The Quarterly Journal of Experimental Psychology
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t716100704

The effectiveness of feedback in multiple-cue probability learning

Ben R. Newell [a]; Nicola J. Weston [a]; Richard J. Tunney [a]; David R. Shanks [a]
[a] University College London, London, UK

First Published on: 17 October 2008

To cite this Article Newell, Ben R., Weston, Nicola J., Tunney, Richard J. and Shanks, David R.(2008)'The effectiveness of feedback in multiple-cue probability learning',The Quarterly Journal of Experimental Psychology,62:5,890 — 908
To link to this Article: DOI: 10.1080/17470210802351411
URL: http://dx.doi.org/10.1080/17470210802351411

THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY
2009, 62 (5), 890–908

Psychology Press
Taylor & Francis Group

The effectiveness of feedback in multiple-cue probability learning

Ben R. Newell, Nicola J. Weston, Richard J. Tunney, and David R. Shanks
University College London, London, UK

How effective are different types of feedback in helping us to learn multiple contingencies? This article attempts to resolve a paradox whereby, in comparison to simple outcome feedback, additional feedback either fails to enhance or is actually detrimental to performance in nonmetric multiple-cue probability learning (MCPL), while in contrast the majority of studies of metric MCPL reveal improvements at least with some forms of feedback. In three experiments we demonstrate that if feedback assists participants to infer cue polarity then it can in fact be effective in nonmetric MCPL. Participants appeared to use cue polarity information to adopt a linear judgement strategy, even though the environment was nonlinear. The results reconcile the paradoxical contrast between metric and nonmetric MCPL and support previous findings of people's tendency to assume linearity and additivity in probabilistic cue learning.

Keywords: Feedback; Judgement; Categorisation; Multiple-cue probability learning.

Multiple-cue probability learning (MCPL) has proved a fertile domain for developing and testing theories about the role of feedback in judgement and decision making (see Juslin, Jones, Olsson, & Winman, 2003a; Newell, Lagnado, & Shanks, 2007, for recent applications, and Balzer, Doherty, & O'Connor, 1989; Brehmer, 1980; Klayman, 1988, for reviews). MCPL at its most fundamental involves learning to predict an outcome on the basis of the values of a number of cues in situations where the relation between the outcome (criterion) and the cues is probabilistic.

Brehmer (1979, 1994) has argued that in such situations there are three components involved in learning how to make accurate judgements. First, the judge needs to learn about the functional relation between each cue and the to-be-predicted criterion. Second, the judge needs to learn the optimal relative weighting to ascribe to different cues. Finally, if multiple cues are involved, the judge has to consider relations among the cues and determine the best way to integrate them. Thus for feedback to be effective, it must, in principle, be presented in such a way as to be informative about one or more of these components.

Correspondence should be addressed to David R. Shanks, Division of Psychology and Language Sciences, University College London, Gower St, London WC1E 6BT, UK. E-mail: d.shanks@ucl.ac.uk

The support of the Economic and Social Research Council (ESRC) and The Leverhulme Trust is gratefully acknowledged. The work was part of the programme of the ESRC Research Centre for Economic Learning and Social Evolution, University College London. We thank Joshua Klayman, Nigel Harvey, David Lagnado, and Henrik Olsson for valuable feedback on an earlier version of this article.

http://www.psypress.com/qjep DOI:10.1080/17470210802351411

Numerous findings support this assertion. When feedback is meagre, for example if only outcome feedback is provided (presentation of the actual criterion value after an estimate has been made), improvements are typically only seen when the environment is very simple (two or three cues that are positively and linearly related to the criterion) and when feedback is combined with a long series of trials (Balzer et al., 1989; Brehmer, 1980; Hammond, 1971; Hammond & Boyle, 1971; Klayman, 1988; Todd & Hammond, 1965, Shanks, Tunney, & McCarthy, 2002). In contrast, provision of cognitive feedback—especially task information, which refers to the "relations between the cues and the criterion, information about the criterion or the cues themselves, or both" (Balzer et al., 1989, p. 412)—generally leads to improved performance even in more complex environments (e.g., Adelman, 1981; Balzer et al., 1989; Schmitt, Coyle, & King, 1976; Schmitt, Coyle, & Saari, 1977). These improvements, relative to outcome feedback, arise presumably because the extra information helps participants "solve" one or more of the three components identified by Brehmer (1979).

However, this conclusion only applies to *metric* MCPL tasks in which participants learn to predict the value of a continuous criterion (e.g., a person's weight) using cue values that are continuous (e.g., the person's age, height, and salary). Findings from studies of *nonmetric* MCPL seem far less conclusive and present a hurdle to theory development. In these, participants learn to make a binary decision (e.g., deciding whether a patient has a disease) using cue values that are discrete (e.g., the presence or absence of different symptoms). Studies of this type (Castellan, 1973, 1974; Castellan & Edgell, 1973; Castellan & Swaine, 1977) have found that no type of feedback enhanced performance over and above simple outcome feedback and indeed that most types resulted in a deterioration in performance, despite the fact that a variety of forms of cognitive feedback were studied. This work in nonmetric MCPL formed an important cornerstone of Kluger and DeNisi's (1996) influential theory of feedback and performance, which attempted to explain the factors determining when feedback is and is not effective.

Why should it be the case that learning and accuracy of judgement is assisted by information over and above simple outcome feedback in metric MCPL but not in nonmetric MCPL? To answer this question we first examine previous experiments in more detail in an attempt to identify any potential shortcomings in the nature of the feedback provided. We then present three experiments that demonstrate that additional feedback (specifically, *probability* feedback, see below) can be effective in nonmetric MCPL provided it is targeted to help solve the problem faced by the judge. These findings are important in both reconciling a theoretical paradox raised by previous research and in shedding light on the nature of the processes by which multiple-cue learning is reinforced.

Castellan (1974) presented participants with a two-cue, binary outcome environment. The two cue dimensions consisted of shape (triangle or square) and direction—whether the shape was built of horizontal or vertical lines. The two outcomes consisted of the symbols " > " or " < ", which were named "right arrow" and "left arrow", respectively. The probability of each outcome occurring on a given trial was .50, and the two cues had validities of .0 and .60 in one condition and .40 and .60 in another. In addition to simple outcome feedback (i.e., information about what outcome actually occurred on the trial), participants were given one of four types of additional feedback: simple percentage correct (e.g., "Over the last X trials you have made XX% correct responses"), cue-outcome validity coefficients (e.g., "Over the last X trials the cue weights have been: Shape YY, Direction ZZ"), cue-response utilization coefficients (e.g., "Over the last X trials you have been weighting the cues as follows: Shape YY, Direction ZZ"), or a combination of cue-outcome and cue-response information.

Before beginning the experiment participants were told that a cue weight of 0 meant that the cue was useless in making a judgement, a weight of 1 meant it would lead to correct predictions on every trial, and that intermediate values

indicated intermediate levels of usefulness. Similar instructions were given about the meaning of the cue-utilization weights (i.e., 0 meant it was not relied on at all, 1 that it was relied on exclusively).

Relative to outcome feedback, such cognitive feedback typically yields improvements in metric MCPL, but in Castellan's (1974) studies it produced decrements. Why might this have happened? Presumably, feedback about the cumulative percentage correct may have helped participants to keep track of their performance without having to memorize a series of individual trial-by-trial outcomes. This reduction in memory load might have allowed more resources to be applied to solving the task, but percentage correct does not contribute information directly relevant to the function, weighting, and combination problems facing the judge.

Presenting cue-utilization information alone does not help solve any of these problems either. Being told how much weight one is assigning to a cue does not provide any help if one does not know what weight should be assigned to the cue. It may increase the accuracy of insight into the individual's judgement process, but it does not necessarily improve knowledge of the ecology. Presentation of this information might have encouraged participants to experiment with different cue weightings, perhaps leading to a decrement in performance, especially when one cue was irrelevant (validity of 0) as it was in one condition (Castellan, 1974).

