Causal model and sampling approaches to reducing base rate neglect

Brett K. Hayes (B.Hayes@unsw.edu.au)
Ben R. Newell (Ben.Newell@unsw.edu.au)
Guy E. Hawkins (G.Hawkins@unsw.edu.au)
School of Psychology, University of New South Wales
Sydney, 2052, Australia

Abstract

Two studies examined how sampling of base rate information and causal explanation of false positives facilitate intuitive probability judgments. Experiment 1a varied these two manipulations factorially. Each had an additive effect on reducing base rate neglect and increasing choice of the normatively correct solution. Experiment 1b showed that description of relevant distributional information produced similar facilitation to sequential sampling. These results indicate that causal and sampling approaches impact on different components of probability judgment.

Keywords: Causal reasoning, Sequential sampling, Base rate neglect, Bayesian judgment, Belief updating

Introduction

One of the most commonly observed biases in human judgment is neglect of relevant base rate information (Eddy, 1982; Gigerenzer & Hoffrage, 1995; Tversky & Kahneman, 1974). For example, when people attempt to solve intuitive probability problems like that in Figure 1 (standard version), they typically ignore the low base rate (p(Cancer) = .01), generating probability estimates that are much higher than the normative Bayesian solution (p(Cancer|Mammogram+) ≈ 0.051, see Appendix for a derivation) .

Previous work has suggested a number of solutions to the problem of base rate neglect. These include the use of frequency rather than probability formats for relevant statistics (Gigerenzer & Hoffrage, 1995), and instructions that clarify set relations between the relevant samples (Barbey & Sloman, 2007; Evans, Handley, Perham, Over, & Thompson, 2000).

Two novel approaches to explaining and reducing base rate neglect have recently been proposed. The first involves consideration of the intuitive causal models that people construct when solving probability problems. Krynski and Tenenbaum (2007) outline a "causal-Bayes" account of probability judgments which assumes that errors arise when the statistics in a given problem do not readily map onto an intuitive causal model. In the standard mammogram problem for example, no causal explanation for the false positive rate (the probability of a positive mammogram in the absence of cancer) is given. According to Krynski and Tenenbaum (2007) this makes it difficult to integrate the false positive rate into Bayesian calculations, leading to inflated probability estimates. The problem can be overcome by providing a causal explanation for the relevant statistics. Krynski and Tenenbaum (2007) found that when such an explanation was supplied (see the causal version in Figure 1) there was a marked increase in the accuracy of probability estimates.

Alternately Hogarth and Soyer (2011) suggest that people are less likely to neglect relevant statistics when they have had "experience" with the relevant sample. Specifically, they suggest that trial-by-trial sampling of the frequency of an event from the relevant probability distribution can lead to more accurate estimates in problems involving low base rates (cf. Lejarraga, 2010; Sedlmeier, 1999). Hence, Hogarth and Soyer (2011) allowed some participants to draw sequential samples of women with a positive mammogram from a distribution with a low base rate of cancer. Sampling led to more accurate probability estimates than when only a description of the base rate was provided.

Mammogram problem

Doctors often encourage women at age 50 to participate in a routine mammography screening for breast cancer.

From past statistics, the following is known:
1% of women had breast cancer at the time of the screening
Of those with breast cancer, 80% received a positive result on the mammogram

[Standard version] Of those without breast cancer, 15% received a positive result on the mammogram

[Causal version] 30% of the women had a benign cyst at the time of screening. Of those with a benign cyst, 50% received a positive test on the mammogram

All others received a negative result

Suppose a woman gets a positive result during a routine mammogram screening. Without knowing any other symptoms, what are the chances she has breast cancer? ___%

Figure 1. The mammogram problem

Combining causal model and sampling approaches

A key motivation for the current work was that the causal model and sampling approaches appear to address different components of intuitive probability problems. Krynski and Tenenbaum (2007), focused on incorporating information about *false positive* rates into a causal model of the problem. In contrast, Hogarth and Soyer's (2011) sequential sampling approach aimed at improving sensitivity to the *low base* *rate*. The major goal of the current research was to combine these two approaches to overcoming base rate neglect. If our analysis is correct, then the causal model and sampling approaches should have additive effects on performance in intuitive probability problems.

