Adaptive Learning Models of Consumer Behavior

Ed Hopkins*
Edinburgh School of Economics
University of Edinburgh
Edinburgh EH8 9JY, UK

January, 2006

Abstract

In a model of dynamic duopoly, optimal price policies are characterized assuming consumers learn adaptively about the relative quality of the two products. A contrast is made between belief-based and reinforcement learning. Under reinforcement learning, consumers can become locked into the habit of purchasing inferior goods. Such lock-in permits the existence of multiple history-dependent asymmetric steady states in which one firm dominates. In contrast, belief-based learning rules must lead asymptotically to correct beliefs about the relative quality of the two brands and so in this case there is a unique steady state.

Journal of Economic Literature classification numbers: C73, D11, D83, L13, M31.

Keywords: learning, consumer behavior, dynamic pricing, behavioral economics, reinforcement learning, market structure.

*I would like to thank Andreas Blume, Pradeep Chintagunta, John Duffy, Teck Ho, Josef Hofbauer, Tatiana Kornienko, Al Roth, Karl Schlag and Jonathan Thomas for helpful discussions. I acknowledge support from the Economic and Social Research Council, award reference RES-000-27-0065. Errors remain my own. E.Hopkins@ed.ac.uk, http://homepages.ed.ac.uk/ehk

1 Introduction

Adaptive learning models attempt to describe the behavior of agents faced with repeated decision problems by assuming they use simple learning rules. These models are used in a number of apparently disparate environments. Economic theorists have analyzed them in abstract settings.[1] They have been fitted to actual choice data both in economic experiments and the quite different context of the empirical analysis of consumer behavior.[2] Despite differences in aims and terminology, some models of dynamic choice found in empirical marketing analysis are essentially the same as those used in economic theory. This research in marketing supports the experimental evidence that even simple adaptive learning models can help to explain human behavior. In the context of econometric work on experimental data, there has been an active debate as whether the more sophisticated belief-based models or very simple reinforcement learning models offer the better fit. Up to now, this has been of interest because it throws light upon human reasoning processes. However, if the same question is considered in the context of consumer choice, there may be significant practical implications to consider as well.

This paper investigates the hypothesis that whether consumer behavior is best described by belief-based or reinforcement learning may have a significant impact on market organization. In particular, we examine a model of dynamic duopoly, where consumers learn about the relative quality of the two different brands. The product is an experience good and so information is partial: consumers only learn the payoff to the good they actually consume. First, we investigate a reinforcement type learning model, where more familiar products have a greater probability of being selected. Consequently, consumers can become locked into inferior choices. Such lock-in permits the existence of multiple history-dependent steady states. When multiple steady states exist, even if the two firms are identical in terms of costs and product quality, the symmetric outcome is unstable: one firm must dominate. This outcome under reinforcement learning is then contrasted with the outcome under belief-based learning. This form of learning leads to correct beliefs about relative quality even under partial information. Firms can influence consumer opinion only in the short run: if consumers' initial estimate of a firm's quality is high (low), it has an incentive to charge above (below) the myopic price in order to slow (speed up) learning. Given the convergence of beliefs to the unique correct outcome, the firms must converge to a unique steady state, where prices are the same as under complete information. This paper, therefore, shows that the small differences in the learning rules, between belief-based and reinforcement learning, can have dramatic effects on market outcomes.

[1] Theoretical papers in this field include Arthur (1993), Rustichini (1999), Börgers and Sarin (2000), Sarin and Vahid (1999), Börgers et al. (2004). A survey of the use of adaptive learning models in games can be found in Fudenberg and Levine (1998).

[2] Empirical work includes Erev and Roth (1998), Camerer and Ho (1999), Erev and Barron (2001) and Blume et al. (2002). Examples of work in marketing are Chintagunta and Rao (1996), Seetharaman and Chintagunta (1998), Ho and Chong (2003).

