Journal of Behavioral Decision Making
J. Behav. Dec. Making, **15**: 233–250 (2002)
Published online in Wiley InterScience 18 March 2002
(www.interscience.wiley.com) **DOI**: 10.1002/bdm.413

A Re-examination of Probability Matching and Rational Choice

DAVID R. SHANKS,* RICHARD J. TUNNEY and JOHN D. McCARTHY
University College London, UK

ABSTRACT

In a typical probability learning task participants are presented with a repeated choice between two response alternatives, one of which has a higher payoff probability than the other. Rational choice theory requires that participants should eventually allocate all their responses to the high-payoff alternative, but previous research has found that people fail to maximize their payoffs. Instead, it is commonly observed that people match their response probabilities to the payoff probabilities. We report three experiments on this choice anomaly using a simple probability learning task in which participants were provided with (i) large financial incentives, (ii) meaningful and regular feedback, and (iii) extensive training. In each experiment large proportions of participants adopted the optimal response strategy and all three of the factors mentioned above contributed to this. The results are supportive of rational choice theory. Copyright © 2002 John Wiley & Sons, Ltd.

KEY WORDS probability matching; maximization; choice; rationality; feedback; payoffs; learning; reinforcement

A striking violation of rational choice theory is commonly observed in simple repeated binary choice tasks in which a payoff is available with higher probability given one response than another. In such tasks people often tend to 'match' probabilities: That is, they allocate their responses to the two options in proportion to their relative payoff probabilities. Thus suppose that a monetary payoff of fixed size is given with probability $p = 0.7$ for choosing left and with probability $1 - p = 0.3$ for choosing right. Probability matching refers to behavior in which left is chosen on about 70% of trials and right on 30%.

Such responding violates rational choice theory because the optimal strategy in such tasks, after an initial period of experimentation and assuming that the payoff probabilities are stationary, is always to select the option associated with the higher probability of payoff. On any trial, the expected payoff for choosing left is higher than the expected payoff for choosing right.[1]

* Correspondence to: David R. Shanks, Department of Psychology, University College London, Gower Street, London WC1E 6BT, UK.
E-mail: d.shanks@ucl.ac.uk

Contract/grant sponsors: United Kingdom Economic and Social Research Council (ESRC); Leverhulme Trust.

[1] In the statistics literature these situations are called *bandit problems* by analogy with slot machines. Berry and Fristedt (1985) review the theory of bandit problems from the perspective of statistical optimization theory. Tustin and Morgan (1985) present a more economically motivated optimization theory for probability learning tasks.

Choice behavior in this sort of game against Nature has been studied in an enormous number of experiments and demonstrations of probability matching are very robust. Extensive reviews are provided by Myers (1976) and Vulkan (2000). For instance, in Neimark and Shuford's (1959) study one response alternative was correct on 67% of trials and the other on 33%, and at the end of 100 trials participants were choosing the former on about 67% of trials. However, there are also many studies reporting 'over-matching', that is, a tendency to choose the option with the higher probability of payoff with probability closer to 1.0. In Edwards' (1961) study, for instance, participants' asymptotic choice probability for a response that had a payoff probability of 0.7 was 0.83.

The fact that participants fail to maximize their payoffs in these choice tasks has attracted the interest of many theorists concerned with the implications of this phenomenon for rational choice theory. Thus Arrow (1958, p. 14) noted that:

> The remarkable thing about this is that the asymptotic behavior of the individual, even after an indefinitely large amount of learning, is not the optimal behavior. . . . We have here an experimental situation which is essentially of an economic nature in the sense of seeking to achieve a maximum of expected reward, and yet the individual does not in fact, at any point, even in a limit, reach the optimal behavior.

What is particularly striking is that participants fail to maximize despite the apparent simplicity of the problem facing them. Keeping track of payoff probabilities across two response options should hardly tax working memory, nor would one expect the comparison of these probabilities to be very demanding. Moreover, unlike many examples of apparently irrational choice behavior, such as preference reversals (see Camerer, 1995) and responding in the Monty Hall problem (Friedman, 1998), participants make repeated choices and receive a steady flow of feedback from their behavior which should provide a strong impetus to help them find the optimal choice strategy.

In response to this somewhat pessimistic perspective, a number of objections can be raised to the conclusion that people inherently behave irrationally in these probability learning tasks. First, many studies used sequences that were not truly random (i.e. not independent and identically distributed (i.i.d.)) and this often means that the optimal strategy is no longer to choose one option with probability 1.0 (see Fiorina, 1971). Second, quite a large number of studies used either non-monetary outcomes or else payoffs of such low monetary value that the difference in expected cumulative earnings from maximizing compared to matching is negligible, and there is some evidence that monetary payoffs promote responding that is more nearly optimal (see Vulkan, 2000). Third, given participants' common suspicion about psychological experiments (Hertwig and Ortmann, 2001), they may be reluctant to believe that the payoff probabilities are constant and may seek sequential dependencies and predictable patterns across trials. Fourth, almost all studies have reported group rather than individual-participant data, with the obvious danger that probability-matching at the group level masks wide variations at the individual-participant level (see Estes, 1956, for a general discussion of this problem).

However, even when these shortcomings have been addressed, there has barely been any convincing evidence that people can reliably maximize expected utility. For example, even under monetary payoffs, asymptotic levels of responding rarely exceed 0.95 for the correct choice alternative (Siegel, 1961; Vulkan, 2000) and participants continue to forfeit payoffs. In a signal detection study by Healy and Kubovy (1981), almost perfect probability matching was obtained despite the presence of performance-contingent payoffs. Moreover, while instructions or other manipulations aimed at emphasizing the randomness of the trial sequence and the stationarity of the payoff probabilities can have some effect (Beach and Swensson, 1967; Braveman and Fischer, 1968), they do not fully overcome participants' tendency to probability match. As Myers (1976, p. 177) concluded: 'One point is evident: while instructional and motivational manipulations can yield increased probability of predicting the more frequent event, subjects consistently fall short of the optimal strategy of always predicting that event.'

Previous research in this area has paid little attention to another potentially important factor, namely the role of feedback in learning. One of the simplest forms of feedback, outcome feedback, refers to the provision of information about the actual outcome on each trial, for example information about the difference between the number of correct choices made and the maximum number of possible correct choices over a block of trials. The literature suggests that outcome feedback can be useful when performing very simple choice tasks (e.g. Balzer *et al.*, 1989). To test this possibility the experiments reported here incorporate the provision of outcome feedback. In the General Discussion we briefly mention other forms of feedback that might also affect performance in probability learning tasks.

One response to the issues raised above is to accept the behavioral evidence for probability matching but to rationalize it theoretically. For instance, it may be conjectured that participants assign negative utility to the boredom associated with making the same response repeatedly, or that there is greater utility in correctly predicting a rare than a common event. Yet even if these possibilities are true, with increasing monetary payoffs and informative feedback there should come a point at which the additional utility achievable from maximizing is sufficient to outweigh these factors. Hence maximizing should eventually be observed.

Multiple-cue choice tasks

In a variant on the classic probability learning task, participants are presented with a cue or set of cues which vary from trial to trial and which signal independent reinforcement probabilities for the choice alternatives. For example, in one condition of a study by Myers and Cruse (1968), one cue signaled that left was correct with probability 0.85 and right with probability 0.15, while another cue signaled probabilities of 0.15 and 0.85 for left and right, respectively. As in the basic probability learning task, participants have difficulty maximizing their proportions of correct choices. Moreover, Castellan (1974) found that various forms of feedback information failed to increase maximizing behavior.

Research on probability matching dwindled considerably after the mid-1970s, but since a seminal article by Gluck and Bower (1988), a growing number of studies have used versions of a multiple-cue probability learning (MCPL) task to examine rational choice in probability learning situations (e.g. Estes *et al.*, 1989; Friedman and Massaro, 1998; Friedman *et al.*, 1995; Kitzis *et al.*, 1998; Myers *et al.*, 1994; Nosofsky *et al.*, 1992; Shanks, 1990, 1991). These studies have uniformly documented sub-optimal responding and in many cases probability matching.