Cue-outcome validity information would appear to be a better candidate for improving knowledge of the ecology and thus accuracy. However, this information was presented on an individual cue basis despite referring to stimuli that were presented as configural wholes (e.g., a triangle made up of vertical lines). Evidence from MCPL studies using similar stimuli to those in the Castellan (1974) study suggests that the format in which information is displayed has a large effect on participants' processing of that information (Edgell & Morrissey, 1992). Although the stimuli varied along psychologically separable dimensions, if participants treated the stimulus as a whole item, then assimilating feedback about separate features of the display (one of which provided no useful information in one condition) may well have proved an extremely difficult task. Furthermore, the environment used by Castellan was configural—that is, cue patterns as well as individual cues contained predictive information. Thus presenting feedback on a cue-by-cue basis when participants needed to learn about the validity of whole patterns would again make their task difficult.

Combining cue-outcome and cue-response utilization coefficients arguably provides a way in which to compare the weighting of cues and the weights that should be assigned to cues. However, given the volume of information contained in this combined form of feedback, it is quite possible that participants were simply "overwhelmed with data and, [are] thus unable to cope with it in an adequate or appropriate manner" (Castellan, 1974, p. 62).

Moreover, it is possible that the highly abstract nature of the task and the environment contributed to the failure of feedback to improve performance. Many studies indicate that increasing the "meaningfulness" of an environment leads to improvements in performance. For example, Muchinsky and Dudycha (1975) showed that participants' performance in a metric MCPL task was significantly superior when cue names were changed from the abstract "Cue 1" and "Cue 2" to meaningful labels such as "average monthly debt" and "average number of creditors". Further improvements are seen if the concrete labels used for cues are congruent with prior hypotheses based on real-world knowledge (e.g., Sniezek, 1986). Even if labels are concrete but do not connect with prior hypotheses, performance is still better than it is with abstract labels (e.g., Adelman, 1981).

In summary, it appears that a number of factors may have contributed to the inability of participants in Castellan's (1974) experiments to relate feedback to the task at hand—or, more specifically, to the task of solving the cue function, weight, and combination problems. It might be the case that variations in the form of feedback simply make no difference in nonmetric MCPL, perhaps due to some intrinsic learning limitation in probabilistic environments (cf. Brehmer, 1980).

Alternatively, the information inherent in the task and provided by the feedback may have been ineffective because it was not suitably targeted at solving the problem faced by the participants (cf. Harries & Harvey, 2000; Todd & Hammond, 1965). As Castellan (1974) noted, the ineffectiveness of feedback in the studies referred to above does not allow us to distinguish between these possibilities.

Experiment 1 readdressed the question of whether richer feedback can be effective in nonmetric MCPL by using a design that overcomes a number of the possible limitations in the Castellan (1974) experiments. Participants were given a task in which the context was meaningful (predicting a change in share price), the additional form of feedback was in a format that was directly relevant to the prescribed task (information about the probability of a change in share price), and feedback pertained to the aspect of the stimulus most useful for performing the task (patterns as opposed to individual cues). We predicted that these changes would lead to the standard metric MCPL finding of an advantage for elaborated over simple outcome feedback, thus allowing progress to be made in developing a combined theoretical framework for both metric and nonmetric MCPL.

Overview of experimental task

The task used in the experiments was a "share market" task in which participants were required to predict whether the share price of a fictional company would increase or decrease based on four pieces of information about the company (e.g., location of company headquarters, employee turnover).

The task structure is summarized in Table 1. The four binary cues C_1, C_2, C_3, and C_4 take on values of 1 or 0, giving rise to 16 possible cue combinations, labelled Patterns 1 to 16. Each distinct pattern was assigned a value indicating the probability that the share price would increase given the pattern. The probability values rose in increments of 1/15 for each pattern such that Pattern 1 had a zero probability of the share price increasing, and Pattern 16 had a probability of 1. The outcome (share price increase or decrease) on each trial was determined by comparing the probability assigned to each pattern with a random number between 0 and 1 generated by the computer. If the random number was greater than the assigned value the share price decreased, and if it was smaller than the assigned value the share price increased.

Table 1. *Stimulus patterns and associated probability of share price increase*

Pattern no.	*Stimulus pattern ($C_1C_2C_3C_4$)*	*Probability of a share price increase*
1	0 0 0 0	0
2	0 0 0 1	.06
3	0 0 1 0	.13
4	0 1 0 0	.20
5	1 0 0 0	.26
6	1 0 0 1	.33
7	1 0 1 0	.40
8	1 1 0 0	.46
9	1 1 0 1	.53
10	1 1 1 0	.60
11	0 0 1 1	.66
12	1 0 1 1	.73
13	0 1 1 0	.80
14	0 1 0 1	.86
15	0 1 1 1	.93
16	1 1 1 1	1

This assignment of probabilities to patterns ensured a nonlinear environment. Although Patterns 1 to 10 are predominantly linear and additive (the addition of a cue value 1 tends to lead to higher probability of share price increase), in Patterns 11–15 the structure is nonadditive—it is no longer the case that a greater number of 1s in the pattern indicates a higher probability of increase. Consider Pattern 14 (0101) and Pattern 9 (1101), in which the absence of a 1 in the first position increases the associated probability from .53 to .86. This nonlinearity in the structure means that in order to perform optimally, participants need to learn about whole patterns (i.e., the value of all four cues and the associated outcome) rather than the contribution of individual cues. Of course, whether or not participants indeed learn such patterns, or instead adopt a simpler linear strategy, is an empirical question.

EXPERIMENT 1

The key manipulation in Experiment 1 was the nature of feedback provided to participants. Half the participants received simple outcome feedback—that is "Share Price Increased (Decreased)" and "Correct (Incorrect)"—whereas the other half saw the additional information: "There was a X% probability that this share price would increase given this pattern"—where X was replaced by the appropriate value from the third column of Table 1. We predicted that this probability feedback would be effective in improving performance, because it provided information directly relevant to solving the task, gave information about the environment, and related to the pattern as an entirety rather than to individual cues in a task with meaningful cue labels. The experiment was divided into a training phase and a test phase; during the latter no feedback of any kind was given to examine whether any advantage acquired from the probability feedback during training persisted when the feedback was removed.

Method

Participants

A total of 24 members of the University College London community took part in the experiment. Eleven were male and 13 female, with a mean age of 23 years (range 18 to 33, $SD = 4.26$). Participants were randomly assigned to either the probability and outcome feedback (PFB) group or the outcome feedback only (OFB) group, to give a total of 12 participants in each.

Procedure

Participants were told that they would play the role of a stockbroker and would be asked to predict on a trial-by-trial basis whether the share price of a fictional company would increase or decrease (N.B., the section in italics was included for the PFB group but not for OFB group):

> Thank you for agreeing to take part in this experiment which takes less than an hour. The experiment examines how difficult it is for people to learn to make accurate stock market predictions. You will initially be presented with information about 240 companies and be asked to predict whether the value of each company's shares will increase or decrease. To help you four pieces of information about each company will be presented. *When you have made your prediction you will be told the probability that the share value would increase, and whether the share value actually increased or decreased.*
>
> All you have to do is to try to learn which combinations of the four pieces of information tend to predict whether the share value is likely to increase or decrease so that you can make as many correct predictions as possible. You will have as much time as you wish to make each prediction. Although this task is different in many ways from what a stock broker actually has to do, it is realistic in that you will see quite a few borderline cases where the information is somewhat ambiguous about whether the company's share value will increase or not. Thus, even when you have become reasonably expert at making accurate predictions, there will still be cases where you make incorrect predictions. Despite this, we would like you to try to make AS MANY ACCURATE PREDICTIONS AS YOU POSSIBLY CAN.
>
> Two things will encourage you in this. First, your payment at the end of the experiment will be determined by how many correct predictions you make. Specifically, for each correct prediction you will receive 5 pence, and for each incorrect prediction you will lose 2 pence. This means that your maximum payment from 240 companies is £12.00. Secondly, the computer will stop after every 60 companies and tell you how well you did, and how well you did compared to our best and our worst participants so far.

On each trial participants were provided with four pieces of information about the companies, all of which had binary values (listed in brackets): *Where are the company headquarters? (1 = London, 0 = New York); Which index are the shares listed on? (1 = NASDAQ, 0 = FTSE); What is the rate of employee turnover? (1 = High, 0 = Low); Is it an established company? (1 = YES, 0 = NO).* For example Pattern 9 (1101) in Table 1 appeared to participants as: London, NASDAQ, Low, Yes. The total amount earned was displayed on the screen throughout the 240 training trials. The order of presentation of trials was random and different for every participant. Participants made their predictions by clicking on an on-screen INCREASE or DECREASE button. After each trial participants in the OFB group simply saw the words "Share price Increased (Decreased)" and "Correct (Incorrect)". Participants in the PFB group saw the additional information:

"There was a X% probability that this share price would increase given this pattern."