The situation to be modelled can be thought of as a consumer going on a regular basis to a supermarket to buy a grocery item and choosing between two competing brands. This type of decision has several aspects which I would like to emphasize. First, the prices for the competing brands are usually clearly marked on the shelves. Thus, the learning the consumer has to undertake is not about prices or their distribution. However, the goods in question are typically experience goods. One has to take them home and consume them before their quality is known. Second, quality in this context is very often subjective and imprecise, for example, whether a food product tastes good. Third, because each successive purchase decision is relatively unimportant to an individual consumer, a model of boundedly-rational behavior may explain actual choices well. Such boundedly-rational agents may have a impression of quality that is ambiguous and difficult to measure against past experience. As a consequence, it may be very difficult to be confident about *relative* quality. For example, I think I like the brand I bought today, but is it clearly better than the one I bought last month? Indeed, in this paper it is assumed that the consumption experience is noisy and memory is imperfect.

The formal model of price competition analyzed here is derived from that of Chintagunta and Rao (1996), who similarly consider a dynamic duopoly with adaptive consumers. Their work is quite distinctive from most of the literature in economics on learning. First, there is the mixture of rational behavior by sellers and reinforcement learning by boundedly rational buyers. Second, while the recent literature on adaptive learning has largely focussed on abstract exogenous environments, Chintagunta and Rao's work is also empirical. The model is fitted to data on actual prices, sales and consumer purchases. They find, for example, that a dynamic specification, taking into account consumers' past purchases, outperforms a static logit model. This result, as I argue in Section 7 of this paper, provides some support for the hypothesis that learning is in fact suboptimal.

Nonetheless, the difference between this current paper and the work of Chintagunta and Rao (1996) is large, and reflects the difference between economics and marketing science. First, the principal question here is one of welfare: do consumers learn to make correct choices and what is the implication that has for the competitiveness of the resulting market structure. In contrast, Chintagunta and Rao's (1996) main objective, as with much marketing analysis, is to predict consumer choice. Second, Chintagunta and Rao do not investigate whether the reinforcement learning rule they specify would lead a consumer to choose the brand which she would prefer in the case of perfect information. We show that frequently this will not be the case. Third, Chintagunta and Rao, in characterizing the dynamic pricing equilibrium, did not identify, as is done here, that there may be multiple steady states. Finally, in their paper only one learning rule is considered. Here, the results under reinforcement learning are contrasted with those resulting under belief based learning, thereby demonstrating that it is familiarity based learning that is responsible for the pathological outcome, and not the situation of experience goods in itself.

This latter point is also what differentiates the current work from an earlier literature, Schmalensee (1978) and Smallwood and Conlisk (1979), that concentrates exclusively on simple forms of reinforcement learning. In these models also, consumers do not necessarily learn which is the highest quality brand. The contribution here is to clarify the conditions on the form of consumer learning under which this is possible. Furthermore, in the last few years, the analysis of experimental data has shown the effectiveness of adaptive learning models in predicting subject behavior. This, combined with the empirical marketing work of Chintagunta and Rao (1996) and Ho and Chong (2003), offers the intriguing prospect of estimating consumer learning models with actual consumer choices. Given the theoretical differences between reinforcement and belief-based learning, in Section 7, I offer two empirical tests that potentially could distinguish between the two models.

The focus on bounded rationality differentiates this paper from most previous literature on dynamic pricing of experience goods that has assumed fully rational consumers.[3] One strand of the existing literature is based on the quality of the good being private information to the seller. The consumer then learns about product quality by making highly sophisticated inferences from the resulting strategic behavior of firms. Depending on the model and/or the parameters of a single model, a seller can signal that the quality of her good is higher than the alternative by charging a price that is either higher or lower than the price that would be myopically optimal (Milgrom and Roberts (1986); Bagwell and Riordan (1991)). Here, consumers can only learn if a brand is of high quality through repeated consumption experience. Bergemann and Välikmäki (1996) is much closer in that it examines the effect of strategic pricing on the rate of information acquisition by a buyer. However, it is quite different in that the buyer's behavior is given by the solution to a stochastic dynamic optimization problem, allowing for an optimal level of experimentation.

