

## Learning to commit or avoid the base-rate error

Adam S. Goodie & Edmund Fantino

Department of Psychology, University of California, San Diego, La Jolla, California 92093, USA

WHEN predicting an event, people neglect overall frequencies (base rates) of various possibilities<sup>1</sup>. We have previously shown<sup>2</sup> that this base-rate occurs not only with word problems, but also in a procedure with repeated trials: a sample cue followed by two choice options, one of which the subject must choose, with feedback regarding the correctness of the choice. Perhaps this base-rate error depends on people's histories of matching physically similar items. In support of this suggestion, we begin by showing that the base-rate error is eliminated with physically unrelated items. We then show that when the relation between items is again arbitrary, but is a relation that subjects already know, the error reappears. Finally we show that teaching subjects new arbitrary relations reintroduces the error in later testing. These experiments demonstrate a fundamental base-rate error dependent on learned relationships: without interference from pre-existing associations between cues and options, predictions may be made more optimally.

When assessing the probability of a future event, people often ignore background information in favour of case-specific evidence, an effect called the base-rate error<sup>1,3</sup>. Consider this problem<sup>4</sup>:

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

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

(b) a witness identified the cab as Blue... The witness correctly identifie[s] each one of the two colors 80% of the time and fail[s] 20% of the time. What is the probability that the cab involved in the accident was Blue rather than Green?

It is more common for the witness to see a green taxicab and mistakenly call it blue ( $0.85 \times 0.20 = 17\%$  of all cases) than for the witness to see a blue taxi and label it correctly

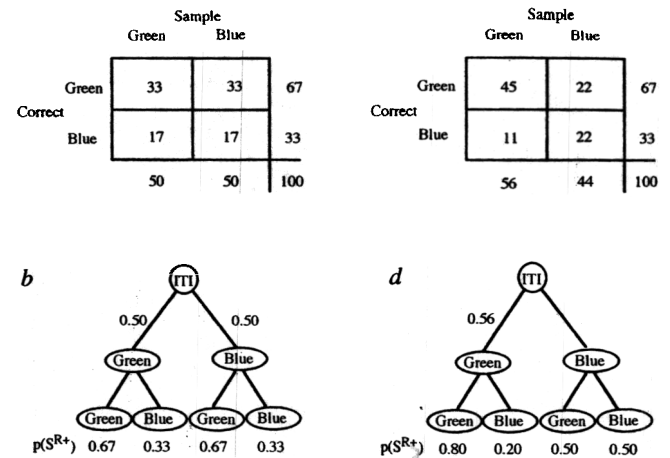


FIG. 1 Procedure used in prior experiments<sup>2</sup>. *a*, An incidence table derived from a taxi problem with a base rate of 67% green and 33% blue, a cue accuracy of 50%, and a sample of 100 total instances. Green is the correct response in 67% of cases, and 50% of those (33) are correctly marked by a green cue, while the other 50% (33) are marked by a blue cue. Of the 33% of all instances where blue is the correct response, 50% (17) are correctly marked by a blue cue, while the other 50% (17) are marked by a green cue. *b*, The conversion of this table to a cue-matching experimental design. Because the green cue (G) appears 50 (33 + 17) times out of every 100, the green cue is presented 50% of the time, and the blue cue (B) appears the remaining 50%, after an inter-trial interval (ITI). When green has appeared, green is correct 33/50 = 67% of the time, and blue is correct the other 33% of the time; these are indicated by their probabilities of reinforcement ( $p(S^{R+})$ ). Hence, after a green cue, choosing green is reinforced 67% of the time, and choosing blue is reinforced 33% of the time. By identical reasoning, when a blue cue appeared, choosing green is reinforced 67% of the time, and choosing blue is reinforced 33% of the time. Note that the 50%-accurate cue is irrelevant to the contingencies of reinforcement. *c*, *d*, The same functions as *a* and *b*, respectively, for the condition where the base rate was still 67%, but where the cue was 67% accurate. Procedural details are as previously described<sup>2</sup>.