In a common version of this task, participants imagine themselves to be medical practitioners making disease diagnoses about a series of patients. Each patient presents with some combination of the presence or absence of each of four conditionally independent symptoms (e.g. stomach cramps, discolored gums)[2] and is either suffering or not suffering from a disease. Denoting the symptom pattern $\mathbf{s} = s_1\, s_2\, s_3\, s_4$, where for $i = 1, \ldots, 4$, $s_i = 1$ if symptom i is present and $s_i = 0$ if it is absent, participants' task is to predict whether the disease is present ($d = 1$) or absent ($d = 0$) for each of many such patients, receiving outcome feedback (the actual value of d) on each trial. The structure of the task is such that for each of the 16 possible symptom patterns there is some fixed probability $P(\mathbf{s}) \in [0, 1]$ that patients with that pattern have the disease and the complementary probability that they have no disease. The standard probability learning experiments reviewed in the last section can be thought of as degenerate cases of this sort of task in which the number of symptoms (cues) is zero.

To maximize the number of correct diagnoses, participants should always choose the outcome (disease or no disease) that has been more frequently associated with that particular symptom pattern, yet participants instead seem to match their choices to the actual outcome probabilities. In an experiment reported by Shanks (1991, Experiment 1), for instance, participants received 160 trials in total, and the correlation between the

[2]That is, whether a given symptom is present does not depend on which other symptoms are present.

actual probability of a disease and the participants' choice probability for that disease across the final 80 trials was 0.89. One particular symptom pattern, for instance, predicted a disease with probability 0.78. To maximize correct diagnoses, participants should predict this disease on every trial, yet the actual choice probability for this symptom pattern was 0.75.

Friedman *et al.* (1995), Kitzis *et al.* (1998), and Friedman and Massaro (1998) have recently reported some important studies using multiple-cue choice tasks of this sort which considerably clarify our understanding of the factors that might drive participants towards more nearly optimal performance. In their initial study (Friedman *et al.*, 1995), very clear probability matching was obtained with correlations between asymptotic choice and actual probabilities in excess of 0.88. In a later study (Friedman and Massaro, 1998; Kitzis *et al.*, 1998), however, the payoff conditions and provision of information about prior relevant cases were systematically varied. Some groups (Score) were provided both trial-by-trial and cumulatively with a score based on their accuracy, others (Pay + Score) were in addition paid on a performance-related basis, and others (No Score) received neither the score nor monetary payoff. An interesting aspect of the score information was that it included information about how well an ideal Bayesian expert would have done, thus allowing participants to see how close or far their performance was from that achievable by the optimal but unspecified algorithm. Orthogonal to the score and payoff manipulation, some groups (History) were able to access on each trial a summary table providing information about the outcomes of previous cases with the same pattern of symptoms as the current patient. Other groups (No History) did not have access to this information.

Friedman and Massaro (1998) found that the provision of history information pushed participants significantly closer to maximizing. The score conditions had a similar beneficial effect, but providing a payoff (somewhat surprisingly) had no such effect. In the present research we do not include a history condition because we suspect that people may adopt very different strategies when a complete record of previous cases is provided: In the extreme case, surely participants will maximize if they are told the complete structure of the task and the payoff probabilities. The effects of providing a score, however, strongly suggest that one reason why probability matching occurs in many situations is because participants have not been adequately motivated to search for the optimal response strategy, and that when appropriate outcome feedback is provided, maximizing might be observed.

Despite these findings, Friedman and Massaro failed to obtain any clear evidence of optimal responding. They concluded (p. 385) that '. . . probability matching in binary choice . . . is less robust than most psychologists seem to believe. Even in our noisier treatments, the subjects tended to overshoot—that is, they chose the more likely alternative more often than in probability matching. The noise reduction treatments, history and score, systematically increased overshooting. However, in all treatments, a substantial gap remained between optimal deterministic behavior (always pick the more likely alternative) and typical subject behavior.' It is possible that the relatively small number of trials (480) in this study contributed to the sub-optimal responding that was observed, an issue to which we return in the General Discussion.

The present research was undertaken in order to challenge the pessimistic conclusion that people's natural behavior in probability learning tasks is sub-optimal. In contrast, we are sympathetic to Friedman's (1998, p. 941) assertion that 'every choice ''anomaly'' can be greatly diminished or entirely eliminated in appropriately structured learning environments'. There is some evidence that nonhuman animals are able roughly to maximize reinforcement under concurrent ratio schedules which deliver reinforcements probabilistically (Douglas and Pribram, 1966; Herrnstein and Loveland, 1975), and plainly the learning environment is quite different for a hungry pigeon whose life depends on an adequate intake of food and a student eager to finish a rather boring experiment. There are limits, of course, in the extent to which one can extrapolate across species, but this contrast does nevertheless provide some encouragement for Friedman's position.

We therefore explored simple probability learning tasks in which we provided large performance-related financial incentives, meaningful and regular feedback, and extensive training in the hope of obtaining evidence that this particular choice anomaly, like others, can be eliminated.

EXPERIMENT 1

In the first study participants trained for 300 trials in a binary choice task in which they gained 5 pence for each correct response and lost 5 pence for an incorrect one (5 UK pence = approx. 8 US cents). This means that the available cumulative payment for a maximizing participant is quite large. After each block of 50 trials participants were provided with informative feedback about their performance.

Method

Participants

Sixteen members of the University College London (UCL) community took part in the experiment. Half were male and half female. They had a mean age of 25.4 (range 19–54; $SD = 8.8$).

Procedure

At the start of the experiment half the participants read the following instructions on the computer display. For the remaining participants the instructions were identical except the word 'choice' was replaced by 'bet':

> During this experiment you will be presented with a series of choices. As a result of your choices you can both win and lose money. Choices are made by clicking one of two choice buttons with the mouse. The left button is labelled 'Left' and the right button 'Right'. Once you have clicked on a button, one of two boxes will be filled in BLUE. If the blue box is on the side you chose, then you win. If the blue box is on the other side, then you lose. Each time you win you gain 5 pence. Each time you lose you lose 5 pence. To practice making a few choices, click on the button below.

There then followed five practice trials identical to the main experimental trials, except that participants experienced two 'lose' and three 'win' trials regardless of their choices. The instructions then continued:

> We will now start the experiment which has 300 rounds. You can play it at your own pace, but we hope it doesn't take longer than 30 mins. This would require that you don't take longer than 3 or 4 seconds for each choice. At the end of the 300 rounds the computer will add up your gains and losses. We will then pay you your winnings. In addition, we will pay you £2.00 for participation. [Note that UK £1.00 = 100 pence].
>
> Before we start there is one more thing we want to tell you. Your payoffs in this experiment are determined by a computer program. This program has been written in advance and we shall not interfere with it during the experiment. The program contains chance elements, so the same choice might give different payoffs on different occasions. Your purpose is to make as many correct predictions as possible, so as to maximize your earnings.
>
> THERE IS NO PATTERN OR SYSTEM YOU CAN USE WHICH WOULD MAKE IT POSSIBLE TO GET ALL YOUR ANSWERS CORRECT. BUT YOU WILL FIND THAT YOU CAN IMPROVE YOUR PERFORMANCE IN THE TEST IF YOU PAY ATTENTION AND THINK ABOUT WHAT YOU ARE DOING. Click below to begin the experiment.

Two buttons marked Left and Right were displayed on the screen horizontally and separated by a gap of about 3 cm. Above these buttons were two adjacent unlabelled grey outcome boxes. To initiate each trial the message '!!CHOOSE!!' appeared. Upon a choice, the correct outcome box filled in blue and one of two messages appeared: 'Win 5p' (above the relevant outcome box) or 'Lose 5p' (below the box). After being displayed for 1 second the message was cleared and the next trial commenced.