A secondary goal was to address a number of methodological limitations of previous work using sequential sampling to overcome base rate neglect. First, Hogarth and Soyer (2011) asked participants to answer the same probability problem on three occasions; after a summary description of the base rate, after sampling experience, and a final estimate. For mammogram problems like that in Figure 1 this led to a complex pattern of results with accuracy increasing when probability problems were solved after sampling, but a marked decrease in accuracy when participants subsequently solved the same problem after reading a description of the base rate. To allow for a more straightforward assessment of the effects of description and experience, Experiment 1a used a between-subjects manipulation in which half the participants provided an answer to the intuitive probability problem after reading a description and having relevant sampling experience, whereas the remainder answered on the basis of the description alone.

Second, Hogarth and Soyer (2011) assessed intuitive probability accuracy using a relatively liberal performance measure (participants had to choose the correct estimate from four options). This is likely to yield higher levels of accuracy than the more conventional method of requesting point estimates of probability. To facilitate comparison of the sampling and causal model approaches we therefore assessed performance using both open-ended estimates (as used by Krynski & Tenenbaum, 2007) and forced choice questions.

Third and most importantly, we aimed to clarify the nature of the information that gives rise to improved base rate representations. Hence, in Experiment 1b participants were provided with a yoked description of sampling outcomes (e.g., out of 4 people observed, 1 person had cancer) to examine whether improved performance was a result of sequential sampling per se or simply the distributional information provided by the sample (cf. Rakow, Demes, & Newell, 2008).

Experiment 1a

This study examined the respective contribution of causal explanation of false positives and sampling experience to performance on the mammogram problem (Figure 1). Each factor was varied factorially and performance was assessed using both point estimate and forced choice methods. Based on the previous work of Krynski and Tenenbaum (2007) and Hogarth and Soyer (2011), we expected that providing causal information and relevant sampling experience would each lead to improved probability judgments. Based on our argument that each of these approaches addresses a different component of the task, we further predicted that these effects would be additive.

Method

Participants. One hundred undergraduate students (M_{AGE} = 20.1 years) participated for course credit. Equal numbers were randomly allocated to the four experimental conditions.

Design and Procedure. The experiment followed a 2 (False positive information: standard vs. causal) x 2 (Base rate presentation: description only vs. description + sampling) design with both factors manipulated between subjects.

All participants were presented with the mammogram problem shown in Figure 1 (cf. Krynski & Tenenbaum, 2007, Experiment 2). The problem was presented in either the standard or causal version, with each version administered to an equal number of participants. In both versions the Bayesian solution to the question about the likelihood of cancer given a positive mammogram was (approximately) 5%.

In all conditions the problem description (the text in normal font in Figure 1) was first presented on a computer screen. In the Description only condition, an open-ended question asking for an estimate of the likelihood of cancer in a woman with a positive mammogram appeared after 15s. As per Krynski and Tenenbaum (2007), the format of this estimate was a % chance of cancer between 0 and 100. Participants were invited to use an on-screen calculator to assist in solving the problem. After a likelihood estimate was entered, the cancer estimation question was repeated together with four alternative "answers that people commonly give to this question" (1%, 5%, 65%, 80%). Participants used a mouse to click on the option they thought was "closest to the correct answer".

Those in the sampling condition received an additional sampling phase between the problem description and the request for a likelihood estimate. In this phase they were told that to assist task completion they would be able to draw samples of women who had received a positive mammogram. Each time a participant clicked a "simulate" button they were told whether or not a sampled woman had cancer. In the standard condition the feedback for cancer-absent cases was "this woman does not have cancer". In the causal condition it was "this woman has a benign cyst". Samples were drawn randomly from a uniform distribution of 10 000 cases[1]. There was no limit on the number of samples that could be drawn. At any time during the sampling process participants could also click an on-screen button to view a running tally of a) samples with cancer; b) samples without cancer; and c) total samples viewed. To familiarize participants with the sampling tool, prior to commencing the main experiment they were shown the outcomes of 10 samples of tossing an unbiased coin. After the sampling phase those in the sampling condition received the same open-ended and multiple choice questions as the

[1] Specifically, each time the simulate button was clicked a random number between 1 and 10 000 was generated. If the number was less than 511 then the woman had cancer.

description only condition, but were not provided with an on-screen calculator. There was no time limit on any part of the procedure.

