

Working Paper 99-51
Economics Series 19
June 1999

Departamento de Economía
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34-91) 624-98-75

**LEARNING TO DEAL WITH RISK: WHAT DOES REINFORCEMENT LEARNING
TELL US ABOUT RISK ATTITUDES? ¹**

Albert Burgos *

Abstract

People are generally reluctant to accept risk. In particular, people overvalue outcomes that are considered certain, relative to outcomes which are merely probable. At the same time, people are also more willing to accept bets when payoffs involve losses rather than gains. I consider how far adaptive learning can go on in explaining these phenomena. I report simulations in which adaptive learners of the kind studied in Roth & Erev (1995, 1998) and Börgers & Sarin (1996, 1997) deal with a problem of iterated choice under risk where alternatives differ by a mean preserving spread. The simulations show that adaptive learning induce (on average) risk averse choices. This learning bias is stronger for gains than for losses. Also, risk averse choices are much more likely when one of the alternatives is a certain prospect. The implications of a learning interpretation of risk taking are examined.

Keywords: Iterated choice, Reinforcement Learning, Risk attitudes.

JEL Clasification: D81, D83

* Burgos, Departamento de Economía, Universidad Carlos III de Madrid. E-mail: burgos@eco.uc3m.es

¹ This research was undertaken when the author was a Visiting Postdoctoral Fellow in Economics at the University of Rochester. The author's thanks go to this institution for its hospitality and to the Fulbright Commission and the MEC in Spain for financial support.

1 Introduction

The idea of modeling people as stimulus-response mechanisms shaped by learning forces has become popular in the last ten years or so. Reinforcement learning models have been used recently in economics either to explain experimental findings in strategic encounters¹ or to give learning theoretical foundations of population dynamics used in evolutionary game theory.² These models consider the adaptive behavior of goal oriented agents which need not to be subjective expected utility maximizers, or indeed maximizers of any sort. In fact, choices are a (stochastic) function of earlier payoffs, while these payoffs are again a function of earlier choices. As a result, one gets a sequence analysis of actions and payoffs in which risk attitudes does not appear explicitly, but only 'between the lines'. Unfortunately, applications of these models to the problem of iterated choice under risk have been confined to noting conditions under which the system moves away from probability matching³ and leads to expected payoff maximization. This paper tries to make explicit the reactions to risk **induced** by such processes. Then, one is in a position to analyze how far adaptive learning provides sufficient structure to tie down the set of possible reactions to risk.

The paper reports the result of computer simulations in which reinforcement learners deal with pairwise choices between risky prospects with the same expected value. Any consistent pattern showing more propensity to choose one alternative rather than the other is interpreted as reflecting risk preference. The central issue here is the possible connection between the attitudes toward risk resulting from adaptive learning and some patterns of actual choices in experiments that have been extensively documented in experimental economics.

In a first block of simulations, learners choose between a certain prospect and an uncertain one with equal expected value. I test for risk aversion over positive and negative payoffs. Experimental and field data studies have shown greater risk aversion for gains than for losses. Reinforcement learners show behavior consistent with this **reflection effect**. The second block of simulations is designed to induce violations of the independence axiom of expected utility theory. These involve a test of the **certainty effect** or Allais ratio paradox (Allais (1953), Kahneman and Tversky (1979)). Simulated learners, too, violate the independence axiom in the Allais-type direction.

¹ See, for example, Erev and Roth (1998), and the references therein.

² Börgers and Sarin (1996).

³ Consider the following situation. There are two risky prospects, P1 and P2. Payoffs realizations are either 0 or 1. The probability that P1 yields 1 is r and the probability that P2 yields 1 is $1 - r$. A prediction that the probability of choosing P1 will approach p through learning is referred to as probability matching. This behavior violates stochastic dominance provided that $r \neq 0.5$. The terminology 'probability matching' derives from the fact that the frequencies with which the decision maker chooses actions match the probabilities with which the actions are successful. For a recent experimental evidence of probability matching in choice under risk see Loomes (1998).

2 Background

Reinforcement models have been developed largely in response to observations by psychologist about human behavior and animal behavior. All of these models are built upon quantifications of Thorndike's (1898) **law of effect**: choices that lead to good (bad) outcomes are more (less) likely to be repeated. Implicit in the law of effect is that choice behavior is probabilistic: In a problem of iterated choice under risk or uncertainty, the agent chooses one alternative at each time, observes its consequence or payoff, and over time updates her choice as a result. Then, a reinforcement learner is characterized by an initial probability vector over alternatives and, subsequently, after each learning trial by a revised probability vector. In this context, 'learning' means updating the probabilities of taking each alternative on the basis of the payoffs or outcomes experienced. Namely, if $p_i(t)$ is a specific learner's probability of choosing alternative i on trial t , a learning model is defined as a rule specifying how $p_i(t)$ is transformed to $p_i(t+1)$.