Every 50 trials there was a pause in which outcome feedback was provided. Participants were told: 'In the last block of trials *x*% of your decisions were correct' where *x* was simply the number of correct responses

expressed as a percentage. Participants were then told: 'Using an optimal strategy you could have got *y*% correct' where *y* was computed by the optimizing algorithm. This algorithm keeps a record of all previous trial outcomes and on the current trial chooses whichever alternative has been rewarded most often in the past (in the case of equal prior payoffs, it chooses Left). At the end of the experiment the following message was presented: 'The experiment is now over. Thank you for participating. Your total winnings are *W* Pence. Your total losses are *L* Pence. Well done, you have won $(W - L)$ Pence.'

Participants were tested individually in a sound-dampened testing room. Each participant was presented with 300 choice trials. The computer was programmed to deliver the outcome with probabilities of 0.7 and 0.3 counterbalanced across participants for left and right responses. These probabilities were strictly independent (i.i.d.) and based on a new random number chosen by the program on each trial.

Results and discussion

Participants earned a mean of £5.29 ($SD = 1.04$) during the experiment, including the £2 turning-up fee.

Exhibit 1 shows the mean proportion of maximizing responses for each block of 10 trials averaged across participants, with the maximizing response defined as the one associated with the higher payoff probability. As is commonly the case, the group data reveal overmatching or overshooting, with a mean proportion in the final block in excess of 0.9. The proportion of maximizing responses increased across blocks, $F(8.7, 130.4) = 7.00$, $p < 0.001$ (with degrees of freedom adjusted according the Greenhouse–Geisser method: Howell, 1992, pp. 446–448).

Exhibits 2 and 3 show individual learning curves for the 16 participants, designated P1–P16, with Exhibit 2 presenting results for the eight participants who received 'choice' instructions and Exhibit 3 the results for the eight participants receiving 'bet' instructions. It is clear that the group data of Exhibit 1 masks considerable variability across participants. If we operationally define maximizing as requiring a run of at least five consecutive blocks (i.e. 50 trials) with *P*(maximizing response) = 1.0, then 6/16 participants (P3, P4, P8, P9, P12, and P13) clearly achieve maximizing.[3] Two further participants (P10 and P16) are borderline,

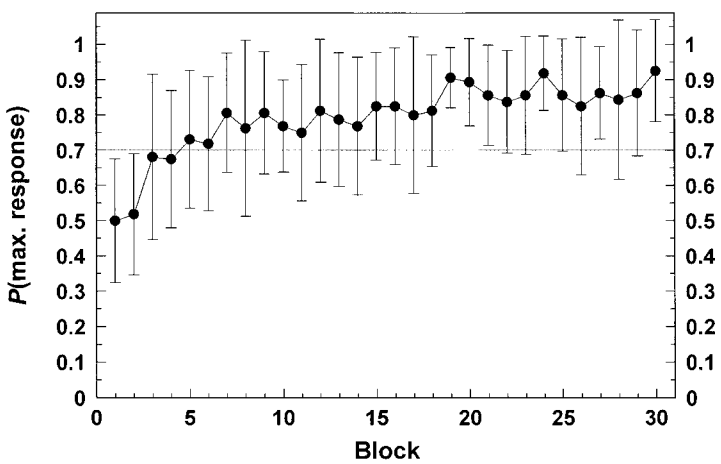

Exhibit 1. Mean proportion of maximizing responses for each block of 10 trials averaged across participants in Experiment 1. The dotted line is the probability matching prediction. Error bars show the standard deviation

[3]Even if a given participant's true asymptotic probability of choosing the maximizing response is as high as 0.9, the likelihood of a run of 50 consecutive choices of that alternative in a block is very low ($p = 0.005$ by a binomial test). Although this depends on the (clearly unrealistic) assumption that the choices are independent, it does indicate that the criterion is quite stringent.

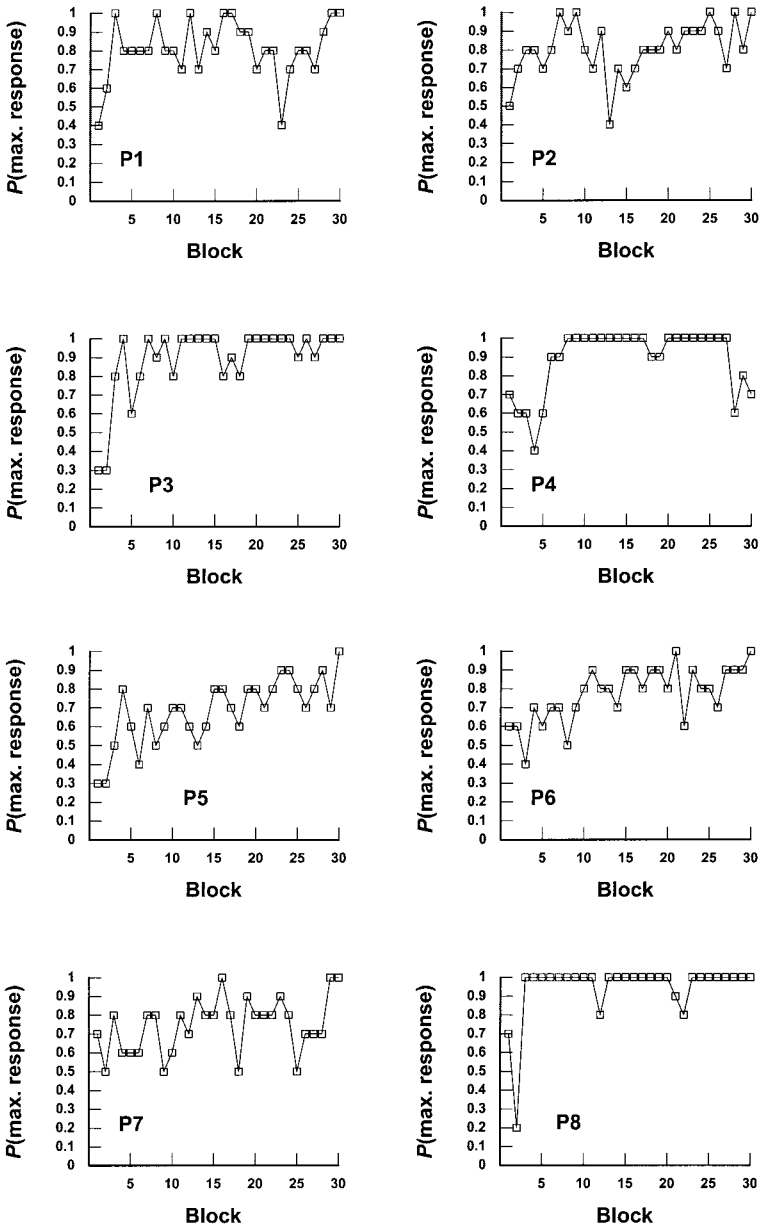

Exhibit 2. Mean proportion of maximizing responses for participants P1–P8 across blocks of 10 trials in Experiment 1

generating runs of at least 30 maximizing responses. The performance of 4/16 participants (P2, P5, P6, and P15) seems not to have reached asymptote, and so their final strategies are indeterminate. Finally, 4/16 participants (P1, P7, P11, and P14) show patterns more consistent with probability matching.

Overall, these results paint a rather more optimistic picture of the rationality of human choice behaviour than is suggested in previous research. If we put aside the four participants whose performance is clearly non-asymptotic, then half (6/12) the remaining participants generated maximizing behaviour.