( $0.15 \times 0.80 = 12\%$  of all cases). If the witness reports seeing a blue taxi, the probability that the cab actually was blue is  $0.12 / (0.17 + 0.12) = 41\%$ . This is one application of Bayes's theorem, which is often held as an optimal model of behaviour. Subjects responding to the 'taxicab problem', however, ignore base rates (how many green and blue taxis operate in the city) and rely on the reliability of what the witness says.

We have found the base-rate error in a different sort of setting<sup>2</sup>. On each of many trials, subjects predicted whether the 'correct' answer to a problem would be green or blue. In 67% of cases, the correct answer was green, with blue being correct the other 33% of the time. On each trial a cue was presented that anticipated the correct answer with either 50% or 67% accuracy. We measured how often subjects matched the cue, guessing green if a green cue had appeared, or blue if a blue one appeared (see Fig. 1).

Two particular data patterns would represent a base-rate error. One is roughly equal matching to the blue and green cues, indicating that subjects treat both cues alike despite differences in their base rates. This failure to distinguish between cue types is a characteristic feature of the base-rate error. The other is predominant matching of the blue cue, which under some conditions is distinctly suboptimal. The results of previous experiments appear in Fig. 2, and show both tell-tale patterns: subjects matched blue more than half the time, and the average difference between matching green and blue was only 7%.

We propose that the base-rate error is a learned phenomenon, and present three supporting experiments. The first two differ from our previous design only in the physical nature of the cues. Experiment 1 used cues that bore no natural or previously learned relation to the choice options. If the base-rate error results from

subjects' histories of matching items that are similar to one another, then this should reduce the base-rate error. The green and blue cues were replaced respectively by vertical and horizontal lines, and the base-rate error disappeared entirely (see Fig. 3a). The 50% condition, in which subjects had previously lost points by preferring blue following a blue cue, now showed a marked preference for green following the horizontal cue. Even in the 67% group, where strategy following a 'blue' cue was irrelevant, the proportion of times it was matched fell considerably, removing the characteristic flatness of slope. The difference between matching 'green' and matching 'blue', averaged across groups, increased from 7% to 28%.

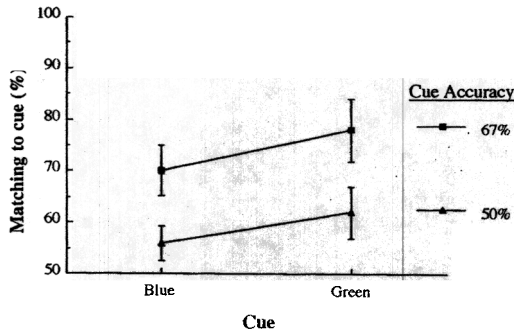


FIG. 2 Data from prior experiment<sup>2</sup>. The base-rate error is represented by two graphical features: the high matching of the blue cue in both groups,

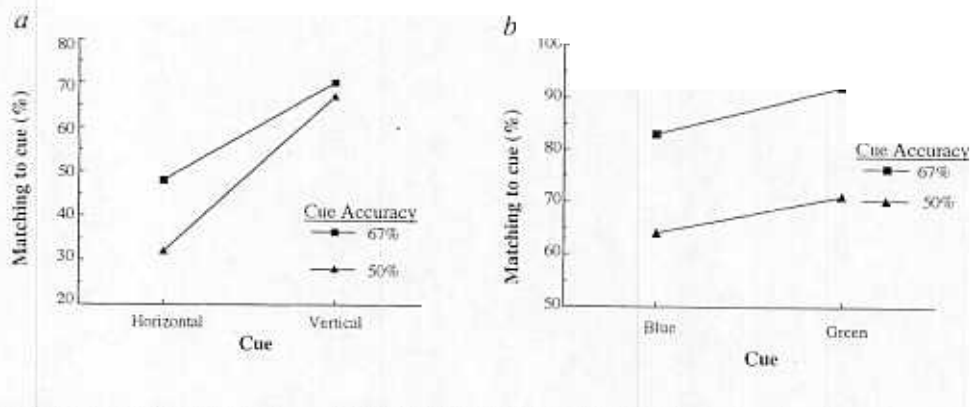
and the relatively flat slope of each line. The high matching of blue is reminiscent of the 'illusory correlation' effect<sup>8,9</sup>, but the flatness of slope is unique to a base-rate error. Perfectly flat lines would represent a perfect base-rate error, but this has not been attained, as subjects responded significantly differently to the green and blue cues ( $F(1, 18) = 6.78$ ,  $P < 0.05$ ). In this experiment, choosing the correct option earned the subject one 'point'. In another experiment, each correct choice earned the subject US\$0.10, but this incentive had no effect on performance. Note that, except for the case where each option is correct 50% of the time, the optimal strategy is to select the more frequently correct option exclusively, and that any deviation from this is an error. It is known that people match their response allocation to the probabilities of being correct<sup>10</sup>. If subjects did not commit the base-rate error, they would still be expected to make an error, namely probability matching (as in Fig. 3a). By committing the base-rate error, however, subjects failed to achieve even the rate of reinforcement expected from probability matching. Interestingly, both these errors might be unique to humans: pigeons maximize rather than match probabilities<sup>11</sup>, and they are properly sensitive to base rates<sup>12</sup>.