Strategic and adaptive models may well be complementary, with different models doing better in different circumstances. For example, Bergemann and Välikmäki (1996), to motivate their model of optimal learning, give the example of a factory manager choosing which production technology to buy. Indeed, such professional decision-makers faced with sharp incentives may well be well-described by optimal learning models. Adaptive models, on the other hand, may do well in those consumer markets where a single purchase represents very small stakes. Indeed, this paper is not the first to apply learning models to consumer behaviour. Weisbuch, Kirman and Herreiner (2000) and Kirman and Vriend (2001), in work that is very close to the present approach, analyze adaptive learning models of consumer behavior and similarly examine conditions under which consumers become loyal to one seller. The principal difference is that here firms are forward looking and price dynamically. Erev and Haruvy (2001) also consider the implications of adaptive learning by consumers but again firms have fixed pricing policies.[4] Ellison and Fudenberg (1995) look at social learning, where agents learn from

[3] Belief-based adaptive learning, although more sophisticated than reinforcement learning is still some way short of full rationality in the traditional sense.

[4] Other work on adaptive learning has had adaptive learning by sellers as well as by buyers (Hopkins

the experience of others as well as from their own. Certainly researchers in consumer behavior have found it plausible that consumers may be prone to a number of cognitive biases, see for example Erdem et al. (1999). One which seems particular relevant in this context is "confirmatory bias". As Rabin and Schrag (1999) discover, there is substantial psychological evidence that once individuals form a hypothesis, they pay greater attention to subsequent evidence that supports that hypothesis than to evidence that is non-supportive. In the current context, this would suggest that consumers may be relatively unwilling to switch away from a favored brand.

This paper highlights a source of bias that is even more basic, but which has significant implications for market organization. Imagine a consumer who initially has greater goodwill towards brand X than the rival brand Y. So, all other factors being equal, she will mostly purchase X, and only rarely sample Y. Now, suppose her choice decision is based on beliefs: estimates of the relative quality of the two brands. Then, the low frequency of purchase of Y will not matter in the long run, as eventually she will accumulate a sufficient number of observations to gain a clear picture of the average quality of Y, and if this is higher than that of X, she will switch allegiance. In contrast, suppose a consumer chooses on the basis of familiarity, a stock of goodwill. Then, in the intervening time between purchases of Y, when he is consuming X, that stock of goodwill towards Y diminishes, he forgets about it. The probability of buying Y falls. He may then never accumulate sufficient positive experience to realize that in fact Y is just as good, or even superior. That is, quite subtle difference in mental attitude toward choices not made, products not consumed, can have quite profound effects on long run outcomes.

The result is that, when consumers are reinforcement learners, possession of a high initial market share is self-reinforcing. There is an extensive literature in industrial organization concerned with the origins of market dominance. Recent theoretical explanations for sustained dominance include network effects, increasing returns to scale and learning by doing. Here none of these factors are present but there is still lock-in. This, however, is broadly consistent with the empirical findings of Sutton (1991) on industries in the food sector, where some outcomes seem history dependent in industries, without network externalities, but where consumer tastes, loyalty and perceptions of quality are important.

In examining dynamic oligopoly, there is the question of which equilibrium concept to use. Open loop equilibrium, as used by Chintagunta and Rao (1996), earlier by Schmalensee (1978) and more recently, for example, by Cellini and Lambertini (1998), has the advantage of analytic simplicity. It is true that many researchers in industrial organization prefer Markov perfect/closed loop equilibrium, despite the fact that, except in simple linear-quadratic models, its complexity precludes analysis except by numerical methods. In contrast, the open loop equilibrium can be analyzed qualitatively, revealing much information such as the number of steady states and their stability. Furthermore, the known disadvantages of open loop equilibria, that it in effect allows commitment to

and Seymour (2002); Harrington and Chen (2003)).

a complete strategy path, are limited here as firms compete on price not quantity. That is, the fact that the open loop equilibrium in the model analysed here has asymmetric steady states cannot be attributed to a Stackleberg phenomenon, where one firm obtains dominance simply by committing to a high level of output. Finally, Markov perfect equilibrium is useful for the analysis of how firms respond to stochastic realisations of demand or other variables. But in the current model, while the evolution of individual consumer behaviour is stochastic, there are no aggregate shocks. So, it is possible that a deterministic approximation, averaging over a large number of consumers, can capture the essentials of consumer behaviour. In turn, open loop equilibrium may be a reasonable approach. This is discussed further in Section 6.

