



ELSEVIER

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

SCIENCE @ DIRECT®

Acta Psychologica xxx (2005) xxx-xxx

acta  
psychologica

[www.elsevier.com/locate/actpsy](http://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,  $\Delta P$ ), 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.

© 2005 Published by Elsevier B.V.

*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](mailto:m.osman@ucl.ac.uk) (M. Osman).

26 **1. Introduction**

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

42 *1.1. Causal induction*

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

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

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

55 By examining the A and B cells of the contingency table, it is possible to determine  
 56 the proportion  $A/(A + B)$ , which is simply the probability of the effect 'e' in the pres-  
 57 ence of the cause 'c' expressed as  $p(e|c)$  in the  $\Delta P$  rule. In contrast,  $p(e|\neg c)$  refers to  
 58 the proportion  $C/(C + D)$ , which is the probability of the effect in the absence of the  
 59 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

60 some indication of the causal strength of the relationship between radiation and but-  
61 terfly mutation.

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

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

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

80 Studies of causal induction suggest that people, although on the whole, good at  
81 judging causal strength according to one or the other rule, tend to exhibit biased  
82 behavior when making inferences from contingency tables and in trial-by-trial learn-  
83 ing tasks (e.g., Allan & Jenkins, 1980; Smedslund, 1963; Vallée-Tourangeau, Hol-  
84 lingsworth, & Murphy, 1998; Ward & Jenkins, 1965). For example, studies show  
85 that people weight cell information, non-normatively, in the order  $A > B > C > D$   
86 (Kao & Wasserman, 1992; Mandel & Lehman, 1998). Hence, people are most sensi-  
87 tive to variations in cell A and tend to overestimate the value of this cell, whereas  
88 they are least sensitive to variations in cell D, often underestimating its value (Arkes  
89 & Harkness, 1983; Vallée-Tourangeau et al., 1998; Wasserman, Dornier, & Koa,  
90 1990). Normatively, the cells should be weighted equally. In addition, when pre-  
91 sented with conditions in which the effect is equally likely in the presence or absence  
92 of the cause ( $\Delta P = 0$ ) but the overall base-rate of the effect increases, people misper-  
93 ceive a contingency that is not there, known as the cause density effect (Buehner,  
94 Cheng, & Clifford, 2003; Perales & Shanks, 2003; Smedslund, 1963; Vallée-Touran-  
95 geau et al., 1998). This is not to say that people are unable to discriminate between  
96 positive, negative, and zero correlations—they can in fact do this well (e.g., Shanks,  
97 1995; Vallée-Tourangeau et al., 1998). However, in zero correlation conditions peo-  
98 ple fail to take sufficient account of the base-rate of the effect and so tend to overes-  
99 timate causal strength.

## 100 1.2. Bayesian decision making

101 We turn now to another type of situation in which people have to predict an out-  
102 come or the probability of an outcome in light of evidence, and in which they tend to

103 show biased behavior when reasoning about base-rate information. Typically, in  
 104 Bayesian decision making tasks, people are asked to judge the likelihood of an event  
 105 having occurred or that will occur. If they respond normatively, they will integrate  
 106 base-rates and the likelihood ratio according to Bayes' rule. In Bayes' rule the prob-  
 107 ability of the hypothesis tested ( $h$ ) is derived by multiplying the likelihood ratio of the  
 108 observed datum ( $d$ ) by the prior probability favoring the focal hypothesis:

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

111 What is summarized in the rule is that the diagnosticity of the likelihood ratio should  
 112 be evaluated independently of the prior odds favoring the focal hypothesis. To do  
 113 this, the rule includes three ratio terms. The far right term refers to the prior odds  
 114 favoring the focal hypothesis. The middle term refers to the likelihood ratio com-  
 115 posed of the probability of the data given the focal hypothesis divided by the prob-  
 116 ability of the data given its mutually exclusive component. The far left term  
 117 represents the posterior odds favoring the focal hypothesis after receipt of the new  
 118 data.

119 Numerous studies show that people tend not to give responses that obey Bayes  
 120 rule; instead, they predominantly make two types of error. First, people routinely ne-  
 121 glect the denominator of the likelihood ratio  $p(d|-h)$ , that is, they show a preference  
 122 for information in which the probability of the datum given the focal hypothesis is  
 123 true rather than false (Beyth-Marom & Fischhoff, 1983; Doherty, Chadwick, Gara-  
 124 van, Barr, & Mynatt, 1996; Einhorn & Hogarth, 1978; Wasserman et al., 1990). To  
 125 illustrate, Doherty and Mynatt (1990) presented participants with a problem in  
 126 which they were asked to determine whether a patient had the disease 'Digirosa'. Par-  
 127 ticipants were asked to select cards which contained information that would be rel-  
 128 evant in making their diagnosis: '% of people with Digirosa'  $p(h)$ , '% of people  
 129 without Digirosa'  $p(-h)$ , '% of people with Digirosa who have a red rash'  $p(d|h)$ ,  
 130 and '% of people without Digirosa who have a red rash'  $p(d|-h)$ . To solve the task  
 131 correctly, the cards  $p(d|h)$ ,  $p(d|-h)$ , and  $p(h)$  corresponding to the terms in the for-  
 132 mula are required;  $p(-h)$  is the complement of  $p(h)$  and so it is not necessary to cal-  
 133 culate the posterior probability.