Applications of reinforcement learning to economic behavior are organized around two models: The first is developed by Börgers and Sarin (1996, 1997). It is a version of the classic Bush and Mosteller's (1955) linear adjustment model in which reinforcements can be either positive or negative, depending on whether the realized payoff is greater or less than the agent's 'aspiration level', which in turn follows a linear adjustment process too. The second is proposed by Roth and Erev (1995, 1998). It is a more highly parametrized version of a quantification of the law of effect by Herrnstein (1961, 1970) based on average returns, and adds to the Börgers and Sarin's model an interesting feature: learning curves get flatter over time, a fact known as the **power law of practice** (Blackburn (1936)). I now present these models in detail for choices involving only two alternatives.

2.1 The Börgers & Sarin (B&S) model

In this model, $p_i(t)$ ($i = 1, 2$) changes as a result of the alternative chosen and the outcome observed at $t - 1$. If outcome exceed the aspiration level, then the probability associated with the action increases. If outcomes fall below the aspiration level, then the probability associated with the alternative chosen decreases. The size of the change in $p_i(t)$ is proportional to the size of the difference between the outcome and the aspiration level. Formally, if $\rho(t)$ denote the aspiration level (or reference point) at time t , probabilities of choosing each alternative evolve in the following way: if payoff at date t is $x \geq \rho(t)$, then the reward associated with x is $R(x, t) = x - \rho(t) > 0$, and for both i ,

$$p_i(t+1) = \begin{cases} [1 - \lambda R(x, t)]p_i(t) + \lambda R(x, t) & \text{if } i \text{ was chosen at } t, \\ [1 - \lambda R(x, t)]p_i(t) & \text{otherwise.} \end{cases}$$

If, however, payoff at t is $x < \rho(t)$, the reward $R(x, t)$ is negative, and

$$p_i(t+1) = \begin{cases} [1 + \lambda R(x, t)]p_i(t) & \text{if } i \text{ was chosen at } t, \\ [1 + \lambda R(x, t)]p_i(t) - \lambda R(x, t) & \text{otherwise.} \end{cases}$$

Parameter λ , which controls the effect of rewards in the $p_i(t+1)$, can assume any value guaranteeing that the absolute value of $\lambda R(x, t)$ always lies (strictly) between zero and one. The greater the value of λ , the faster the adaptation.⁴

In addition to the probability vector, also the aspiration level is adjusted in the direction of the outcome experienced. Thus,

$$\rho(t+1) = (1 - \beta)\rho(t) + \beta x,$$

where $0 \leq \beta < 1$. Initial values for $\rho(1)$ can be set to any value. Thus, the model is controlled by the initial probabilities and three parameters: the aspiration level's parameters ($\rho(1)$ and β) and the learning parameter λ .

2.2 The Roth and Erev (R&E) model

In R&E's model, $p_i(t)$ changes on the basis of the history of returns that have been obtained from the two alternatives. This memory of the average return from each alternative is modified by the effect of reference points, forgetting (or recency), and experimentation, yielding what R&E call 'propensities'. Again, let $R(x, t) = x - \rho(t)$ be the reward associated with x . Thus, if at date t the decision maker receives a payoff of x , then the propensity to play each alternative is updated by setting

$$q_i(t+1) = \begin{cases} \max\{\nu, (1 - \phi)q_i(t) + (1 - \varepsilon)R(x, t)\} & \text{if } i \text{ was chosen at } t, \\ \max\{\nu, (1 - \phi)q_j(t) + \varepsilon R(x, t)\} & \text{otherwise.} \end{cases}$$

Parameter ν represents a small "cutoff" value guaranteeing that propensities remain positive. Parameters ϕ and ε have behavioral meaning: ϕ (recency) slowly reduces the importance of past experience, and ε (experimentation) prevents the probability of choosing any alternative from going to zero.

⁴ In fact, Börgers & Sarin (1997) interpret payoffs as parametrizations of agents' responses to their experiences, not as physical payoffs. Thus, they implicitly assume $\lambda = 1$. For the purposes of this paper, however, this assumption would have the undesired effect of making the speed of adaptation only dependent of the speed of adjustment of the aspiration level. For instance, in the case of a fixed, exogeneous aspiration level, all learners would be equally fast.

The probability of choosing alternative i in period t is proportional to past average propensities, i.e. for both i ,

$$p_i(t) = \frac{q_i(t)}{q_1(t) + q_2(t)}.$$

Thus, the ratio of the probabilities of choosing each alternative equals the ratio of their propensities. The task of determining initial propensities is reduced by setting $S(1) = q_1(1) + q_2(1)$.⁵

Then initial propensities follow from the initial choice probabilities and $S(1)$, the strength parameter, which controls for the weight of initial tendencies. Notice that when $S(1)$ is high (i.e. initial propensities are strong) learning will be lower than when $S(1)$ is low.

Finally, two additional parameters, ω^- and ω^+ , control the adjustment of the reference point following negative and positive rewards, respectively:

$$\rho(t+1) = \begin{cases} (1 - \omega^-) \rho(t) + (\omega^-) x & \text{if } x < \rho(t), \\ (1 - \omega^+) \rho(t) + (\omega^+) x & \text{if } x \geq \rho(t). \end{cases}$$

R&E calibrate their model against experimental data on two-player matrix games with mixed-strategy equilibria. The calibration of the general seven-parameter model appears in Erev and Roth (1996). Calibrations of one-parameter ($S(1)$), and three-parameter ($S(1)$, ϕ , and ϵ) reduced versions of the model appear in the published version of the paper, Erev and Roth (1998).⁶ Table 1 summarizes these results.

