Individual differences in causal learning and decision making

Osman, M; Shanks, David R

ACTPSY 1038
27 May 2005 Disk Used
ARTICLE IN PRESS
No. of Pages 20, DTD = 5.0.1

Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Acta Psychologica xxx (2005) xxx–xxx
acta psychologica
www.elsevier.com/locate/actpsy

Individual differences in causal learning and decision making

Magda Osman *, David R. Shanks

University College London, Department of Psychology, Gower Street, London WC1E 6BT, England, UK

Received 5 November 2004; received in revised form 18 April 2005; accepted 19 April 2005

Abstract

In judgment and decision making tasks, people tend to neglect the overall frequency of base-rates when they estimate the probability of an event; this is known as the base-rate fallacy. In causal learning, despite people's accuracy at judging causal strength according to one or other normative model (i.e., Power PC, ΔP), they tend to misperceive base-rate information (e.g., the cause density effect). The present study investigates the relationship between causal learning and decision making by asking whether people weight base-rate information in the same way when estimating causal strength and when making judgments or inferences about the likelihood of an event. The results suggest that people differ according to the weight they place on base-rate information, but the way individuals do this is consistent across causal and decision making tasks. We interpret the results as reflecting a tendency to differentially weight base-rate information which generalizes to a variety of tasks. Additionally, this study provides evidence that causal learning and decision making share some component processes.

PsycINFO classification: 2343; 2340

Keywords: Learning and memory; Cognitive processes

* Corresponding author. Tel.: +44 20 7679 7572.
E-mail address: m.osman@ucl.ac.uk (M. Osman).

0001-6918/$ - see front matter © 2005 Published by Elsevier B.V.
doi:10.1016/j.actpsy.2005.04.003

1. Introduction

There are two research domains in which people are explicitly required, on the basis of some evidence, to evaluate the association between two events [X (cause) and Y (effect)], and predict from this the likelihood of event Y given event X: causal induction and Bayesian decision making. In one, the task environment typically involves gathering evidence on a trial by trial basis (causal induction task), that is, people actually experience the relationship between the events across time. In the other, people are merely presented summarized data in the form of a one-shot problem (Bayesian decision making task). For both types of task, an accurate response involves integrating two forms of probabilistic information: the background data (base-rate) and the indicant or diagnostic information (likelihood ratio). Typically, what has been found is that people are insufficiently sensitive to base-rate information and fail to adequately incorporate it in their decision making and reasoning. The aim of this article is to examine what, if any, are the relations between causal induction and decision making with particular emphasis on people's use of base-rate information in causal and decision making tasks.

1.1. Causal induction

When people are asked to judge the relationship between two binary variables, they should normatively consider four different sources of evidence, that is, the frequency with which the two variables co-occurred (Cell A), the frequency with which each variable occurred in the absence of the other (Cells B and C), and the frequency with which both were absent (Cell D). The contingency table (see Table 1) summarizes the frequencies with which the various events occur.

For example, in order to determine the extent to which one type of radiation causes butterflies to mutate, a simple way of calculating the degree of contingency between the putative cause (e.g., radiation 'X') and its effect (e.g., mutation) is to use the ΔP rule (Allan, 1980) where

$$\Delta P = p(e|c) - p(e|\neg c) \quad (1)$$

By examining the A and B cells of the contingency table, it is possible to determine the proportion A/(A + B), which is simply the probability of the effect '*e*' in the presence of the cause '*c*' expressed as $p(e|c)$ in the ΔP rule. In contrast, $p(e|\neg c)$ refers to the proportion C/(C + D), which is the probability of the effect in the absence of the cause. Intuitively, we can see that the extent to which $p(e|c)$ exceeds $p(e|\neg c)$ gives

Table 1
Representation of information in a contingency table

Candidate cause	Effect	
	Present	Absent
Present	A	B
Absent	C	D

some indication of the causal strength of the relationship between radiation and butterfly mutation.

Alternatively, causal strength can be calculated by using the Power PC rule (e.g., Buehner & Cheng, 1997; Cheng, 1997):

$$P = \frac{p(e|c) - p(e|\neg c)}{1 - p(e|\neg c)} \tag{2}$$

This rule is an alternative normative description of causal strength that seeks to differentiate causation from covariation. To estimate the causal strength of a candidate cause to produce an effect, the model takes into account alternative candidate causes of the same effect. This is done by integrating ΔP and the base-rate of the effect $p(e|\neg c)$. The main prediction that follows from Eq. (2) is that if two candidate cause–effect pairings result in equal ΔP but different values of $p(e|\neg c)$, then the causal judgments will be different, and these will vary in accordance with $p(e|\neg c)$: as the latter increases (but is not equal to 1) so does the judged generative power of the cause. Differences between the contingency and power rule become evident once the probability of the effect in the absence of the cause is greater than 0. In the present study, we do not take any position on the relative merits of these two rules or of the claims each of them can make to being normative. This issue has been widely discussed elsewhere (see Shanks, 2004, for a review).

Studies of causal induction suggest that people, although on the whole, good at judging causal strength according to one or the other rule, tend to exhibit biased behavior when making inferences from contingency tables and in trial-by-trial learning tasks (e.g., Allan & Jenkins, 1980; Smedslund, 1963; Vallée-Tourangeau, Hollingsworth, & Murphy, 1998; Ward & Jenkins, 1965). For example, studies show that people weight cell information, non-normatively, in the order A > B > C > D (Kao & Wasserman, 1992; Mandel & Lehman, 1998). Hence, people are most sensitive to variations in cell A and tend to overestimate the value of this cell, whereas they are least sensitive to variations in cell D, often underestimating its value (Arkes & Harkness, 1983; Vallée-Tourangeau et al., 1998; Wasserman, Dorner, & Koa, 1990). Normatively, the cells should be weighted equally. In addition, when presented with conditions in which the effect is equally likely in the presence or absence of the cause ($\Delta P = 0$) but the overall base-rate of the effect increases, people misperceive a contingency that is not there, known as the cause density effect (Buehner, Cheng, & Clifford, 2003; Perales & Shanks, 2003; Smedslund, 1963; Vallée-Tourangeau et al., 1998). This is not to say that people are unable to discriminate between positive, negative, and zero correlations—they can in fact do this well (e.g., Shanks, 1995; Vallée-Tourangeau et al., 1998). However, in zero correlation conditions people fail to take sufficient account of the base-rate of the effect and so tend to overestimate causal strength.