134 Doherty and Mynatt (1990) found that, consistent with much of the judgment lit-  
 135 erature, few participants (11%) demonstrated an understanding of Bayesian reason-  
 136 ing by selecting the correct information. The least popular card choices were the  
 137 prior probability  $p(h)$  and  $p(d|-h)$ . To evaluate a target hypothesis, alternative  
 138 hypotheses must be considered, and Doherty and Mynatt proposed that participants  
 139 adopting a good hypothesis testing strategy would select the card  $p(d|-h)$  because it  
 140 indicates an awareness of alternative hypotheses. A later study by Stanovich and  
 141 West (1998) reported that participants choosing  $p(d|-h)$  in Doherty and Mynatt's  
 142 (1990) disease problem scored higher on tests of cognitive ability and a battery of  
 143 reasoning tasks (e.g., syllogisms, conditional reasoning tasks, probability based  
 144 problems) compared with those that had excluded this card from their choices.

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

147 Tversky & Kahneman, 1982). For example, in Kahneman and Tversky's (1973) clas-  
148 sic task participants are presented a short cover story: 85% of cabs in a particular city  
149 are green and the remainder are blue. A witness identifies a cab involved in an acci-  
150 dent as blue. Under tests, the witness correctly identifies both blue and green cabs on  
151 80% of the occasions. Participants are then asked: What is the probability that the  
152 cab was in fact blue? The posterior probability is in fact 0.41, however, few respond  
153 with this answer, tending instead to give estimates that range between 0.70 and 0.90.  
154 This highly robust finding has been taken as evidence of peoples' reliance on errone-  
155 ous intuitions such as the degree of correspondence between a sample and a popu-  
156 lation (the "representativeness" heuristic). Thus, people are sensitive to the  
157 diagnosticity of the descriptions in the cover story, but disregard the fact that the dif-  
158 ferent sub-classes are of different sizes (e.g., 85% green cabs vs. 15% blue cabs).

159 Bar-Hillel's (1980) alternative interpretation of Kahneman and Tversky's results  
160 suggests that the fallacy is the result of misperceiving the relevance of such informa-  
161 tion. There is evidence to suggest that base-rate information can be made more rel-  
162 evant when framed in such a way that it has a direct causal relation to the target  
163 information (Ajzen, 1977; Tversky & Kahneman, 1980). In tasks like the cab prob-  
164 lem base-rate information is presented as incidental to the main focus of the prob-  
165 lem, whereas contexts that increase the causal efficacy of base-rate information  
166 and therefore its status in the problem helps to attenuate base-rate neglect. Bar-Hillel  
167 (1980) claimed that such contexts clarify the relation between the base-rate and a  
168 target case enabling both types of information to become integrated. Formally the  
169 versions that Bar-Hillel used in her study were the same as Kahneman and Tversky's,  
170 but used causal contexts. Students were presented with a cover story which discussed  
171 suicide rates: A study was done on causes of suicide among young adults (aged 25–  
172 35). It was found that the percentage of suicides is three times larger among single  
173 than married people. In this age group, 80% are married and 20% are single. In  
174 one version of this task students were simply asked to estimate the likelihood of sui-  
175 cide in a given sub-population in which the posterior probability was 0.43. Bar-Hillel  
176 found that through various modifications to the framing of this task base-rate  
177 neglect could be reduced from 85% of responses to 25%. Changes to the framing in-  
178 cluded varying the base rate information and likelihood ratio. However, Bar-Hillel's  
179 study demonstrates that it is not causality per se that reduces base-rate neglect, but  
180 rather the relevancy it adds to this type of information, and so other contexts that do  
181 this are also able to attenuate base-rate neglect.

182 Evidence of deviations from Bayesian reasoning, such as base-rate neglect, have  
183 been the cause of much debate, raising questions about the appropriateness of tasks  
184 studying people's probabilistic reasoning (Kohler, 1996) and whether people are able  
185 to reason rationally (Kahneman & Tversky, 1996; Shafir, 1993). Similarly, in causal  
186 induction it is unclear why people should differentially weight the cells of a contin-  
187 gency table. Some have argued that this in fact implies an underlying bias for posi-  
188 tive or confirmatory evidence (Klayman & Ha, 1987; Mandel & Lehman, 1998).  
189 However, others suggest that the biases that have been found are inflated by the par-  
190 ticular choice of framing in which a task is couched or the phrasing of causal ques-

191 tions, rather than being an unavoidable property of people's causal judgments (e.g.,  
192 Beyth-Marom, 1982; Crocker, 1981; Perales & Shanks, 2003; Vallée-Tourangeau  
193 et al., 1998; Waldmann, 2001; White, 2003). These mixed findings can also be seen  
194 as representing a broader controversy between prescriptive (or normative) and  
195 descriptive explanations of non-normative behavior. That is, are deviations from  
196 normative models (e.g., Bayes rule,  $\Delta P$  rule, Power PC model) examples of biased  
197 information processing behavior, or the product of a cognitive system with limited  
198 computational capacity?