2 Adaptive Learning, Hypothetical Reasoning and Suboptimal Learning

This section introduces the models of adaptive learning used in this analysis and reviews the evidence as to whether adaptive learning can lead to optimal choices. A crucial aspect will be how capable an agent is of hypothetical reasoning, in particular, how she treats the question, "How well would I have done, if I had chosen differently from the choice I actually made?" This matters, as in many choice situations, including that of experience goods, one only sees the payoff to the choice actually consumed, leaving one to speculate if one could have done better by having chosen differently. The danger is that since the value of unchosen alternative is less clear, one will become convinced that one's own choice is superior, simply because it was the one chosen, not because of any objective superiority. As we will see, there is both theoretical and empirical evidence that this can be an important phenomenon.

Models of learning have been employed both to explain behavior in games and in single person decision making. There is now considerable evidence that they can explain actual choice behavior (Erev and Roth, 1998; Camerer and Ho, 1999; Erev and Barron, 2001). In games, payoffs are determined by the choices of one's opponents and in decision problems, by an exogenous random process, but in both cases, learning rules have three components. First, a decision maker is endowed with propensities (alternative terms are assessments or weights or scores), one for each of the possible actions in her action set. The propensities of a representative agent we denote as $\theta = (\theta_1, \theta_2, ..., \theta_n) \in I\!R^n$, when the agent must choose from n actions. Second, there is a choice rule that chooses an action as a function of current propensities. A general principle is that actions with higher propensities are chosen with higher probability. Finally, there is an updating rule, which changes the propensities in response to experience.

The choice rule that has attracted the most attention is the logit or exponential rule

$$x_i(t) = \frac{\exp(\beta\theta_i(t))}{\sum_{j=1}^{n} \exp(\beta\theta_j(t))}. \tag{1}$$

where $x_i(t)$ is the probability that the agent chooses action i at time t. In the exponential choice rule, the parameter β represents the degree of optimization. At high levels of β the agent will choose the action with the highest propensity with very high probability.

There have been two commonly used differing assumptions about what information is available when updating propensities. If an agent in each period can observe the return to all possible actions, including the "foregone" payoffs to actions that were not taken, this is "full" information in the terminology of Rustichini (1999). If, however, the agent can only see the payoff to the action actually taken this is "partial" information. A crucial assumption here is that a consumer choosing amongst experience goods is in a situation of partial information: she only finds out about the good that she actually chooses. Therefore, we look at adaptive learning models that only use information about actual payoffs.

The payoff that an agent receives at any given time will be random in two senses. First, this is because given a choice rule such as the logit above, the action she chooses will be random. Second, in addition, it is assumed that experience is stochastic. Specifically, if at time t, conditional on taking the action i an agent receives a payoff $\tilde{u}_i(t)$ which is a random variable with mean u_i.

Our first updating rule can be called reinforcement learning. Upon receiving a payoff, the agent then updates his propensities

$$\begin{aligned} \theta_i(t+1) &= (1-\delta)\theta_i(t) + \delta\tilde{u}_i(t) \\ \theta_j(t+1) &= (1-\delta)\theta_j(t), \text{ for all } j \neq i \end{aligned} \tag{2}$$

where $1 \geq \delta > 0$ is a "recency" parameter. If δ is equal to one, then only the very last experience is remembered. With δ close to zero, experience from long ago may still have a significant weight in current beliefs. Crucially, this rule responds only to realized payoffs. No information about payoffs to actions not taken is utilized. This assumption is found in the reinforcement learning models put forward by Arthur (1993) and Erev and Roth (1998). This may be because the payoff to other actions at that time was not observed, there is partial information, or the learner is boundedly rational. Another model that uses only partial information has recently been proposed by Sarin and Vahid (1999).[5] The rule is, if action/good i is chosen at time t, then

$$\begin{aligned} \theta_i(t+1) &= (1-\delta)\theta_i(t) + \delta\tilde{u}_i(t) \\ \theta_j(t+1) &= \theta_j(t) \text{ for all } j \neq i. \end{aligned} \tag{3}$$