1.2. Bayesian decision making

We turn now to another type of situation in which people have to predict an outcome or the probability of an outcome in light of evidence, and in which they tend to show biased behavior when reasoning about base-rate information. Typically, in Bayesian decision making tasks, people are asked to judge the likelihood of an event having occurred or that will occur. If they respond normatively, they will integrate base-rates and the likelihood ratio according to Bayes' rule. In Bayes' rule the probability of the hypothesis tested (h) is derived by multiplying the likelihood ratio of the observed datum (d) by the prior probability favoring the focal hypothesis:

$$\frac{p(h|d)}{p(\neg h|d)} = \frac{p(d|h)}{p(d|\neg h)} \times \frac{p(h)}{p(\neg h)}$$

What is summarized in the rule is that the diagnosticity of the likelihood ratio should be evaluated independently of the prior odds favoring the focal hypothesis. To do this, the rule includes three ratio terms. The far right term refers to the prior odds favoring the focal hypothesis. The middle term refers to the likelihood ratio composed of the probability of the data given the focal hypothesis divided by the probability of the data given its mutually exclusive component. The far left term represents the posterior odds favoring the focal hypothesis after receipt of the new data.

Numerous studies show that people tend not to give responses that obey Bayes rule; instead, they predominantly make two types of error. First, people routinely neglect the denominator of the likelihood ratio $p(d|\neg h)$, that is, they show a preference for information in which the probability of the datum given the focal hypothesis is true rather than false (Beyth-Marom & Fischhoff, 1983; Doherty, Chadwick, Garavan, Barr, & Mynatt, 1996; Einhorn & Hogarth, 1978; Wasserman et al., 1990). To illustrate, Doherty and Mynatt (1990) presented participants with a problem in which they were asked to determine whether a patient had the disease 'Digirosa'. Participants were asked to select cards which contained information that would be relevant in making their diagnosis: '% of people with Digirosa' $p(h)$, '% of people without Digirosa' $p(\neg h)$, '% of people with Digirosa who have a red rash' $p(d|h)$, and '% of people without Digirosa who have a red rash' $p(d|\neg h)$. To solve the task correctly, the cards $p(d|h)$, $p(d|\neg h)$, and $p(h)$ corresponding to the terms in the formula are required; $p(\neg h)$ is the complement of $p(h)$ and so it is not necessary to calculate the posterior probability.

Doherty and Mynatt (1990) found that, consistent with much of the judgment literature, few participants (11%) demonstrated an understanding of Bayesian reasoning by selecting the correct information. The least popular card choices were the prior probability $p(h)$ and $p(d|\neg h)$. To evaluate a target hypothesis, alternative hypotheses must be considered, and Doherty and Mynatt proposed that participants adopting a good hypothesis testing strategy would select the card $p(d|\neg h)$ because it indicates an awareness of alternative hypotheses. A later study by Stanovich and West (1998) reported that participants choosing $p(d|\neg h)$ in Doherty and Mynatt's (1990) disease problem scored higher on tests of cognitive ability and a battery of reasoning tasks (e.g., syllogisms, conditional reasoning tasks, probability based problems) compared with those that had excluded this card from their choices.

The second type of error people make is to neglect or underweight base-rate information (Bar-Hillel, 1980; Doherty & Mynatt, 1990; Fischhoff & Bar-Hillel, 1984;

Tversky & Kahneman, 1982). For example, in Kahneman and Tversky's (1973) classic task participants are presented a short cover story: 85% of cabs in a particular city are green and the remainder are blue. A witness identifies a cab involved in an accident as blue. Under tests, the witness correctly identifies both blue and green cabs on 80% of the occasions. Participants are then asked: What is the probability that the cab was in fact blue? The posterior probability is in fact 0.41, however, few respond with this answer, tending instead to give estimates that range between 0.70 and 0.90. This highly robust finding has been taken as evidence of peoples' reliance on erroneous intuitions such as the degree of correspondence between a sample and a population (the "representativeness" heuristic). Thus, people are sensitive to the diagnosticity of the descriptions in the cover story, but disregard the fact that the different sub-classes are of different sizes (e.g., 85% green cabs vs. 15% blue cabs).

Bar-Hillel's (1980) alternative interpretation of Kahneman and Tversky's results suggests that the fallacy is the result of misperceiving the relevance of such information. There is evidence to suggest that base-rate information can be made more relevant when framed in such a way that it has a direct causal relation to the target information (Ajzen, 1977; Tversky & Kahneman, 1980). In tasks like the cab problem base-rate information is presented as incidental to the main focus of the problem, whereas contexts that increase the causal efficacy of base-rate information and therefore its status in the problem helps to attenuate base-rate neglect. Bar-Hillel (1980) claimed that such contexts clarify the relation between the base-rate and a target case enabling both types of information to become integrated. Formally the versions that Bar-Hillel used in her study were the same as Kahneman and Tversky's, but used causal contexts. Students were presented with a cover story which discussed suicide rates: A study was done on causes of suicide among young adults (aged 25–35). It was found that the percentage of suicides is three times larger among single than married people. In this age group, 80% are married and 20% are single. In one version of this task students were simply asked to estimate the likelihood of suicide in a given sub-population in which the posterior probability was 0.43. Bar-Hillel found that through various modifications to the framing of this task base-rate neglect could be reduced from 85% of responses to 25%. Changes to the framing included varying the base rate information and likelihood ratio. However, Bar-Hillel's study demonstrates that it is not causality per se that reduces base-rate neglect, but rather the relevancy it adds to this type of information, and so other contexts that do this are also able to attenuate base-rate neglect.

Evidence of deviations from Bayesian reasoning, such as base-rate neglect, have been the cause of much debate, raising questions about the appropriateness of tasks studying people's probabilistic reasoning (Kohler, 1996) and whether people are able to reason rationally (Kahneman & Tversky, 1996; Shafir, 1993). Similarly, in causal induction it is unclear why people should differentially weight the cells of a contingency table. Some have argued that this in fact implies an underlying bias for positive or confirmatory evidence (Klayman & Ha, 1987; Mandel & Lehman, 1998). However, others suggest that the biases that have been found are inflated by the particular choice of framing in which a task is couched or the phrasing of causal questions, rather than being an unavoidable property of people's causal judgments (e.g., Beyth-Marom, 1982; Crocker, 1981; Perales & Shanks, 2003; Vallée-Tourangeau et al., 1998; Waldmann, 2001; White, 2003). These mixed findings can also be seen as representing a broader controversy between prescriptive (or normative) and descriptive explanations of non-normative behavior. That is, are deviations from normative models (e.g., Bayes rule, ΔP rule, Power PC model) examples of biased information processing behavior, or the product of a cognitive system with limited computational capacity?