In addition to the trial-by-trial feedback, participants were provided with outcome feedback at the end of each block of 60 trials during training. This block outcome feedback told the participants how many correct decisions (out of 60) they made in the preceding block, and it was used both to provide a break in trials and in an attempt to motivate participants (see Shanks et al., 2002, for details). Blocked outcome feedback was not provided in Experiment 3.

The procedure at test was the same for both groups. All participants saw each of the 16 patterns four times in a different random order for a total of 64 presentations. On each trial they were required to make a prediction but no feedback of any kind was given, and the total earned display was removed from the screen. Participants were informed that the computer would keep track of performance and that they would be paid accordingly at the end of the experiment.

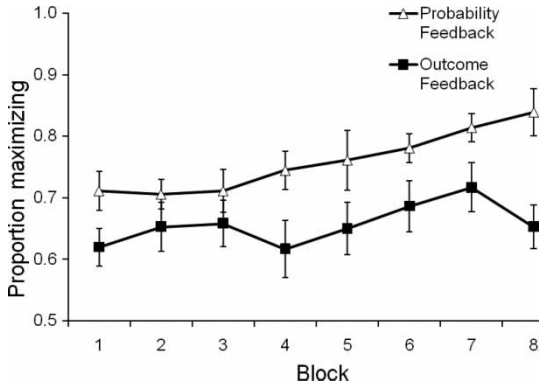

Figure 1. *Experiment 1: Maximizing performance for the probability feedback and outcome feedback groups during training.*

Table 2. *Mean proportion of maximizing responses in the probability feedback and outcome feedback groups across the 240 training trials and 64 test trials in Experiments 1 and 2*

	Group	
	PFB	*OFB*
Experiment 1		
Training	.76	.66
Test		
Block 1	.82	.65
Block 2	.83	.67
Experiment 2		
Training	.80	.77
Test		
Block 1	.83	.78
Block 2	.80	.77

Note: PFB = probability feedback. OFB = outcome feedback. Test trials: 64 (two blocks of 32).

Results

Performance during training was analysed using a maximum performance score. This "max" score is a measure of the proportion of trials on which participants made a response in the correct direction according to the probability of an increase associated with the pattern (i.e., if the probability of the share price increasing for that pattern is >.50 then the response should be "increase"; if it is <.50 it should be "decrease"). The training data are shown divided into eight 30-trial blocks in Figure 1. The mean performance across the 240 training trials for both the PFB and OFB groups are also displayed in the first row of Table 2.

Maximum performance during training was compared using a repeated measures analysis of variance with group as a between-subjects factor. Results showed a reliable main effect of block, $F(7, 154) = 3.31$, $p < .01$, indicating an improvement in performance throughout the eight blocks of trials; a main effect of group, $F(1, 22) = 8.22$, $p < .01$, indicating that participants in the PFB group performed more accurately than those in the OFB group; but no reliable interaction between the two variables, $F < 1.5$, $p = .34$.

Table 2 displays the mean maximum performance for both groups in the first and second 32 trials of the test phase (Blocks 1 and 2, respectively). Two things are apparent: The main effect of group indicates that the advantage conferred by the provision of probability feedback during training is maintained throughout the test, $F(1, 22) = 14.23$, $p < .01$. Secondly, the absence of an effect of block indicates that this advantage

does not diminish when feedback is removed, $F(1, 22) = 0.46$, $p > .50$.

Discussion

The simple message in these results is that probability feedback can be effective in comparison to simple outcome feedback in nonmetric MCPL. This finding implies that poor performance in previous studies was not due to any inherent limitation in feedback-driven learning in nonmetric MCPL, but rather that the particular form of feedback and experimental environment used were not conducive to helping participants solve the problems they faced. We now consider why the type of feedback we gave was effective.

Figure 1 indicates that the group difference occurred very early in training. The main effect of block does indicate that feedback leads to incremental improvements in performance, but these improvements were not large.[1] Furthermore, the nonsignificant interaction between block and group indicates that the advantage for the PFB group remains constant throughout training. This pattern of results is consistent with the idea that participants learned rapidly from the probability feedback at an early stage in training and then used this information almost immediately to improve their performance.

What information are participants acquiring so rapidly? One explanation for this early advantage is that the probability information gives participants an insight into the direction in which each cue points (*cue polarity*). For example, seeing a pattern associated with 0% or 100% probability of an increase would immediately allow the participant to infer the polarity of each cue (e.g., is NASDAQ or FTSE associated with a price increase?). Imagine, for the sake of illustration, that the first two trials observed by a participant are the patterns 0000 (New York, FTSE, Low, No) and 1111 (London, NASDAQ, High, Yes), with the former being paired with a share price decrease and a probability of 0% and the latter with a share price increase and a probability of 100%. From this information alone, the participant can infer the polarity of each cue (e.g., company headquarters: London = increase, New York = decrease). Contrast this with the equivalent trials for a participant in the OFB group. These two patterns would be accompanied by a criterion of DECREASE and INCREASE, respectively. In the absence of probability information, however, the participant cannot infer the polarity of any of the cues. For example, the cue "company headquarters = London" may truly reduce rather than increase the likelihood of a share price increase, but be combined in the "London, NASDAQ, High, Yes" pattern with 3 cues that all increase share price and outweigh it. The participant has no reason to eliminate this possibility from consideration.

With this hypothesized strategy, participants might then simply respond by referring to the number of "increase-indicating cues" present in a pattern. Such a "cue-tallying" heuristic (often referred to as Dawes' rule; see Gigerenzer, Todd, and the ABC Research Group, 1999) could be learned rapidly and used equally well at test when feedback was no longer received. Use of such a heuristic is thus consistent with the almost immediate advantage for the PFB group (present by Block 1 and remaining constant throughout training) and the maintenance of that advantage throughout the test trials. (Note in Table 2 that test performance was the same in the first and second blocks of 32 test trials.) In fact, given the linearity in Patterns 1–10 and 16 (in which the addition of a cue tends to increase the probability of a share price increase), a simple cue-tallying strategy, which responded "decrease" when one or fewer increase-indicating cues were present, "increase" when three or more were present, and guessed when two were present,

[1] The slope of the line for the PFB group in Figure 1 suggests that asymptotic performance had not been reached. To investigate this possibility we conducted an experiment with extended training (three sessions of 240 trials over a 3-day period). Although a marginally significant effect indicating improvement across sessions was found, there was little suggestion that performance exceeded the .80 level achieved at the end of 240 trials in Experiment 1.

would give rise to accuracy of .81. This value reflects the requirement to guess on the six patterns that are "tied" in terms of the number of increase and decrease indicating cues present. Assuming that on average three of these six patterns were guessed correctly, a cue-tallying strategy would make 13 out of 16 correct predictions or .81 maximizing score (see Table 1 for clarification of pattern composition). As shown in Figure 1, this value is the level achieved by the PFB group by the end of training and maintained throughout the test.

Is there any indication in the binary predictions made at test of an influence of the probability feedback? Close analysis suggests that probability feedback did lead to more accurate predictions, but also reveals that the modal strategy was slightly more sophisticated than the simple cue-tallying one that we have described. Consider Patterns 6, 7, 8, 11, 13, and 14, which all appeared on the screen as two increases and two decreases (e.g., Pattern 8 was "INCREASE, INCREASE, DECREASE, DECREASE"). For a simple cue-tallying strategy these patterns constituted a "tie" with equal probability of a share price increase or decrease. However, as shown in Table 1, for the first three of these patterns the probability of an increase was, in fact, less than .50 whereas for the second three it was greater than .50. Were participants sensitive to these differences or did they simply guess?