An on-screen version of the 4-item Berlin Numeracy Test (Cokely, Galesic, Schulz, Ghazal & Garcia-Retamero, 2012) was also administered. Numerical ability ($M_{\text{CORRECT}} = 2.44$) did not differ across experimental conditions ($p > .35$).

Results and Discussion

As a preliminary step we examined behavior in the sampling condition. The number of samples drawn ranged from 3 to 50 ($M_{\text{SAMPLES}} = 17.26$, SD = 12.22). A majority of participants experienced no positive cases of cancer (42%) or only one positive case (34%). The mean number of samples did not differ between the causal or standard versions of the sampling condition (p's > 0.5).

Intuitive probability – Open-ended estimates. Estimates of the likelihood of cancer were analyzed by computing the simple deviation of the estimate from the normative solution (5.1%, see Figure 2). To examine group differences in estimate accuracy, deviation scores were entered into a 2(description vs. sampling) x 2(standard vs. causal version) analysis of variance (ANOVA). Estimates in the sampling condition ($M_{DEVIATION} = 25.69$) were closer to the normative solution than those in the description only condition ($M_{DEVIATION} = 39.91$), $F(1, 96) = 4.78$, $p = .03$, Cohen's $d = 0.43$. There was a non-significant trend for estimates in the causal condition ($M_{DEVIATION} = 26.87$) to be closer to the normative solution than those in the standard condition ($M_{DEVIATION} = 38.74$), $F(1, 96) = 3.33$, $p = .07$, $d = 0.36$. There was no interaction between base rate presentation and causal factors, $p = .45$. [2]

As per Krynski and Tenenbaum (2007), we also tallied the frequency of estimates that could be classified as correct (estimates in the range 4%-6%) or as base rate neglect (estimates >= 65%). Binary logistic regression showed that "neglect" estimates were less common in sampling than description only (24% vs. 44% of responses in the respective conditions), Wald (1) = 4.36, $p = .04$, and less common in causal than the standard condition (24% vs. 44%), Wald (1) = 4.36, $p = .04$. However, the frequency of estimates classified as normatively correct (M = 12%) did not differ across conditions. No interactions between the sampling and causal factors were found (p's > .4).

Intuitive probability – Forced choice. These responses were classified as correct (a choice of 5%), base rate overuse (1%), or base rate neglect (a choice of 65 or 80%). The proportion of responses in each category within each condition is given in Figure 3. Logistic regression was again used to examine changes in the proportion of each type of response across conditions. Figure 3 shows that selection of the correct response was more common in the sampling than the description only condition, Wald (1) = 7.24, $p = .007$, and in the causal than the standard condition, Wald (1) = 8.96, $p = .003$. The interaction between these factors was not significant ($p > .35$). Choice of the base rate neglect options was less common in the causal than the standard condition, Wald (1) = 4.87, $p = .03$. These choices were unaffected by the sampling manipulation and there was no interaction with causal version (p's > .15). Neither manipulation affected selection of the base rate overuse option, (p's > .15).