Note that the first and second learning rule differ from each other in an important sense. The second can be thought of as a "belief-based" learning rule, in that each θ_i is an estimate of the payoff to each action. In contrast, our first rule is best described as a reinforcement or stimulus-response type learning rule. Here, each θ_i cannot be

[5] Fudenberg and Levine (1998, Chapter 4) have a similar model, as do Kirman and Vriend (2001). But it is also true that this form of learning rule had already been studied for some time in the artificial intelligence field, see Sutton and Barto (1998).

interpreted as a belief; it is rather a stock of positive feeling. With a belief-based model, the consumer is assumed to have a belief, albeit adaptively formed, of the quality of each of the different brands. With a reinforcement model, a propensity potentially incorporates a much wider set of feelings, such as familiarity or recognition. For example, when a consumer in a hurry grabs a product off a shelf maybe it is not because he believes it offers the best value for money but because it is the only product he recognizes.

What turns out to be the crucial difference between reinforcement and belief-based learning is the treatment of the propensities of actions not taken. With reinforcement based on familiarity, rule (2), the propensity for actions not chosen naturally decreases as familiarity with those actions/products declines. In contrast, under the belief-based rule (3), the propensity for actions unchosen remains unaltered, as there is no new information about the action/product with which to update one's quality estimate.

Most attempts to distinguish empirically between reinforcement and belief-based learning models have attempted to see whether information on foregone payoffs is in fact used. That is, the attempt has been to distinguish between the reinforcement rule (2) and the rule

$$\theta_i(t+1) = (1-\delta)\theta_i(t) + \delta\tilde{u}_i(t) \text{ for } i = 1, 2, ..., n. \tag{4}$$

This last rule is only applicable under full information as it uses information about all possible actions to update simultaneously all propensities. However, as here we concentrate on experience goods, where there is only partial information, differences in reaction to information about foregone payoffs will not be important, as this information is not available.

Do different updating rules lead to different outcomes? In particular, do these rules lead an agent to optimal choices? For example, suppose each $\tilde{u}_i(t)$ is an independent draw from a fixed distribution with mean u_i. Then, if these means were known, the optimal action would clearly be to choose always the action with the highest mean payoff. Can agents learn to do this without any prior information about payoffs? With the belief-based Sarin and Vahid rule (3), at least asymptotically an agent's choices will be close to optimal. Informally, if δ is "small", then asymptotically each θ_i will be close to u_i. That is, in the long run an agent will have correct estimates of the return to each strategy. This implies that using, for example, the logit choice rule (1) for a high β, asymptotically the agent will place a very high probability on the optimal action (see Sarin and Vahid (1999)). Similar results can also be shown to hold for the rule (4).

But, importantly as Rustichini (1999) points out, if one combines the reinforcement updating rule (2) with the logit choice rule (1), optimality is not always achieved.[6] The asymptotic result is history dependent, with the agent likely to become locked into choosing the strategy she initially favors independent of whether it is optimal. That is,

[6] The first "lock-in" result under adaptive learning is found in the pioneering work of Arthur (1993). His results were in a slightly different context, however. See Hopkins and Posch (2005) for a fuller discussion of the issues involved.

the exponential choice rule in a situation of partial information can be interpreted as a form of overconfidence. With a high value of β the action that seems the best will be chosen with a high probability. Therefore, under partial information, the agent may never find out that another action would actually give a higher payoff.

What evidence is there that this might happen in practice? Erev and Barron (2001) report on a large number of single person decision experiments. First, Erev and Barron claim that a reinforcement learning model similar to a combination of the learning rule (2) and the exponential choice rule (1), identified by Rustichini (1999) as non-optimal, is a better fit to actual behavior than the model of Sarin and Vahid, that is optimal in this context. Second, this is perhaps because, as the individual data reveals, many subjects, even when there are only two possible actions, end up choosing the inferior action. Lastly, Chintagunta and Rao (1996) and Ho and Chong (2003) fit reinforcement type learning models on actual consumer choices. It is argued in Section 7 that their findings give some support for the hypothesis that learning is suboptimal.

