AN ECONOMIST'S PERSPECTIVE ON PROBABILITY MATCHING

Nir Vulkan

Economics Department, University of Bristol
Centre for Economic Learning and Social Evolution,
University College London.

Abstract. The experimental phenomenon known as 'probability matching' is often offered as evidence in support of adaptive learning models and against the idea that people maximise their expected utility. Recent interest in dynamic-based equilibrium theories means the term re-appears in Economics. However, there seems to be conflicting views on what is actually meant by the term and about the validity of the data.

The purpose of this paper is therefore threefold: First, to introduce today's readers to what is meant by probability matching, and in particular to clarify which aspects of this phenomenon challenge the utility-maximisation hypothesis. Second, to familiarise the reader with the different theoretical approaches to behaviour in such circumstances, and to focus on the differences in predictions between these theories in light of recent advances. Third, to provide a comprehensive survey of repeated, binary choice experiments.

Keywords. Probability matching; Stochastic learning; Optimisation

1. Introduction

Do people make choices that maximise their expected utility? By and large, economists believe that they do, especially in those cases where the underlying decision situation is simple and is repeated often enough. Somehow people learn how to choose optimally. In introductory courses we teach our students that equilibrium is reached by a process of bayesian-style belief updating, or a process of imitation, or reinforcement-type learning, or even by the replicator dynamics. However, with the notable exception of Cross (1973), it is only recently that economists have started to study seriously these models and have attempted to explain behaviour in terms of the underlying dynamic process which may, or may not, lead to equilibrium. Experimental game theory, behavioural economics, and evolutionary economics all focus on the learning process and on the effects it might have on behaviour in the steady state. An explosion of models in which an individual learner is faced with an uncertain environment had recently been developed (e.g. McKelvey & Palfrey, 1995, Fudenberg & Levine, 1997, Erev &

0950-0804/00/01 0101–18 JOURNAL OF ECONOMIC SURVEYS Vol. 14, No. 1

Roth, 1997, Roth & Erev, 1995, Camerer & Ho, 1996, Chen & Tang, 1996, and Tang, 1996). A one-player decision problem with a move of Nature often provides a simple case where the basics of these learning models can be expressed and tested. In this paper I restrict attention to this seemingly simple case.

Naturally, the study of human learning falls within the scope of another social science, psychology. Furthermore, mathematical learning theories date back to the early 1950s (starting with Estes' seminal 1950 paper), when, for about two decades, they have played a major role in the research agenda of experimental and theoretical psychology. During the 1950s and 1960s a large data set was collected about the behaviour of humans, rats and pigeons in repeated choice experiments. At a typical experiment each subject at each trial has to predict whether a light would appear on his left or his right, whether the next card will be blue of yellow, or any other of many mutually-exclusive choice situations. Which light actually appears (or card etc.) depends on a random device operating with fixed probabilities which are independent of the history of outcomes and of the behaviour of the subject. The experiment then continues for many trials. In some experiments subjects received (small) monetary rewards for making the correct prediction, and in a few of those experiments, had to pay a small penalty for making the wrong prediction.

A striking feature of this data is that subjects match the underlying probabilities of the two outcomes. Denote by $p(L)$ the (fixed) probability with which the random device picks the option Left (alternatively, if the sequence of outcomes is predetermined, $p(L)$ denotes the proportion of Left), then after a period of learning, subjects choose Left in approximately $p(L)$ of the trials. Notice that matching suggests that subjects had learnt the underlying probabilities. But if those probabilities are known then the strategy that maximises expected utility is to always choose the side which is chosen with probability greater than one half. Now if people are not able to maximise utility in this simple setting, can we reasonably expect them to do so in more complicated situation? This was recognised as a challenge to economists already in 1958 by Kenneth Arrow who wrote: '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 behaviour. I suggest that this result points out strongly the importance of learning theory, not only in the greater understanding of the dynamics of economic behaviour, but even in suggesting that equilibria maybe be different from those that we have predicted in our usual theory.' (Arrow, 1958, p. 14). Moreover, stochastic learning theories (like Bush and Mosteller, 1955) which assume only that a person is more likely to choose an option in the future if he receives a positive feedback (the so called 'Law of Effect'), *do* predict this kind of behaviour (the 'probability matching theorem' of Estes and Suppes, 1959). So who is right?

The purpose of this survey is threefold: First, to introduce today's readers to what is meant by probability matching, and in particular to clarify which aspects of this phenomenon challenge the utility-maximisation hypothesis. Second, to familiarise the reader with the different theoretical approaches to behaviour in such circumstances, and to focus on the differences in predictions between these theories in light of recent advances (such as Borgers & Sarin, 1993, 1995 and Borgers, Morales and Sarin, 1997). Third, to provide a comprehensive survey of repeated, binary choice experiments. Although these experiments have been surveyed before, my goal is to provide a complete and unbiased 'survey of surveys' of these results (previous surveys, like Edwards, 1956, Fiorina, 1971 and Brackbill & Bravos, 1962 focus only on specific set of experiments, and within the context of their own theories).

With respect to my first objective, I note that the terms 'probability matching' and 'matching law' are sometimes confused: The behaviour known as probability matching is explained in details in sections 2 and 3. The 'matching law' and other types of behaviour associated in the literature with the term 'matching', but which do not conflict with the assumption of utility (or reward) maximisation, are described in some detailed in the appendix.

The results, in section 3, show that if the experiment is repeated often enough and/or if subjects are paid enough, they tend to asymptotically chose the side which maximises their expected reward, although humans appear to be very slow learners. Moreover, looking at the group's average behaviour over a relatively small number of trials is likely to generate results supporting the matching hypothesis. Probability matching is therefore not a robust prediction of asymptotic behaviour in these settings.