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

214 Like Stanovich and West, we also emphasize the relevance of individual differ-  
215 ences in relation to causal induction and Bayesian decision making by exploring  
216 the possible connection between people's use of base-rate information in both do-  
217 mains. The evidence of non-normative behavior in both research domains suggests  
218 that people encounter problems in tasks where they should incorporate base-rate  
219 information and that, particularly in decision making tasks, individuals vary accord-  
220 ing to whether or not they integrate such information.

221 Thus far, there has, to our knowledge, been no empirical work that compares cau-  
222 sal contingency judgments with responses to decision making tasks. However, one  
223 connection between causal learning and decision making that has been explored is  
224 in the context of discounting (Kelley, 1973; Morris & Larrick, 1995; Oppenheimer,  
225 2004; Reeder, Vonk, Ronk, Ham, & Lawrence, 2004) which refers to the phenome-  
226 non in which people show biased behavior when making a causal attribution in light  
227 of new information. Despite the fact that this work is based on the discounting prin-  
228 ciple and its common application in causal and decision making domains, there is no  
229 empirical comparison of how people use this principle in each of the domains.

230 In the present study, we investigated whether there are individual differences in the  
231 use of base-rate information in causal learning and how these relate to the use of  
232 base-rate information in Bayesian decision making. We used a causal learning task  
233 which is a modified version of a task described by Shanks (2004) and standard  
234 Bayesian decision making tasks: two probabilistic estimation problems (Kahneman

235 and Tversky's Cab problem, causal and non-causal versions), and two base-rate  
236 inference tasks (Doherty and Mynatt's Disease problem, causal and non-causal  
237 versions).

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

## 255 2. Method

### 256 2.1. Participants and apparatus

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

### 270 2.2. Design and procedure

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

Table 2

Cell frequencies, contingency ( $\Delta P$ ), power ( $P$ ) and values of  $p(e|c)$  and  $p(e|-c)$  in each condition

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

273 the first of these referring to the value of  $\Delta P$  and the second to the value of the power  
 274 measure  $P$ . Presented in the two rightmost columns are the values of  $p(e|c)$  and  
 275  $p(e|-c)$ , respectively, which are based on the cell frequencies in Columns 4–7, and  
 276 which were used to calculate  $\Delta P$  and  $P$ . In conditions 1 and 4 the cell frequencies  
 277 were exactly the same and they were used to generate a low value for  $\Delta P$  and a high  
 278 value for  $P$ . The rationale for incorporating two identical conditions was to examine  
 279 the consistency of people’s causal judgments. In the remaining two conditions the  
 280 values of  $\Delta P$  and  $P$  were similar; in condition 2,  $\Delta P$  and  $P$  were low, and in condition  
 281 3 they were both high. Varying the values of  $\Delta P$  and  $P$  in the four conditions allowed  
 282 us to estimate base-rate usage for each participant via a method which will be de-  
 283 scribed shortly. Participants were presented all four conditions, but the order of pre-  
 284 sentation of the conditions was counterbalanced across participants according to a  
 285 Latin square design.

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

### 300 2.3. Weightings of causal judgments

301 To examine the relationship between judgments of causal strength and judgments  
 302 in the four decision making tasks, we weighted the power PC model and the  $\Delta P$   
 303 model, and participants’ mean weights from each model were then correlated with  
 304 performance in the decision making tasks.



305 The procedure used is as follows. In the case of the  $\Delta P$  model, for each condition<sup>1</sup>  
306 [1 (low  $\Delta P$ ), 3 (high  $\Delta P$ ), and 4 (low  $\Delta P$ )] we added a weight ranging between 0 and 1  
307 (in increments of 0.05) to the value of  $p(e|\neg c)$ , and calculated a new value of  $\Delta P$   
308 according to the equation:

$$\Delta P = p(e|c) - wp(e|\neg c) \quad (3)$$

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

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

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

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

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

340 Specifically, we conducted a similar procedure as described above using weighted  
341  $\Delta P$  and  $P$ , but this time  $p(e|c)$  was weighted. The minimized absolute weights from  
342 these calculations were also used as a control in later analyses when correlating re-  
343 sponses from decision making tasks with the causal learning task. In order to dem-  
344 onstrate a genuine relationship between individuals' usage of base-rate information  
345 in decision making and causal learning tasks, we would not expect to find correla-  
346 tions between responses to decision making tasks and weights associated with  $p(e|c)$ .

---

<sup>1</sup> 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.