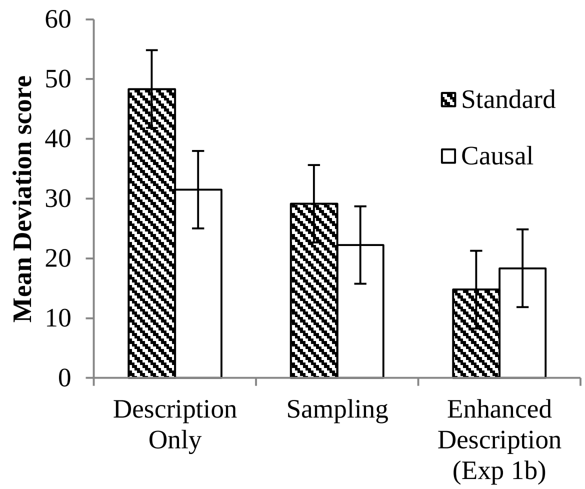
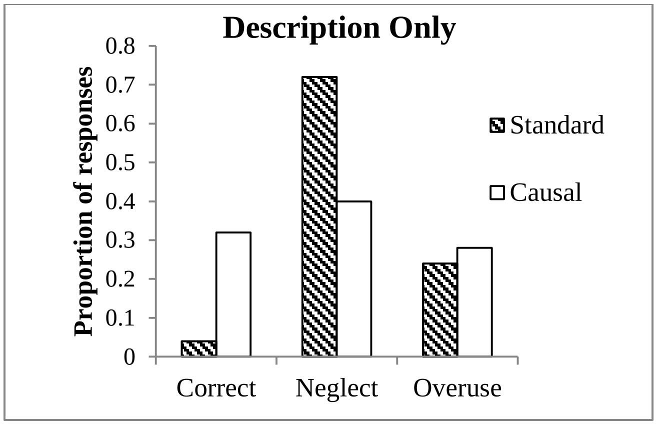


Figure 2. Deviation scores for probability estimates (with standard error bars).

Additional analyses. In the description only condition, accessing the on-screen calculator during testing was positively correlated with the likelihood of giving the correct estimate on the open-ended test, $r(49) = 0.28$, $p = .04$, and with selection of the correct alternative in forced choice, $r(49) = 0.35$, $p = .01$. In the sampling condition, no sampling statistics (number of samples drawn, number of cancer positive cases observed, proportion of cancer positive cases observed) were correlated with any performance measures (all p's > 0.1). However, the frequency with which the summary tally was accessed was positively correlated with the likelihood of providing a correct estimate, $r(49) = 0.32$, $p = .02$.

Summary. The accuracy of judgments of cancer probability was facilitated by an opportunity to sample instances with a positive mammogram and by causal explanation of false positives. Although correct probability estimates were rare, both causal and sampling manipulations led to a downward shift in estimates in the direction of the normative solution. Both manipulations increased choice of the correct estimate and decreased choice of the neglect option. Notably these effects were additive, supporting the view that the causal and sampling manipulations affect different components of intuitive probability judgment.

[2] These qualitative results remained unchanged when deviation scores in the sampling condition were recomputed against a normative solution that replaced the stated base rate of 1% with the base rate implied by the sample drawn by each participant (i.e. the observed proportion of positive cancer trials).



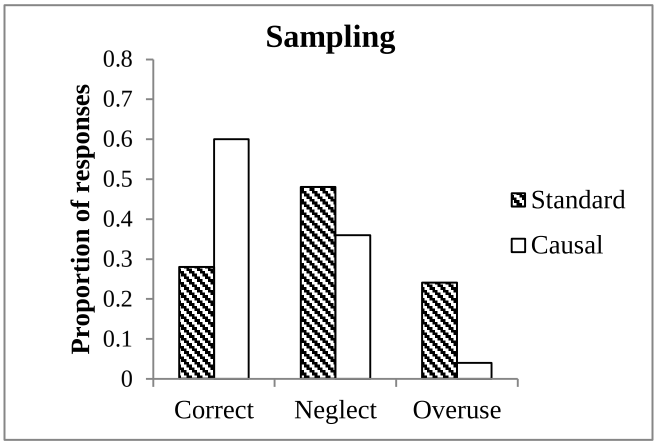


Figure 3. Experiment 1a. Proportion of forced choices.

Experiment 1b

The beneficial effects of sampling found in Experiment 1a and in past work (Hogarth & Soyer, 2011) could arise from a range of mechanisms. Hogarth and Soyer (2011) suggest that "across time, a person observes sequences of outcomes that can be used to infer the characteristics of the data generating process" (p. 435). However it is unclear whether sampling experience per se is critical here. Sequential sampling may be just one of many methods of obtaining information about the distribution of positive and negative cases. Other methods such as description of a frequency distribution (cf. Gigerenzer & Hoffrage, 1995) could convey the same information, and hence may also reduce base rate neglect. Some support for this view comes from the Experiment 1 finding that use of a summary tally was correlated with estimate accuracy.