I show that this experimental setting is not as simple as one might think. First, we do not have a learning theory which predicts optimisation with probability 1 in this setting — impatient, but rational, decision makers could end up choosing the wrong side forever. Second, although the data supports the fact that subjects condition their behaviour on the outcomes of the last trial, it also suggests that they condition it on additional features, like the outcomes of the trial before last. Subjects have no problems learning, for example, the pattern *Left*, *Left*, *Left*, *Right*, *Left*, *Left*, *Left*, *Right*, etc. which requires a memory of at least 4 trials back. However, stochastic learning theories restrict attention to memories of size one, hence ruling out front the possibility of any such pattern matching. In sections 3 and 4 I look in some more details on what exactly is being reinforced.

Moreover, subjects do not like to believe outcomes occur in random (for example, they are more likely to guess *Left* after 3 consecutive *Rights*). Subjects try to look for patterns, even in situations when there are not any. This is supported by experiments showing that behaviour changes when subjects actually observe the random selection process. When these types of behaviour are averaged over a whole group, the matching hypothesis could artificially appear to outperform the utility maximisation hypothesis.

The rest of the paper is organised in the following way. Section 2 provides the theoretical background. Section 3 surveys many of the known experiments. In section 4 I discuss the results and some of their implications. Section 5 concludes.

2. Theoretical background

For simplicity, I refer to the two options as *Left* and *Right* throughout this paper. Denote by $p(L)$ the fixed probability with which *Left* is chosen by the random device. If we normalise to zero the utility of making the wrong prediction, then the expected payoff from choosing *Left* with probability (or frequency) p^* is $p^* \cdot p(L) \cdot U(R_L) + (-p^*) \cdot (1 - p(L)) \cdot U(R_R)$, where $U(R_i)$ is the utility of the reward received from correctly predicting i. If the utility of both rewards is constant, then the expression is maximised by $p^* = 0$ or $p^* = 1$, depending on which of the two expected rewards is greater. This is a static decision rule (no learning). Other static decision rules relevant to this experimental setting include Edwards' *Relative Expected Loss Minimisation rule* (Edwards, 1961, 1962) and Simon's *Minimal Regret* (Simon, 1976), where the decision maker either maximises, or minimises an expression based on his regret (rather than payoff) matrix (see Savage, 1972). In the setting considered in this paper, the predictions of all three static rules are identical.

Static rules neglect any effect that the learning process might have. Even if subjects learn and understand that *Left* and *Right* are chosen randomly and independently, they still have to learn the value of $p(L)$. We can, therefore, go one step 'down' in our rationality assumptions, and look at the learning process of a mathematician who perfectly understands the structure of the problem and who is trying to maximise his expected utility. The length of the learning process will depend on how time is discounted, and its outcome will depend on the actual outcomes of the trials. These types of maximisation problems are known as the bandit problems (where a bandit is a nickname for a slot machine). If $p(L) + p(R) = 1$ then the decision maker is, in effect, estimating only one probability. Hence, this is the *one-armed bandit problem*. If $p(L) \neq 1 - p(R)$ then this becomes the *two-armed bandit problem*. In general, the solution to these problems involves a period of experimenting followed by convergence to one side (this is sometimes known as the *Gittins indices*; see Gittens, 1989). The multi-arm bandit problem was first introduced to Economics by Rothschild (1974), who pointed out that it is possible that a rational, but impatient decision maker will end up choosing the wrong side forever. To see why, consider a setting where $p(L) = 0.7$ and a subject who's impatience leads her to experiment for only three periods. Then with probability 0.216 she will end up choosing *Right* (because this is the probability that *Right* was chosen by the random device at least twice).

From a descriptive point of view, the Gittins indices imply that the decision maker had somehow figured out the structure of the problem, that she experiments and that she keep statistics of all the outcomes of these experiments. These are obviously very strong assumptions. An alternative route was taken by mathematical psychologists (and some economists) which makes only a minimal assumption — that people are more likely to repeat a certain action if it proved successful in some sense in the past (what Erev & Roth call the 'law of effect' and which dates back at least to Thorndike, 1898). More specifically, the decision maker is characterised in every given moment in time by a distribution over her strategy space (which represents the probability with which each strategy will be played in the next stage), and by an updating rule based on reinforcement. Theories which follow this general structure, where no beliefs, or beliefs-updating rules, are specified, are known as *stochastic learning theories*. These theories differ only with respect to the specific of this updating rule. Predictions can now be made with regards to transitory behaviour and to behaviour in the limit. An attractive feature of stochastic learning theories is that, under certain conditions, they are equivalent to the replicator dynamic,[1] another favourite metaphor of modern economists.

A typical stochastic learning model is Bush and Mosteller (1955), where learning is assumed to be linear in the reinforcement. More specifically, assuming that the decision maker *a priori* prefers choosing *Left* (alt. *Right*) when the outcome is *Left* (alt. *Right*), the transition rule can be written as: $p_L(n+1) = (1-\phi_1)\cdot p_L(n) + \phi_1$, when the outcome is *Left*, or $p_L(n+1) = (1-\phi_2)\cdot p_L(n)$ otherwise, where ϕ_1 and ϕ_2 are learning constants. In the limit,

$$p_L(\infty) = \frac{p(L)}{p(L) + (1-p(L)).\dfrac{\theta_2}{\theta_1}}$$

(see Bush & Mosteller (1955) for the exact conditions under which this limit exists). Notice that if the ratio of the Tetas is close to one, the model predicts probability matching in the limit. This is no coincidence: all stochastic learning models predict matching, under some conditions (different conditions for the different models).

In general these models can be divided to two broad classes:

1. Models where players always play a pure strategy (e.g. *Right*) but use a probabilistic updating rule (as in Suppes, 1960 or Suppes and Atkinson, 1960). For example, 'start with *Left*; stick with your strategy when you made a correct predictions; otherwise switch with probability ε', and
2. Models in which agents use a deterministic updating rule to choose between the set of all mix strategies (like the Bush-Mosteller model mentioned above).