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

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

#### 364 2.4. Decision making tasks

365 Participants were given a booklet with four decision making tasks. Although no  
366 time restrictions were imposed, participants were told not to spend too long on each  
367 task; the mean time spent on each task was approximately 2 min. Each of the two  
368 sets of tasks (probability estimates, probability inference) included a non-causal  
369 and causal version. The original instructions from Kahneman and Tversky's  
370 (1973) non-causal and Bar-Hillel's (1980) causal problem were used for the probab-  
371 ility estimate tasks (see Appendix A: Probability estimate problems). In both tasks  
372 probability estimates were given on a scale between 0 and 100. Doherty and Mynatt's  
373 (1990) causal base-rate inference task was used along with a non-causal version (see  
374 Appendix A: Probability inference problems).

### 375 3. Results

#### 376 3.1. Causal learning task: causal judgments

377 Starting with the judgment data first, Fig. 1 presents the mean ratings for each  
378 condition after 40 and 80 trials, and indicates that judgments did not change between  
379 these stages. This trend was confirmed using an ANOVA with condition (conditions  
380 1–4)  $\times$  block (40, 80 trials) as within-subject factors, which revealed no significant  
381 main effect of block and no block  $\times$  condition interaction,  $F < 1$ .

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

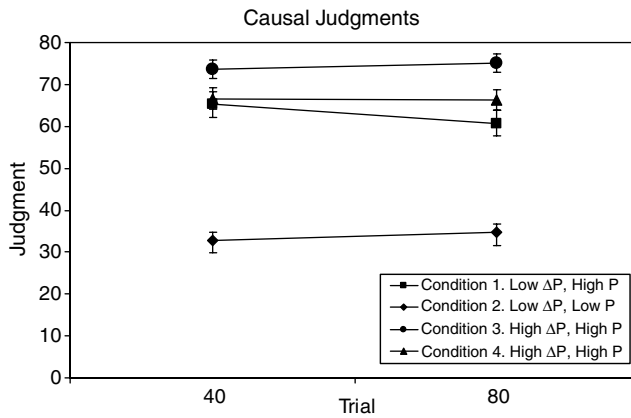


Fig. 1. Mean causal judgments ( $\pm$ SE) at both judgment periods for each condition in the causal learning task.

385  $p < 0.0005$ . Paired sample  $t$ -tests revealed that there were significant differences in  
386 judgments between each pair of conditions ( $p < 0.05$ ), with the exception of condi-  
387 tions 1 (low  $\Delta P$ /high  $P$ ) and 4 (low  $\Delta P$ /high  $P$ ) which are identical ( $t < 1$ ). These find-  
388 ings are consistent with those from experiments described by Shanks (2004) on which  
389 this task was based.

390 3.2. Causal learning task: confidence ratings

391 Fig. 2 presents the mean confidence ratings for each condition after 40 and 80 tri-  
392 als and shows that these ratings did not change between these blocks. A one-way  
393 ANOVA comparing confidence ratings in the final trial block for each condition re-  
394 vealed no significant difference in ratings between the four conditions,  $F(3, 204) =$   
395 1.32,  $p = 0.27$ .

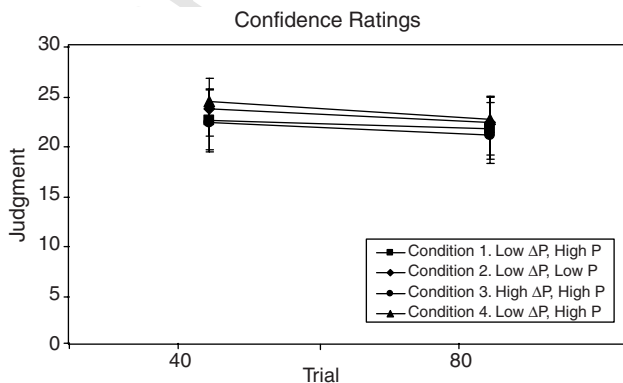


Fig. 2. Mean confidence ratings ( $\pm$ SE) at both judgment periods for each condition in the learning task.

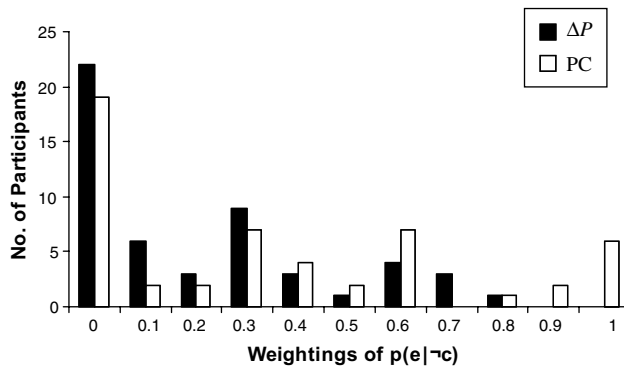


Fig. 3. Frequency of weightings of  $p(e|\neg c)$  averaged across conditions 1, 3 and 4 for the Power PC and  $\Delta P$  models.