⁵ This definition of the strength parameter is taken from Roth and Erev (1995). In Erev and Roth (1996, 1998), $S(1)$ is chosen to be $(q_1(1) + q_2(1))/X$, where X is the mean return associated with the problem given uniformly distributed choice probabilities. This formulation, however, would imply a negative value for $S(1)$ when the average payoff is negative.

⁶ In these reduced versions, the initial reference point is set to be the minimum payoff in the game, and hence, all rewards are positive. The remaining parameters of the general model that are not included in the reduced version are constrained to be zero.

TABLE 1—CALIBRATIONS OF THE R&E MODEL

Parameters:	$S(1)$	ϕ	ε	$\rho(1)$	ν	ω^-	ω^+
best fit (1 parameter):	54						
best fit (3 parameters):	9	0.1	0.2				
best fit (general model):	3	0.001	0.2	0	0.0001	0.01	0.02

Sources: Erev and Roth (1996, 1998).

3 Testing for Risk Taking with Gains and Risk Taking with Losses

3.1 Exercising the models

A common result in experiments conducted to elicit certainty equivalents for lotteries is that revealed risk preferences show greater risk aversion when payoffs are positive than when payoffs are negative. For the most part, studies in this tradition have treated positive payoffs as **gains** and negative payoffs as **losses**, therefore assuming that the relevant reference point is zero. Learning models as those presented earlier allow, however, for the reference point to reflect aspirations, expectations, or targets. From this point of view, a gain is a return that falls above the reference point, and a loss is a return that falls below.

The central issue here is whether adaptive learning generates risk biases in the direction of showing more risk aversion in a domain of positive payoffs than in a domain of negative payoffs, as experiments show. This gives rise to the following hypothesis.

Hypothesis 1: *Adaptive learners exhibit greater risk aversion for gains than for losses.*

To examine this question, I simulate the choice situation between two pairs of alternatives. Each pair consists of one lottery, denoted S for “safer,” which pays qx with certainty, and a mean preserving spread of S , denoted R , for “riskier,” which pays x with probability q and nothing otherwise. Measure for evaluating Hypothesis 1 consists of the comparison of the ex-post probability of choosing S over M when x is positive (gains) and when x is negative (losses) starting from a situation in which both choices are equiprobable.⁷ Figure 1 shows the average

⁷ An alternative measure would be the relative frequency of choosing prospect S over R . I have also tried this measure, and the results are substantially the same.

results involving 1000 simulated learners over the first 500 trials for the calibrated R&E model. The simulations assume that $q = 0.5$, and the figures compare (i) a case of negative outcomes in which $x = -1$ with (ii) a case of positive outcomes in which $x = 1$.

The simulations clearly support Hypothesis 1 for the case of the calibrated R&E model. Although the model ultimately yields equal probabilities of choosing each alternative, the amount of experience required to reach such a result when $q = 0.5$ is quite substantial—about 200 trials. Risk aversion for gains is weak, and learning ultimately result in risk neutrality, but in the short run adaptive learning produces risk seeking when the returns lie in the negative domain.

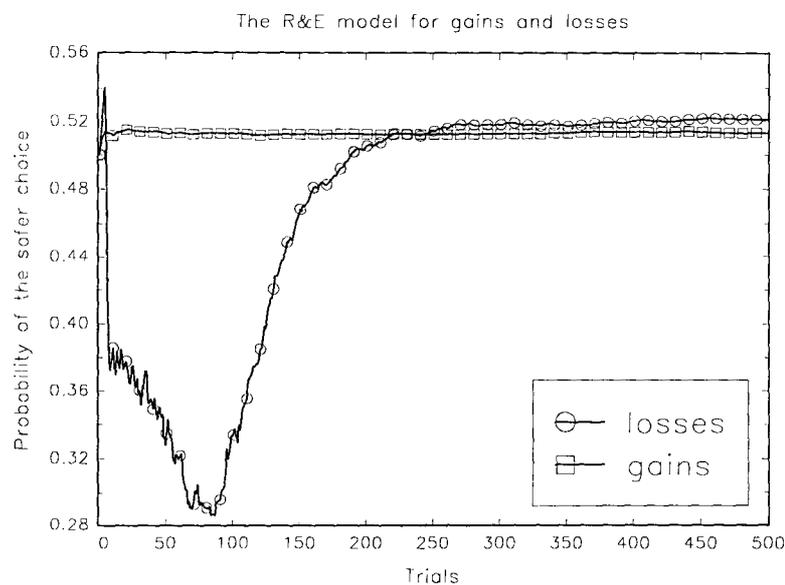


FIGURE 1. COMPARING RISK PREFERENCES FOR GAINS AND LOSSES, CALIBRATED R&E MODEL

Figure 2 shows the results of the same simulation for the B&S model when $\lambda = 0.1$, $\rho(1) = 0$, and $\beta = 0.01$.⁸ Again choosing the less risky alternative is more likely when possible outcomes lie in the positive domain than when they are negative. However, unlike the case of the R&E model, learning that conforms to the B&S model produces risk averse behavior both for losses and for gains, and maintains this behavior in the long run.