In a recent paper, Borgers, Morales and Sarin (1997) show that no learning (updating) rule specified for class (a) can lead to optimal choice, and conjecture that a similar result holds for models in class (b). Leaving aside for the moment the question whether such models provide a realistic description of human learning, their result serves as an important benchmark for any theorems which might be proven in such settings. To be blunt, if we start with a rule that does not converge to the optimal behaviour in a simple setting, we should not be surprised when it does not converge in more complicated settings, like multiplayer games.

In experimental settings, we can only guess what exactly is being reinforced. The typical approach, implicit to our discussion so far, is (a) that the set of strategies consists of one-shot strategies only, i.e. what to chose next, bearing in mind that these could be mixed, and (b) that the strength of the reinforcement is directly related to the payoffs (typically linear). Despite their intuitive appeal, these two assumptions are very strong and their experimental validity remains, still today, in doubt. As for the first assumption, it was repeatedly shown that subjects are able (quite easily) to respond differently to events which are four or five trials back in the sequence (see, for example, Anderson 1960). To this, Goodnow (1955), and Nicks (1959) suggested that subjects do not react to the outcome of the last trial, but instead, to runs of consecutive Lefts or Rights. These idea was further developed by Restle (1961). As for the second assumption, several effects (like the framing effect, and the negativity effect) have been identified in probability learning experiments. There is no simple solution to these problems. Several attempts have been made recently to account for the second set of effects (for example, Erev & Roth, 1997, Tang, 1996, Chen & Tang, 1996) with some success.

Some experimenters suggested that subjects get bored with always choosing the same option, therefore switching between *Left* and *Right* throughout the trials. The most formal attempt is Brackbill & Bravos's (1962) model where subjects receive a greater utility by guessing correctly the outcome of the less frequent option. In these types of models utility-maximising individuals will not choose one strategy with probability 1 in the limit. Under some such utility structures, subjects may optimally end up matching $p(L)$.

An even more daring explanation is that subjects believe in the existence of some sort of regularities, or patterns, in the sequence of outcomes. Such a belief is the reason why they disregard their own experience and keep looking for rules and patterns. Of course, if a pattern exists, it is worth spending some time trying to find it. Once it is found, the subject can get 100% of the rewards (compared to only $p(L)$ in the static optimal rule described above). Restle (1961) discusses some of the typical attitudes of subjects suggesting that '... the subject seems to think that he is responding to patterns. Such attempts are natural. The subject has no way of knowing that the events occur in random, and even if he is told that the sequence is random he does not understand this information clearly, nor is there any strong reason for him to believe it.' (Restle, 1961, p. 109). A theory which accounts for pattern matching is clearly an attractive idea. Unfortunately, it is also an extremely hard idea to formalise, because of the size of the set of all patterns. Restle's own theory (Restle, 1961), which only accounts for one class of patterns (namely for consecutive runs), already becomes very complicated analytically when he considers the behaviour of subjects who get paid, or those who face decisions with more than two choices.

3. Survey of experimental results

Subjects: Subjects in most experiments are undergraduate students (mostly psychology students). In Neimark (1956), Gardner (1958) and Edwards (1956,

1961) subjects are army recruits. Children were the subjects of Derks & Paclisanu (1967), Brackbill *et al.* (1962) and Brackbill & Bravos (1962) experiments.

Apparatus: The most popular setting is the light guessing experiment: Subjects face two lights, their task being to predict which of the two would illuminate at the end of the trial. Otherwise, pre-prepared multiple choice answer sheets were used (as in Edwards, 1961). Here, subjects choose one of two options (*Left* or *Right*) and then revealed a third column to find out whether it matched their choice. Sheets are prepared in advanced according to fixed probabilities. Finally, in the setting of Mores and Randquist (1960), subjects collectively observed a random event after individually predicting its outcome.

Instructions: Subjects were instructed to maximise the number of correct predictions. In most experiments they were told that their actions could not affect the outcome of the next trials (this was obvious when pre-prepared sheets were used). Otherwise, instructions vary: some mention probabilities and others did not. I tried to exclude those experiments were subjects knew 'too much', for example, those experiments where they were told that the probabilities are fixed.

Experimental Design: The important features (see discussion below) are: group size, number of trials, size of the last block of trials (where asymptotic behaviour is measured) and the size of reward(s). These details, whenever available, are provided in the tables below.

Tables: The first table summarises results from experiments where subjects did not receive any payoff, but were still informed about the outcome of the trial. Table 2 lists the results of those experiments with monetary payoffs. The payoffs, in cents, appear in the fourth column where the leftmost number describes the payoff. For example, $(1, 0)$ means that subjects receive I cent for each correct guess, and 0 otherwise. $p(L)$ is as before, and the group means are obtained by taking the group's average frequency of choosing *Left* over the last block of trials. The third table contains some individual results, taken from Edwards (1961) where the mean was measured over the last 80, out of 1000, trials. Each column contains the results for one of four groups which faced different $p(L)$'s. For example, in the group which faced $p(L) = 0.7$ two subjects chose *Left* in all of the last 80 trials. Five (out of 20) chose *Left* 70% of the time or less.

The fourth table summarises the results of Brackbill, Kappy & Starr (1962), and Barckbill & Bravos (1962). The left most column describes the ratio between the two rewards: for correctly predicting M (the most frequent event, with $p(M) = 0.75$), and L. The main difference between this table and the previous three is that here the frequencies of choosing *Left* ($p(L) = 0.75$ throughout) in the nth trial are given as a function of the prediction and outcome in the $n-1$th trial. For example, if subjects predicted M in the $n-1$th trial and the actual outcome of that trial was L, then the mean frequency with which M was chosen in the nth trial is given in the ML column.