### 396 3.3. Weightings

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

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

### 406 3.4. Decision making tasks: probability estimate problems

407 Participants performed poorly in both the Cab and Suicide problems, with only  
408 21% of participants giving correct estimates of 41 (+/−10) in the cab problem, and  
409 13% estimating 43 (+/−10) in the suicide problem. Thus, the causally framed version  
410 did not attenuate base-rate neglect. The modal estimate (80) given by 35% of partic-  
411 ipants in response to the cab problem was consistent with that reported in Kahneman  
412 and Tversky's (1973) original study. For the suicide problem the modal estimate was  
413 75 and was made by 33% of participants, consistent with Bar-Hillel's (1980) study.

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

## 423 3.5. Decision making tasks: probability inference problem

424 Participants tended to perform poorly in both inference tasks, with only 6% of par-  
 425 ticipants choosing the correct card combination [ $p(h)$ ,  $p(d|h)$ ,  $p(d|\neg h)$ ] in the non-cau-  
 426 sal bird problem, and only 10% responding correctly in the causal disease problem.  
 427 This also suggests that, as with the probability estimate tasks, the causal version did  
 428 not facilitate correct performance. However, in contrast to the probability estimate  
 429 tasks, the modal response in the two inference tasks differed. In the non-causal prob-  
 430 lem, approximately half (52%) of the participants chose the combination  $p(h)$  and  
 431  $p(d|h)$ , while the next most popular response was the selection of  $p(d|h)$  made by  
 432 14% of participants, followed by  $p(d|h)$  and  $p(d|\neg h)$  which was made by 12% of partic-  
 433 ipants. Choices in the causal version converged with those reported by Stanovich and  
 434 West (1998) and Doherty and Mynatt (1990). The modal response in the causally  
 435 framed disease problem was  $p(d|h)$  and  $p(d|\neg h)$  and this accounted for 44% of partic-  
 436 ipants' choices in the present study (cf. 36% of Stanovich and West's sample and 30% of  
 437 Doherty and Mynatt's). The second most frequent choice was  $p(d|h)$  made by 19% of  
 438 participants, followed by  $p(h)$  and  $p(d|\neg h)$  made by 11.5% of participants. To better  
 439 illustrate the difference in card choices between the two inference tasks, Table 3 shows,  
 440 for each of the three key cards  $p(h)$ ,  $p(d|h)$  and  $p(d|\neg h)$ , the proportion of participants  
 441 who responded with a card combination that included that card.

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

456 The scores based on card choices in both inference tasks were correlated with par-  
 457 ticipants' weights in the causal learning task. A significant relationship was found be-  
 458 tween 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

459  $p(e|-c)$  was weighted in both the  $\Delta P$  and  $P$  models. Specifically, there was a strong  
460 positive correlation between  $\Delta P$  and the selection of the  $p(d|-h)$  card in the disease  
461 problem,  $r(52) = 0.38$ ,  $p = 0.006$ , and  $\Delta P$  and the selection of the  $p(d|-h)$  card in the  
462 bird problem,  $r(52) = 0.30$ ,  $p = 0.03$ . There was also a strong positive correlation be-  
463 tween  $P$  and the selection of the  $p(d|-h)$  card in the bird problem,  $r(52) = 0.37$ ,  
464  $p = 0.007$ , and between  $P$  and the selection of the  $p(d|-h)$  card in the disease prob-  
465 lem,  $r(52) = 0.29$ ,  $p = 0.04$ . No correlations were found between card choices and  
466  $p(e|c)$  weights.

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

#### 471 4. General discussion

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

481 So, what does the evidence imply about peoples' ability to use base rates in gen-  
482 eral? To begin, we must stress that this study is exploratory in nature, and as such,  
483 the conclusions are drawn with some caution. However, we were able to show that  
484 people consistently varied in their use of base-rate information across causal learning  
485 and decision making tasks. We are, however, tentative in suggesting that the tasks  
486 used in the present study index peoples' general ability to use base-rate information,  
487 since there are issues surrounding the framings of the classic Bayesian decision mak-  
488 ing tasks used here (e.g., Kohler, 1996; Maachi, 1995). This is clearly illustrated by  
489 comparing the modal responses to the probability inference tasks, in which the  
490 underlying structure of the tasks were identical, and only the context they referred  
491 to differed. Although framed in a causal context, there was more evidence of base-  
492 rate neglect in the Disease problem as compared with the Bird problem. This is at  
493 odds with Bar-Hillel's (1980) claim that framing standard decision making tasks in  
494 causal contexts can help to overcome base-rate neglect. However, Bar-Hillel  
495 (1980) also claimed that people's inability to integrate base-rate information is  
496 apparent in tasks where the indicant and base-rate information is not made relevant  
497 to the reasoner, and causal contexts are only one example in which their relevancy  
498 can be increased. Consistent with this, we were able to show that in a non-causal  
499 framing of a probability inference task in which the indicant and base-rate informa-  
500 tion were made relevant, base-rate neglect was attenuated.