Stanovich and West's (2000) work on individual differences attempts to answer this question. They showed that people's performance deviates systematically from that which is prescribed by normative models (i.e., logic, probability calculus, expected utility theory). They proposed that the underlying basis for these deviations has strong implications for the way in which the relationship between descriptive and normative models is understood. One is that there are instances in which people's behavior is far from optimal, and that poor performance on reasoning tasks provides evidence of irrational tendencies inherent in human behavior (e.g., Nisbett & Ross, 1993; Tversky & Kahneman, 1974). Alternatively, individuals may simply fail to perform well because of cognitive constraints such as resource limitations of the human cognitive apparatus (e.g., Baron, 1985; Oaksford & Chater, 1993). Finally, individuals' performance might be consistent with a different normative model to that prescribed by the experimenter (e.g., Kohler, 1996), or the normative model used to assess responses to a particular task might be inappropriately applied (e.g., Hilton, 1995; Schwarz, 1996).

Like Stanovich and West, we also emphasize the relevance of individual differences in relation to causal induction and Bayesian decision making by exploring the possible connection between people's use of base-rate information in both domains. The evidence of non-normative behavior in both research domains suggests that people encounter problems in tasks where they should incorporate base-rate information and that, particularly in decision making tasks, individuals vary according to whether or not they integrate such information.

Thus far, there has, to our knowledge, been no empirical work that compares causal contingency judgments with responses to decision making tasks. However, one connection between causal learning and decision making that has been explored is in the context of discounting (Kelley, 1973; Morris & Larrick, 1995; Oppenheimer, 2004; Reeder, Vonk, Ronk, Ham, & Lawrence, 2004) which refers to the phenomenon in which people show biased behavior when making a causal attribution in light of new information. Despite the fact that this work is based on the discounting principle and its common application in causal and decision making domains, there is no empirical comparison of how people use this principle in each of the domains.

In the present study, we investigated whether there are individual differences in the use of base-rate information in causal learning and how these relate to the use of base-rate information in Bayesian decision making. We used a causal learning task which is a modified version of a task described by Shanks (2004) and standard Bayesian decision making tasks: two probabilistic estimation problems (Kahneman and Tversky's Cab problem, causal and non-causal versions), and two base-rate inference tasks (Doherty and Mynatt's Disease problem, causal and non-causal versions).

The first objective of this study was to identify patterns in the causal judgments people gave in the four conditions of the causal learning task. This was based on the extent to which judgments were influenced by base-rate information [i.e., the probability of the effect in the absence of the cause, $p(e|\neg c)$]; the precise details of the procedure used are presented below in the results section headed 'weightings of causal judgments'. From this, the second objective was to examine whether participants who incorporated base-rate information into their probabilistic estimates, or who made inferences that involved base-rate information, also gave causal judgments that reflect a greater influence of $p(e|\neg c)$. Conversely, participants who gave probabilistic estimates that suggested base-rate neglect and who drew inferences in which the base-rate information was ignored were, in turn, expected to give causal judgments that indicated that they had not been influenced by this information when making estimates of causal covariation. Finally, the inclusion of causal and non-causal versions of typical decision making tasks enables a further hypothesis to be tested. Bar-Hillel (1980) claimed that causal versions of decision making tasks such as those devised by Kahneman and Tversky can facilitate performance, as compared with standard non-causal versions, and we aimed to test this conjecture.

2. Method

2.1. Participants and apparatus

Fifty-two students from University College London volunteered to take part in the experiment and were paid £5 for their involvement. Of the students that took part, fifteen were first year undergraduates studying psychology, and each was screened for prior experience with the tests included in the study. Participants were tested individually and were presented with the causal learning task first, which was run on Dell Optiplex computers. The experimental programme used was adapted from studies described in Shanks (2004) and was written in Visual Basic 6.0. Although, we did not counterbalance the order of presentation of the causal and decision making tasks, the requirements and context of the learning task were sufficiently different from the paper and pencil tasks for this not to be a serious concern. However, the order of presentation of the four remaining paper and pencil decision making tasks was randomized for each participant because the structure of the tasks was similar.

2.2. Design and procedure

The causal learning task included four conditions (1–4) each of which was 80 trials long (see Table 2). In the second and third column of Table 2, are two numbers,

Table 2
Cell frequencies, contingency (ΔP), power (P) and values of $p(e|c)$ and $p(e|\neg c)$ in each condition

Condition	Model		Cell frequencies				Model term	
	ΔP	P	A	B	C	D	$p(e\|c)$	$p(e\|\neg c)$
1. Low ΔP, high P	0.35	0.78	36	4	22	18	0.9	0.55
2. Low ΔP, low P	0.35	0.35	14	26	0	40	0.35	0.0
3. High ΔP, high P	0.70	0.78	32	8	4	36	0.8	0.1
4. Low ΔP, high P	0.35	0.78	36	4	22	18	0.9	0.55

the first of these referring to the value of ΔP and the second to the value of the power measure P. Presented in the two rightmost columns are the values of $p(e|c)$ and $p(e|\neg c)$, respectively, which are based on the cell frequencies in Columns 4–7, and which were used to calculate ΔP and P. In conditions 1 and 4 the cell frequencies were exactly the same and they were used to generate a low value for ΔP and a high value for P. The rationale for incorporating two identical conditions was to examine the consistency of people's causal judgments. In the remaining two conditions the values of ΔP and P were similar; in condition 2, ΔP and P were low, and in condition 3 they were both high. Varying the values of ΔP and P in the four conditions allowed us to estimate base-rate usage for each participant via a method which will be described shortly. Participants were presented all four conditions, but the order of presentation of the conditions was counterbalanced across participants according to a Latin square design.

In the initial phase participants were presented with a set of instructions (see Appendix A: Causal learning instructions) along with five practice trials. In each trial participants were presented with a graphic image denoting the presence or absence of radiation, after which they would respond using mouse activated buttons either "YES, the mutation is going to occur", or "NO, the mutation is not going to occur". An image of a mutated or non-mutated butterfly then appeared together with the word "Yes" or "No" indicating its actual state. After 40 and 80 trials participants were asked "To what extent does radiation cause mutation?" Responses to this question were given on a 0–100 scale, the extreme ends of which were labeled "Radiation does not cause mutation" and "Radiation causes mutation" with the center point being labeled "Radiation is a moderate cause of mutation". In addition, participants were asked to give a confidence rating of their judgment on a scale ranging from "Not at all confident" to "Mildly confident" to "Very confident".