⁸ These parameters produce slower initial learning than in the calibrated R&E model. However, since the B&S model does not employ intermediate propensities, the speed of learning quickly exceeds that of the R&E model.

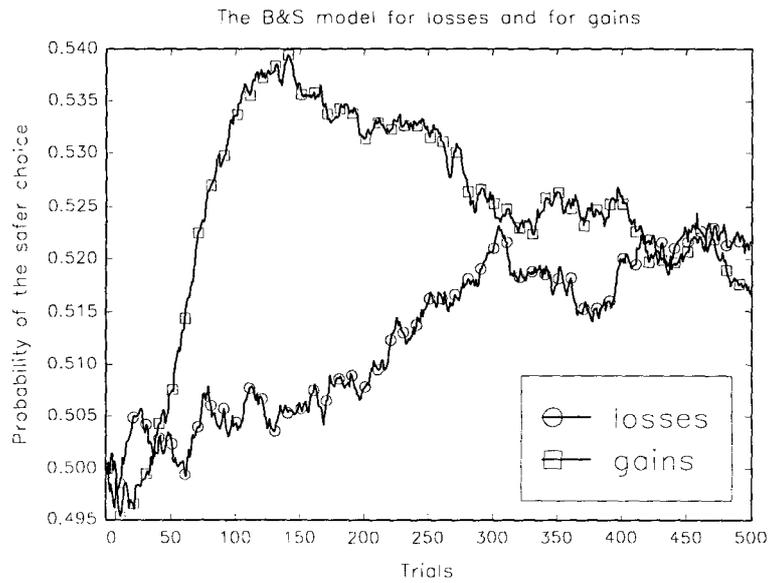


FIGURE 2. COMPARING RISK PREFERENCES FOR GAINS AND LOSSES, B&B MODEL ($\lambda = 0.1$, $\rho(1) = 0$, $\beta = 0.01$)

3.2 Dependence on q . Effect of $S(1)$ and λ

It might be expected from the reduced difference in riskiness associated with the two alternatives, the difference between learned risk preferences for gains and learned risk preferences for losses decreases as q increases. This is indeed the case for the R&E model. Figure 3 shows the relation between q and the difference between gains and losses in the probability of choosing the safer alternative after 75 trials.⁹ The B&S model, however, is quite insensitive to changes in q . As Figure 4 shows, the difference between gains and losses in the probability of choosing the safer alternative after 150 trials is not monotonic.

⁹ This was the amount of experience at which the difference between $p(t)$ for gains and for losses attained its maximum when $q = 0.5$.