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

513 We also demonstrated that weightings of participants' causal judgments according  
514 to the  $\Delta P$  model more accurately tracked their use of base-rate information in deci-  
515 sion making tasks compared with the PC model. The evidence showed that absolute  
516 weightings of  $p(e|\neg c)$  according to the  $\Delta P$  model were lower than the PC model, sug-  
517 gesting that base-rates were undervalued consistent with the findings from the deci-  
518 sion making tasks. The  $\Delta P$  model is an expression of covariation whereas the PC  
519 model normalizes  $\Delta P$  by the base-rate of the effect to express causation. Because  
520 of the normalization procedure, the PC model restricts the range of values between  
521  $P$  (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  
522 example, in conditions 1 and 4, when  $w = 1$  for  $p(e|\neg c)$  the value of  $P = 0.78$  and  
523  $\Delta P = 0.35$  and when  $w = 0$  for  $p(e|\neg c)$  the value of  $P = 0.90$  and  $\Delta P = 0.90$ . Thus,  
524 according to the weighting of  $p(e|\neg c)$  of the  $\Delta P$  model the range of values it accom-  
525 modates is wider and so it is more sensitive than the PC model to the range of weigh-  
526 tings of base-rate information in causal and decision making tasks. Another possible  
527 reason for the better tracking of the  $\Delta P$  model to decision making behavior is that it  
528 might more accurately reflect the process of human covariation judgment than the  
529 PC model. There has been extensive debate about the relative merits of these models.  
530 Although arguments can be presented favoring each, there are good reasons to ques-  
531 tion the empirical and theoretical status of the PC model (Perales & Shanks, 2003;  
532 Shanks, 2004; see Buehner et al., 2003, for a contrasting view).

#### 533 4.1. Future directions

534 The evidence from this study strongly suggests that Tbase-rate information in a  
535 cause-effect learning task and decision making problems is treated in the same way.  
536 Currently, Bayesian reasoning is coming to the fore in research on causal structure  
537 learning in which there are multiple cause-effect relationships (e.g., is insomnia the  
538 cause of stress and depression or is depression a common effect of stress and insom-  
539 nia?) (Cooper & Herskovits, 1992; Steyvers, Tenenbaum, Wagenmakers, & Blum,  
540 2003; Tenenbaum & Griffiths, 2001). Formal models (Pearl, 2000; Spirtes, Glymour,  
541 & Scheines, 2000) have been developed to capture the probabilistic dependencies pres-  
542 ent in a set of data and their relation to causal structures that could have generated

543 that data. Many recent studies show that people generally take advantage of information  
544 about causal structure when making probability estimations or causal strength  
545 judgments (Waldmann & Hagmayer, 2001). Moreover, they benefit most by making  
546 interventions that disrupt causal chains rather than passively observing trials in which  
547 information about different causal structures is presented (Lagnado & Sloman, 2004).

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

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

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

## 580 Acknowledgements

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



584 Lagnado for his advice and suggestions, and also Robert Hartsuiker and two anon-  
585 ymous reviewers for their helpful comments on an earlier draft of this article.

## 586 **Appendix A**

### 587 *Causal learning instructions*

588 Imagine you are working in a laboratory and you want to find out whether certain  
589 types of radiation cause or prevent a specific genetic mutation in butterflies' DNA.  
590 During this task you will see laboratory records from four studies. In each study,  
591 you will see information about administering one type of radiation to one species  
592 of insect. In one study, *Gonepteryx Formosana* were irradiated with U256 nuclear  
593 radiation, *Ixias Pyrene* were irradiated with P290, in a third *Catopsilia Scylla* were  
594 irradiated with Z210, and in a forth study *Calliithea Leprieuri* were irradiated with  
595 N235. In each study, some butterflies received nuclear radiation and some did not. In  
596 a test given 5 min later, the butterflies were examined for a specific genetic mutation  
597 at a particular DNA locus. Of course, mutations sometimes occur spontaneously in  
598 insects not exposed to nuclear radiation. What you must decide is whether and how  
599 strongly the radiation can independently cause this particular mutation. There are 80  
600 butterflies in each study. The likelihood that mutations occur on their own (without  
601 radiation) is the same in all 80 butterflies in each study. Half of the butterflies in each  
602 study were randomly assigned to a group receiving nuclear radiation and half to a  
603 group not receiving any radiation. Each record tells you whether the butterfly was  
604 exposed to the relevant nuclear radiation or not. You will then be asked to predict  
605 whether or not the butterflies' DNA will show a genetic mutation in the test given  
606 5 min later. When you have made your prediction you will be told whether the muta-  
607 tion was found or not. Use this feedback to try to find out whether the radiation  
608 really causes mutations. Although initially you will have to guess, by the end you will  
609 be an expert! At regular intervals during each study you will be asked to estimate the  
610 degree to which the radiation causes mutations, and to state how confident you are  
611 in your estimate. Further instructions will explain at the appropriate time how to  
612 make these estimates. You can now try some practice trials. GOOD LUCK!

### 613 *Probability estimate problems*

#### 614 *Cab problem*

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

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

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

627 Please write your probability estimate in the box below

628 Estimate between 0% and 100%

629 *Suicide problem*

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

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

635 Please write your estimate in the box below

636 Estimate between 0% and 100%

637 *Bare-rate inference problems*

638 *Disease problem*

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

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

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

646 1. Percentage of people without Digirosa who have a red rash.