2.3. Weightings of causal judgments

To examine the relationship between judgments of causal strength and judgments in the four decision making tasks, we weighted the power PC model and the ΔP model, and participants' mean weights from each model were then correlated with performance in the decision making tasks.

The procedure used is as follows. In the case of the ΔP model, for each condition[1] [1 (low ΔP), 3 (high ΔP), and 4 (low ΔP)] we added a weight ranging between 0 and 1 (in increments of 0.05) to the value of $p(e|\neg c)$, and calculated a new value of ΔP according to the equation:

$$\Delta P = p(e|c) - wp(e|\neg c) \tag{3}$$

For example, in condition 1 (low ΔP) the value of $p(e|\neg c)$ is 0.55 (see Table 2), hence a weight of 0 changed the value of ΔP to 0.9 while a weight of 1 changed it to 0.35. If a participant gave a judgment of 90 in condition 1 (low ΔP), then their weighting of $p(e|\neg c)$ would be 0. For each participant the judgment they gave for condition 1(low ΔP) was compared with the range of predicted judgments for that condition according to Eq. (3). An optimal weight was selected that minimized the discrepancy between their judgment and the prediction of Eq. (3). The same procedure was then repeated for judgments in conditions 3 (high ΔP) and 4 (low ΔP). Thus, each participant was assigned an optimal weight for each of the three conditions, and these weights were then averaged to give a final minimized absolute weight which was used in later analyses as an estimate of base-rate sensitivity.

To find the weightings of participants' judgments in the three conditions [1(high P), 3 (high P), and 4 (high P)] according to the PC model, we used the following equation:

$$P = \frac{p(e|c) - wp(e|\neg c)}{1 - wp(e|\neg c)} \tag{4}$$

Using the same procedure as that used for comparing judgments according to weighted ΔP, each participant's judgments were compared with weighted P to find the closest fit between actual and predicted judgments. Each participant's three weights corresponding to the three conditions were again averaged to give a final minimized absolute weight which was also used in later analyses.

In the causal learning task, we included a condition [condition 2 (low ΔP/low P)] in which the value of $p(e|\neg c)$ is equal to 0 (see Table 2); adding weights to $p(e|\neg c)$ in condition 2 (low ΔP/high P) does not change the value of ΔP or P. Therefore, the reason, we included condition 2 was to permit an estimate of the weighting of $p(e|c)$ which we predicted would not correlate with base-rate usage in the decision making tasks.

Specifically, we conducted a similar procedure as described above using weighted ΔP and P, but this time $p(e|c)$ was weighted. The minimized absolute weights from these calculations were also used as a control in later analyses when correlating responses from decision making tasks with the causal learning task. In order to demonstrate a genuine relationship between individuals' usage of base-rate information in decision making and causal learning tasks, we would not expect to find correlations between responses to decision making tasks and weights associated with $p(e|c)$.

[1] Condition 2 was not included because the actual value of $p(e|\neg c)$ for both models equalled 0, and so it is meaningless to ask how participants weighted base-rate information in this condition; however, we did include this condition in a different analysis discussed later in this section.

One might ask why we do not predict that weightings of $p(e|c)$ (according to either normative model) should correspond with responses in decision making tasks; for instance, the $p(d|h)$ option in the probability inference problems is equivalent to $p(e|c)$. Predicting a correspondence between $p(e|c)$ and responses to decision making task rests on the assumption that people fully incorporate base-rate information but vary according to the extent they weight $p(e|c)$. This is at odds with evidence showing that people actually vary according to the extent that they neglect base-rate information (e.g., Bar-Hillel, 1980; Doherty & Mynatt, 1990; Fischhoff & Bar-Hillel, 1984; Tversky & Kahneman, 1982). It is for this reason, that we only predict a correspondence between the weighting of $p(e|\neg c)$ in both causal models and performance in the decision making tasks.

Thus for each participant an optimal weight was computed so as to minimize the discrepancy between judgments and the predictions of Eq. (3), and this procedure was then repeated with Eq. (4). Finally, weights were calculated again according to these equations, but with weightings on $p(e|c)$ rather than $p(e|\neg c)$. The four minimized absolute weights were used in later correlation analyses with responses from the decision making tasks.

2.4. Decision making tasks

Participants were given a booklet with four decision making tasks. Although no time restrictions were imposed, participants were told not to spend too long on each task; the mean time spent on each task was approximately 2 min. Each of the two sets of tasks (probability estimates, probability inference) included a non-causal and causal version. The original instructions from Kahneman and Tversky's (1973) non-causal and Bar-Hillel's (1980) causal problem were used for the probability estimate tasks (see Appendix A: Probability estimate problems). In both tasks probability estimates were given on a scale between 0 and 100. Doherty and Mynatt's (1990) causal base-rate inference task was used along with a non-causal version (see Appendix A: Probability inference problems).

3. Results

3.1. Causal learning task: causal judgments

Starting with the judgment data first, Fig. 1 presents the mean ratings for each condition after 40 and 80 trials, and indicates that judgments did not change between these stages. This trend was confirmed using an ANOVA with condition (conditions 1–4) $\times$ block (40, 80 trials) as within-subject factors, which revealed no significant main effect of block and no block $\times$ condition interaction, $F < 1$.

All remaining analyses of judgment data are based on the average of the ratings given after 40 and 80 trials. A one-way ANOVA indicated that there was a highly significant difference between judgments in the four conditions, $F(3, 204) = 40.84$,

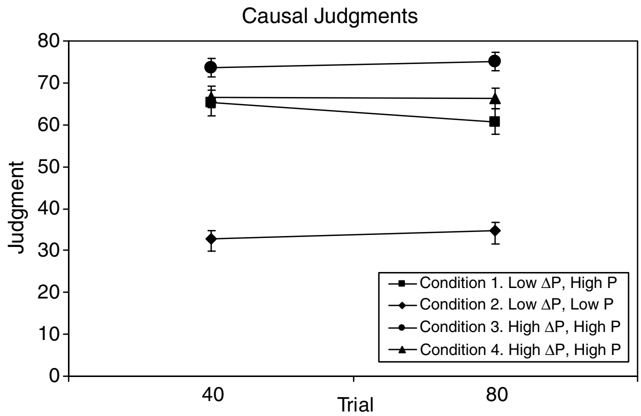

Fig. 1. Mean causal judgments (±SE) at both judgment periods for each condition in the causal learning task.