This tendency to become locked into one particular choice, simply because of an initial preference is reminiscent of "confirmatory bias" which has been well-documented in the psychology literature (see, for example, Rabin and Schrag (1999)). This can be defined as the tendency to interpret new evidence as supporting an existing belief even if it is not truly favorable. This trait is to some extent captured by the learning rule (2), in which one's opinion of the action not chosen continuously deteriorates. However, in this context there is a further problem. By repeatedly choosing only one option, an agent can avoid seeing any evidence that is favorable to the alternative. The likelihood of being locked into one's initial choice therefore would seem to be even higher.

3 A Model of Dynamic Duopoly

In this section, the dynamic duopoly model with learning by consumers is introduced. This is similar to the earlier model of Chintagunta and Rao (1996) (hereafter, "CR"), though as discussed in the Introduction, there are differences in approach and in the results obtained. There are two firms that produce a product at constant marginal cost. Marginal cost for both brands is normalized to zero. Prices are given by $p = (p_1, p_2)$. We also use the difference in prices $q = p_1 - p_2$. For simplicity, we consider a single representative consumer. Reasons why this may serve as a reasonable approximation of the more realistic case of a large population of consumers are given in Section 6. In any case, this consumer has goodwill for the two brands equal to $\theta = (\theta_1, \theta_2)$. θ_1 can be thought of as the consumer's estimate of quality of the first firm's product and θ_2 the estimate of the quality of the alternative. We will also use $\eta = \theta_1 - \theta_2$, the relative goodwill toward the first brand.

At each point in time the consumer seeks to buy one unit of the good, either from firm 1 or firm 2. The consumer uses a decision rule of the logit form. This rule has been extensively used in the literature on learning and is given in its usual form in (1). Here it is modified to take into account that the decision has two aspects, price as well as the utility of consumption. The expected utility of purchasing the first brand is $\theta_1 - p_1$ and the utility of purchasing the alternative is $\theta_2 - p_2$. Therefore, the logit rule will give the probability of purchasing from the first firm as

$$x_1(\theta, p) = \frac{\exp(\beta(\theta_1 - p_1))}{\exp(\beta(\theta_1 - p_1)) + \exp(\beta(\theta_2 - p_2))} = \frac{\exp(\beta(\eta - q))}{\exp(\beta(\eta - q)) + 1} \tag{5}$$

where $\beta > 0$ is a parameter measuring the sensitivity to goodwill and prices.[7] The probability of purchasing the second brand is $x_2(\theta, p) = 1 - x_1$. Clearly, here if prices are equal, and if β is large, the consumer will purchase the brand with the higher associated θ with a probability close to one. Let $x(\theta, p)$ denote the vector of market shares (x_1, x_2).

The consumer's goodwill will change over time in response to her consumption experience. If she consumes good i at time t, then she receives a utility of $\tilde{u}_i(t)$. It is assumed that $\tilde{u}_i(t)$ is an independent draw from a constant distribution with mean $u_i > 0$. Let $u^* = u_1 - u_2$, that is, the actual expected quality premium of the first brand. A *learning rule* in this context will be a way of updating goodwill θ in response to the consumption experience $\tilde{u}_i(t)$. This paper uses several different learning rules, as set out in the previous section, each representing different behavioral assumptions, and each having differing predictions.

Each firm is assumed to know the learning rule of the consumer and we can assume also that each is able to observe purchase patterns and therefore should at any given time have a good estimate of the consumer's goodwill. The actual evolution of goodwill will be follow the consumer's consumption experience and will be stochastic. There are various methods applicable for stochastic dynamic optimization. In effect, it is assumed that each firm uses stochastic approximation theory, which predicts the stochastic evolution of goodwill by the solution of an associated differential equation. The first step is to calculate an expected change in θ which we can write as

$$E[\theta_i(t+1)|\theta(t)] - \theta_i(t) = \delta f_i(\theta, x(\theta, p)),$$

where the exact form of $f_i(\cdot)$ depends on which of the two learning rules, (2) or (3), that is currently under analysis. Stochastic approximation, which has been widely used in the recent literature on learning, shows that if δ is small, the solution of the original stochastic difference equation to the differential equation (6) will be closely approximated by the solution to the following parallel continuous time system

$$\dot{\theta}_i = f_i(\theta, x(\theta, p)). \tag{6}$$

We assume that this approximation is close enough for the purposes of the firms, and

