorporate them into a judgment. In that case, learning the correct extensional set structure of the environment

is more important; for example, that cases of those who test positive for a disease and have the disease are a subset of those who test positive (Sloman & Over, in press). In such a case, the causal modeling hypothesis predicts that causal models will fail to predict probability judgment.

Although the restriction to causally-relevant evidence has a convincing normative justification in the determination of guilt and blame (one should not be accountable for thy father's sins), morally (and legally) ambiguous cases certainly exist. The issue of profiling is a case in point. Should police be allowed to detain those who fit a racial or ethnic profile for a crime even without direct evidence causally linking the individual to the crime? Such detention can be discriminatory and violate individual rights. Of course, if the profile is statistically valid, it can also help to keep criminals (and terrorists) off the streets. The problem of airport screening brings the issue close to home for frequent flyers. Soon after the terrorist incidents of September 11, 2001, the United States Secretary of Transportation ruled out profiling at airport security gates despite a near universal concern over a particular terrorist profile. Such a rule may have reduced discrimination, but it certainly did not optimize the effectiveness of airport security.

Causal relevance comes in degrees. The proportion of cabs involved in accidents in a city is somewhat causally relevant to a particular accident but not as causally relevant as, say, the degree of inebriation of a driver or the state of a cab's brakes. Decisions seem to require a higher threshold of causal relevance than judgments of probability. In Wells's (1992) study, even base rates concerning the proportions of accidents caused by vehicles of a certain type were deemed irrelevant by his American student participants in determining blame for an accident. Apparently, causal relevance is not a sufficient condition for the ascription of guilt. When it is the only evidence available, Americans (at least) seem to require that data speak to the specific set of events leading to the specific effect under scrutiny in order to convict someone. It may be that, even though causal relevance sometimes increases the likelihood that people will consider data relevant to a probability judgment (Ajzen, 1977), many people may have a higher threshold for using that evidence to convict.

Acknowledgments

This work was funded by NASA grant NCC2-1217 and an American Philosophical Society sabbatical fellowship. I thank Henry Parkin for collecting and analyzing data and Jean Baratgin, Dave Lagnado, and Ed Wisniewski for valuable discussion.

References

Ajzen, I. (1977). Intuitive theories of events and the effects of base-rate information on prediction. *Journal of Personality and Social Psychology, 35*, 303-314.

Birnbaum, M. H. & Mellers B. A. (1983). Bayesian inference: Combining base rates with opinions of sources who vary in credibility. *Journal of Personality & Social Psychology, 45*, 792-804.

Casscells, W., Schoenberger, A., & Grayboys, T. (1978). Interpretation by physicians of clinical laboratory results. *New England Journal of Medicine, 299*, 999-1000.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*, 237-251.

Koehler, J. J. (1996). The base rate fallacy reconsidered: Descriptive, normative and methodological challenges. *The Behavioral and Brain Sciences, 19*, 1-53.

Pearl, J. (2000). *Causality*. Cambridge: Cambridge University Press.

Sloman, S.A., & Lagnado, D. (2002a). Counterfactual undoing in deterministic causal reasoning. *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society, Maryland*.

Sloman, S.A., & Lagnado, D. (2002b). Do we "do"? Manuscript submitted for publication.

Sloman, S. A. & Over, D. (in press). Probability judgment: from the inside and out. To appear in D. Over (Ed.) *Evolution and the psychology of thinking: The debate* (Psychology Press).

Spirtes, P., Glymour, C. & Scheines, R. (1993). *Causation, prediction, and search*. New York: Springer-Verlag.

Tversky, A., & Kahneman, D. (1980). Causal schemas in judgments under uncertainty. In Fishbein, M. (Ed.), *Progress in Social Psychology*. Hillsdale, N.J.: Lawrence Erlbaum.

Tversky, A., & Kahneman, D. (1982). Evidential impact of base rates. In Kahneman, D., Slovic, P., & Tversky, A. (Eds.), *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.

Wells, G.L. (1992). Naked statistical evidence of liability: Is subjective probability enough? *Journal of Personality and Social Psychology, 62*, 739-752.