Figure 2 plots the mean proportion of increase responses for the "$<.50$" and the "$>.50$" patterns across the 64 test trials. The figure clearly shows that the PFB group were better than the OFB group at distinguishing between the two sets of patterns. There was no effect of group, $F(1, 70) = 0.20$, $p > .6$, but there was a significant main effect of pattern, $F(1, 70) = 43.80$, $p < .001$, and a significant interaction between group and pattern, $F(1, 70) = 10.73$, $p < .01$. Paired-sample t tests confirmed that the difference in proportion of increase responses for the two sets of patterns was significant in both the PFB group, $t(35) = 6.26$, $p < .001$, and the OFB group, $t(35) = 2.72$, $p < .05$.

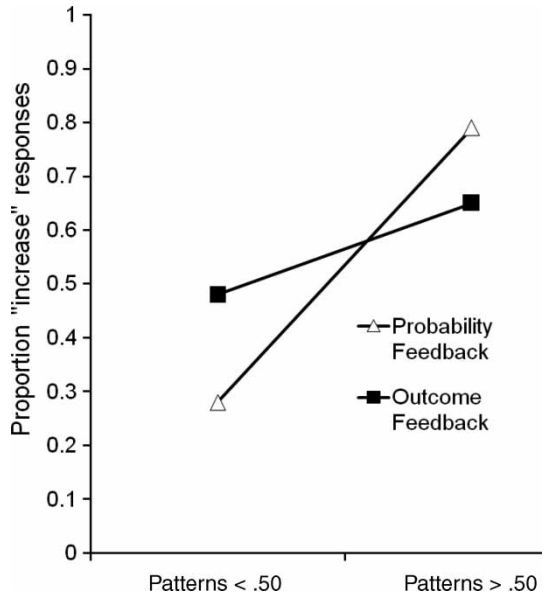

Figure 2. *Experiment 1: Proportion of increase responses for tied patterns containing two INCREASE and two DECREASE cue labels. Note: Patterns <.50 are patterns 6, 7, 8, and patterns >.50 are 11, 13, and 14 (see Table 1).*

Examining the task structure in Table 1 sheds some light on what might be responsible for the greater sensitivity of the PFB group to the difference between the two sets of patterns. Table 1 shows that both values of Cue 1 (INCREASE, coded as a 1; and DECREASE, coded as a 0) are associated with a share price increase (a probability above .50 in Table 1) as often as with a share price decrease (a probability below .50). Thus, Cue 1 on its own provides no predictive information about the outcome (its cue validity is .50). In contrast, for Cues 2–4, the cue value INCREASE (1) is associated with an actual increase on 3/4 of occasions and a decrease on 1/4—thus providing useful predictive information. More formally, the cue validities of Cues 2–4 are identical (.75). The principal difference in the composition of the "$<.50$" and the "$>.50$" patterns is that the "$>.50$" patterns all contain two INCREASE values associated with a combination of Cues 2 to 4, whereas the "$<.50$" patterns contain two INCREASE

values but one is associated with Cue 1. Thus it seems that participants in the PFB group (and to a significantly lesser extent those in the OFB group) adopted a slightly more sophisticated strategy—akin to multiple linear regression—in which the values associated with Cues 2, 3, and 4 were weighted more heavily than that associated with Cue 1. This strategy made more accurate predictions when a simple tally rule resulted in a tie.

EXPERIMENT 2

Brehmer (1979, 1987) argued that the first step in learning from feedback is to learn the functional form. In a nonmetric environment this amounts to learning cue polarity—that is, which value of a cue is associated with which outcome. The data of Experiment 1 suggest that a major effect of the probability feedback that we used was to help participants get over this first hurdle of inferring cue polarities.

If this is indeed how the feedback acted, then providing cue polarity information "up-front" to participants should eliminate the advantage seen for the PFB group in Experiment 1. This hypothesis is pursued in the remaining two experiments. Experiment 2 tested this prediction by explicitly telling participants the direction (increase or decrease) indicated by each cue. The data of Experiment 1 also suggest that participants used a relatively simple (OFB group) or slightly more sophisticated (PFB group) cue-tallying heuristic to solve the task. To investigate this possibility further, in Experiment 2 we asked participants during the test to provide numerical estimates of the probability that the share price would increase. We predicted that if participants relied on a simple cue-tallying strategy, the estimates would be commensurate with the number of increase-indicating cues present in the pattern (i.e., 1 cue 25%, 2 cues 50%, etc.), rather than the veridical values (i.e., the values shown in Table 1). In Experiment 3 we asked whether the polarity information that we hypothesize to be the important feature conveyed by probability feedback is indeed sufficient to enhance performance.

The key modification in Experiment 2 was that the cues were now all assigned the values INCREASE or DECREASE. The instructions were altered to reflect this change and now said that the answer to each piece of information had been used to make a prediction about the likelihood of a share price increase and that the participants' task was to use these predictions to make predictions of their own. For instance, the information "Where are the company's headquarters? = INCREASE" was to be read as stating that the headquarters location (wherever it was) implied a share price increase. If the key property of probability feedback is that it facilitates learning about cue polarity, then the beneficial effect of such feedback should now disappear. Participants in the OFB condition should be able to extract polarity information from a different source, namely the cues themselves. Thus the pattern 1100 (Pattern 8), which would have appeared to participants as "London, NASDAQ, Low, No" in Experiment 1, now appeared as "INCREASE, INCREASE, DECREASE, DECREASE" and accordingly eliminated the requirement for participants to learn which values for each cue (e.g., London/New York for company headquarters) predicted an increase in share value and which a decrease.

Method

Participants

A total of 24 members of the University College London community took part in the experiment. Sixteen were male and 8 female, with a mean age of 23.8 years (range 19 to 29, $SD = 3.10$). Participants were randomly assigned to either the Probability and Outcome Feedback (PFB) group or the Outcome Feedback Only (OFB) group, to give a total of 12 participants in each.

Procedure

The design of the training phase was identical to that in Experiment 1 with the exception that the pieces of information about each company were described in terms of anticipated share value increases or decreases, rather than with the labels

used in Experiment 1. For example, participants were told that based on the location of the company's headquarters the share value would increase, or that based on the index on which the shares were listed the share value would decrease, and so on. Participants completed 240 training trials and 64 test trials as in Experiment 1. However, at test, in addition to making a prediction, participants from both groups were asked to make a numerical prediction (a number between 0 and 100) of the probability that the share price would increase.

Results and discussion

Figure 3 displays the mean performance across blocks for both groups during training. Table 2 displays the mean overall training performance for both groups. The means are higher than they were in Experiment 1. Unlike in Experiment 1 there is little difference between the groups. Performance across Blocks 1 to 8 was compared using a repeated measures analysis of variance with group as a between-subjects factor. Results showed no significant main effects nor a significant interaction (all Fs < 1). The absence of a group effect is in direct contrast with Experiment 1. The group difference seems to have disappeared because performance in the OFB group has improved to a level comparable with that in the PFB group, as predicted if cue polarity information now available to the OFB group is psychologically equivalent to the enhancing effect (when polarity has to be learned) of probability feedback.

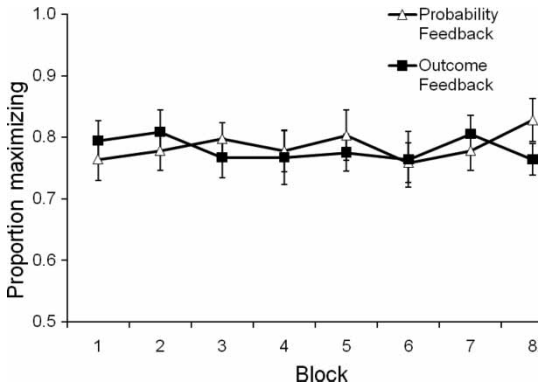

Figure 3. *Experiment 2: Maximizing performance for the probability feedback and outcome feedback groups during training.*

Table 2 displays the mean test performance for both groups in Blocks 1 and 2 of the test trials. Whereas performance in the PFB group is similar to that seen in Experiment 1, performance in the OFB group is much higher. There were no effects of group, $F(1, 22) = 0.894$, $p > .35$, or block, $F(1, 22) = 0.288$, $p > .50$. The absence of a group effect on test scores is again in direct contrast with Experiment 1.