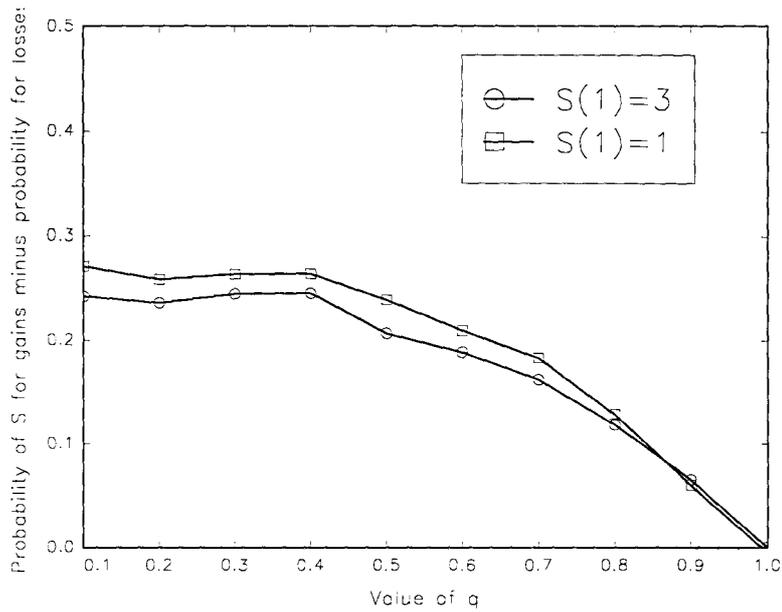


FIGURE 3. DIFFERENCE BETWEEN RISK PREFERENCES FOR GAINS AND LOSSES FOR THE R&E MODEL AFTER 75 TRIALS

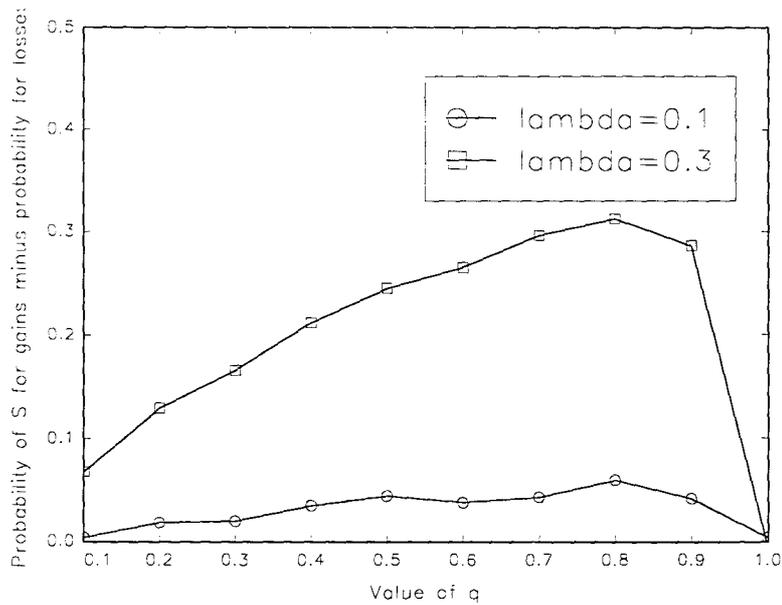


FIGURE 4. DIFFERENCE BETWEEN RISK PREFERENCES FOR GAINS AND LOSSES FOR THE B&S MODEL AFTER 150 TRIALS ($\rho(1) = 0, \beta = 0.01$)

Figures 3 and 4 above contain also information regarding the effect on risk taking of the learning rate. The learning parameters $S(1)$ and λ are the main determinants of the rate at which learning takes place. Low values of $S(1)$ and high values of λ are associated with fast learning. As might be expected, fast learning accentuates the tendency towards risk aversion in

the positive domain, slow learning reduces it—provided the initial probabilities of choosing the various alternatives are (as in this case) equal.

4 Testing for the “certainty effect”

Among the experimental challenges to expected utility theory, perhaps the most systematically observed violation of the theory refers to the systematic ‘over-valuation’ of outcomes that are considered certain, relative to outcomes which are merely probable. This anomaly (from the expected utility theory point of view), appears in a class of systematic violations of expected utility theory known as the certainty effect or Allais ratio paradox (see for instance, Kahneman and Tversky (1979 p. 266)). In this section I test for this effect in adaptive learning.

A simple test of the certainty effect involves choices between a prospect S , with r chance of y , or $1 - r$ chance of 0, and a prospect R , with qr chance of x , or $1 - qr$ chance of 0. Probability q is kept fixed, and experiments measure the effect of r on choices. The null hypothesis is that choices are independent of r , as the independence axiom of expected utility theory predicts. In the classic experiments x and y are positive, and y is chosen to be equal to (or slightly smaller than) qx . Thus, any risk averse expected utility maximizer would chose S for all r (or else R for all r if the subject were a risk seeker). The certainty effect occurs when S is chosen in problems with $r = 1$ and R is chosen in problems with $r < 1$. Starting as early as in Allais (1953), researchers have found considerable evidence that this effect is a systematic property of risk attitudes, for a range of parameter values. This provides the basis for the following hypothesis.

Hypothesis 2: *When faced with choice problems which mimic experimental tests of the certainty effect, adaptive learners exhibit Allais-type behavior.*

I employed the same simulation procedure as in the previous section. Figure 5 illustrates the certainty effect using the calibrated R&E model. It shows the effect of r on probability of choice of the safer alternative, S , over 500 trials, where $x = 1$, $q = 0.5$, and $y = qx$. As the hypothesis states, certainty accentuates the tendency toward risk aversion, though the impact on choices probability is low.¹⁰

¹⁰ The certainty effect is quite higher if, instead of using the R&E model with seven parameters, we restrict ourselves to the three parameter model in which the aspiration level is set equal to the minimum payoff.

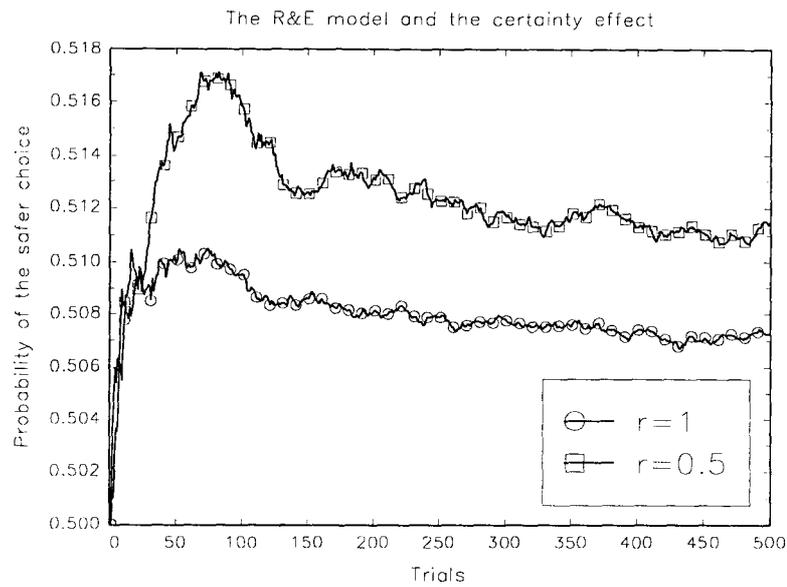


FIGURE 5. COMPARING RISK PREFERENCES FOR CERTAIN AND UNCERTAIN OUTCOMES, CALIBRATED R&E MODEL

Figure 6 shows the much stronger certainty effect associated with the B&S model. The relation between r and the difference between the probability of choosing S and the probability of choosing R for two different values of the learning parameter λ is shown in Figure 7.