[7] In Chintagunta and Rao's original specification, allowance was made for β to take different values for the two goods whereas the utility from consumption of any good was normalized to unity. I opt for a convention which is closer to the learning literature where β is fixed but the average consumption utility u_i varies across the brands. See also Section 7.

that each attempts to solve the deterministic continuous time optimization problem implied by (6).[8]

Note that here, just as in CR's original model, price only affects the evolution of goodwill through the probability of purchase $x(\theta, p)$ and not directly. There are arguments for and against this modelling choice. The model is intended to represent the situation of a consumer choosing between products in a supermarket, where prices are clearly displayed. In that sense, because current prices are easily available, the consumer's choice may not be affected by her knowledge of past prices. On the other hand, a consumer may not check prices again each time he shops. In which case, his choice in a particular period may be determined by his impression about which of the brands is the least expensive, an impression formed by past prices.[9]

Given the assumption of a representative consumer, firm i's instantaneous profits will be $p_i(t)x_i(t)$. Each firm seeks to maximize

$$\int_{t=0}^{\infty} e^{-rt} p_i x_i(\theta, p)\, dt \text{ subject to } \dot{\theta} = f(\theta, x(\theta, p)), \tag{7}$$

where $r > 0$ is the firms' common discount rate. This in turn gives rise to a current-value Hamiltonian for each firm,

$$H_i = p_i x_i(\theta, p) + \mu_i f_i(\theta, x(\theta, p)) + \nu_i f_j(\theta, x(\theta, p)), \tag{8}$$

where μ_i, ν_i are firm i's costate variables, and $f_j(\cdot)$ gives the expected motion of the other firm's goodwill. The dynamics of the costate variables are given by $\dot{\mu}_i = -\partial H_i / \partial \theta_i + r\mu_i$ and $\dot{\nu}_i = -\partial H_i / \partial \theta_j + r\nu_i$. If each firm maximizes its Hamiltonian at each point in time treating its opponent's price as fixed, this constitutes a Nash equilibrium in open loop strategies (see, for example, Kamien and Schwartz (2000, Chapter 23)). An open loop strategy is a path for prices $p_i(t)$ as a function of initial goodwill $\theta(0)$ and calendar time t only. Open loop equilibrium, therefore, assumes that the firms choose such strategies simultaneously and independently at the beginning of the game. They are then committed to the resultant price path for whole rest of the game. We characterize these equilibrium strategies in the next section and discuss the applicability of open loop equilibrium in this context in Section 6.

It will be useful to contrast the optimal policy derived from the above dynamic equilibrium. with the myopic policy, where each firm seeks to maximize instantaneous profits, which from (5), can be written $p_i x_i(\eta, q)$. Then, each firm charges the static duopoly price, which I will write as $\hat{p}(\eta)$ as it will depend on the current level of relative goodwill. That is, $\hat{p}(\eta)$ solves the simultaneous equations $x_i(\eta, p) + p_i \partial x_i(\eta, p)/\partial p_i = 0$

[8] An introduction to stochastic approximation is given in Fudenberg and Levine (1998, Ch. 4). Benaïm (1999) provides a more extensive survey. Benaïm (1998) considers the case we consider here, where δ is constant.

[9] The model could be modified to incorporate such effects. However, while it would lead to lower prices, as each firm has an additional dynamic incentive to lower prices to build goodwill, I hypothesize it would not lead to substantial qualitative changes in behavior.

for $i = 1, 2$ or equivalently, it solves

$$p_i = \frac{1}{\beta(1 - x_i(\eta, p))} \tag{9}$$

again for $i = 1, 2$ and $p = (p_1, p_2)$.[10] One of the principal questions of this analysis will be when will the sellers have an incentive to charge a price above or below $\hat{p}(\eta)$.