The fifth and final table is reproduced from Derks and Paclisanu (1967). It examines the relationship between probability matching and age (this is a part of a more general study into the relationship between cognitive development and

Table 1. Experiments with no monetary payoffs

Experimenter(s)	Group Size	Trials	p(L)	Group Mean
Grant *et al.* (51)	37	60	0.25	0.15
			0.75	0.85
Jarvik (51)	29	87	0.60	0.65
	21		0.67	0.70
	28		0.75	0.80
Hake & Hyman (53)	10	240	0.75	0.80
Burke *et al.* (54)	72	120	0.9	0.87
Estes & Straughan (54)	16	240	0.30	0.25
		120	0.15	0.13
Gardner (58)	24	450	0.60	0.62
			0.70	0.72
Engler (58)	20	120	0.75	0.71
Neimark & Shuford (59)	36	100	0.67	0.63
Rubinstein (59)	37	100	0.67	0.78
Anderson & Whalen (60)	18	300	0.65	0.67
			0.80	0.82
Suppes & Atkinson (60)	30	240	0.60	0.59
Edwards (61)	10	1000	0.30	0.11
			0.40	0.31
			0.60	0.70
			0.70	0.83
Myers *et al.* (63)	20	400	0.60	0.62
			0.70	0.75
			0.80	0.87
Friedman *et al.* (64)	80	288	0.80	0.81

decision making). 200 trials were used and the group average was measured over the last 100. $p(L)$ equals 0.75 for all groups.

Finally, consider Morse and Rundquist (1960) experiment, where 16 subjects are instructed to guess whether a small rod dropped to the floor would intersect with a crack in the floor. Then, the same subjects went through a standard light guessing experiment, were the sequence of *Lefts* and *Rights* was determined by the outcomes of the first part of the experiment (that is, the same sequence as before, with 'Left' replacing the outcome 'No intersection' in the first round). In the second stage subjects, who are not aware of how the second sequence had been generated, are not able to watch the random move. In the first stage Morse and Rundquist reported that 5 subjects adopt a 'maximising' strategy, and the group average was much higher than predicted by the probability matching hypothesis. Matching behaviour was observed in the second stage.

3.1. *Comments on the quality of the experiments*

First important observation is that the sequences of Left's and Right's in some experiments were not statistically independent from the outcomes of the previous trials. For example, in some places randomisation took place within small blocks

Table 2. Experiments with monetary payoffs

Experimenter(s)	Group Size	Trials	Payoffs	p(L)	Group Mean
Goodnow (55)	14	120	(−1, 1)	0.70	0.82
				0.90	0.99
Edwards (56)	24	150	(10, −5)	0.30	0.19
				0.80	0.96
	6	150	(4, −2)	0.70	0.85
				0.80	0.96
	6	150	(4 or 12, −2)[2]	0.70	0.46
				0.90	0.95
Nicks (59)	144	380	(1, 0)	0.67	0.71
	72	380	(1, 0)	0.75	0.79
Siegel & Goldstein (59)	4	300	(0, 0)	0.75	0.75
			(5, 0)	0.75	0.86
			(5. −5)	0.75	0.95
Suppes & Atkinson (60)	24	60	(1, 0)	0.60	0.63
			(5, −5)	0.60	0.64
			(10, −10)	0.60	0.69
Siegel (61)	20	300	(5, −5)	0.65	0.75
				0.75	0.93
Myers *et al.* (63)	20	400	(1, −1)	0.60	0.65
				0.70	0.87
				0.80	0.93
			(10, −10)	0.60	0.71
				0.70	0.87
				0.80	0.95
Berby-Meyer & Erev (97)[3]	42	500	(0, −4)	0.70	0.89
			(4, 0)	0.70	0.85
			(2, −2)	0.70	0.95

of trials: say 7 of every 10 consecutive trials were *Lefts* and the other 3 *Rights*. Also common practice was to exclude from the experiment three or more (or four or more) consecutive *Lefts*. Although I excluded most obvious forms of such violation of the non-contingency condition (which is assumed by the theoretical discussion so far) from the above tables, the sequences used by Grant *et al.*, Jarvik, Gardner, Anderson & Whalen, Goodnow, and Galanter & Smith are not i.i.d. either. This is partially the fault of the technology that was available in those years for generating random sequences and partially because some experimenters did not appreciate the importance of such considerations. Of course, it then becomes possible that attentive subjects noticed the contingencies and act accordingly. For example if 7 out of each 10 trials are *Lefts* and in the first 8 you have counted 7 *Lefts*, it is optimal to guess *Right* in the remaining two trials. In general, optimising subjects will, in such circumstances, sometime guess the less frequent option. For similar considerations Fiorina (1971) concluded that the whole psychological literature on probability matching should be disregarded, and that the gambler's fallacy might not be a fallacy after all. I leave it for the

Table 3. Individual asymptotics in 1000 trials (Edwards 1961)

$\pi = 0.7$	$\pi = 0.6$	$\pi = 0.4$	$\pi = 0.3$
100	91	49	26
100	90	48	26
97	85	47	20
96	81	46	20
95	77	43	17
93	76	43	13
91	74	43	13
88	74	40	13
88	71	35	13
87	70	31	12
85	69	31	11
85	66	29	11
80	64	26	8
80	64	22	8
75	63	21	4
70	61	20	4
65	61	16	0
60	59	15	0
58	56	11	0
56	46	0	0
$\mu = 83$	$\mu = 70$	$\mu = 0.31$	$\mu = 0.11$

Table 4. Probability of guessing 'Left' as a function of last outcome and last guess (Brackbill, Kappy Starr, 1962, and Brackbill Bravos, 1962).