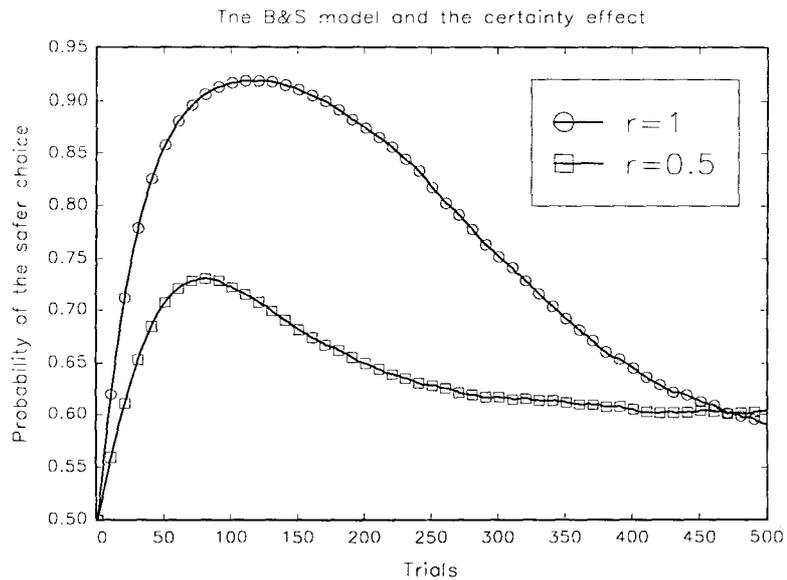


FIGURE 6. COMPARING RISK PREFERENCES FOR CERTAIN AND UNCERTAIN OUTCOMES, B&S MODEL ($\lambda = 0.1, \rho(1) = 0, \beta = 0.01$)

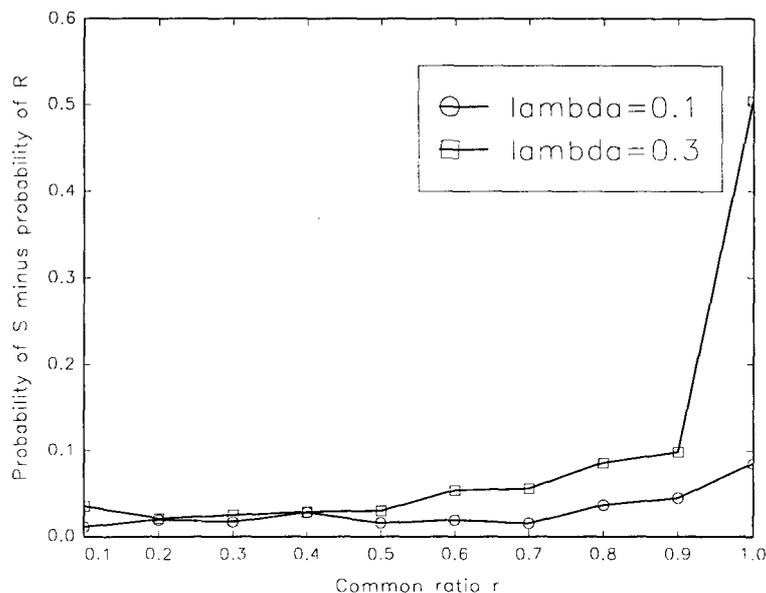


FIGURE 7. EFFECT OF r IN RISK PREFERENCES AFTER 150 TRIALS, B&S MODEL ($\rho(1) = 0$, $\beta = 0.01$)

5 Relevance to modelers?

As long as risk taking is interpreted within a framework of preference maximization, the understanding of risk attitudes is couched in terms of features of utility functions. According to this interpretation, a considerable amount of time and effort has gone into the quest for models of preferences that try to accommodate at least the most widely observed behavioral regularities. Nowadays, however, the idea that there is a simple and coherent model of preferences that accommodates all experimental anomalies (from an expected utility point of view) is appearing more and more like an unattainable “Holy Grail” for choice theory under risk. See for instance, Selten, Sadrieh, and Abbink (1995) and Loomes (1998).

A learning perspective provides a different view of the sources of risk aversion and risk seeking. Risk preference can be interpreted as a learned response, rather than a consequence of prior nonlinear utilities. We should expect that the behavior of inexperienced individuals (such as typical experimental subjects) is driven by norms that are triggered by the framing of the problem. As these individuals gain experience, we expect from them to select payoff maximizing actions. However, as we have seen, the **velocity** of this process depends, among other things, on whether the choices are about gains or losses, and payoffs are certain or uncertain. The R&E model is particularly successful in explaining differences in risk preferences for gains and for losses, whereas the B&S model provides a more robust learning interpretation of the certainty

effect.¹¹ To illustrate this, I offer predictions of these dynamic models starting from initial conditions observed in experiments.