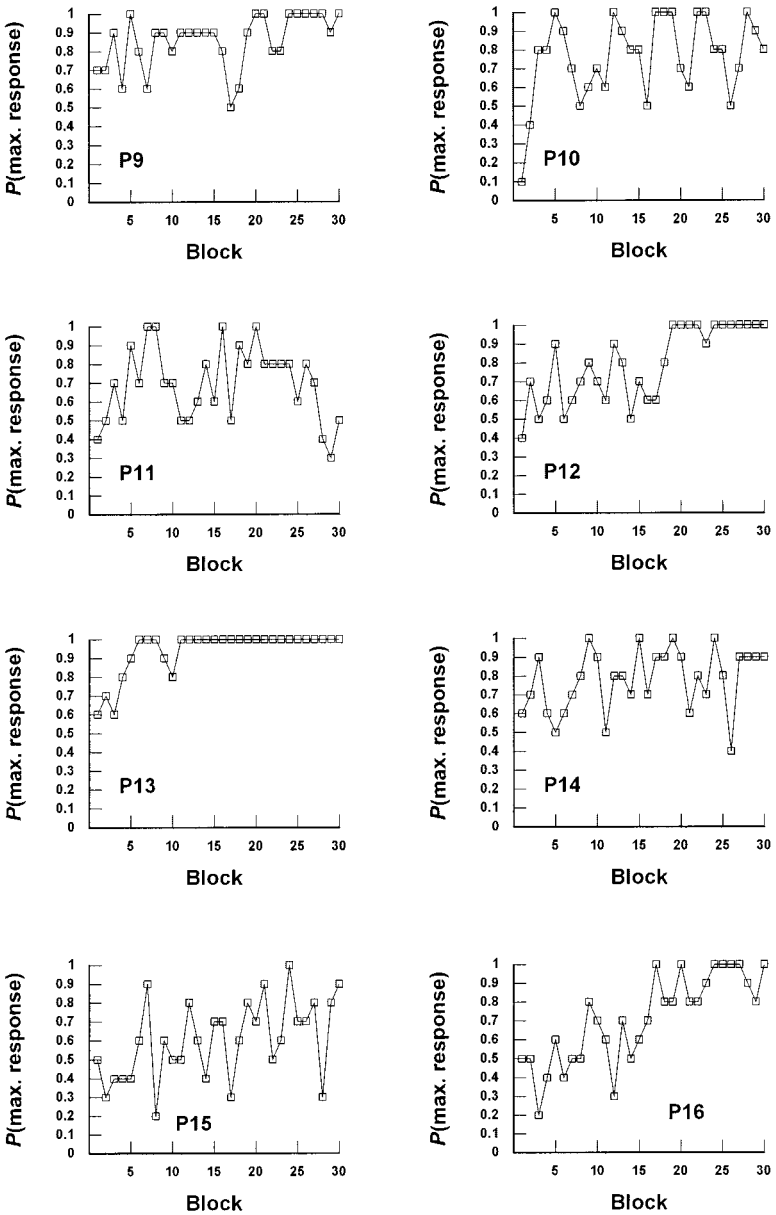

Exhibit 3. Mean proportion of maximizing responses for participants P9–P16 across blocks of 10 trials in Experiment 1

EXPERIMENT 2

Because the responding of several participants in Experiment 1 was plainly not at asymptote, it is natural to predict that longer training will increase the proportion of participants who appear to be maximizing. Therefore in Experiment 2 we ran participants for six times as long in the task. Except where specifically mentioned, this experiment was exactly like Experiment 1 except that participants were given 1800 trials, distributed across six testing sessions, instead of 300.

Method

Participants

Twelve members of the UCL community took part in the experiment. Seven were male and 5 female. None had taken part in Experiment 1. They had a mean age of 28.3 years (range 22–42; $SD = 6.9$).

Procedure

This was the same as in Experiment 1, except that participants took part in six sessions each of 300 trials. Each participant completed two sessions per day over a period of three consecutive days. All participants received the 'choice' rather than the 'bet' instructions.

Results and discussion

Across the six sessions of the experiment, earnings increased steadily. The means (including the £2 turning-up fee) were £5.27 ($SD = 1.40$), £6.10 ($SD = 2.15$), £6.55 ($SD = 1.37$), £7.11 ($SD = 1.23$), £7.01 ($SD = 1.18$), and £7.52 ($SD = 1.06$), respectively, across sessions 1–6. The greatest total amount earned by a participant across the whole experiment was £63.60.

Exhibit 4 shows the mean proportion of maximizing responses across blocks of 50 trials averaged across participants, with the maximizing response again defined as the one associated with the higher payoff probability. As in Experiment 1 the group data reveal overmatching, with a mean proportion in the final block in excess of 0.9. The proportion of maximizing responses increased across blocks, $F(5.1, 56.2) = 8.97$, $p < 0.001$.

Exhibits 5 and 6 show individual learning curves for the 12 participants (designated P1–P12). It is clear that the group data of Exhibit 4 again mask considerable variability across participants. If we operationally define maximizing as requiring at least one 50-trial block with P(maximizing response) = 1.0, then 8/12 participants (P1, P4, P5, P7, P8, and P10–P12) clearly achieve maximizing. As expected, this proportion is higher than that achieved in Experiment 1. Note, moreover, that this criterion is somewhat stricter than the one adopted in Experiment 1 because here we require the 50 consecutive trials to all fall within one block whereas in Experiment 1 we included any sequence of five 10-trial blocks instead. Perhaps the most impressive responding is that observed in participant P5 who during the second half of the experiment (900 trials) only chose the low-probability response on one occasion.

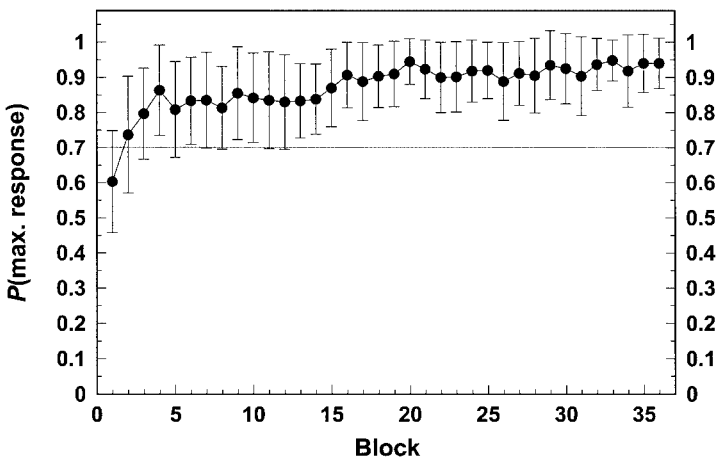

Exhibit 4. Mean proportion of maximizing responses for each block of 50 trials averaged across participants in Experiment 2. The dotted line is the probability matching prediction. Error bars show the standard deviation

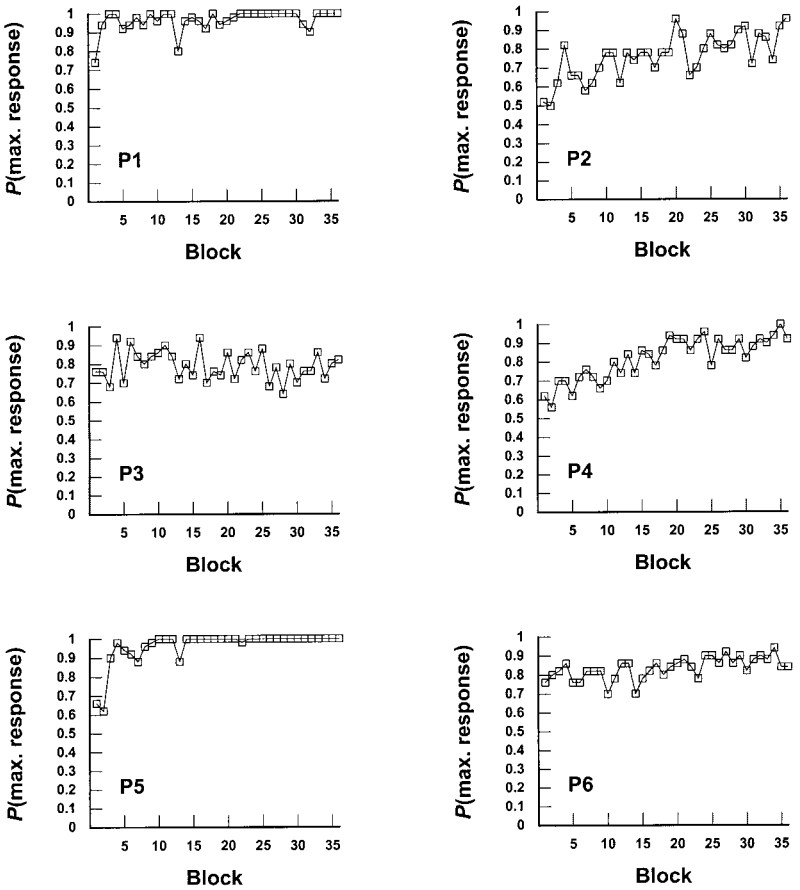

Exhibit 5. Mean proportion of maximizing responses for participants P1–P6 across blocks of 50 trials in Experiment 2

Of the remaining 4/12 participants (P2, P3, P6, and P9) all but one (P2) seem to have reached asymptote and show patterns more consistent with probability matching.