Experiment 1b examined this possibility by presenting all participants with a summary tally of positive and negative cases of cancer from a sample of women with a positive mammogram. This 'enhanced description' presents the same base rate information that was present in the sampling condition of Experiment 1a, but without trial-by-trial sampling. To allow for close matching of the statistical information presented to participants, the sampling tallies used in this study were yoked to the outcomes of sequential sampling in Experiment 1a (see Rakow et al., 2008, for a related manipulation). If sampling experience is crucial for gaining a more accurate representation of the problem, then the Experiment 1a sampling condition should yield superior probability estimates to enhanced description. If the critical issue is the generation of a representative distribution of positive and negative cases, then enhanced description should do as well as sequential sampling. As in Experiment 1a, descriptions of the problem included a standard or causal explanation of the false positives.

Method

Participants

Fifty undergraduate students (M_{AGE} = 19.3 years) participated for course credit. Equal numbers were randomly allocated to causal and standard conditions.

Procedure

The general procedure was similar to the causal and standard description only conditions in Experiment 1a, with the important exception that all participants were given an on-screen tally of positive and negative cases of cancer from samples of women with a positive mammogram. Fifty tallies were generated based on sampling outcomes in the sampling condition of Experiment 1a. For example, if a participant in the earlier study drew 20 samples containing 1 cancer positive and 19 negative instances, then a tally containing the same information was constructed for an enhanced description participant. An on-screen calculator was available to assist in answering the problem.

Results and Discussion

Intuitive probability – Open-ended estimates. Estimation performance was again examined by calculating the deviation of estimates from the normative solution (see Figure 2). Accuracy as measured by deviation scores was not affected by causal explanation, $F(1, 49) = 0.2$, $p = .66$.

The more important issue was how estimation performance compared with the sampling and description conditions in Experiment 1a. Inspection of Figure 2 suggests that the pattern of deviation scores in enhanced description was more similar to the sampling than the description condition from the earlier study. These trends were examined using a cross-experimental task (description only, sequential sampling, enhanced description) x causal framing ANOVA. Planned comparisons compared performance in the enhanced description condition with the description only and sampling conditions respectively. The analysis confirmed that estimates in the enhanced description condition ($M_{DEVIATION}$ = 16.56) were more accurate than in the description only condition, $F(1, 144) = 14.04$, $p < .001$, $d = 0.74$, but did not differ from estimates in the sampling condition, $F(1, 144) = 2.24$, $p = .14$. No significant influence of causal framing was found. [3]

[3] These qualitative results remained unchanged when deviation scores were recalculated using individual cancer base rates implied by the sampling information instead of the stated rate of 1%.

Intuitive probability – Forced choice. Forced choice responses in the enhanced description group are given in Figure 4. Binary logistic regression found no significant differences between the enhanced description and sampling groups for any type of response, and no interactions with causal framing, *p*'s > .06. Correct responses were more common in enhanced description than in the description only conditions, Wald (1) = 9.66, *p* = .002, and neglect responses were less common, Wald (1) = 7.93, *p* = .005. Across the enhanced and description only conditions, causal framing led to a higher rate of correct responding than standard framing, Wald (1) = 4.06, *p* = .04, but this effect was stronger in the description only condition, Wald (1) = 4.87, *p* = .03. The enhanced description and description only groups did not differ in base rate overuse.