Reward	N	School Grade	Trials	MM	LM	ML	LL
None	4	5	321–400	0.77	0.71	0.56	0.68
None	12	3	101–200	0.74	0.82	0.38	0.56
(1, 1)	12	3	101–200	0.80	0.83	0.65	0.85
(3, 3)	12	3	101–200	0.82	0.87	0.67	0.73
(5, 5)	12	3	101–200	0.89	0.87	0.64	0.78
(1, 4)	10	4	121–200	0.76	0.79	0.50	0.75
(1, 3)	10	4	121–200	0.83	0.70	0.68	0.76
(2, 3)	10	4	121–200	0.80	0.84	0.58	0.85
(1, 4)	10	12	121–200	0.74	0.57	0.71	0.61
(1, 3)	10	12	121–200	0.72	0.62	0.63	0.76
(2, 3)	10	12	121–200	0.84	0.88	0.70	0.72

reader to draw his own conclusions from the above results and the discussion below.

Second, asymptotic behaviour is estimated by taking the group average over the last block of trials. For this to be justified, it is required, at a minimum, that individual behaviour has already stabilised. Once again, I excluded those experiments where this was clearly not happening, but I suspect that the individual learning curves have not yet converges in Estes & Straughan (1954) and

Table 5. Matching and age (Derks and Paclisanu, 1967).

Group	Over-match	PM	Under-match	Total
Nursery	22	3	4	29
Kindergarten	5	3	21	29
First Grade	5	5	10	20
Second Grade	4	8	8	20
Third Grade	3	10	7	20
Fifth Grade	2	13	5	20
Seventh Grade	2	13	3	20
College	4	13	3	20

Neimark & Shuford (1959) experiments, where behaviour seems to still be changing in the last block of trials.

4. Discussion

I first compare the matching hypothesis with that of extreme behaviour (where subjects choose the same side always, in the limit). Note first that the matching hypothesis starts off with a better chance of being closer to the observed asymptotic behaviour, because of the following reasons: First, if behaviour still changes in the last block of trials, it is always in favour of the matching hypothesis. Second, taking the group mean as an estimate is appropriate only if the distribution of the individual results is approximately binomial with most of the mass concentrated around the mean. Otherwise, it will be biased in favour of the less extreme hypothesis, probability matching (for example, from a group of subjects using Gittens' method, a small number will end up choosing *Right* from some point onwards, and the group's mean will be biased towards matching). Finally, as commented by Edwards (1961), 'Obtaining an estimate... and testing the null hypothesis that that estimate is not significantly different from $p(L)$ is widespread in the probability learning literature. Such a procedure constitutes attempting to prove a null hypothesis; the smaller the amount of data or the greater its variability, the more likely it is that such a procedure will "confirm" the matching hypothesis. This is why the small but consistent disagreements with the matching hypotheses relevant by most probability learning experiments have not been noticed'.

Results listed in Tables 1 and 2 suggest that probability matching is not only a theoretical phenomenon. However, they also suggest that it is not robust: when monetary payoffs are introduced (in Table 2), the average means mostly exceeds $p(L)$. In general, matching decreased with size of the reward (Edwards (1956), Siegel & Goldstein (1959), Siegel (1960), Suppes & Atkinson (1960), Atkinson (1962), Myers *et al.* (1963), and Brackbill, Kappy & Srarr (1962)). It is also decreasing with the number of trials (Edwards (1961), Bereby-Meyer & Erev (1997)), although humans are very slow learners. In fact, as pointed out by Restle (1961), all sequence effects may disappear after 1000 trials. This could serve as supporting evidence in favour of the idea that matching is a by-product of sequence effects by optimising learners (who give-up the idea of finding patterns after many trials).

A closer examination of individual behaviour (Table 3) shows that very few subjects (7 out of 80) chose the more rewarding side in all of the last 80 trials. None of the subjects chose the less rewarding side, as one might expect from the Gittins theory. The distribution also suggests that looking at groups' averages may not be appropriate. Only 16 (out of 80) subjects chose *Left* with probability smaller than or equal to $p(L)$, which suggest that the outcomes are not conclusive in favour of any of the two hypothesis, but more in line with the predictions of the extreme behaviour hypothesis.

Results summarised in Table 4 show that subjects' behaviour is (partially) contingent on what happened in the last trial. This is compatible with the predictions of stochastic learning theories. However, the results show that the best indicator for behaviour is the *outcome* of the last trial and not the *difference between the prediction and the outcome*. More specifically notice that LL exceeds ML in 10 out of 12 cases, and that the mean value for LL exceeds that of ML in 14 out of 16 cases. This was offered by the authors as evidence in favour of the hypothesis that subjects receive a greater utility from guessing the less frequent outcome. Even if one remains sceptical about this idea, the results still suggest that the structure of reinforcement might be different from what is prescribed by models of stochastic learning. This is particularly relevant today when economists simply replace last period payoffs with reinforcement (as in McKelvey & Palfrey 1995, Fudenberg & Levine 1997 and Camerer & Ho 1996). Whereas in the past economists were able to get away with assuming that payoffs represent some vague notion of satisfaction, this is no longer so simple once explicit learning is considered. The properties of the chosen updating rule should be consistent with human behaviour (see also the discussion in Erev & Roth 1997). Remember also that people look for patterns and are well able to find those which exist. Why should reinforcement learning models imply that the decision maker is a simpleton? A model of reinforcement learning with sophisticated players (i.e. who are able to remember several states back, and to identify simple patterns) might be analytically complex, but seems unavoidable. The 'unified approach' of Camerer and Ho (1997), where behaviour is an weighted average of reinforcement and belief-updating, could be seen as a first attempt in that direction.