The results from Experiment 2 are fairly clear in demonstrating probability maximizing in a sizable proportion (two-thirds) of participants. This led to extremely stable responding on a single choice option across large numbers of trials. Nevertheless, about a third of participants continued to distribute their responses across both choice options despite the fact that this lost them money.

EXPERIMENT 3

In Experiment 3 we examine more directly the effects of feedback and reinforcement on how participants learn to allocate responses. Participants in the *Payoff* conditions received payoffs that were contingent upon their allocation of choices: 3 pence for a 'correct' response and −3 pence for an 'incorrect' response, slightly less than the amount (±5 pence) used in Experiment 2. Participants who matched their responses to the payoff probabilities could expect to earn approximately £3.60 for each session, while those who allocated all of their responses to the 'correct' alternative could expect approximately £9.00 for each session. Half of these participants received feedback (*Payoff + Feedback*) at the end of each 50-trial block that indicated how many pence they had won relative to how many pence they could have won (approximately 60 pence), calculated

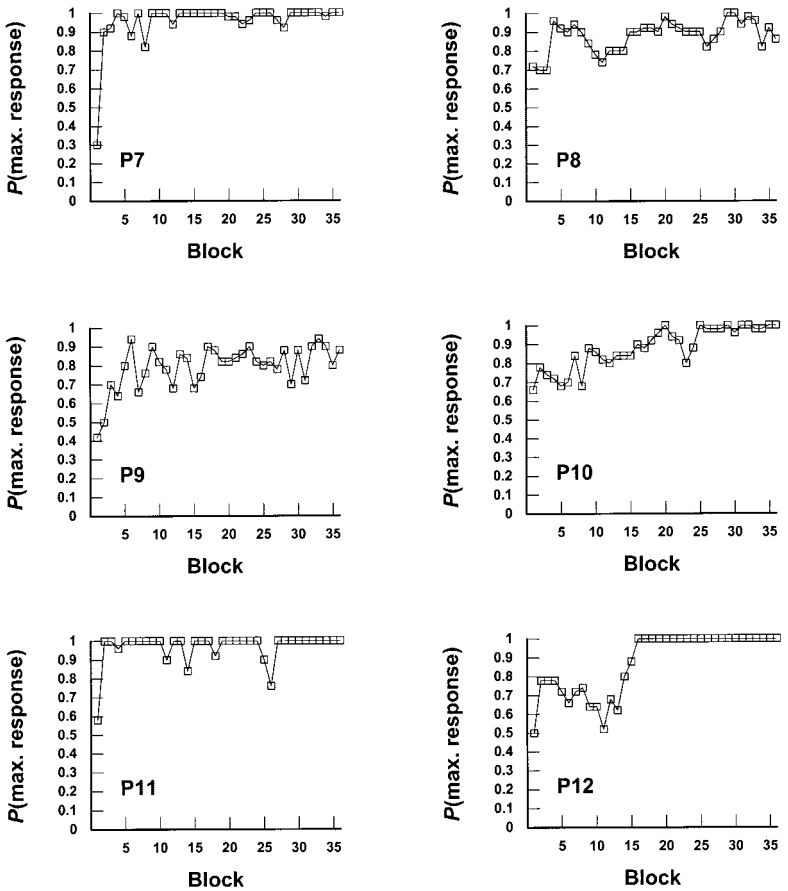

Exhibit 6. Mean proportion of maximizing responses for participants P7–P12 across blocks of 50 trials in Experiment 2

as before by the optimizing algorithm. This group was identical to those in Experiments 1 and 2. Participants in the *Payoff + No Feedback* group received contingent payoffs but did not receive any additional feedback.

Participants in the *No Payoff* conditions received a flat payment of £3.00 for each session. Throughout the experiment the word 'points' was substituted for 'pence'. One group (*No Payoff + Feedback*) received feedback that indicated how many points they had won relative to how many points they could have won. The remaining group (*No Payoff + No Feedback*) received no additional feedback.

Method

Participants

Forty-eight members of the UCL community took part in this experiment. None had previously taken part in any similar experiment. Twenty were male and 28 female. They had a mean age of 22.8 years (range 17–60; $SD = 7.9$).

Design

This was a factorial design with two between-subject factors, payoff (*Payoff* versus *No Payoff*) and feedback (*Feedback* versus *No Feedback*). Participants trained for 1500 trials (slightly less than the 1800 presented in Experiment 2) distributed across 2 sessions.

Procedure

The procedure was the same as in Experiments 1 and 2, with the exception that participants took part in two sessions each of 750 trials separated by 7 days. For the participants in the non-contingent payoff condition 'pence' was substituted for 'points' in both the instructions and throughout the experiment.

Results and discussion

The number of points scored by participants in the *No Payoff* conditions increased across sessions: 520.7 ($SD = 174.2$) and 594.0 ($SD = 222.1$) on Sessions 1 and 2, respectively, for the *No Feedback* group, and 563.4 ($SD = 178.2$) and 680.4 ($SD = 181.3$) on Sessions 1 and 2 for the *Feedback* group. Similarly the number of pence earned by participants in the *Payoff* conditions increased across sessions (here we have converted pence into pounds): £5.20 ($SD = 1.74$) and £5.94 ($SD = 2.22$) on Sessions 1 and 2 for the *No Feedback* group, and £5.63 ($SD = 1.78$) and £6.80 ($SD = 1.61$) on Sessions 1 and 2 respectively for the *Feedback* group. The greatest amount earned by any one participant across the experiment was £18.30.

Exhibit 7 shows the mean proportion of maximizing responses across each block of 100 trials. These data were entered into separate ANOVAs for each session. The degrees of freedom for within-subjects effects and interactions were adjusted according the Greenhouse-Geisser method.

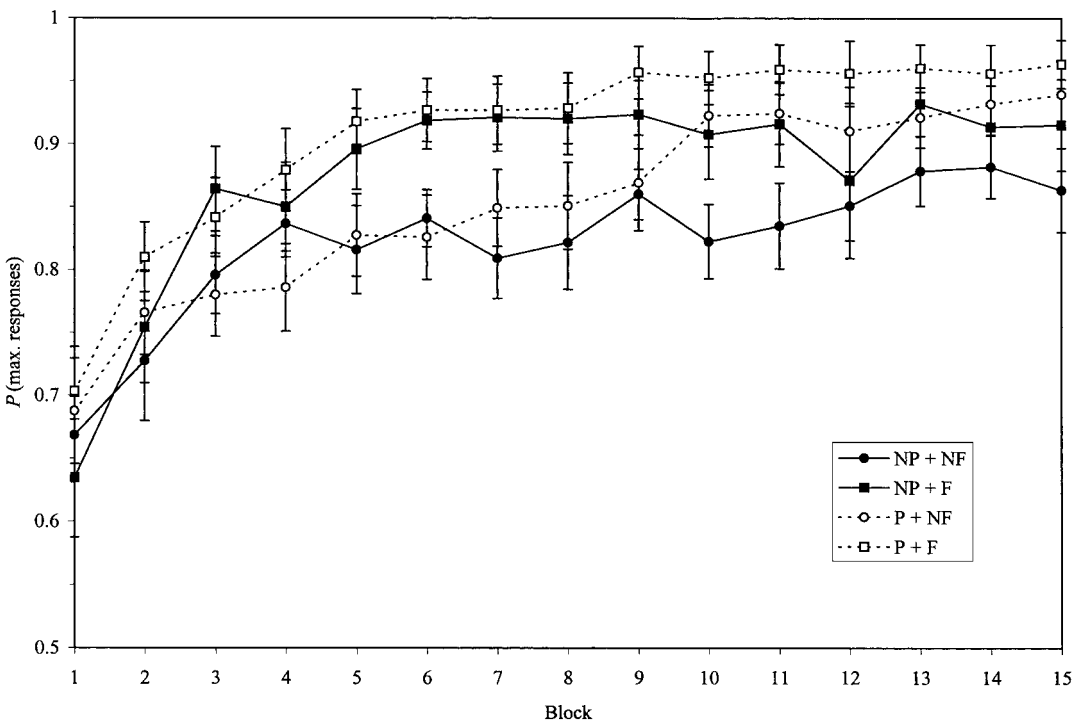

Exhibit 7. Mean proportion of maximizing responses for each group across 100-trial blocks in Experiment 3. Key: NP + NF = No Payoff + No Feedback, NP + F = No Payoff + Feedback, P + NF = Payoff + No Feedback, P + F = Payoff + Feedback. Error bars show standard errors