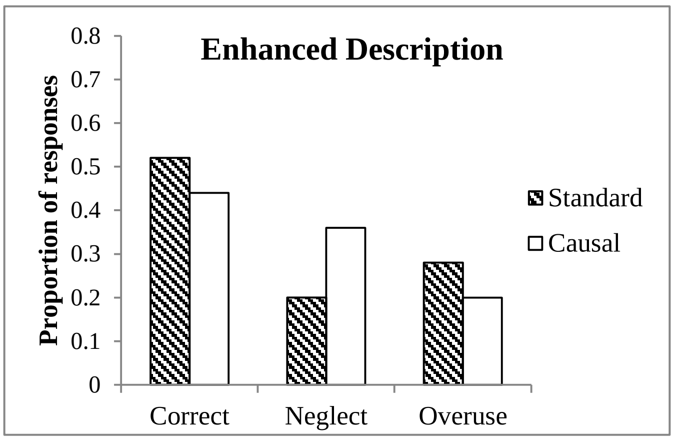


Figure 4. Experiment 1b. Proportion of forced choices.

Summary. This study examined whether sampling experience is necessary to reduce base rate neglect in intuitive probability. When people were given a description of the relevant sampling information they performed as well as those in the sampling condition of Experiment 1a, and better than those in description only. It appears that what is crucial is having relevant information about the distribution of positive and negative cases; this can be obtained through sampling or a description of the frequency distribution.

A puzzling finding was that causal framing, which had a positive effect on probability estimates in Experiment 1a, had little impact on enhanced description estimates. This may have been due to the accuracy of intuitive probability estimates in the standard version of enhanced description being higher than the standard conditions in the earlier study (see Figure 1). In other words, estimates may have been approaching ceiling in the enhanced description standard group, reducing the likelihood of finding further facilitation due to causal explanation.

General Discussion

These studies examined how providing sampling information about base rates and a causal explanation of false positives can improve intuitive probability judgments. Experiment 1a found that these manipulations led to a shift in probability judgments toward the normative response, and away from inflated estimates that would usually be classified as base rate neglect. Moreover, each manipulation increased choice of the normative solution.

The results replicate and extend previous findings of a positive effect of causal framing (Krynski & Tenenbaum, 2007) and sampling experience (Hogarth & Soyer, 2011) on intuitive probability judgment. Experiment 1a, however, was the first study to combine these manipulations. An important result was that effects of sampling and causal explanation were additive. This is consistent with the view that these manipulations address different aspects of probability judgment. The sampling and enhanced description manipulations helped establish greater sensitivity to the base rate. The causal manipulation facilitated the incorporation of false positives into the problem solution.

Experiment 1b clarified the role of sampling experience in improving probability judgment. Contrary to the views of Hogarth and Soyer (2011), we found that sequential sampling was not essential for reducing base rate neglect. A similar level of facilitation was obtained when the relevant statistical information was conveyed by a description of sampling outcomes. This is consistent with other findings in the judgment and decision-making literature which show that detailed descriptions of statistical information can produce equivalent effects to sequential sampling (e.g., Rakow et al., 2008).

The causal facilitation effects in these studies are consistent with the broader perspective on probability judgments outlined by Krynski and Tenenbaum (2007). This "causal Bayesian" view suggests that encoding the relevant statistics in an intuitive probability problem will not lead to accurate judgments, unless the statistics can be incorporated into a causal model of the problem. In the current studies both standard and causal groups were given equivalent statistical information about false positives but only the latter were supplied with a cause. According to Krynski and Tenenbaum (2007) this allows those in the causal condition to construct an intuitive model with two generative nodes that provide alternative explanations for positive mammograms. More broadly, these findings are consistent with the idea that people often fail to spontaneously consider alternative causes for probabilistic outcomes but can do so when prompted (e.g., Fernbach, Darlow, & Sloman, 2011).

It is notable that although both causal explanation and sampling shifted open-ended probability estimates in the right direction, neither manipulation increased the rate of normatively correct estimation. Similar results have been reported in previous work on base rate neglect. Krynski and Tenenbaum (2007) found that although causal explanation of false positives reduced base rate neglect, most participants in the causal condition still failed to produce a normative probability estimate. Likewise, although Gigerenzer and Hoffrage (1995) found that frequency formats for relevant statistics improved the accuracy of probability estimates, the majority of participants still gave normatively incorrect answers to the mammogram problem. This raises the question of what additional barriers need to be overcome to produce normative probability estimates.