Training and test performance are consistent with the hypothesis that probability information helped participants in Experiment 1 to infer cue polarities, but that when the cue labels provided this information (Experiment 2), both groups were able to perform at similar levels. The absence of an effect of block and group suggests that there was simply less learning to do in this experiment as one of the key problems faced by participants in MCPL tasks—inferring cue polarity—had already been solved for them. Participants readily understood and utilized the information provided by the cue labels. This rapid and early learning is also consistent with the almost immediate use of a cue-tallying strategy, which predicts performance levels of approximately .80 (because of the need to guess for tied patterns). Importantly, both groups reached approximately this level by the first block of training and maintained it throughout training and test.

Analysis of the probability estimates provided further evidence for the use of a cue-tallying strategy in both groups. Figure 4 plots the pattern number along the x-axis and the mean estimated probability of an increase along the y-axis. The objective probabilities lie along the identity line. Also plotted is a line produced by a "tally-the-cues" strategy. This strategy estimates 0 for a pattern with no INCREASEs, .25 for a pattern with one, .50 for a pattern with two, .75 for a pattern with three, and 1.0 for a pattern with four. It is clear that this strategy maps onto both groups' responses very well, as indicated by their step-like form. The closeness of the lines for the

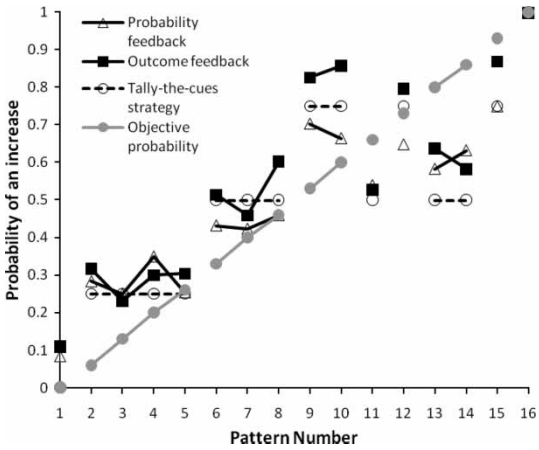

Figure 4. *Experiment 2: Mean estimated probability of a share price increase for Patterns 1–16 for the probability feedback and outcome feedback groups. Note: Objective probability is the value associated with each pattern (see Table 1). The "tally-the-cues strategy" gives an estimate of an increase ranging from 0 to 1.0 depending on the number of increase-indicating cues present in the pattern (0 = 0, 1 = .25, 2 = .50, etc.) See text for clarification.*

PFB and OFB groups and the cue tallying heuristic indicate that there was little difference between the estimates provided by the two groups and those implied by the heuristic. This pattern strongly suggests that participants were using a simple cue-tallying strategy for making their estimates.

To obtain more fine-grained measures of the accuracy of estimates we calculated two independent components of overall accuracy—calibration and discrimination indices. Our calibration index (CI) is a measure of how well the estimated probabilities match the actual probability of the target event occurring. An individual's judgements are well calibrated to the extent that those judgements match the proportion of times the target event actually occurs. The discrimination index (DI) measures the extent to which the estimates participants make when the event happens (share price increase) are different from the estimates they make when the event does not happen.

To calculate a calibration index participants' estimates were compared with the actual probability of the event occurring—in other words, a mean squared deviation was computed between estimated and objective probabilities. Calibration scores were calculated for each participant, summed, averaged over trials, and then averaged across participants. The mean calibration index for the OFB group was .08 (.05), and for the PFB group it was .05 (.03), where standard deviations are in parentheses. A one-way analysis of variance found a marginally significant difference between the two groups, $F(1, 23) = 3.09$, $p = .09$. A lower number indicates better calibration, suggesting that the provision of probability feedback led to slight improvements in calibration at test.

To calculate the discrimination index we used a method from Yates (1990, chap. 3):

$$\begin{aligned} \mathrm{DI} &= \Sigma \mathrm{DI}_j / N \\ \text{where } \mathrm{DI}_j &= N_j (d_j - d)^2 \end{aligned} \tag{1}$$

In the first step, each pattern that is associated with an objective probability of share price increase above .50 is assigned a 1, and each with an objective probability below .50 is assigned a 0. These probabilities were then averaged (d_j) for each given estimate. In Step 2, for every given numerical estimate (on the 0–100 scale) of a share price increase provided by the participant the average "score" (0–1) of the patterns for which that estimate was given is calculated. For example, if a participant gave an estimate of 60% (.60) 10 times across the 64 test trials, and on 8 of these occasions it was for a pattern assigned a 1 in Step 1, and on 2 occasions it was for a pattern assigned a 0 in Step 1, then the average score for the estimate of 60% would be 8 "correct" estimates out of 10, or .80.

The base rate (d) for this task was .5. This is because there is an equal number of patterns with a greater than .50 and a less than .50 chance of an increase. So the mean squared deviation between this base rate and the value calculated in Step 2 is now calculated and then multiplied by the number of times the particular estimate was given (N_j). In our example this gives $10[(.80 - .50) \times (.80 - .50)] = .90$. Hence the DI for the estimate of 60% chance of an

increase is .90. Lastly, individual DIs are then calculated for each separate estimate given by each participant, summed, and averaged across the 64 trials and then averaged across all participants to give the group DI.

Perfect discrimination is equal to .25.[2] The mean discrimination index for the OFB group was .14 ($SD = .04$), and for the PFB group it was .19 ($SD = .04$). A one-way analysis of variance found a significant difference between the two groups, $F(1, 23) = 6.81$, $p < .05$. A higher number indicates better discrimination, suggesting that the provision of probability feedback during training led to an improvement in discrimination at test.

Although overall maximizing performance was equal across the two groups, analysis of the calibration and discrimination indices suggested that the probability feedback did confer a slight advantage on the accuracy of probability estimates—especially with regard to discriminating between when the target event would happen and when it would not. As with Experiment 1, we can ask whether there is evidence from the test choices that is consistent with this conclusion. Close examination of the test data suggests that the groups did differ in their cue weightings, as in Experiment 1. Figure 5 plots the mean proportion of increase responses for the "<.50" and the ">.50" patterns across the 64 test trials. The figure clearly shows that the PFB group did distinguish the two sets of patterns, but the OFB group did not. There was no effect of group, $F(1, 70) = 0.97$, $p > .3$, but there was a significant main effect of pattern, $F(1, 70) = 9.42$, $p < .002$, and a significant interaction between group and pattern, $F(1, 70) = 4.48$, $p < .05$. Paired-sample

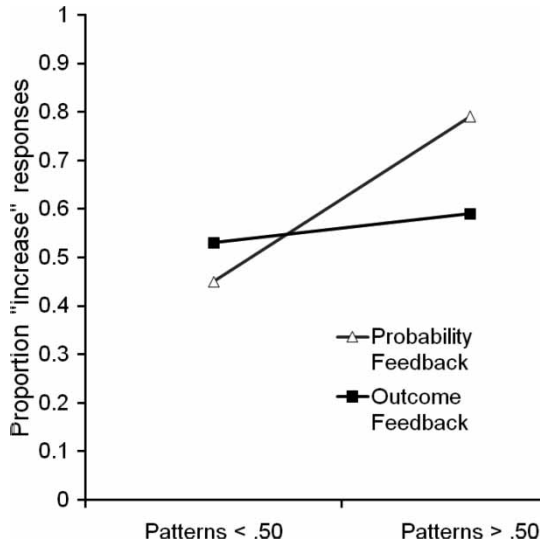

Figure 5. *Experiment 2: Proportion of increase responses for tied patterns containing two INCREASE and two DECREASE cue labels. Note: Patterns <.50 are patterns 6, 7, 8, and patterns >.50 are 11, 13, and 14 (see Table 1).*

t tests confirmed that the difference in proportion of increase responses for the two sets of patterns was significant in the PFB group, $t(35) = 3.88$, $p < .001$, but not in the OFB group, $t(35) = 0.63$, $p > .5$. Thus the PFB group were able to employ a cue-tallying strategy with weights varying across the cues. The OFB group, in contrast, do not seem in these circumstances to have been able to move beyond a simple tallying strategy. Compared to their counterparts in Experiment 1, it seems that the up-front presentation of polarity information in Experiment 2 detracted from learning about differential cue weights.