5.1 Simulations and experimental data

For the case of risk aversion and risk seeking for gains and for losses, I shall follow data in Tversky and Kahneman (1992). This study used a certainty-equivalent task to elicit risk preferences for gains and losses. Tversky and Kahneman estimated the median cash equivalents (in dollars) for a set of two-outcome prospects. For a prospect offering a 0.5 chance to win \$100 or 0.5 chance of zero, the median cash equivalent was \$36, indicating risk aversion. When the prospect offered a 0.5 chance to lose \$100 or 0.5 chance of zero, the median cash equivalent was -\$42, indicating risk seeking.

Starting from these initial conditions, Figure 8 displays the evolution of risk preferences as agents become experienced. At $t = 1$, prospects are assumed to be indifferent, and therefore initial choice probabilities are set equal to 0.5. As learners gain experience, the prospect offering a certain positive payoff of \$36 becomes less and less attractive in comparison with a 0.5 chance of \$100, showing convergence to risk neutrality. In the cases of losses, however, learning accentuates risk seeking in the short run, and it takes about 80 trials to reverse this pattern.

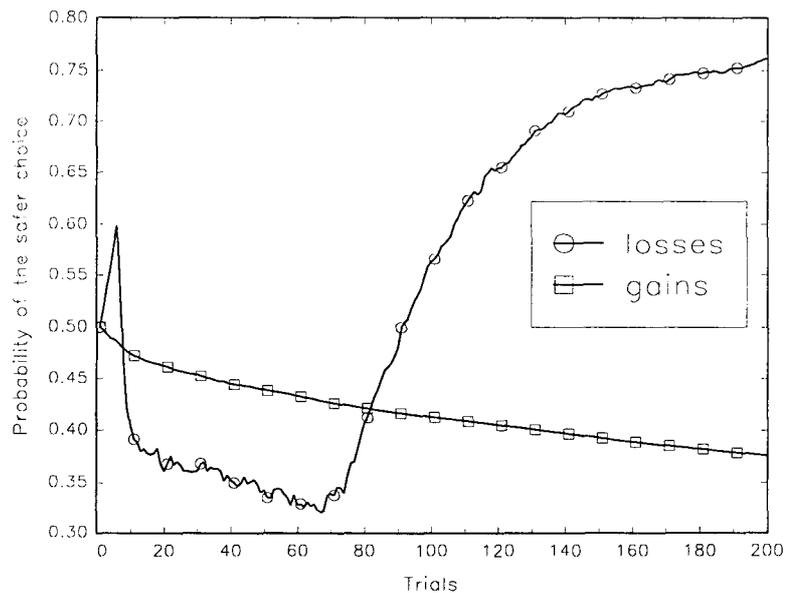


FIGURE 8. SIMULATION OF RISK PREFERENCES FOR GAINS AND LOSSES STARTING FROM EXPERIMENTAL DATA, CALIBRATED R&E MODEL

¹¹ This raises an interesting question. Do individuals use different adaptive processes for different problems? This might be the case if individuals, when presented with any particular problem, construct their preferences using available rules of thumb in conjunction with salient basic principles to process the task in hand.

In order to check the effect of learning in the evolution of choices in Allais-type problems, I take again experimental data from Kahneman and Tversky. In a now classical experiment (Kahneman and Tversky (1979), p. 266) found that 80 percent of their subjects preferred a sure 3,000 Israeli Shekels (IS) to a 0.8 chance of winning 4,000 IS or a 0.2 chance of nothing, while only 35 percent preferred a 0.25 chance of winning 3,000 IS or a 0.75 chance of nothing to a 0.2 chance of winning 4,000 IS or a 0.8 chance of nothing. Thus, after normalizing payoffs to $x = 1$ and $y = 0.75$, if $q = 0.8$, Kahneman and Tversky's estimates of the probability of choosing the safer alternative are 0.8 when $r = 1$ and 0.35 when $r = 0.25$. Figure 9 portrays the evolution of learning that conforms to the B&S model starting from these initial conditions. According to expected utility maximization, we should expect the probability of choosing the safer alternative to approach a value independent of r . Such a result is found in the long run. In the short run, however, when $r = 1$ the probability of choosing the safer alternative rises to near 1, and then drops slowly. The asymptote appears to be somewhere around 0.6, indicating that risk aversion persists indefinitely in the B&S model.