647 2. Percentage of people with Digirosa.

648 3. Percentage of people without Digirosa.

649 4. Percentage of people with Digirosa who have a red rash.

650

651 *Bird problem*

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

659 1. Percentage of Blue Bellied Chaffinch without pink speckled eggs.

660 2. Percentage of Blue Bellied Chaffinch with pink speckled eggs.

661 3. Percentage of Blue Bellied Chaffinch in the area.

662 4. Percentage of Blue Bellied Chaffinch not in the area.

665 **References**

- 666 Ajzen, I. (1977). Intuitive theories of events and the effects of base rate information on prediction. *Journal*  
667 *of Personality and Social Psychology*, 35, 303–314.
- 668 Allan, L. G. (1980). A note on measurement of contingency between two binary variables in judgment  
669 task. *Bulletin of the Psychonomic Society*, 15, 147–149.
- 670 Allan, L. G., & Jenkins, H. M. (1980). The judgment of contingency and the nature of response  
671 alternatives. *Canadian Journal of Psychology*, 34, 1–11.
- 672 Arkes, H. R., & Harkness, A. R. (1983). Estimates of contingency between two dichotomous variables.  
673 *Journal of Experimental Psychology: General*, 112, 117–135.
- 674 Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica*, 44, 211–233.
- 675 Baron, J. (1985). *Rationality and intelligence*. Cambridge, England: Cambridge University Press.
- 676 Beyth-Marom, R. (1982). Perception of correlation reexamined. *Memory and Cognition*, 10, 511–519.
- 677 Beyth-Marom, R., & Fischhoff, B. (1983). Diagnosticity and pseudodiagnosticity. *Journal of Personality*  
678 *and Social Psychology*, 45, 1185–1195.
- 679 Buehner, M. J., & Cheng, P. (1997). Causal induction: The power PC theory versus the Rescorla–Wagner  
680 model. In M. G. Shafto & P. Langley (Eds.), *Proceedings of the nineteenth annual conference of the*  
681 *cognitive science society* (pp. 55–60). Mahwah, NJ: Erlbaum.
- 682 Buehner, M. J., Cheng, P., & Clifford, D. (2003). From covariation to causation: a test of the assumption of  
683 causal power. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 1119–1140.
- 684 Cheng, P. (1997). From covariation to causation: a causal power theory. *Psychological Review*, 104, 367–405.
- 685 Cooper, G. F., & Herskovits, E. (1992). A Bayesian method for the induction of probabilistic networks  
686 from data. *Machine Learning*, 9, 309–347.
- 687 Crocker, J. (1981). Judgment of covariation by social perceivers. *Psychological Bulletin*, 90, 272–292.
- 688 Doherty, M. E., Chadwick, R., Garavan, H., Barr, D., & Mynatt, C. R. (1996). On people's understanding  
689 of the diagnostic implications of probabilistic data. *Memory and Cognition*, 24, 644–654.
- 690 Doherty, M. E., & Mynatt, C. (1990). Inattention to  $P(H)$  and to  $P(D/-H)$ : a converging operation. *Acta*  
691 *Psychologica*, 75, 1–11.
- 692 Einhorn, H. J., & Hogarth, R. M. (1978). Confidence in judgment: persistence of the illusion of validity.  
693 *Psychological Review*, 85, 395–416.
- 694 Fischhoff, B., & Bar-Hillel, M. (1984). Diagnosticity and the base-rate effect. *Memory and Cognition*, 24,  
695 402–410.
- 696 Hilton, D. J. (1995). The social context of reasoning: conversational inference and rational judgment.  
697 *Psychological Bulletin*, 118, 248–271.
- 698 Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80, 237–251.
- 699 Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review*, 103,  
700 582–591.
- 701 Kao, S.-F., & Wasserman, E. A. (1992). Assessment of information integration account of contingency  
702 judgment with examination of subjective cell importance and method of information presentation.  
703 *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 19, 1363–1386.
- 704 Kelley, H. H. (1973). The process of causal attribution. *American Psychologist*, 28, 107–128.
- 705 Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing.  
706 *Psychological Review*, 94, 211–228.
- 707 Kohler, J. J. (1996). The base rate fallacy reconsidered: descriptive, normative and methodological  
708 challenges. *Behavioral & Brain Sciences*, 19, 1–53.
- 709 Lagnado, D., & Sloman, S. (2004). The advantage of timely intervention. *Journal of Experimental*  
710 *Psychology: Learning, Memory, & Cognition*, 30, 856–876.
- 711 Maachi, L. (1995). Pragmatic aspects of the base-rate fallacy. *Quarterly Journal of Experimental*  
712 *Psychology*, 48A, 188–207.
- 713 Mandel, D. R., & Lehman, D. R. (1998). Integration of contingency information in judgments of cause,  
714 covariation, and probability. *Journal of Experimental Psychology: General*, 127, 269–285.
- 715 Morris, M. W., & Larrick, R. P. (1995). When one cause casts doubt on another—a normative analysis of  
716 discounting in causal attribution. *Psychological Review*, 102, 331–355.