$p < 0.0005$. Paired sample *t*-tests revealed that there were significant differences in judgments between each pair of conditions ($p < 0.05$), with the exception of conditions 1 (low ΔP/high P) and 4 (low ΔP/high P) which are identical ($t < 1$). These findings are consistent with those from experiments described by Shanks (2004) on which this task was based.

3.2. Causal learning task: confidence ratings

Fig. 2 presents the mean confidence ratings for each condition after 40 and 80 trials and shows that these ratings did not change between these blocks. A one-way ANOVA comparing confidence ratings in the final trial block for each condition revealed no significant difference in ratings between the four conditions, $F(3, 204) = 1.32$, $p = 0.27$.

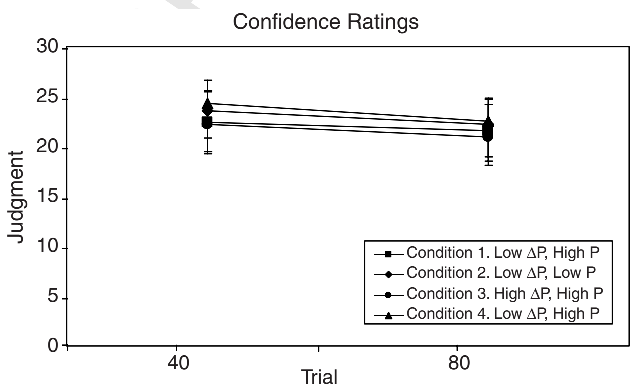

Fig. 2. Mean confidence ratings (±SE) at both judgment periods for each condition in the learning task.

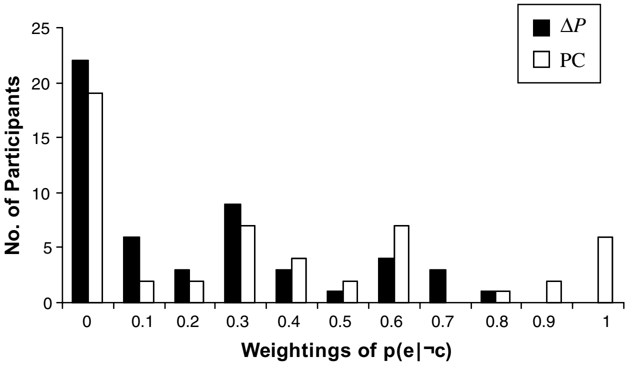

Fig. 3. Frequency of weightings of $p(e|\neg c)$ averaged across conditions 1, 3 and 4 for the Power PC and ΔP models.

3.3. Weightings

Fig. 3 presents the frequency of participants' final minimized absolute weighting of $p(e|\neg c)$ according to the ΔP and Power PC models.

For both of the models, a weight of 1 indicates that participants are consistent with the (unweighted) normative model. The figure also shows that most participants deviated from the normative models showing a tendency to underweight $p(e|\neg c)$. Fig. 3 suggests that the distribution of weights differed between the models, with weightings according to the ΔP model skewed towards the lower end of the scale. A Wilcoxon signed ranks test revealed a significant difference between the absolute weightings of the ΔP and PC models, $t(51) = 5.67$, $p < 0.0005$.

3.4. Decision making tasks: probability estimate problems

Participants performed poorly in both the Cab and Suicide problems, with only 21% of participants giving correct estimates of 41 (+/−10) in the cab problem, and 13% estimating 43 (+/−10) in the suicide problem. Thus, the causally framed version did not attenuate base-rate neglect. The modal estimate (80) given by 35% of participants in response to the cab problem was consistent with that reported in Kahneman and Tversky's (1973) original study. For the suicide problem the modal estimate was 75 and was made by 33% of participants, consistent with Bar-Hillel's (1980) study.

A correlation analysis between estimates given in both tasks revealed a significant relationship, suggesting that participants responded similarly to them, $r(52) = 0.41$, $p < 0.005$. Participants' estimates from the cab and suicide problems were then correlated with their weights in the causal learning task. The analysis revealed that estimates according to weighted ΔP correlated positively with actual estimates in the suicide problem, $r(52) = 0.28$, $p = 0.046$. No other correlations between probability estimates and weights estimated from power approached significance. As expected, there were no correlations between probability estimates and weights assigned to $p(e|c)$.

3.5. Decision making tasks: probability inference problem

Participants tended to perform poorly in both inference tasks, with only 6% of participants choosing the correct card combination [$p(h)$, $p(d|h)$, $p(d|\neg h)$] in the non-causal bird problem, and only 10% responding correctly in the causal disease problem. This also suggests that, as with the probability estimate tasks, the causal version did not facilitate correct performance. However, in contrast to the probability estimate tasks, the modal response in the two inference tasks differed. In the non-causal problem, approximately half (52%) of the participants chose the combination $p(h)$ and $p(d|h)$, while the next most popular response was the selection of $p(d|h)$ made by 14% of participants, followed by $p(d|h)$ and $p(d|\neg h)$ which was made by 12% of participants. Choices in the causal version converged with those reported by Stanovich and West (1998) and Doherty and Mynatt (1990). The modal response in the causally framed disease problem was $p(d|h)$ and $p(d|\neg h)$ and this accounted for 44% of participants' choices in the present study (cf. 36% of Stanovich and West's sample and 30% of Doherty and Mynatt's). The second most frequent choice was $p(d|h)$ made by 19% of participants, followed by $p(h)$ and $p(d|\neg h)$ made by 11.5% of participants. To better illustrate the difference in card choices between the two inference tasks, Table 3 shows, for each of the three key cards $p(h)$, $p(d|h)$ and $p(d|\neg h)$, the proportion of participants who responded with a card combination that included that card.

The most striking trend suggested in Table 3 is the sharp increase in the choice of base-rate $p(h)$ information in the non-causal problem compared with the causal problem, and the decrease in the inclusion of the $p(d|h)$ card in the non-causal version. There is little to distinguish these tasks with the exception of the actual context, and this suggests that the context did have a dramatic effect on the way base-rate information was perceived. A simple scoring scheme was used to classify card selection: for each of the cards $p(h)$, $p(d|h)$, $p(d|\neg h)$ participants scored '1' for its selection and '0' for its exclusion, and from this, we examined patterns in the card choices made in both inference tasks. Because the variables are binary, we use a φ (phi) coefficient which is equivalent to Pearson's R but is the appropriate correlation coefficient for dichotomous variables. The φ coefficient revealed a significant negative correlation between the selection of the $p(h)$ card in the two tasks, $\varphi(1) = -0.36$, $p < 0.0005$, and a positive correlation between the selection of the $p(d|\neg h)$ card, $\varphi(1) = 0.34$, $p < 0.0005$.