Experiment 2 used cues that again bore no natural relation but have a learned relation to the choice options. We used the words 'green' and 'blue' as we had once used the colours green and blue. Given the presumed history of matching the word 'green' to green things and 'blue' to blue things, we predicted that the base-rate error would reappear; it did (see Fig. 3b) as strongly as with identical cues. Subjects matched the blue cue a great deal in both conditions, and the average difference between matching 'green' and matching 'blue' was only 8%.

In Experiment 2 we presumed that our subjects had a particular learning history. In Experiment 3, we gave them such a history: perfect correspondence between a horizontal cue and blue being

and the relatively flat slope of each line. The high matching of blue is reminiscent of the 'illusory correlation' effect<sup>8,9</sup>, but the flatness of slope is unique to a base-rate error. Perfectly flat lines would represent a perfect base-rate error, but this has not been attained, as subjects responded significantly differently to the green and blue cues ( $F(1, 18) = 6.78$ ,  $P < 0.05$ ). In this experiment, choosing the correct option earned the subject one 'point'. In another experiment, each correct choice earned the subject US\$0.10, but this incentive had no effect on performance. Note that, except for the case where each option is correct 50% of the time, the optimal strategy is to select the more frequently correct option exclusively, and that any deviation from this is an error. It is known that people match their response allocation to the probabilities of being correct<sup>10</sup>. If subjects did not commit the base-rate error, they would still be expected to make an error, namely probability matching (as in Fig. 3a). By committing the base-rate error, however, subjects failed to achieve even the rate of reinforcement expected from probability matching. Interestingly, both these errors might be unique to humans: pigeons maximize rather than match probabilities<sup>11</sup>, and they are properly sensitive to base rates<sup>12</sup>.

FIG. 3 Data from Experiments 1 and 2. a, The base-rate error disappeared in Experiment 1: matching of the 'blue' cue was much lower than it has been in previous work<sup>2</sup>, and the slopes of lines increased markedly, being significantly greater than zero ( $F(1, 15) = 28.09$ ,  $P < 0.05$ ). (The 'blue' cue was in fact not blue but a white horizontal line. We continue to call this matching 'blue' to clarify the relationship of this experiment to previous ones<sup>2</sup> and to Experiments 2 and 3: horizontal lines replace blue cues here, and vertical lines replace green cues. In all experiments the choice options were blue and green patches, presented side-by-side on a computer monitor.) The increase in slope is partly obscured because the ordinate axis is altered to accommodate the increased range of values. b, The base-rate error reappeared in Experiment



2. Matching of the 'blue' cue returned to levels at or above those observed in previous studies<sup>2</sup>, and slopes returned to near-zero, although they were once again significant ( $F(1, 10) = 6.28$ ,  $P < 0.05$ ).