Session 1

The ANOVA revealed an effect of Block indicating that the proportion of maximizing responses increased across the session, $F(8.0, 349.6) = 33.79$, $p < 0.01$, an effect of Feedback due to the fact that participants who received feedback made more maximizing responses than those who did not, $F(1, 44) = 6.89$, $p = 0.01$, and an interaction between Block and Feedback suggesting that participants learned faster when feedback was provided, $F(8.0, 349.6) = 2.23$, $p < 0.03$. However, there was no effect of Payoff indicating that participants who received contingent payoffs did not make any more maximizing responses than those who did not ($F < 1$). There was no interaction between Payoff and Block, $F(8.0, 349.6) = 1.19$, $p > 0.05$: hence Payoff did not affect learning rate. Similarly, Payoff did not interact with Feedback ($F < 1$). The three-way interaction between Block, Payoff, and Feedback was not significant, $F(8.0, 349.6) = 1.23$, $p > 0.05$.

Session 2

The corresponding ANOVA on Session 2 data revealed that participants had reached asymptote as there was no effect of Block, $F(7.0, 308.1) = 1.84$, $p > 0.05$, nor any interactions involving Block, largest $F(7.0, 308.1) = 1.35$, $p > 0.05$. Main effects of both Payoff and Feedback indicated that each determined asymptotic performance, $F(1, 44) = 6.30$, $p < 0.02$; and $F(1, 44) = 6.81$, $p < 0.01$, respectively. However, Block and Feedback did not interact, $F < 1$.

The histograms in Exhibit 8 show how the participants in each group are classified according to their proportions of maximizing responses in the final block of 50 trials of the experiment. It is clear that the fewest maximizers (i.e. participants who choose the maximizing response on every trial of the final block) are in the *No Payoff + No Feedback* group and the most in the *Payoff + Feedback* group, with the other groups in

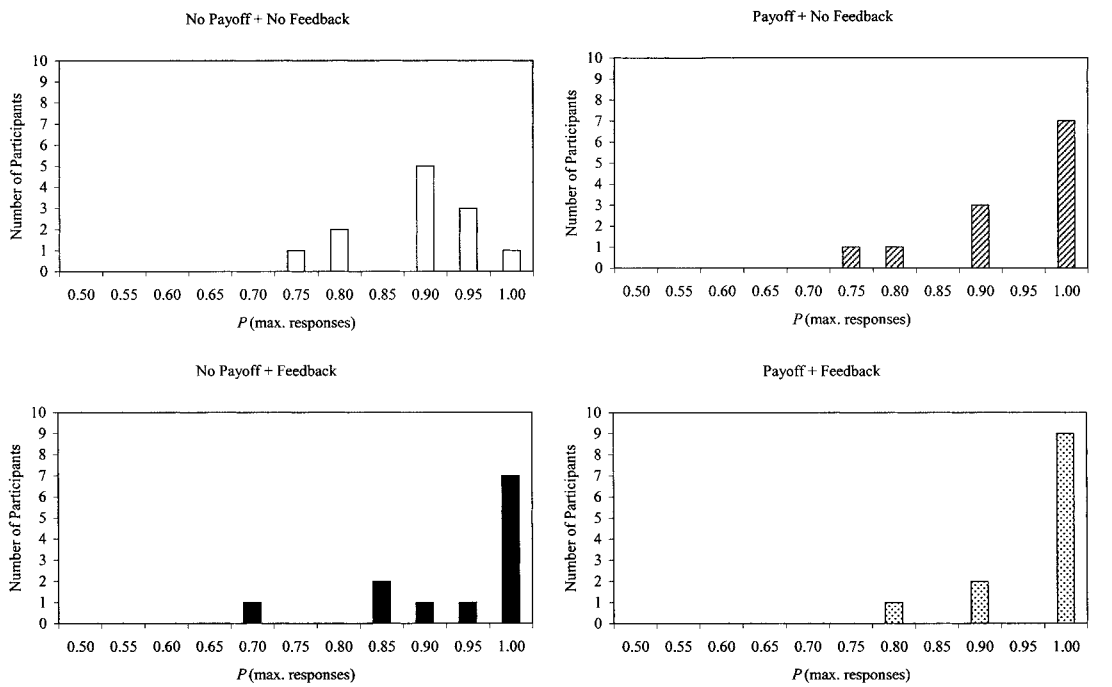

Exhibit 8. Classification of participants in each group (n = 12) according to their proportion of maximizing responses in the final block of 100 trials in Experiment 3

between. In the *Payoff + Feedback* group, 9/12 (75%) participants were maximizers which closely replicates the figure in Experiment 2 (66%), despite the slight reduction in the number of trials and magnitude of payoff. Combining these data we obtain a mean percentage of 71% with 95% confidence limits of ±18%. Thus a conservative estimate from our data of the proportion of maximizers is slightly more than half (53%) and the true proportion may be as high as 89%.

Lastly, the results confirm that maximizing is the optimal thing to do in the sense that it increases the amount of money earned. For participants in the *Payoff* groups, maximizers earned a mean of £14.14 ($SD = 2.70$) in the experiment while the remaining participants earned a mean of £12.64 ($SD = 1.97$).

These results combine to confirm that the provision of feedback and monetary payoffs are important determinants of the level of asymptotic performance participants can achieve in probability matching experiments. Moreover, in Session 1 feedback (but not payoff) affected the rate of learning. We conclude that the relatively high prevalence of maximizing in the previous experiments was in part due to these factors.

GENERAL DISCUSSION

Friedman (1998, p. 941) has asserted that virtually every choice 'anomaly' can be greatly diminished or entirely eliminated in appropriately structured learning environments. In the present research we sought to establish if this is the case for a simple choice situation in which, according to a wealth of previous research, people's natural behavior is sub-optimal. We therefore conducted probability learning experiments in which we provided (i) large financial incentives, (ii) meaningful and regular feedback, and (iii) extensive training in the hope of providing evidence that this particular choice anomaly can be eliminated. The results are fairly clear in demonstrating that large proportions of participants (71% ± 18% across Experiment 2 and the *Payoff + Feedback* group of Experiment 3) can maximize their payoffs when maximizing is defined as a run of at least 50 consecutive choices of the optimal response. Many participants quite comfortably exceeded this criterion. Each of the three factors mentioned above contributed to participants' ability to maximize. Using a factorial design, Experiment 3 demonstrated that both feedback and payoffs affected asymptotic choice levels. Similarly, the data reveal that stable performance is not reached until after many hundreds of trials.

Although outcome feedback appears to affect asymptotic performance, previous research has suggested that compared to other forms of information outcome feedback is quite limited in its usefulness, particularly in comparison to cognitive feedback (Balzer *et al.*, 1989). Cognitive feedback refers to information about the relations between responses and outcomes (functional validity information), the relations between responses and cue values (cognitive information), or a summary of the relations between cues and outcomes (task information). In MCPL studies, task information appears to be particularly useful (Balzer *et al.*, 1992; Friedman and Massaro, 1998). It would hence be interesting in future studies to see whether varieties of cognitive feedback have a beneficial effect in experiments of the sort reported here over and above outcome feedback.