The scores based on card choices in both inference tasks were correlated with participants' weights in the causal learning task. A significant relationship was found between selection of the $p(d|\neg h)$ card in the inference tasks and causal estimates when

Table 3
Percentage of the sample in each of the probability inference tasks that included $p(h)$, $p(\neg h)$, $p(d/h)$, and $p(d/\neg h)$ in their choices

Options	$p(h)$	$p(\neg h)$	$p(d/h)$	$p(d/\neg h)$
Disease	28.8	9.6	88.8	69.2
Bird	65.4	0	78.8	34.6

$p(e|\neg c)$ was weighted in both the ΔP and P models. Specifically, there was a strong positive correlation between ΔP and the selection of the $p(d|\neg h)$ card in the disease problem, $r(52) = 0.38$, $p = 0.006$, and ΔP and the selection of the $p(d|\neg h)$ card in the bird problem, $r(52) = 0.30$, $p = 0.03$. There was also a strong positive correlation between P and the selection of the $p(d|\neg h)$ card in the bird problem, $r(52) = 0.37$, $p = 0.007$, and between P and the selection of the $p(d|\neg h)$ card in the disease problem, $r(52) = 0.29$, $p = 0.04$. No correlations were found between card choices and $p(e|c)$ weights.

Overall, our analyses suggest that participants' weighting of base-rate information is consistent across causal learning and decision making tasks; thus, we were able to demonstrate an underlying relationship between these domains, which until now had remained unexplored.

4. General discussion

The evidence from this study can be summarized as follows: first, we found that individuals differ according to the way they weight base-rate information in a causal learning task and that the way they do this corresponds to performance in decision making tasks. Second, as a control we demonstrated that only the weighting of $p(e|\neg c)$ in both models corresponded to performance in the decision making tasks, and not the weighting of $p(e|c)$. Third, weighting ΔP corresponded more closely to participants' performance in the decision making tasks than Power PC. Fourth, decision making tasks that were framed in a causal context did not facilitate correct performance as compared with standard non-causal versions.

So, what does the evidence imply about peoples' ability to use base rates in general? To begin, we must stress that this study is exploratory in nature, and as such, the conclusions are drawn with some caution. However, we were able to show that people consistently varied in their use of base-rate information across causal learning and decision making tasks. We are, however, tentative in suggesting that the tasks used in the present study index peoples' general ability to use base-rate information, since there are issues surrounding the framings of the classic Bayesian decision making tasks used here (e.g., Kohler, 1996; Maachi, 1995). This is clearly illustrated by comparing the modal responses to the probability inference tasks, in which the underlying structure of the tasks were identical, and only the context they referred to differed. Although framed in a causal context, there was more evidence of base-rate neglect in the Disease problem as compared with the Bird problem. This is at odds with Bar-Hillel's (1980) claim that framing standard decision making tasks in causal contexts can help to overcome base-rate neglect. However, Bar-Hillel (1980) also claimed that people's inability to integrate base-rate information is apparent in tasks where the indicant and base-rate information is not made relevant to the reasoner, and causal contexts are only one example in which their relevancy can be increased. Consistent with this, we were able to show that in a non-causal framing of a probability inference task in which the indicant and base-rate information were made relevant, base-rate neglect was attenuated.

The evidence from this study suggests that people may in fact be performing optimally according to their understanding of decision making tasks, or given the cognitive limitations under which they are working. Moreover, recent research on causal induction also shows that, in this case, the framing of the question used to elicit causal judgments can have marked affects on the types of responses given (e.g., Buehner et al., 2003; Shanks, 2004). Therefore, the evidence from the present study can also be viewed in the context of the three departures from normative standards identified by Stanovich and West (2000). We are inclined towards the position that given the potential ambiguity of the framing of many of the tasks used in this study, the findings suggest that people vary in their construal of the task requirements and the way in which they weight base-rate information, but they are consistent as to how they use this information across different task domains.

We also demonstrated that weightings of participants' causal judgments according to the ΔP model more accurately tracked their use of base-rate information in decision making tasks compared with the PC model. The evidence showed that absolute weightings of $p(e|\neg c)$ according to the ΔP model were lower than the PC model, suggesting that base-rates were undervalued consistent with the findings from the decision making tasks. The ΔP model is an expression of covariation whereas the PC model normalizes ΔP by the base-rate of the effect to express causation. Because of the normalization procedure, the PC model restricts the range of values between P (when $p(e|\neg c)$ has a weight of 1) and $p(e|c)$ (when $p(e|\neg c)$ has a weight of 0). For example, in conditions 1 and 4, when $w = 1$ for $p(e|\neg c)$ the value of $P = 0.78$ and $\Delta P = 0.35$ and when $w = 0$ for $p(e|\neg c)$ the value of $P = 0.90$ and $\Delta P = 0.90$. Thus, according to the weighting of $p(e|\neg c)$ of the ΔP model the range of values it accommodates is wider and so it is more sensitive than the PC model to the range of weightings of base-rate information in causal and decision making tasks. Another possible reason for the better tracking of the ΔP model to decision making behavior is that it might more accurately reflect the process of human covariation judgment than the PC model. There has been extensive debate about the relative merits of these models. Although arguments can be presented favoring each, there are good reasons to question the empirical and theoretical status of the PC model (Perales & Shanks, 2003; Shanks, 2004; see Buehner et al., 2003, for a contrasting view).

4.1. Future directions

The evidence from this study strongly suggests that Tbase-rate information in a cause–effect learning task and decision making problems is treated in the same way. Currently, Bayesian reasoning is coming to the fore in research on causal structure learning in which there are multiple cause–effect relationships (e.g., is insomnia the cause of stress and depression or is depression a common effect of stress and insomnia?) (Cooper & Herskovits, 1992; Steyvers, Tenenbaum, Wagenmakers, & Blum, 2003; Tenenbaum & Griffiths, 2001). Formal models (Pearl, 2000; Spirtes, Glymour, & Scheines, 2000) have been developed to capture the probabilistic dependencies present in a set of data and their relation to causal structures that could have generated that data. Many recent studies show that people generally take advantage of information about causal structure when making probability estimations or causal strength judgments (Waldmann & Hagmayer, 2001). Moreover, they benefit most by making interventions that disrupt causal chains rather than passively observing trials in which information about different causal structures is presented (Lagnado & Sloman, 2004).