Results in Table 5 relate probability matching to human cognitive development. It shows an inverse relationship between behaviour and age in a repeated binary choice situation (the 'ignorance is bliss' effect: looking at these experiments across species it seems as though rates and young children do best, and human adults fare worst). Derks and Paclisanu (1967) explain that very young children use a simple maximisation strategy (that is, always choose the same side, which is consistent with the observation that the children who under matched typically choose *Right* throughout the trials). Around the ages of 5–7, children develop skills of learning by associating events and outcomes, which support more sophisticated learning rules. For economists, this can be seen as one more piece of evidence that sometimes simple strategies outperforms sophisticated rules.

In similar experiments where subjects had three or more outcomes to choose from, the matching hypothesis failed to predict asymptotic behaviour. Instead, subjects tend to choose the most rewarding option with asymptotic probability higher than the one determining its success (Gardner, 1957, 1958, Cotton & Rechtschaffen, 1958, McCormack, 1959). This can still consistent with Bush–Mosteller type learning models, but can also suggest that the binary choice situation might be a special case:[4] Say, bounded rational players who believe in some form of contingencies, or patterns, in the sequence of events are more likely to abandon these beliefs and choose the most frequent outcome, in three-or-more choice situations, when the complexity of the set of possible patterns increases exponentially. The idea that matching is a by-product of 'over thinking' is supported by the Morse & Rundquist (1960) experiment, where subjects faced with a clearly random device chose (almost) optimally, but were closer to probability matching when that random choice mechanism was unobserved. More generally, this suggests that the decision maker can be in one of two modes — a reinforcement learning mode (when it becomes clear to her that nothing else will work, for example, when the random device is observed), and a belief based reasoner (when the decision maker is trying to find patterns, or rules which determine the sequence of outcomes) — and that observed behaviour crucially depends on his or her current mode.

5. Conclusions

Decision problems are games with no strategic affects, and are therefore assumed to be a good place to start testing the predictions of equilibrium theories. However, this is no longer true if we wish to focus on the *process* by which players learn to play the equilibrium. In most experiments where subjects play games with each other and find their way to equilibrium, they do so fairly quickly. However, in the seemingly simple experiment described here, they do not. The experiment is proving problematic — at least from the point of view of neo-classical economics — because subjects do not understand the nature of random, independent sequences of outcomes. For example, we know that after observing three 'Tail's more people are likely to guess 'Head', and there is no reason why we should not expect similar sequence effects in probability learning experiments. Though it might seem perfectly natural (at least for those of us who are trained in analytical thinking) that some sort of simple adaptive process takes place when no information is given, most subjects prefer to start by using heuristics and rules of thumb (such as: Guess 'Left' if 3 or 4 consecutive 'Right's are observed).

What can we learn from these experiments? First, not all hope is lost for the neoclassical approach to economics: 'matching' is not robust, and some people do end up maximising their expected utility. Second, that the asymptotics depend on experience, size and direction of payments, age, and the observability of the random device. Third, that the rate of learning depends on experience and size of payments but that the exact nature of this relationship is not at all trivial. Reflecting on the discussions above, understanding the exact nature of this relationship is likely to be the key issue in our quest for providing dynamics-based equilibrium theories.

Appendix: Other matching behaviours

In this section I describe some of the other behaviours associated in the literature as with the name 'probability matching'. The findings described below are do not conflict the assumption that animals maximise their access to food (the nearest measurable indicator to utility maximisation). Instead the observed behaviour can be seen as optimal under the circumstances.

Experiments with groups: Subjects are either a group of fish in a tank, or ducks in a pond. Food is being offered from both ends of the tank (or pond). The food at one side is given twice as often as food at the other side. After a few seconds a pattern was formed where 2/3 of the fish located themselves in the more frequently rewarded side, while the remaining 1/3 went to the other side. Behaviour here is optimal in the sense that it constitute a Nash equilibrium (in pure strategies) of the game where each player has to choose one side.

A somewhat similar idea was used by Fretwell and Lucas to explain the hunting behaviour of great tits: observations showed that the hunters would choose a certain hunting area with probability equal to that of succeeding in finding prey there. They showed that the hunters matching behaviour is compatible with the mixed strategy Nash equilibrium of the game where hunters' success is determined by the presence of prey *and* the number of other hunters present. Moreover, they showed that in that game, the mixed Nash equilibrium strategies are the only evolutionary stable strategies.

Experiments with Pigeons: To many psychologists, the term probability matching is associated with Herrnstein and his pigeons. In this famous set of experiments Herrnstein placed hungry pigeons in a box with keys on both side. On each side food appeared according to a VI schedule (a concurrent VI schedule is a program that generates a sequence of random time intervals with a given mean. For example, a VI-24 schedule operating a key means that, on the average, it will be armed every 24 seconds). Herrnstein (1960) compared the recorded number of pecks on each key with the ratio of the mean delay times of the two schedules. His finding is known as the 'direct matching law': Let $R(i)$ be the accumulated reinforcement and $p(i)$ as before, then:

$$\frac{p(L)}{p(L)+p(R)} \approx \frac{R(L)}{R(L)+R(R)} \text{ or } \frac{p(L)}{p(R)} \approx \frac{R(L)}{R(R)}$$

Small, but consistent deviations from Herrnstein's model were found by Baum and Richlin (1969) and led Baum (1974) to his 'generalised matching law':

$$\frac{p(L)}{p(R)} \approx a \times \left(\frac{R(L)}{R(R)}\right)^{b}$$

where a is approximately 1 and b is approximately 0.9.

Neither versions of the matching law were successful in predicting pigeons behaviour once the VI schedules were replaced with VR (in a concurrent variable – ratio (VR) schedule, the program advances to the next stage as a function of the number of responses made by the subject. It does so using a random sequence of numbers with a given mean. For example, using a VR−45 means that the pigeon will, on the average, get food every 45 pecks it makes). Instead, after a learning period, pigeons spent almost all of their time on the most profitable side (Herrnstein and Loveland, 1975). An interesting observation is that pigeons faced with two VR schedules with the same mean, chose one of the sides and stayed there. In contrast, when faced with two VI schedules with the same mean time, pigeon spent about half of their time in each of the sides. This result seems odd at first since behaviour is seemingly affected by the particulars of the mechanism governing the delivery rate, despite it being unobservable.