One might think that there is something curious about needing several hundred trials to reach asymptotic levels of performance in these experiments. Surely it is rare in real life to encounter the same decision so many times under identical conditions? A couple of points need to be made concerning the ecological validity of experiments using such large amounts of practice. First, trials in the present experiments were massed into a very short period of time yet people are known to learn much faster—often *very* much faster—with distributed than with massed trials (Dempster, 1996). Hence if only one choice was made each day, as might be more representative of real-world decision making, we would expect a much faster learning rate across trials. Second, apart from the simplicity of the decision environment, the present results are consistent with a vast amount of evidence showing that learning is indeed quite slow in many circumstances. Many perceptual classification tasks require hundreds of trials before they are mastered (e.g. Ashby and Maddox, 1992; Biederman and Shiffrar, 1987). Thus the slowness of learning in the present studies is not as surprising as

it might at first appear, nor does this slowness necessarily imply that the evidence we have presented here for maximizing is irrelevant to real-world decision making.

These experiments used monetary payoffs quite considerably larger than those used in previous research. For example, the average total amount earned by participants in Experiment 2 was almost £40 (= approx. US$64). Although Friedman and Massaro (1998) failed to find any clear effect of monetary payoffs in their study, the probability learning literature does document many positive effects of payoffs (Siegel and Goldstein, 1959; Vulkan, 2000) and the results of Experiment 3 confirm that the maximizing behavior seen in the present experiments is at least in part attributable to the financial incentives we provided. An unexpected finding of Experiment 3 was that payoff only affected performance in later trials, with no significant effect in earlier trials. This may suggest another reason for the apparent inconsistency between the results of Experiment 3 and those of Friedman and Massaro: their study only used 480 trials, compared to 1500 in Experiment 3. The inconsistency therefore disappears if we compare their results to those of Session 1 in Experiment 3: in both cases, payoff had no effect. It would be interesting to repeat Friedman and Massaro's experiment to see whether a payoff effect emerges with more choice trials.

We were not able to obtain maximizing in all participants, however, and it remains an interesting question for future research why this is. We offer three reasons why a given participant might not maximize and which would not involve giving up rational choice theory:

(1) Our results show very clearly that many trials are needed to learn the optimal strategy. Rates of learning differ considerably across individuals and hence nonmaximizers may be ones who tend to learn more slowly (and this in turn may relate to psychometric variables such as intelligence and memory capacity).
(2) Our results also show that monetary payoff affects performance. Although we have not shown that the magnitude of payoff matters, it seems quite likely that it does. However, it also seems likely that individuals differ in their utility functions for money: a millionaire is likely to care less about the trial payoffs than a pauper. Nonmaximizers may therefore be individuals who, in terms of rational choice theory, get less utility from the monetary payoffs available than from other sources of utility such as predicting the infrequent outcome or relieving boredom.
(3) Feedback also affects performance, and individuals may be differentially sensitive to the motivating properties of feedback.

Although the question that motivated this study is an empirical one, it is nonetheless worth briefly considering what the theoretical implications of our findings are, and in particular what bearing the data have on some of the current models of choice behavior. We will consider three such models. The first is the melioration account of the matching law developed by Herrnstein (1997). According to this account, organisms allocate their responses so as to equate the reinforcement rate per unit of consumption across the various choice options. Melioration achieves this outcome by proposing that at any moment the alternative with the highest reinforcement rate is chosen until such time as the other alternative acquires a higher reinforcement rate or the alternatives are exhausted. Since the maximizing response always has the highest momentary reinforcement rate in the sorts of probability learning tasks we have studied here, melioration and the matching law predict maximizing at asymptote (Herrnstein and Loveland, 1975). (The term 'matching law' is distinctly confusing since it does not predict probability matching in the sort of task used here. Instead it refers to the organism's matching of the relative frequency of choosing each alternative to the relative frequency of reinforcement). Thus, although there is some doubt about the validity of the matching law and its instantiation via the melioration mechanism as a general theory of human choice (see Savastano and Fantino, 1994; Tunney and Shanks, in press; Williams, 1994), it comfortably accounts for the behavior of those participants who maximize in the present experiments. On the other hand, it is not clear on this account why feedback and monetary payoff affect asymptotic levels of performance.

The melioration model is not a theory of learning: It does not describe how the individual comes to know which choice alternative has the higher likelihood of reinforcement on the current trial and it says nothing about how beliefs are updated. A second account that is consistent with the data but which deals explicitly with belief-updating is the reinforcement learning model of Cross (1973), Roth and Erev (1995), Erev and Roth (1998), and Bereby-Meyer and Erev (1998). According to the basic version of this model, a choice that is reinforced receives a fixed increment in its response 'propensity' q. This learning rule is combined with a probabilistic response rule according to which the probability of choosing, say, left (P_L) is a function of its relative propensity:

$$P_L = \frac{q_L}{q_L + q_R}$$

Like the matching law, this model says that the choice probabilities match the ratio of the accumulated reinforcements. At asymptote, this choice probability approaches 1.0 for the maximizing response under the conditions of the present experiments.

As with the melioration account, however, the effects of feedback and monetary payoff on asymptotic levels of performance are not explained.

The final theoretical approach comprises a family of connectionist learning models in which connections between representations of cues and outcomes are updated via an error-correcting rule (Friedman *et al.*, 1995; Gluck and Bower, 1988; Shanks, 1990). For Experiment 1, for example, we can think of left and right as separate cues each of which forms an association with reinforcement. On each trial t the connection weights $w_i (i = \{L, R\})$ are updated according to the rule:

$$w_i^t = w_i^{t-1} + \lambda\left(d - w_i^{t-1}\right)s_i^{t-1}$$

where λ is a learning rate parameter, d is the reinforcement (1 if the payoff is positive, 0 otherwise), and s_i ($=1$ if i is the chosen alternative, 0 otherwise) is an indicator which restricts weight changes to the chosen alternative. At asymptote the weights will have mean values of 0.7 for the alternative with the higher payoff probability and 0.3 for the low payoff one. The weights are converted to response probabilities via the logistic rule $P(R_i) = 1/[1 + e^{-\theta(w_L - w_R)}]$, where θ is a scaling parameter. If θ is small (<1), responding is predicted to be quite close to probability matching (see Friedman and Massaro, 1998, Fig. 2), but provided θ is sufficiently large, maximizing is predicted since $P(R_i)$ follows a step function with the step at 0.5.

Although this family of connectionist models is known to have some shortcomings as a general account of probability learning (López *et al.*, 1998), previous studies have generally obtained good fits to data collected from participants performing the medical diagnosis task (Friedman *et al.*, 1995; Kitzis *et al.*, 1998; Nosofsky *et al.*, 1992). Another significant attraction compared to the other models is that the effects of feedback and monetary payoff on asymptotic levels of performance can be modeled via appropriate changes in the parameters λ, d, and θ. For instance, a larger payoff would translate into a larger value of d, feedback would increase θ, and individual differences in learning rate would be captured by variations in λ.

In sum, there are a variety of models which can account for the evidence we have obtained here of probability maximizing, although the effects of feedback and monetary payoff remain poorly understood from a theoretical point of view. We conclude that probability matching is another example of a choice anomaly which is heavily context-dependent and which can be made to disappear under appropriate conditions of task structure, training, motivation, and feedback. However, further research will be needed to develop and distinguish between the various models capable of accounting for choice behavior in these conditions.

ACKNOWLEDGEMENTS

The research described here was supported by grants from the United Kingdom Economic and Social Research Council (ESRC) and the Leverhulme Trust. The work is part of the programme of the ESRC Centre for Economic Learning and Social Evolution, University College London. We thank Ken Binmore, Dan Friedman, Nigel Harvey, and Ben Newell for their helpful comments.