Like the evidence reported in this study, there is evidence to suggest that Bayesian reasoning is incorporated in causal structure learning and that people differ in the way they do this. For example, Steyvers et al. (2003) recently presented an account of peoples' inferences of causal structure from a Bayesian perspective. They propose that as Bayesian hypothesis testers, people have a set of possible causal models to explain a particular observation in the world and that depending on their prior knowledge or particular biases they have, they evaluate relevant hypotheses for the one that best explains the observed data. From this they can make inferences about the causal structure and where best to alter some aspect of the structure in order to understand the cause–effect links. In their study they report three different strategies that people used when learning to discriminate between two causal structures. Those that used a strategy that integrated information across trials reliably made the optimal decision as dictated by the likelihood ratio. The next best were 'one trial Bayesians' because, although sensitive to the likelihood ratio, they failed to integrate information across trials. The worst performers failed to give judgments consistent with the likelihood ratio or examining data across trials. The findings give some indication of the variability of peoples' Bayesian reasoning in a causal inference task, but in particular, this evidence suggests that for some, the kinds of heuristics or deliberative Bayesian strategies that they employ come surprisingly close to those of a rational statistical inference model.

Given the strong Bayesian reasoning component that is implicated in learning about causal structures discussed here, it is plausible that the different kinds of strategies that people develop in complex cause–effect learning tasks should also correspond to their behavior in single cause-event learning tasks and decision making tasks. One possibility then, following the findings of the present study, would be to investigate whether people are consistent in the way they weight base-rate information across causal structure learning, single cause–effect learning tasks, and decision making tasks.

In conclusion, this study examined reasoning across related cognitive domains and has provided evidence that people vary as to how they weight the probability of the effect in the absence of the cause in a causal learning task, and that this is also indicative of the way in which they make probability estimates and inferences from Bayesian decision making tasks.

Acknowledgements

The support of the Economic and Social Research Council (ESRC) is gratefully acknowledged. The work was part of the programme of the ESRC Center for Economic Learning and Social Evolution. The authors would like to thank David

Lagnado for his advice and suggestions, and also Robert Hartsuiker and two anonymous reviewers for their helpful comments on an earlier draft of this article.

Appendix A

Causal learning instructions

Imagine you are working in a laboratory and you want to find out whether certain types of radiation cause or prevent a specific genetic mutation in butterflies' DNA. During this task you will see laboratory records from four studies. In each study, you will see information about administering one type of radiation to one species of insect. In one study, Gonepteryx Formosana were irradiated with U256 nuclear radiation, Ixias Pyrene were irradiated with P290, in a third Catopsilia Scylla were irradiated with Z210, and in a forth study Calliithea Leprieuri were irradiated with N235. In each study, some butterflies received nuclear radiation and some did not. In a test given 5 min later, the butterflies were examined for a specific genetic mutation at a particular DNA locus. Of course, mutations sometimes occur spontaneously in insects not exposed to nuclear radiation. What you must decide is whether and how strongly the radiation can independently cause this particular mutation. There are 80 butterflies in each study. The likelihood that mutations occur on their own (without radiation) is the same in all 80 butterflies in each study. Half of the butterflies in each study were randomly assigned to a group receiving nuclear radiation and half to a group not receiving any radiation. Each record tells you whether the butterfly was exposed to the relevant nuclear radiation or not. You will then be asked to predict whether or not the butterflies' DNA will show a genetic mutation in the test given 5 min later. When you have made your prediction you will be told whether the mutation was found or not. Use this feedback to try to find out whether the radiation really causes mutations. Although initially you will have to guess, by the end you will be an expert! At regular intervals during each study you will be asked to estimate the degree to which the radiation causes mutations, and to state how confident you are in your estimate. Further instructions will explain at the appropriate time how to make these estimates. You can now try some practice trials. GOOD LUCK!

Probability estimate problems

Cab problem

A cab was involved in a hit and run accident at night. Two cab companies, the Green and the Blue, operate in the city. You are given the following data:

(a) Although the two companies are roughly equal in size, 85% of the cabs accidents in the city involve Green cabs and 15% involve Blue cabs.
(b) A witness identified the cab as Blue. The court tested the reliability of the witness under the same circumstances that existed on the night of the accident and concluded that the witness correctly identified each one of the two colours 80% of the time and failed 20% of the time.

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

Please write your probability estimate in the box below

Estimate between 0% and 100%

Suicide problem

A study was done on causes of suicide among young adults (aged 25–35). It was found that the percentage of suicides is three times larger among single people than among married people. In this age group, 80% are married and 20% are single.

Of 100 cases of suicide among people aged 25–35, how many of the people would you estimate were single?

Please write your estimate in the box below

Estimate between 0% and 100%

Bare-rate inference problems

Disease problem

Imagine you are a doctor. A patient comes to you with a red rash on his fingers. What information would you want in order to diagnose whether the patient had the disease “Digirosa”?

Below are four pieces of information that may or may not be relevant to the diagnosis.

Please indicate by ticking the boxes below the piece/pieces of information that are necessary to make the diagnosis, but <u>only</u> tick information that is necessary to do so.

1. Percentage of people without Digirosa who have a red rash.
2. Percentage of people with Digirosa.
3. Percentage of people without Digirosa.
4. Percentage of people with Digirosa who have a red rash.

Bird problem

You are a bird watcher and have found a nest with pink speckled eggs. You are trying to find out whether they belong to the Blue Bellied Chaffinch. You need to consult your pocket guidebook to help you make the classification that the eggs do belong to the Blue Bellied Chaffinch. Below are four pieces of information that may or may not be relevant to make your classification. Please tick the piece/pieces of information that are necessary to make your classification, but <u>only</u> tick information that is necessary to do so.

1. Percentage of Blue Bellied Chaffinch without pink speckled eggs.
2. Percentage of Blue Bellied Chaffinch with pink speckled eggs.
3. Percentage of Blue Bellied Chaffinch in the area.
4. Percentage of Blue Bellied Chaffinch not in the area.

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.

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

Allan, L. G., & Jenkins, H. M. (1980). The judgment of contingency and the nature of response alternatives. *Canadian Journal of Psychology, 34*, 1–11.

Arkes, H. R., & Harkness, A. R. (1983). Estimates of contingency between two dichotomous variables. *Journal of Experimental Psychology: General, 112*, 117–135.

Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica, 44*, 211–233.

Baron, J. (1985). *Rationality and intelligence*. Cambridge, England: Cambridge University Press.

Beyth-Marom, R. (1982). Perception of correlation reexamined. *Memory and Cognition, 10*, 511–519.

Beyth-Marom, R., & Fischhoff, B. (1983). Diagnosticity and pseudodiagnosticity. *Journal of Personality and Social Psychology, 45*, 1185–1195.