The causal Bayesian perspective suggests one answer. According to this view the solution of probability problems proceeds in three stages. The first involves constructing an intuitive causal model of the problem. The second involves encoding the relevant statistics and mapping these onto the various nodes of the causal model. The third stage uses Bayesian inference to update beliefs in the light of the observed statistics. Arguably, the causal and sampling manipulations in the current studies impacted on the first two stages. The finding that a majority still do not produce normative estimates suggests that people may need further assistance with the final stage of implementing Bayes' rule (cf. Sedlmeier & Gigerenzer, 2001).

A final caveat is that although an opportunity to draw samples and description of sampling outcomes facilitated performance, sampling should not be regarded as a panacea for the problem of representing base rates. It is important to note that in the current studies and in Hogarth and Soyer (2011), samples were conditionalized on a woman having a positive mammogram. This ensured that with sufficient draws, a representative base rate was observed. However, samples outside the laboratory are not always constrained in this way. Sampling based on incorrect conditionalization (e.g., drawing samples of women with cancer and seeing whether they have a positive mammogram) can actually lead to more biased intuitive probability estimates (e.g., Fiedler, Brinkmann, Betsch, & Wild, 2000).

These studies suggest that the causal Bayesian approach represents a useful framework for analyzing the sub-components of intuitive probability problems, and intervening on these components to improve judgment accuracy. Our findings show that using experienced or described samples can reduce base rate neglect, and that supplying a cause for false positives increases the likelihood that these will be considered in probability judgments.

Acknowledgments

This work was supported by Australian Research Council Discovery Grant DP120100266 to the first two authors and a Future Fellowship FT110100151 to the second author. We thank Chris Moore and Ann Martin for help with programming and running the experiments.

References

Barbey, A. K., & Sloman, S. A. (2007). Base rate respect: From ecological rationality to dual processes. *Behavioral and Brain Sciences*, *30*, 241-297.

Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring risk literacy: The Berlin numeracy test. *Judgment and Decision Making, 7*, 25-47.

Eddy, D. M. (1982). Probabilistic reasoning in clinical medicine. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 249–267). Cambridge, UK: CUP.

Evans, J. S. B. T., Handley, S. J., Perham, N., Over, D. E., & Thompson, V. A. (2000). Frequency versus probability formats in statistical word problems. *Cognition*, *77*, 197-213.

Fernbach, P. M., Darlow, A., & Sloman, S. A. (2011). Asymmetries in predictive and diagnostic reasoning. *Journal of Experimental Psychology: General, 140*, 168-185.

Fiedler, K., Brinkmann, B., Betsch, T., & Wild, B. (2000). A sampling approach to biases in conditional probability judgments: Beyond base rate neglect and statistical format. *Journal of Experimental Psychology: General, 129*, 399-418.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review, 102*, 684-704.

Hogarth, R. M., & Soyer, E. (2011). Sequentially simulated outcomes: Kind experience versus nontransparent description. *Journal of Experimental Psychology: General, 140*, 434-463.

Krynski, T. R., & Tenenbaum, J. B. (2007). The role of causality in judgment under uncertainty. *Journal of Experimental Psychology: General, 136*, 430-450.

Lejarraga, T. (2010). When experience is better than description: Time delays and complexity. *Journal of Behavioral Decision Making, 23*,100–116.

Rakow, T. R., Demes, K., & Newell, B. R. (2008). Biased samples not mode of presentation: Re-examining the apparent underweighting of rare events in experience-based choice. *Organizational Behavior and Human Decision Processes, 106*, 168-179.

Sedlmeier, P. (1999). *Improving statistical reasoning: Theoretical models and practical applications.* Mahwah, NJ: Erlbaum.

Sedlmeier, P., & Gigerenzer, G. (2001). Teaching Bayesian reasoning in less than two hours. *Journal of Experimental Psychology: General, 130*, 380-400.

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

Appendix

The normative probability of cancer (C) given a positive mammogram (M) is given by:

$$p(C|M) = \frac{p(C)*p(M|C)}{p(C)*p(M|C) + p(\neg C)*p(M|\neg C)}$$

$$= \frac{0.01*0.80}{0.01*0.80 + 0.99*0.15}$$

$$\approx 0.051$$