REFERENCES

Arrow KJ. 1958. Utilities, attitudes, choices: A review note. *Econometrica*, **26**: 1–23.

Ashby FG, Maddox WT. 1992. Complex decision rules in categorization: contrasting novice and experienced performance. *Journal of Experimental Psychology: Human Perception and Performance*, **18**: 50–71.

Balzer WK, Doherty ME, O'Connor R. 1989. Effects of cognitive feedback on performance. *Psychological Bulletin*, **106**: 410–433.

Balzer WK, Sulsky LM, Hammer LB, Sumner KE. 1992. Task information, cognitive information, or functional validity information: which components of cognitive feedback affect performance? *Organizational Behavior and Human Decision Processes*, **53**: 35–54.

Beach LR, Swensson RG. 1967. Instructions about randomness and run dependency in two-choice learning. *Journal of Experimental Psychology*, **75**: 279–282.

Bereby-Meyer Y, Erev I. 1998. On learning to become a successful loser: a comparison of alternative abstractions of learning processes in the loss domain. *Journal of Mathematical Psychology*, **42**: 266–286.

Berry DA, Fristedt B. 1985. *Bandit Problems: Sequential Allocation of Experiments*. Chapman and Hall: London.

Biederman I, Shiffrar MM. 1987. Sexing day-old chicks: a case study and expert systems analysis of a difficult perceptual-learning task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **13**: 640–645.

Braveman NS, Fischer GJ. 1968. Instructionally induced strategy and sequential information in probability learning. *Journal of Experimental Psychology*, **76**: 674–676.

Camerer C. 1995. Individual decision making. In *The Handbook of Experimental Economics* (pp. 587–703), Kagel JH, Roth AE (eds). Princeton University Press: Princeton, NJ.

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

Cross JG. 1973. A stochastic learning model of economic behavior. *Quarterly Journal of Economics*, **87**: 239–266.

Dempster FN. 1996. Distributing and managing the conditions of encoding and practice. In *Memory* (pp. 317–344), Bjork EL, Bjork RA (eds). Acadamic Press: San Diego.

Douglas RJ, Pribram KH. 1966. Learning and limbic lesions. *Neuropsychologia*, **4**: 197–220.

Edwards W. 1961. Probability learning in 1000 trials. *Journal of Experimental Psychology*, **62**: 385–394.

Erev I, Roth AE. 1998. Predicting how people play games: reinforcement learning in experimental games with unique, mixed strategy equilibria. *American Economic Review*, **88**: 848–881.

Estes WK. 1956. The problem of inference from curves based on group data. *Psychological Bulletin*, **53**: 134–140.

Estes WK, Campbell JA, Hatsopoulos N, Hurwitz JB. 1989. Base-rate effects in category learning: a comparison of parallel network and memory storage-retrieval models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **15**: 556–571.

Fiorina MP. 1971. A note on probability matching and rational choice. *Behavioral Science*, **16**: 158–166.

Friedman D. 1998. Monty Hall's three doors: construction and deconstruction of a choice anomaly. *American Economic Review*, **88**: 933–946.

Friedman D, Massaro DW. 1998. Understanding variability in binary and continuous choice. *Psychonomic Bulletin & Review*, **5**: 370–389.

Friedman D, Massaro DW, Kitzis SN, Cohen MM. 1995. A comparison of learning models. *Journal of Mathematical Psychology*, **39**: 164–178.

Gluck MA, Bower GH. 1988. From conditioning to category learning: an adaptive network model. *Journal of Experimental Psychology: General*, **117**: 227–247.

Healy AF, Kubovy M. 1981. Probability matching and the formation of conservative decision rules in a numerical analog of signal detection. *Journal of Experimental Psychology: Human Learning and Memory*, **7**: 344–354.

Herrnstein RJ. 1997. *The Matching Law: Papers in Psychology and Economics*, Rachlin H, Laibson DI (eds). Harvard University Press: Cambridge, MA.

Herrnstein RJ, Loveland DH. 1975. Maximizing and matching on concurrent ratio schedules. *Journal of the Experimental Analysis of Behavior*, **24**: 107–116.
Hertwig R, Ortmann A. 2001. Experimental practices in economics: a methodological challenge for psychologists? *Behavioral and Brain Sciences*, **24**: 383–451.
Howell DC. 1992. *Statistical Methods for Psychology.* Duxbury Press: Belmont, CA.
Kitzis SN, Kelley H, Berg E, Massaro DW, Friedman D. 1998. Broadening the tests of learning models. *Journal of Mathematical Psychology*, **42**: 327–355.
López FJ, Shanks DR, Almaraz J, Fernández P. 1998. Effects of trial order on contingency judgments: a comparison of associative and probabilistic contrast accounts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **24**: 672–694.
Myers JL. 1976. Probability learning and sequence learning. In *Handbook of Learning and Cognitive Processes: Approaches to Human Learning and Motivation* (pp. 171–205), Estes WK (ed.). Erlbaum: Hillsdale, NJ.
Myers JL, Cruse D. 1968. Two-choice discrimination learning as a function of stimulus and event probabilities. *Journal of Experimental Psychology*, **77**: 453–459.
Myers JL, Lohmeier JH, Well AD. 1994. Modeling probabilistic categorization data: exemplar memory and connectionist nets. *Psychological Science*, **5**: 83–89.
Neimark ED, Shuford EH. 1959. Comparison of predictions and estimates in a probability learning situation. *Journal of Experimental Psychology*, **57**: 294–298.
Nosofsky RM, Kruschke JK, McKinley SC. 1992. Combining exemplar-based category representations and connectionist learning rules. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **18**: 211–233.
Roth AE, Erev I. 1995. Learning in extensive-form games: experimental data and simple dynamic models in the intermediate term. *Games and Economic Behavior*, **8**: 164–212.
Savastano HI, Fantino E. 1994. Human choice in concurrent ratio-interval schedules of reinforcement. *Journal of the Experimental Analysis of Behavior*, **61**: 453–463.
Shanks DR. 1990. Connectionism and the learning of probabilistic concepts. *Quarterly Journal of Experimental Psychology*, **42A**: 209–237.
Shanks DR. 1991. A connectionist account of base-rate biases in categorization. *Connection Science*, **3**: 143–162.
Siegel S. 1961. Decision making and learning under varying conditions of reinforcement. *Annals of the New York Academy of Sciences*, **89**: 766–783.
Siegel S, Goldstein DA. 1959. Decision-making behavior in a two-choice uncertain outcome situation. *Journal of Experimental Psychology*, **57**: 37–42.
Tunney RJ, Shanks DR. A re-examination of melioration and rational choice. *Journal of Behavioral Decision Making*, in press.
Tustin RD, Morgan P. 1985. Choice of reinforcement rates and work rates with concurrent schedules. *Journal of Economic Psychology*, **6**: 109–141.
Vulkan N. 2000. An economist's perspective on probability matching. *Journal of Economic Surveys*, **14**: 101–118.
Williams BA. 1994. Reinforcement and choice. In *Animal Learning and Cognition* (pp. 81–108), Mackintosh NJ (ed.). Academic Press: San Diego, CA.

Authors' biographies:

David Shanks (PhD 1985, University of Cambridge) is Professor of Experimental Psychology at University College London and Scientific Director of the ESRC Centre for Economic Learning and Social Evolution. His research interests include human learning, the implicit–explicit distinction, computational modeling, especially with neural network models, economic psychology, and game theory.

Richard Tunney (DPhil 1999, University of York) is Senior Research Fellow at the ESRC Center for Economic Learning and Social Evolution. His interests include human learning, judgment and decision making, and economic psychology.

John McCarthy (PhD 1996, University College London) is a Research Fellow in the Computer Science Department at University College London working on the British Telecom-funded Higherview Project. His interests include distributed computing architectures, collaborative filtering, attention, and quality of service.

Authors' address:

David Shanks, **Richard Tunney**, and **John McCarthy**, Department of Psychology, University College London, Gower Street, London WC1E 6BT, UK.