Herrnstein and Loveland (1975), and later Myerson and Miezin (1980), suggested models that explain both behaviours. Matching and optimising are parts of a more general rule of behaviour: the matching law for observed reward. The main idea is that pigeons do not learn the relationship between their actions and the appearance of food, but rather maximise somewhat differently the amount of reward they observed. Either way, Herrnstein's and Baum's matching laws do not imply probability matching. In fact, in most settings, it predicts that learning leads to optimal choice. The interested reader is referred to Davison and McCarthy (1988), for more details and a complete survey of the relevant literature.

Acknowledgements

I would like to thank Tilman Borgers, Ido Erev and Pasquale Scaramozzino for their helpful comments.

Notes

1. Some work is needed before comparison can be made between the two interpretations: In a learning model the decision maker can choose between a continuum of strategies. In the biological model there is a continuum of agents, each with a fixed rule of behaviour. See Borgers and Sarin (1993) for more details.
2. Asymmetric payoffs here: Subjects received 12 cents for correctly predicting the right light, 4 cents for correctly predicting the left light, and −2 cents otherwise.
3. Payoffs in Israeli agorot (In 1997, 1 Israeli Agora = 0.01 was approximately equal to \$0.003).
4. This is consistent with connectionism models (also known as neural networks). See, for example, Rescorla & Wagner (1972), and Shanks (1990).

References

Anderson, N. H. (1960) Effects of first-order conditional probability in a two-choice learning situation, *Journal of Experimental Psychology*, 59, 73–93.

Anderson, N. H., and Whalen, R. (1960) Likelihood judgements and sequential effects in a two choice probability learning situation, *Journal of Experimental Psychology*, 60, 111–120.

Arrow, K. (1962) Utilities, attitudes, choices: A review note, *Econometrica*, 26, 1–23.

Atkinson, R. C. (1962) Choice behaviour and monetary payoffs. In *Mathematical methods in small group processes*, Stanford University Press.

Baum, W. M. and Richlin, (1969) Choice as time allocation, *Journal of the Experimental Analysis of Behaviour*, 12.

Baum, W. M. (1974) On Two Types of deviation from the Matching Law: Bias and Undermatching, *Journal of Experimental Analysis of Behavior*, 22.

Bereby-Meyer, Y. and Erev, I. (1997) On learning to become successful loser: A comparison of alternative abstractions of learning processes in the loss domain, mimeo, Technion — Israel Institute of Technology.

Borgers, T. and Sarin, R. (1993) Learning through Reinforcement and Replicator Dynamics, Forthcoming, *Journal of Economic Theory*.

Borgers, T. and Sarin, R. (1995) Naive Reinforcement Learning With Endogenous Aspirations, mimeo, University College London.

Borgers, T., Morales, A. and Sarin, R. (1997) Simple Behaviour Rules Which Lead to Expected Payoff Maximising Choices, mimeo, University College London.

Brackbill, N. and Bravos, A. (1962) Supplementary report: The utility of correctly predicting infrequent events, *Journal of Experimental Psychology*, 64, 648–649.

Brackbill, N., Kappy, M. S. and Starr, R. H., (1962) Magnitude of Reward and Probability Learning, *Journal of Experimental Psychology*, 1, 32–35.

Brunswik, (1939) Probability as a determiner of rat behavior, *Journal of Experimental Psychology*, 25.

Burke, C. J., Estes, W. K. and Hellyer, S. (1954) Rate of verbal conditioning in relation to stimulus variability, *Journal of Experimental Psychology*, 48, 153–161.

Bush, R. and Mosteller, F. (1955) *Stochastic models for learning*. New York: Wiley.

Camerer, C. F. and Ho, T. (1996) Experience weighted attraction learning in games: A unified approach, mimeo, California Institute of Technology.

Chen, Y. and Tang, F. F. (1996) Learning and incentive compatible mechanism for public provision, mimeo, University of Michigan.

Cotton, J. and Rechtschaffen, A. (1958) Replication report: Two-and-three-choice verbal conditioning phenomena, *Journal of Experimental Psychology*, 56, 96.

Cross, J. G. (1973) Stochastic Learning Model of Economic behaviour, *Quarterly Journal of Economics*, 87.

Davison, M. and McCarthy, D. (1988) *The Matching Law: A Research Review*, Hillsdale, NJ: Lawrence Elbaum Associates.

Derks, P. L. and Paclisanu, M. I. (1967) Simple Strategies in Binary Prediction by Children and Adults, *Journal of Experimental Psychology*, 2, 278–285.

Edwards, W. (1956) Reward probability, amount, and information as determiners of sequential two-alternative decision, *Journal of Experimental Psychology*, 52, 177–188.

Edwards, W. (1961) Probability Learning in 1000 trials, *Journal of Experimental Psychology*, 4, 385–394.

Edwards, W. (1962) Dynamic decision making and probabilistic information processing, *Human Factors*.

Engler, J. (1958) Marginal and conditional stimulus and response probabilities in verbal conditioning, *Journal of Experimental Psychology*, 55, 303–317.

Erev, I. and Roth, A. E. (1997) On the need for low rationality, cognitive game theory: Reinforcement learning in experimental games with unique, mixed strategy equilibria, mimeo, University of Pittsburgh.

Estes, W. (1950) Towards a statistical theory of learning, *Psychological Review*, 57.

Estes, W. (1957), Theory of learning with constant, variable or contingent probabilities of reinforcement, *Psychometrika*, 22.

Estes, W. K. and Strughan, J. H. (1954) Analysis of a verbal conditioning situation in terms of statistical learning theory, *Journal of Experimental Psychology*, 47, 225–234.