Buehner, M. J., & Cheng, P. (1997). Causal induction: The power PC theory versus the Rescorla–Wagner model. In M. G. Shafto & P. Langley (Eds.), *Proceedings of the nineteenth annual conference of the cognitive science society* (pp. 55–60). Mahwah, NJ: Erlbaum.

Buehner, M. J., Cheng, P., & Clifford, D. (2003). From covariation to causation: a test of the assumption of causal power. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 1119–1140.

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

Cooper, G. F., & Herskovits, E. (1992). A Bayesian method for the induction of probabilistic networks from data. *Machine Learning, 9*, 309–347.

Crocker, J. (1981). Judgment of covariation by social perceivers. *Psychological Bulletin, 90*, 272–292.

Doherty, M. E., Chadwick, R., Garavan, H., Barr, D., & Mynatt, C. R. (1996). On people's understanding of the diagnostic implications of probabilistic data. *Memory and Cognition, 24*, 644–654.

Doherty, M. E., & Mynatt, C. (1990). Inattention to $P(H)$ and to $P(D/\neg H)$: a converging operation. *Acta Psychologica, 75*, 1–11.

Einhorn, H. J., & Hogarth, R. M. (1978). Confidence in judgment: persistence of the illusion of validity. *Psychological Review, 85*, 395–416.

Fischhoff, B., & Bar-Hillel, M. (1984). Diagnosticity and the base-rate effect. *Memory and Cognition, 24*, 402–410.

Hilton, D. J. (1995). The social context of reasoning: conversational inference and rational judgment. *Psychological Bulletin, 118*, 248–271.

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

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review, 103*, 582–591.

Kao, S.-F., & Wasserman, E. A. (1992). Assessment of information integration account of contingency judgment with examination of subjective cell importance and method of information presentation. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 19*, 1363–1386.

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

Klayman, J., & Ha, Y-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94*, 211–228.

Kohler, J. J. (1996). The base rate fallacy reconsidered: descriptive, normative and methodological challenges. *Behavioral & Brain Sciences, 19*, 1–53.

Lagnado, D., & Sloman, S. (2004). The advantage of timely intervention. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 30*, 856–876.

Maachi, L. (1995). Pragmatic aspects of the base-rate fallacy. *Quarterly Journal of Experimental Psychology, 48A*, 188–207.

Mandel, D. R., & Lehman, D. R. (1998). Integration of contingency information in judgments of cause, covariation, and probability. *Journal of Experimental Psychology: General, 127*, 269–285.

Morris, M. W., & Larrick, R. P. (1995). When one cause casts doubt on another—a normative analysis of discounting in causal attribution. *Psychological Review, 102*, 331–355.

Nisbett, R. E., & Ross, L. (1993). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.

Oaksford, M., & Chater, N. (1993). Reasoning theories and bounded rationality. In K. Manktelow & D. Over (Eds.), *Rationality: Psychological and philosophical perspectives* (pp. 31–60). London: Routledge.

Oppenheimer, D. M. (2004). Spontaneous discounting of availability in frequency judgment tasks. *Psychological Science, 15*, 100–105.

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

Perales, J. C., & Shanks, D. R. (2003). Normative and descriptive accounts of the influence of power and contingency on causal judgment. *Quarterly Journal of Experimental Psychology, 56A*, 977–1007.

Reeder, G., Vonk, R., Ronk, M., Ham, J., & Lawrence, M. (2004). Dispositional attribution: multiple inferences about motive-related traits. *Journal of Personality and Social Psychology, 86*, 530–544.

Schwarz, N. (1996). *Cognition and communication: Judgmental biases, research methods, and the logic of conversation*. Mahweh, NJ: Lawrence Erlbaum Associates.

Shafir, E. (1993). Intuitions and rationality and cognition. In K. Manktelow & D. Over (Eds.), *Rationality: Psychological and philosophical perspectives* (pp. 260–283). London: Routledge.

Shanks, D. R. (1995). Is human learning rational. *Quarterly Journal of Experimental Psychology, 48A*, 257–279.

Shanks, D. R. (2004). Judging covariation and causation. In D. J. Koehler & N. Harvey (Eds.), *Handbook of judgment and decision making* (pp. 220–239). Oxford: Blackwell.

Smedslund, S. (1963). The concept of correlation in adults. *Scandinavian Journal of Psychology, 4*, 165–173.

Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, prediction, and search* (2nd ed.). New York, NY: MIT Press.

Stanovich, K. E., & West, R. F. (1998). Who uses base rates and $P(D/\neg H)$. An analysis of individual differences. *Memory and Cognition, 26*, 161–179.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: implications for the rationality debate. *Behavioral & Brain Sciences, 22*, 645–665.

Steyvers, M., Tenenbaum, J. B., Wagenmakers, E. J., & Blum, B. (2003). Inferring causal networks from observations and interventions. *Cognitive Science, 27*, 453–489.

Tenenbaum, J. B., & Griffiths, T. L. (2001). Structure learning in human causal induction. In T. K. Leen, T. G. Dietterich, & V. Tresp (Eds.), *Advances in neural information processing systems* (pp. 59–65). Cambridge, MA: MIT Press.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science, 185*, 1124–1131.

Tversky, A., & Kahneman, D. (1980). Causal schemata in judgments under uncertainty. In M. Fishbein (Ed.), *Progress in social psychology* (pp. 49–72). Hillsdale, NJ: Erlbaum.

Tversky, A., & Kahneman, D. (1982). Evidential impact of base rates. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgments under uncertainty: Heuristics and biases* (pp. 153–160). Cambridge: Cambridge University Press.

Vallée-Tourangeau, F., Hollingsworth, L., & Murphy, R. (1998). Attentional Bias in correction judgments Smedslund (1963) revisited. *Scandinavian Journal of Psychology, 39*, 221–233.

Waldmann, M. R. (2001). Predictive versus diagnostic causal learning: evidence from an overshadowing paradigm. *Psychonomic Bulletin and Review, 8*, 600–608.

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

Ward, W. D., & Jenkins, H. M. (1965). The display of information and the judgment of contingency. *Canadian Journal of Psychology, 58*, 231–241.

Wasserman, E. A., Dorner, W. W., & Koa, S.-F. (1990). Contributions of specific cell information to judgments of interevent contingency. *Journal of Experimental Psychology: Learning, Memory and Cognition, 16*, 509–521.

White, P. A. (2003). Effects of wording and stimulus format on the use of contingency information in causal judgment. *Memory and Cognition, 31*, 231–242.