- 717 Nisbett, R. E., & Ross, L. (1993). *Human inference: Strategies and shortcomings of social judgment*.  
718 Englewood Cliffs, NJ: Prentice-Hall.
- 719 Oaksford, M., & Chater, N. (1993). Reasoning theories and bounded rationality. In K. Manktelow & D.  
720 Over (Eds.), *Rationality: Psychological and philosophical perspectives* (pp. 31–60). London: Routledge.
- 721 Oppenheimer, D. M. (2004). Spontaneous discounting of availability in frequency judgment tasks.  
722 *Psychological Science*, 15, 100–105.
- 723 Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge University Press.
- 724 Perales, J. C., & Shanks, D. R. (2003). Normative and descriptive accounts of the influence of power and  
725 contingency on causal judgment. *Quarterly Journal of Experimental Psychology*, 56A, 977–1007.
- 726 Reeder, G., Vonk, R., Ronk, M., Ham, J., & Lawrence, M. (2004). Dispositional attribution: multiple  
727 inferences about motive-related traits. *Journal of Personality and Social Psychology*, 86, 530–544.
- 728 Schwarz, N. (1996). *Cognition and communication: Judgmental biases, research methods, and the logic of  
729 conversation*. Mahwah, NJ: Lawrence Erlbaum Associates.
- 730 Shafir, E. (1993). Intuitions and rationality and cognition. In K. Manktelow & D. Over (Eds.), *Rationality:  
731 Psychological and philosophical perspectives* (pp. 260–283). London: Routledge.
- 732 Shanks, D. R. (1995). Is human learning rational. *Quarterly Journal of Experimental Psychology*, 48A,  
733 257–279.
- 734 Shanks, D. R. (2004). Judging covariation and causation. In D. J. Koehler & N. Harvey (Eds.), *Handbook  
735 of judgment and decision making* (pp. 220–239). Oxford: Blackwell.
- 736 Smedslund, S. (1963). The concept of correlation in adults. *Scandinavian Journal of Psychology*, 4,  
737 165–173.
- 738 Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, prediction, and search* (2nd ed.). New York,  
739 NY: MIT Press.
- 740 Stanovich, K. E., & West, R. F. (1998). Who uses base rates and  $P(D/-H)$ . An analysis of individual  
741 differences. *Memory and Cognition*, 26, 161–179.
- 742 Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: implications for the rationality  
743 debate. *Behavioral & Brain Sciences*, 22, 645–665.
- 744 Steyvers, M., Tenenbaum, J. B., Wagenmakers, E. J., & Blum, B. (2003). Inferring causal networks from  
745 observations and interventions. *Cognitive Science*, 27, 453–489.
- 746 Tenenbaum, J. B., & Griffiths, T. L. (2001). Structure learning in human causal induction. In T. K. Leen,  
747 T. G. Dietterich, & V. Tresp (Eds.), *Advances in neural information processing systems* (pp. 59–65).  
748 Cambridge, MA: MIT Press.
- 749 Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science*, 185,  
750 1124–1131.
- 751 Tversky, A., & Kahneman, D. (1980). Causal schemata in judgments under uncertainty. In M. Fishbein  
752 (Ed.), *Progress in social psychology* (pp. 49–72). Hillsdale, NJ: Erlbaum.
- 753 Tversky, A., & Kahneman, D. (1982). Evidential impact of base rates. In D. Kahneman, P. Slovic, & A.  
754 Tversky (Eds.), *Judgments under uncertainty: Heuristics and biases* (pp. 153–160). Cambridge:  
755 Cambridge University Press.
- 756 Vallée-Tourangeau, F., Hollingsworth, L., & Murphy, R. (1998). Attentional Bias in correction judgments  
757 Smedslund (1963) revisited. *Scandinavian Journal of Psychology*, 39, 221–233.
- 758 Waldmann, M. R. (2001). Predictive versus diagnostic causal learning: evidence from an overshadowing  
759 paradigm. *Psychonomic Bulletin and Review*, 8, 600–608.
- 760 Waldmann, M. R., & Hagmayer, Y. (2001). Estimating causal strength: the role of structural knowledge  
761 and processing effort. *Cognition*, 82, 27–58.
- 762 Ward, W. D., & Jenkins, H. M. (1965). The display of information and the judgment of contingency.  
763 *Canadian Journal of Psychology*, 58, 231–241.
- 764 Wasserman, E. A., Dorner, W. W., & Koa, S.-F. (1990). Contributions of specific cell information to  
765 judgments of interevent contingency. *Journal of Experimental Psychology: Learning, Memory and  
766 Cognition*, 16, 509–521.
- 767 White, P. A. (2003). Effects of wording and stimulus format on the use of contingency information in  
768 causal judgment. *Memory and Cognition*, 31, 231–242.
- 769