Estes, W. and Suppes, P. (1959) Foundation of linear models. In R. R. Bush and W. Estes (eds), *Studies in Mathematical Learning Theory*. Stanford University Press.

Estes, W. (1964) Probability Learning. In A. W. Melton (ed.), *Categories of Human Learning*.

Fiorina, P. M. (1971) Critique and Comments: A Note on Probability Matching and Rational Choice, *Behavioral Science*, 16, 158–166.

Friedman, M. P., Padilla, G. and Gelfand, H. (1964) The learning of choice between bets, *Journal of Mathematical Psychology*.

Fudenberg, D. and Levine, D. (1997) Theory of learning in games, mimeo, University of California, Los Angeles.

Galanter, E. H. and Smith, A. S. (1958) Some experiments on a simple thought problem, *American Journal of Psychology*, 71, 359–366.

Gardner, R. (1957) Probability learning with two and three options, *American Journal of Psychology*, 70, 174–185.

Gardner, R. (1958) Multi-choice decision behavior, *American Journal of Psychology*, 71.

Gittins, J. (1989) *Allocation Indices for Multi-Armed Bandits*. London: Wiley.

Goodnow, J. (1955) Determinants of choice distribution in two choice situations, *American Journal of Psychology*, 68, 106–116.

Grant, D. A., Hake, H. W. and Hornseth, J. P. (1951) Acquisition and extinction of verbal expectations in situation analogous to conditioning, *Journal of Experimental Psychology*, 42, 1–5.

Hake, H. W. and Hyman, R. (1953) Perception of the statistical structure of a random series of binary symbols, *Journal of Experimental Psychology*, 45, 64–74.

Herrnstein R. J. (1961) Relative and absolute strength of response as a function of frequency of reinforcement, *Journal of Experimental Analysis of Behavior*, 4.

Herrnstein, R. J. (1974) Formal Properties of the Matching Law, *Journal of Experimental Analysis of Behavior*, 21.

Herrnstein, R. J. and Loveland, J. (1975) Maximising and Matching on concurrent ratio schedules, *Journal of Experimental Analysis of Behavior*, 24.

Humphreys, L. G. (1939) Acquisition and extinction of verbal expectations in a situation analogous to conditioning, *Journal of Experimental Psychology*, 25.

Jarvik, M. E. (1951) Probability learning and a negative recency effect in a serial anticipation of alternative symbols, *Journal of Experimental Psychology*, 41, 291–297.

Luce, D. and Suppes, P. (1965) Preference, Utility, and Subjective Probabilities. In D. Luce, R. Bush and E. Galanter (eds), *Editors Handbook of Mathematical Psychology*, 111, New York: Wiley.

McCormack, P. (1959) Spatial generalization and probability learning in five-choice situation, *American Journal of Psychology*, 72.

McKelvey, R. and Palfrey, T. (1995) Quantal response equilibria for normal form games, *Game and Economic Behaviour*, 10, 6–38.

Morse, E. and Rundquist, W. (1960) Probability matching with an unscheduled random sequence, *American Journal of Psychology*, 73.

Myers, J. L., Fort, J. G., Katz, L. and Suydam, M. (1963) Differential memory gains and losses and event probability in a two-choice situation, *Journal of Experimental Psychology*, 66, 521–522.

Myerson and Miezin (1980) The kinetics of choice: An operant system analysis, *Psychological Review*, 87.

Neimark, E. (1956) Effect of type of non reinforcement and number of alternative responses in two verbal conditioning situations, *Journal of Experimental Psychology*, 52, 209–220.

Neimark, E. and Shufod, E. (1959) Comparison of predictions and estimates in a probability learning situation, *Journal of Experimental Psychology*, 57, 294–298.

Nicks, D. C. (1959) Prediction of sequential two-choice decisions from event runs, *Journal of Experimental Psychology*, 57, 105–114.

Rescorla, R. A. and Wagner, A. R. (1972) A theory of pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black and W. F. Prokasy (eds), *Classical conditioning II: Current theory and research*. New York: Appleton-Century-Crofts.

Restle, F. (1961) *Psychology of judgement and choice: a theoretical essay*. New York: Wiley.

Roth, A. E. and Erev, I. (1995) Learning in extensive-form games: Experimental data and simple dynamic models in intermediate term, *Games and Economic Behaviour*, 8, 164–212.

Rotschild, M. (1974) A Two-Armed Bandit Theory of Market Pricing, *Journal of Economic Theory*, 9, 185–202.

Rubinstein, I. (1959) Some factors in probability matching, *Journal of Experimental Psychology*, 57, 413–416.

Shanks, D. R. (1990) Connectionism and the learning of probabilistic concepts, *Quarterly Journal of Experimental Psychology*, 42A, 209–237.

Siegel, S. and Goldstein, D. A. (1959) Decision making behaviour in a two-choice uncertain outcome situation, *Journal of Experimental Psychology*, 57, 37–42.

Siegel, S. (1960) Decision Making and Learning under Varying Conditions of Reinforcement, *Annals of the New York Academy of Sciences*, 89, 766–783.

Simon, H. (1976) Rational Choice and the Structure of the Environment, *Psychological Review*, 63, 129–138.

Suppes, P. and Atkinson, R. C. (1960) *Markov learning models for multi person interactions*. Stanford: Stanford University Press.

Suppes, P. (1961), Behavioristic foundations of utility, *Econometrica*, 29, 186–202.

Tang, F. (1996) Anticipatory learning in two-person games: An experimental study, mimeo, University of Bonn.

Thorndike, E. L. (1898) Animal intelligence: An experimental study of associative processes in animals, *Psychological Monographs*, 2.

Tversky, A. and Edwards, W. (1965) Information versus Reward in Binary Choices, *Journal of Experimental Psychology*, 71.