[2] To see why, imagine that a participant gave an estimate of 25% (.25) on all 32 trials on which the presented pattern had a probability of increase below .50 (and thus assigned a 0 in Step 1 above). This would mean the participant made a correct estimate (below .50) on all the patterns on which the objective probability was indeed below .50, giving a score of 32/32 or 1.0. The calculation in Step 3 would then be $32[(1 - .5) \times (1 - .5)] = 8$. Now imagine the same participant assigned the estimate 75% (.75) on all 32 trials on which the presented pattern had a probability of increase above .50 (thus assigned a 1 in Step 1). By the same logic as above we would have the identical calculation in Step 3, giving a DI of 8 for the 75% estimate. The summed DI for this participant would be 16, which when divided by the number of trials (64) gives the perfect mean DI of .25. Note that any estimate that is consistently below 50% for all patterns with an objective probability below .50 and consistently above 50% for an objective probability above .50 will lead to the same final DI of .25. This is why the measure reveals *discrimination* between occasions when the share price will go up or not, as compared to *calibration* of the participant either side of that "binary" prediction.

EXPERIMENT 3

Experiment 1 found clear evidence for a beneficial effect of probability feedback, and we have suggested a polarity hypothesis to explain this effect, whereby a major property of such feedback is that it serves to enhance the speed of learning about the polarity of the different cues. In Experiment 2 feedback had no enhancing effect in the learning stage because, on our hypothesis, placing cues and criterion on the same scale (scale here refers to using compatible labels, e.g., "INCREASE" for both cues and criterion) enabled more useful expectations about the internal structure of the task than when cues and criterion are on different scales (e.g., "NASDAQ" and "INCREASE", respectively). Thus performance of participants in the OFB group of Experiment 2 was substantially higher than that of their counterparts in Experiment 1. Experiment 3 sought to replicate this key cross-experiment contrast within a single experiment while also ruling out two alternative explanations of the results.

We have hypothesized that probability feedback is qualitatively different from simple outcome feedback in that one of its major effects is to enhance learning of cue polarity. But participants in the PFB group in Experiment 1 also received more feedback overall than those in the OFB group as they received both outcome and probability information. Perhaps the advantage conferred by probability feedback was a side effect of the provision of a greater amount of information. In Experiment 3 we compare two groups who receive equivalent amounts of feedback, but under circumstances where only one of them should benefit from polarity information.

To test both the polarity and the scale hypotheses in a single experiment, Experiment 3 compared performance in three groups in total. Feedback was held constant—only outcome feedback was given to all groups. The task environment for the scale incompatible group was the same as that for the OFB group in Experiment 1: Cue values were presented as, for example, NASDAQ or FTSE, and the criterion as an increase or decrease in share price. For the scale compatible/direction compatible group the task environment was the same as that for the OFB group in Experiment 2—cue values were presented as INCREASE or DECREASE, as was the criterion. Consistent with Experiment 2, cue values of 1 were coded as "INCREASE" and 0 as "DECREASE"—thus the polarity information conveyed by the cues was compatible with the criterion. However, unlike Experiment 2 participants were not told explicitly that if the "answer" to a question (e.g., Invest in new projects?) was INCREASE that this implied an increase in the share price. The omission of this instruction was necessary for the inclusion of a third group, the scale compatible/direction incompatible group in which participants were presented with cues and criterion on the same scale, but for two of the cues a 1 was coded as a DECREASE and 0 as an INCREASE, while for the other two cues 1s were INCREASES, and 0s were DECREASES. In other words, two cues were positively correlated with the criterion, and two were negatively correlated.

We expected to replicate the pattern from the first two experiments, with polarity information in the scale compatible/direction compatible group leading to better performance than that in the scale incompatible group. If cue polarity information drives the improvement then performance in the scale compatible/direction compatible group should also be superior to that in the scale compatible/direction incompatible group. This is because cue direction in the scale compatible/direction compatible group is "given" inasmuch as it is compatible with what we assumed participants' expectations about the internal structure of the task would be and with the cue criterion relation. In the scale compatible/direction incompatible condition the polarity for two of the cues is incompatible both with the criterion and with participants' (probable) prior expectations about task structure. This incompatibility of direction forces participants to learn cue direction. In the scale incompatible group, cue direction has to be learned because no directional information is conveyed either by instructions or by the task structure (as in Experiment 1).

In contrast, if the effects observed in Experiments 1 and 2 are due to a difference in the amount of feedback provided, then we should see no difference between the scale compatible/direction compatible and scale compatible/direction incompatible groups, as they are equated in this regard.

Method

Participants

A total of 30 members of the University College London community took part in the experiment. Eight were male and 22 female, with a mean age of 22 years (range 18 to 50, $SD = 5.94$). Participants were randomly assigned to the scale incompatible (SI), scale compatible/direction compatible (SC/DC), and scale compatible/direction incompatible (SC/DI) groups to give a total of 10 participants in each group.

Procedure

The procedure was similar to that in Experiments 1 and 2. The instructions were modified to reflect the assignment of cue values to cue labels. In Experiment 2 participants were told that, for example, based on the index on which the shares were listed the share value would decrease, or based on the location of the company headquarters the share price would increase. In Experiment 3 we changed the cue labels slightly so that the cue values could be coded as INCREASE or DECREASE without the added assumption that this necessarily implied an increase or decrease in the share value. The cue labels and values were as follows: SI condition: employee turnover (1 = low, 0 = high), invest in new projects (1 = yes, 0 = no), largest market share (1 = south, 0 = north), index shares listed on (1 = NASDAQ, 0 = FTSE); SC/DC condition: employee turnover (1 = INCREASE, 0 = DECREASE), invest in new projects (1 = INCREASE, 0 = DECREASE), size of market share (1 = INCREASE, 0 = DECREASE), number of indices that shares were listed on (1 = INCREASE, 0 = DECREASE). For the SC/DI condition labels and values were the same as those for the SC/DC but the coding of increase and decrease was reversed for the invest in new projects and number of indices labels.

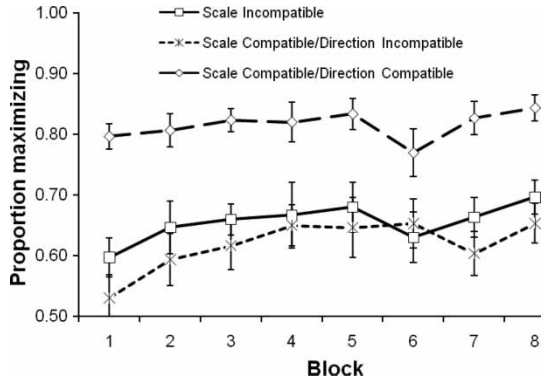

Figure 6. *Experiment 3: Maximizing performance for the scale incompatible, scale compatible/direction compatible, and scale compatible/direction incompatible groups during training.*

Results and discussion

Figure 6 displays performance for the three groups during training. Table 3 displays the mean values averaged across the eight training and two test blocks. A 3 (group) × 8 (block) analysis of variance on the training trial data revealed a main effect of block, $F(7, 189) = 3.27$, $p > .01$, and a main effect of group, $F(2, 27) = 17.00$, $p < .001$. The interaction between block and group was

Table 3. *Mean proportion of maximizing responses in the scale incompatible, scale compatible/direction compatible, and scale compatible/direction incompatible groups in the 240 training and 64 test trials of Experiment 3*

	Group		
	SI	*SC/DC*	*SC/DI*
Training	.66	.82	.62
Test			
Block 1	.72	.85	.59
Block 2	.75	.88	.63

Note: SI = scale incompatible. SC/DC = scale compatible/direction compatible. SC/DI = scale compatible/direction incompatible.

not significant, $F < 1$, $p > .7$. Simple main effects analyses revealed significant differences between performance in the SC/DC and SI groups, $F(1, 19) = 25.38$, $p < .001$, and between the SC/DC and SC/DI groups, $F(1, 19) = 32.18$, $p < .001$, but not between the SI and SC/DI groups, $F(1, 19) = 0.827$, $p > .$ 1. A 3 (group) × 2 (block) analysis of variance on the test trial data revealed a main effect of group, $F(1, 27) = 14.21$, $p < .001$. The effect of block approached significance, $F(1, 27) = 3.79$, $p = .062$, but the interaction was not significant, $F < 1$, $p > .9$. Consistent with the training data, simple main effects analyses revealed that the difference between the SC/DC group and the SI group was significant, $F(1, 19) = 16.73$, $p < .001$; however, in the test phase the difference between the SI and the SC/DI group was also significant, $F(1, 19) = 5.01$, $p < .05$.