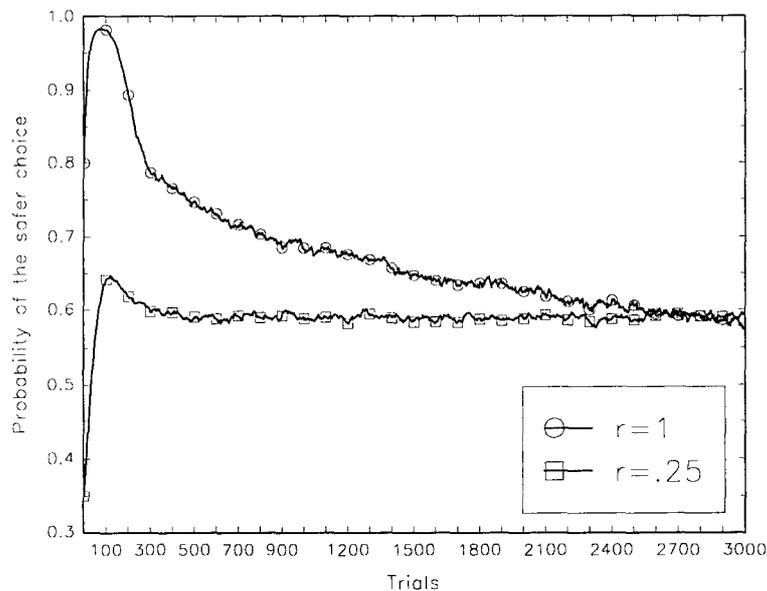


FIGURE 9. SIMULATION OF RISK PREFERENCES FOR CERTAIN AND UNCERTAIN OUTCOMES STARTING FROM EXPERIMENTAL DATA, B&S MODEL ($\lambda = 0.1$, $\rho(1) = 0$, $\beta = 0.01$)

6 Concluding Remarks

A traditional argument favoring the use of expected utility theory (and its special case, risk neutrality) in economics has been an appeal—often implicit—to learning forces.¹² If individuals are provided with sufficient opportunity for trial-and-error learning, they will eventually behave in accordance with the expected utility postulates. Thus, there is a spectrum of experience and expertise, with novices at one extreme and solid expected utility maximizers at the other. Be that as it may, there is a very large gulf between the two extremes, and little, if anything, is presently known about how to place a given choice problem along this spectrum, or about how to divide this spectrum into the portion on which expected utility theory applies and the portions when other theories are more appropriate.

Economics has now passed through the phase of providing evidence of where expected utility breaks down. The reflection effect and the Allais paradox are two of the most widely documented examples. In this paper, I study how they may evolve or be modified in response to feedback or experience. I regard the dynamics presented here as a crude model of such evolution. The simulations show that adaptive learning generates risk taking biases *in the direction* predicted by experiments. In particular, the fact that fast learning (i.e. in which behavior is modified rapidly in response to feedback) is especially likely to produce these biases may give pause to the easy assumption that fast learning is a sure route to maximizing behavior.

¹² One can appeal to evolutionary forces too (see Robson (1995) for a recent reference). In these models, behavior is determined by genes, and Darwinian mutation and selection mechanisms determine which behavior genes survive. However, since we are surrounded by uncertainty, it seems odd that people in experiments should be so bad at assessing probabilities in experiments. Neo-Darwinian anthropologists have an explanation for this (as they do for almost everything), essentially to do with the fact that abstract logical thought was of little use on the savannah.

References

- ALLAIS, M. (1953): "Le Comportement de l'Homme Rationnel devant le Risque, Critique des Postulats et Axiomes de l'École Americaine," *Econometrica*, **21**, 503–546.
- BLACKBURN, J.M. (1936): "Acquisition of Skill: An Analysis of learning Curves," IHRB Report No. 73.
- BÖRGERS, T., AND R. SARIN (1996): "Naive reinforcement learning with Endogenous Aspirations," mimeo, University College London. (To appear in *International Economic Review*.)
- (1997): "Dynamic Consistency and Non-Expected Utility Models of Choice Under Uncertainty," *Journal of Economic Theory*, **77**, 1-14.
- BUSH, R., AND F. MOSTELLER (1955): *Stochastic Models for Learning*, New York, NY: John Wiley & Sons.
- EREV, I., AND A.E. ROTH (1996): "On the Need for Low Rationality, Cognitive Game Theory: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria," mimeo, University of Pittsburgh. (Earlier version of Erev and Roth (1998).)
- (1998): "Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria," *American Economic Review*, **2**, 225-243.
- HERRNSTEIN, R.J. (1961): "Relative and Absolute Strength of Response as a Function of Frequency of Reinforcement," *Journal of the Experimental Analysis of Behavior*, **4**, 267-272.
- (1970): "On the Law of Effect," *Journal of the Experimental Analysis of Behavior*, **13**, 244-266.
- KAHNEMAN, D., AND A. TVERSKY (1979): "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, **47**, 263–291.
- LOOMES, G. (1998): "Probabilities vs. Money: A Test of Some Fundamental Assumptions about Rational decision Making," *Economic Journal*, **108**, 477–489.
- ROTH, A.E., AND I. EREV (1995): "Learning in Extensive-Form games: Experimental Data and Simple Dynamic Models in the Intermediate Term." *Games and economic Behavior*, Special Issue: Nobel Symposium, **8**, 164-212.
- SELTEN, R., A. SADRIEH, AND K. ABBINK (1995): "Money does not induce risk neutral behavior, but binary lotteries do even worse," Discussion Paper B-343, University of Bonn.

- THORNDIKE, E.L. (1898): "Animal Intelligence: An Experimental Study of the Associative Processes in Animals," *Psychological Review Monographs Supplement*, **2**.
- TVERSKY, A., AND D. KAHNEMAN (1992): "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty*, **5**, 297-323.