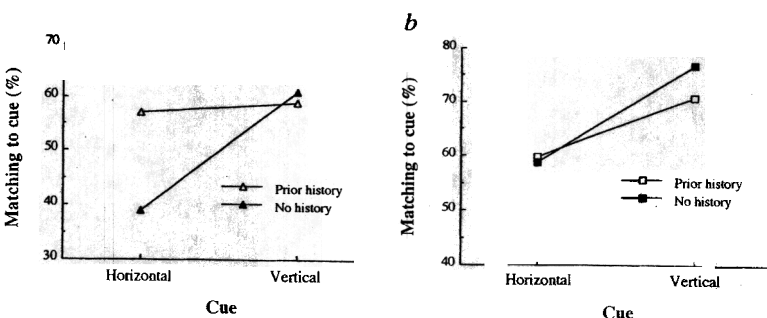


FIG. 4 Data from Experiment 3. Subjects in the 'Prior history' condition were first exposed to 200 trials in which green was always correct following a vertical cue, and blue was always correct following a horizontal cue. For these subjects, a horizontal line was thus established as equivalent to a blue cue, and so choosing blue following a horizontal line continues to be called matching 'blue'. Those in the 'No history' condition saw only the 200 trials of the test phase, which was identical to Experiment 1. Subjects were divided into two levels of cue accuracy, 50% and 67%, depicted in a and b, respectively. The 'Prior history' groups show substantially flatter slopes, resulting in cross-over effects at both levels of cue accuracy, which are statistically significant ( $t(22) = 2.04$ ,  $P < 0.05$ ).

correct, and between a vertical cue and green being correct. This induced our subjects to a greater flatness of slope than those without prior training, a distinct trend towards the base-rate error (see Fig. 4). The average difference between matching 'green' and matching 'blue' was 20% for subjects for no prior training, but only 6% for those with prior training.

Base-rate neglect has previously been obtained<sup>5-7</sup> in experimental settings. However, this has generally involved cues and choices with some prior relationship, such as symptom and disease, but not in as dependable a relationship as identity. By using elements with either no prior relation or a lock step relation of identity, we have isolated the effects of what is previously known (Experiments 2 and 3) from that which is learned during an experimental session (Experiments 1 and 3). These experiments demonstrate a base-rate error that depends on learned probabilistic relationships: pre-existing associations between cue and outcome can prevent learning from experience. □

Received 28 August 1995; accepted 8 January 1996.

1. Bar-Hillel, M. *Acta psychol.* **44**, 211-233 (1980).
2. Goodie, A. S. & Fantino, E. *Psychol. Sci.* **6**, 101-106 (1995).
3. Kahneman, D. & Tversky, A. *Psychol. Rev.* **80**, 237-251 (1973).
4. Tversky, A. & Kahneman, D. in *Judgment under Uncertainty: Heuristics and Biases* (eds Kahneman, D., Slovic, P. & Tversky, A.) 153-160 (Cambridge Univ. Press, 1982).
5. Gluck, M. A. & Bower, G. H. *J. exp. Psychol., gen.* **117**, 227-247 (1988).
6. Medin, D. L. & Edelson, S. M. *J. exp. Psychol., gen.* **117**, 63-85 (1988).
7. Nosofsky, R. M., Kruschke, J. K. & McKinley, S. C. *J. exp. Psychol., Learn. Memory Cogn.* **18**, 211-233 (1992).
8. Chapman, L. J. & Chapman, J. P. *J. abnorm. Psychol.* **72**, 193-204 (1967).
9. Chapman, L. J. & Chapman, J. P. *J. abnorm. Psychol.* **74**, 271-280 (1969).
10. Myers, J. L. in *Handbook of Learning and Cognitive Processes* Vol. 3 (ed. Estes, W. K.) 171-205 (Erlbaum, Hillsdale, NJ, 1976).
11. Herrnstein, R. J. & Loveland, D. H. *J. exp. Analysis Behav.* **24**, 107-116 (1975).
12. Hart, J. A. & Fantino, E. *J. exp. Analysis Behav.* (in the press).

ACKNOWLEDGEMENTS. We thank J. Otsuka for assistance in collecting data. This work was supported by a grant from the NSF.