The pattern of results strongly suggests that it is information about cue polarity that drives the improvement in performance when only outcome feedback is given. The SC/DC group performed at a similarly high level to the equivalent group in Experiment 2. Participants achieved a score of approximately .80 from the first block of trials onwards. In fact this level of performance was found in the first 10 trials: Performance for these trials was .82. In contrast, in the two groups in which cue polarity had to be learned, performance in the first block was only .53 and .59 for the SC/DI and SI groups, respectively. Performance for the latter group mirrors that seen for the equivalent group in Experiment 1.

Performance in the SC/DI group suggests that, at least initially, having cues and criterion on the same scale was disadvantageous. It appears that participants in both the SC/DI and the SC/DC groups assumed initially that that a cue value labelled INCREASE led to an increase in the share price. For the latter group, in which this expectation was consistent with the structure of the environment, participants performed at a high level. For the SC/DI group in which the expectation was inconsistent (due to the two negatively correlated cues) it hindered learning and, at least initially, pulled it down to a level below the SI group. It is interesting to note that in the test phase, when no feedback was provided, performance in the SC/DI group fell back to a level significantly below that in the SI group, perhaps reflecting the difficulty in remembering cue polarity in the absence of feedback.

The size of the learning effect was small, but significant. Although the interaction was not significant, most of the improvements across blocks seem to have been in the two groups in which cue direction needed to be learned (mean improvement of .11 across blocks, compared with .04 for the SC/DC group). The pattern is consistent with the small learning effects seen in Experiment 1, in which cue direction was not provided, and the absence of an effect in Experiment 2, in which it was.

In summary, we conclude that the presence of a probability feedback effect seen in Experiment 1, and its absence in Experiment 2, were due to the absence and presence, respectively, of directional polarity information conveyed by the cues and criterion being placed on a common scale.

GENERAL DISCUSSION

The starting point for these experiments was the observation that the standard effect of richer forms of feedback on performance found in metric MCPL tasks has not previously been found in nonmetric tasks (e.g., Castellan, 1974; Castellan & Swaine, 1977). This finding, if general, would create a theoretical puzzle. We questioned whether existing findings were due to elaborated feedback per se simply making no difference in nonmetric MCPL or whether the reported findings reflect the ineffectiveness of the particular forms of feedback employed. The results strongly suggest that the latter interpretation is more likely.

Experiment 1 clearly demonstrated that provision of probability feedback led to an advantage in comparison with simple outcome feedback. This finding is consistent with the effect of cognitive feedback found in numerous metric-MCPL tasks (e.g., Balzer et al., 1989; Schmitt et al.,

1976; Schmitt et al., 1977). We hypothesized that one of the main reasons why probability feedback enhanced performance was because it helped participants infer the direction pointed to by each cue. From this it follows that if such polarity information is provided by an alternative source, then the beneficial effect of probability feedback should be eliminated. Experiment 2 supported this hypothesis. Note that Experiment 2 is a conceptual replication of the finding that elaborated feedback does not enhance performance in nonmetric MCPL as variation in feedback led to no difference in performance. Finally, in Experiment 3 we demonstrated directly that cue polarity information is sufficient to generate the sort of enhancement effect seen in Experiment 1 (the better performance of the PFB than of the OFB group) and Experiment 2 (the OFB group in comparison to the equivalent group from Experiment 1).

The standard MCPL approach, outlined with reference to Brehmer (1979) in the introduction, suggests that training leads to the learning of cue–criterion functions, the abstraction of the relative weights of individual cues in the patterns, and then integration of cues to make a judgement about the current pattern. Why do we believe that probability feedback principally relates to the first of these task requirements? There are three answers to this question. First, our logic has been to predict that if such feedback does indeed aid in the learning of cue polarity, then providing the very same information via a different source (cue labels) should abolish the beneficial feedback effect, as indeed observed in Experiment 2. Secondly, from a conceptual point of view, whereas it is easy to see how it can enhance function learning, probability information provides no obvious source of information about cue weighting or integration and no obvious explanation for why such requirements should be changed by the manipulation of cue labels in Experiment 2. Finally, the fact that the feedback effect occurred so early in training in Experiments 1 and 3 is consistent with a rapid cue-polarity explanation. Any effects on cue weighting or integration would be expected to take longer to manifest themselves.

It is true that we did obtain some evidence that probability feedback can also relate to the second of Brehmer's (1979) task requirements, cue weighting. In Experiment 1 such feedback improved accuracy of predictions on test trials comprising conflicting cue values. In Experiment 2 such feedback improved performance in the test stage, even though it had no detectable impact on performance in the learning stage. Specifically, probability feedback improved the calibration of probability estimates and also improved accuracy of predictions on conflicting cue test trials. These findings suggest that, in addition to learning about cue polarity, participants were able to use feedback to improve their profile of cue weights—at the very least, to assign Cue 1 a lower weight than Cues 2–4.

In addition to providing evidence to help understand a perplexing paradox in the use of feedback in metric and nonmetric MCPL, our results also provide an insight into the way in which judges attempt to solve MCPL tasks. Recent studies have focused on two general approaches. One class of models assumes linear cue addition. The general form of such models is:

$$\mathrm{P(INCREASE}\,|\,\mathrm{S}) = \mathrm{C}_1 w_1 + \mathrm{C}_2 w_2 + \mathrm{C}_3 w_3 + \mathrm{C}_4 w_4 + k, \quad (2)$$

where S is a cue pattern, C_n is the value of cue n (coded as 0 or 1), w_n is the weight of cue n, and k is a constant. When the weights are derived by multiple linear regression, this model is optimal for the task environment (see Lagnado, Newell, Kahan, & Shanks, 2006). However, numerous alternative strategies emerge from versions of the model in which the weights are constrained. For instance, when the weights are restricted to values of $+1$ or -1, and $k = 0$, the model is Dawes' rule, the cue-tallying rule that we have considered here at length. Alternatively, the "one-cue" strategies of Gluck, Shohamy, and Myers (2002; see also Meeter, Radics, Myers, Gluck, & Hopkins, 2008) emerge when all but one of the cue weights are zero.

An alternative approach, connecting MCPL with extensive research on categorization, suggests

that judgements about a current cue pattern (or exemplar) can be made on the basis of similarity to previously seen patterns, which are stored in memory (e.g., Juslin et al., 2003a; Juslin, Olsson, & Olsson, 2003b; Nosofsky, 1986). Juslin and his colleagues have noted the conceptual similarity between MCPL decision tasks and those used in the study of category learning and have reported several studies attempting to bridge these domains both at the task level and theoretically. One conclusion of these studies is that nonlinear cue–criterion relations can be exceptionally hard to learn, and that participants often find themselves "trapped in persistent and futile attempts to abstract the cue–criterion relations" (Olsson, Enkvist, & Juslin, 2006, p. 1371) characterized by Equation 2 (see also Nosofsky & Bergert, 2007).

In our experiments, the environment was nonlinear, and feedback was provided on a pattern rather than a cue basis. These are factors that arguably should promote the adoption of an exemplar memorization strategy (cf. Ashby, Maddox, & Bohl, 2002; Edgell & Morrissey, 1992). Yet despite this, participants appeared to quickly discover the linear component of the environment and adopted a strategy that treated cues independently and additively. Evidence for this came from finding that performance failed to exceed a level predicted by cue tallying, and that numerical estimates of the probability of an increase mapped onto those predicted by cue tallying, rather than the veridical values. We also found some evidence that probability feedback led participants to weight cues correctly by distinguishing which ones were more predictive than others. This was apparent in both the ability to discriminate between share price increases and decreases and in the predictions about "tied" patterns. Thus, under appropriate circumstances (such as elaborated feedback), participants can outperform the elementary cue-tallying (Dawes' rule) heuristic. Other research has similarly highlighted people's capacity to approximate the optimal linear strategy (e.g., Lagnado et al., 2006; Speekenbrink, Channon, & Shanks, 2008; White & Koehler, 2007).

The effect of such strong assumptions about linearity and additivity on performance has been noted in a number of previous MCPL studies (Brehmer, 1980, 1987; Juslin, Karlsson, & Olsson, 2008). If the functions relating the cues to the criterion are negative, or nonlinear, then learning is often impeded (Deane, Hammond, & Summers, 1972; Slovic, 1974). Furthermore, if the cues themselves are intercorrelated, then learning is typically disrupted (Lindell & Stewart, 1974; Schmitt & Dudycha, 1975). Because there was a strong linear component in our environment, a simple additive strategy afforded an acceptable level of accuracy (approximately 80%). It is plausible that participants were not motivated enough to improve further by learning the anomalous nonlinear patterns, especially given that they were told they could only achieve "reasonable" but not "perfect" accuracy in the instructions. It is interesting to note that we did not find improvements above the .80 level even after 720 training trials (see Footnote 1).

An important question for future studies is to determine the extent of training and the nature of feedback required to eliminate participants' assumptions about linearity and additivity and to promote optimal responding even in nonlinear environments (cf. Shanks et al., 2002). Despite the pessimistic claims made in MCPL studies about judges' inability to learn in nonlinear environments, there is ample evidence from the categorization literature demonstrating accurate performance in nonlinear environments (e.g., Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Ashby et al., 2002). The categorization area is rich with computational models of how people represent and combine probabilistically related information, which are directly relevant to the questions often asked in MCPL research. Increasing cross-fertilization between the MCPL and categorization literatures (as urged by some 20 years ago, e.g., Klayman, 1988) should provide further valuable advances in understanding how people learn from feedback in both linear and nonlinear probabilistic environments.

Original manuscript received 2 May 2006
Accepted revision received 15 July 2008
First published online 17 October 2008

REFERENCES

Adelman, L. (1981). The influence of formal, substantive, and contextual task properties on the relative effectiveness of different forms of feedback in multiple-cue probability learning tasks. *Organizational Behavior and Human Performance*, *27*, 423–442.

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple-systems in category learning. *Psychological Review*, *105*, 442–481.

Ashby, F. G., Maddox, W. T., & Bohl, C. J. (2002). Observational versus feedback training in rule-based and information-integration category learning. *Memory & Cognition*, *30*, 666–677.

Balzer, W. K., Doherty, M. E., & O'Connor, R. (1989). Effects of cognitive feedback on performance. *Psychological Bulletin*, *106*, 410–433.

Brehmer, B. (1979). Preliminaries to a psychology of inference. *Scandinavian Journal of Psychology*, *20*, 193–210.

Brehmer, B. (1980). In one word: Not from experience. *Acta Psychologica*, *45*, 223–241.

Brehmer, B. (1987). Note on subjects' hypotheses in multiple-cue probability learning. *Organizational Behavior and Human Decision Processes*, *40*, 323–329.

Brehmer, B. (1994). The psychology of linear judgment models. *Acta Psychologica*, *87*, 137–154.

Castellan, N. J. (1973). Multiple-cue probability learning with irrelevant cues. *Organizational Behavior and Human Performance*, *9*, 16–29.

Castellan, N. J. (1974). The effect of different types of feedback in multiple-cue probability learning. *Organizational Behavior and Human Performance*, *11*, 44–64.

Castellan, N. J., & Edgell, S. E. (1973). An hypothesis generation model for judgment in nonmetric multiple-cue probability learning. *Journal of Mathematical Psychology*, *10*, 204–222.

Castellan, N. J., & Swaine, M. (1977). Long term feedback and differential feedback effects in nonmetric multiple-cue probability learning. *Behavioral Sciences*, *22*, 116–128.

Deane, D. H., Hammond, K. R., & Summers, D. A. (1972). Acquisition and application of knowledge in complex inference tasks. *Journal of Experimental Psychology*, *92*, 20–26.

Edgell, S. E., & Morrissey, J. M. (1992). Separable and unitary stimuli in nonmetric multiple cue probability learning. *Organizational Behavior and Human Decision Processes*, *51*, 118–132.

Gigerenzer, G., Todd, P. M., & the ABC Research Group. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.

Gluck, M. A., Shohamy, D., & Myers, C. (2002). How do people solve the "weather prediction" task? Individual variability in strategies for probabilistic category learning. *Learning & Memory*, *9*, 408–418.

Hammond, K. R. (1971). Computer graphics as an aid to learning. *Science*, *172*, 901–908.

Hammond, K. R., & Boyle, P. J. R. (1971). Quasi-rationality, quarrels and new conceptions of feedback. *Bulletin of the British Psychological Society*, *24*, 103–113.

Harries, C., & Harvey, N. (2000). Taking advice, using information and knowing what you are doing. *Acta Psychologica*, *104*, 399–416.

Juslin, P., Jones, S., Olsson, H., & Winman, A. (2003a). Cue abstraction and exemplar memory in categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 924–941.

Juslin, P., Karlsson, L., & Olsson, H. (2008). Information integration in multiple cue judgment: A division of labor hypothesis. *Cognition*, *106*, 259–298.

Juslin, P., Olsson, H., & Olsson, A.-C. (2003b). Exemplar effects in categorization and multiple-cue judgment. *Journal of Experimental Psychology: General*, *132*, 133–156.

Klayman, J. (1988). On the how and why (not) of learning from outcomes. In B. Brehmer & C. R. B. Joyce (Eds.), *Human judgment: The SJT view* (pp. 115–160). North-Holland: Elsevier.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*, 254–284.

Lagnado, D. A., Newell, B. R., Kahan, S., & Shanks, D. R. (2006). Insight and strategy in multiple cue learning. *Journal of Experimental Psychology: General*, *135*, 162–183.

Lindell, M. K., & Stewart, T. R. (1974). The effects of redundancy in multiple cue probability learning. *American Psychologist*, *87*, 393–398.

Meeter, M., Radics, G., Myers, C. E., Gluck, M. A., & Hopkins, R. O. (2008). Probabilistic categorization: How do normal participants and amnesic patients do it? *Neuroscience and Biobehavioral Reviews*, *32*, 237–248.

Muchinsky, P. M., & Dudycha, A. L. (1975). Human inference behavior in abstract and meaningful

environments. *Organizational Behavior and Human Performance*, *13*, 377–391.

Newell, B. R., Lagnado, D. A., & Shanks, D. R. (2007). Challenging the role of implicit processes in probabilistic category learning. *Psychonomic Bulletin & Review*, *14*, 505–511.

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39–57.

Nosofsky, R. M., & Bergert, F. B. (2007). Limitations of exemplar models of multi-attribute probabilistic inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 999–1019.

Olsson, A. C., Enkvist, T., & Juslin, P. (2006). Go with the flow: How to master a nonlinear multiple-cue judgment task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 1371–1384.

Schmitt, N., Coyle, B. W., & King, L. (1976). Feedback and task predictability as determinants of performance in multiple-cue probability learning tasks. *Organizational Behavior and Human Performance*, *16*, 388–402.

Schmitt, N., Coyle, B. W., & Saari, B. B. (1977). Types of task information feedback in multiple-cue probability learning. *Organizational Behavior and Human Performance*, *18*, 316–328.

Schmitt, N., & Dudycha, A. (1975). A reevaluation of the effect of cue redundancy in multiple-cue probability learning. *Journal of Experimental Psychology: Human Learning and Memory*, *1*, 307–315.

Shanks, D. R., Tunney, R. J., & McCarthy, J. D. (2002). A re-examination of probability matching and rational choice. *Journal of Behavioral Decision Making*, *15*, 233–250.

Slovic, P. (1974). Hypothesis testing in the learning of positive and negative linear functions. *Organizational Behavior and Human Decision Processes*, *11*, 368–376.

Sniezek, J. (1986). The role of variable labels in cue probability learning tasks. *Organizational Behavior and Human Decision Processes*, *38*, 141–161.

Speekenbrink, M., Channon, S., & Shanks, D. R. (2008). Learning strategies in amnesia. *Neuroscience and Biobehavioral Reviews*, *32*, 292–310.

Todd, P. M., & Hammond, K. R. (1965). Differential effects of feedback in two multiple-cue probability learning tasks. *Behavioral Science*, *10*, 429–435.

White, C. M., & Koehler, D. J. (2007). Choice strategies in multiple-cue probability learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 757–768.

Yates, J. F. (1990). *Judgment and decision making.* Englewood Cliffs, NJ: Prentice Hall.