4 Equilibrium under Reinforcement by Familiarity

In this section, the change in goodwill for the consumer follows what is now the standard reinforcement learning updating rule. This is CR's original model with slight modifications to the definition of the choice function (5) as noted in Section 3. This learning rule has been popularized in the field of economics by Erev and Roth (1998) and was given in Section 2 as rule (2). What is crucial about this rule is that the consumer's opinion of the brand not chosen deteriorates, with the consequence that the consumer becomes progressively more convinced that he has chosen correctly, even if that choice is not in fact optimal.

Moving to the expected motion and continuous time, one obtains (for convenience, in this and in what follows, I suppress the dependence of the market shares x_1, x_2 on goodwill θ and prices p)

$$\dot{\theta}_1 = x_1 u_1 - \theta_1, \quad \dot{\theta}_2 = x_2 u_2 - \theta_2. \tag{10}$$

Remember as stated in Section 2, we cannot interpret the θ parameters as beliefs. Rather, they are measures of goodwill. Furthermore, we will see that they tend to diverge to extreme values: the consumer will become loyal to one product alone. In this case, it is easier to replace the two goodwill variables (θ_1, θ_2) with the single variable giving relative goodwill $\eta = \theta_1 - \theta_2$. One can calculate that

$$\dot{\eta} = x_1 u_1 - x_2 u_2 - \eta \tag{11}$$

The Hamiltonian for firm i becomes

$$H_i = p_i x_i + \xi_i (x_1 u_1 - x_2 u_2 - \eta) \tag{12}$$

where ξ_i is the new costate variable replacing μ_i and ν_i. Differentiating each H_i with respect to p_i and setting to zero, given that $\partial x_i / \partial p_i = -\beta x_i (1 - x_i)$ prices satisfy

$$p_1 = \frac{1}{\beta x_2} - \xi_1 (u_1 + u_2), \; p_2 = \frac{1}{\beta x_1} + \xi_2 (u_1 + u_2). \tag{13}$$

[10] It has been established by Caplin and Nalebuff (1991) that Nash equilibrium in oligopoly with logit demand functions exists and is unique.

The dynamics of the costate variables ξ_1, ξ_2 can be derived from the basic formula $\dot{\xi}_i = -\partial H_i/\partial\eta + r\xi_i$. Substituting in from (13) the resulting differential equations can be written,

$$\dot{\xi}_1 = \xi_1(1+r) - x_1,\ \dot{\xi}_2 = \xi_2(1+r) + x_2. \tag{14}$$

First, it is interesting to compare prices in the dynamic duopoly with the myopic level $\hat{p}(\eta)$. As is common in dynamic models, prices are set at a level below static duopoly levels, as there is an incentive to price low to build goodwill.

Proposition 1 *Let $p^*(\eta)$ solve (13). Then on any optimal price path, each firm's price $p^*(\eta)$ is always less than the myopic level $\hat{p}(\eta)$.*

Proof: In the Appendix. ▮

Turning to the steady states of the duopoly, in any such steady state by definition $\dot{\eta} = 0$, $\dot{\xi}_1 = 0$ and $\dot{\xi}_2 = 0$. This combined with the first order condition (13) gives us the following equations

$$q + \frac{(u_1+u_2)(2x_1(\eta,q)-1)}{1+r} = \frac{1}{\beta x_2(\eta,q)} - \frac{1}{\beta x_1(\eta,q)}, \quad \eta = x_1(\eta,q)u_1 - x_2(\eta,q)u_2, \tag{15}$$

where $q = p_1 - p_2$ is the relative price. These are nonlinear simultaneous equations, with possible multiple solutions. The nonlinearity arises from the nonlinearity of the demand function $x_i(\eta,q)$ and the degree of its nonlinearity depends on the optimization parameter β. For example, if β is very small then $x_1(\eta,q) \approx \hat{x}_1(\eta,q) = 1/2 + \beta(\eta - q)/4$. If one were to replace x_1 with the linear approximation $\hat{x}_1$, the steady state equations (15) themselves become linear, and a single solution would be guaranteed. However, for higher levels of β, the demand function x_1 is extremely nonlinear, and we do indeed have multiple steady states.

In particular, the equations (15) in fact can be consistent with three distinct steady states. For example, with $u_1 = u_2 = u$, the two products are in effect identical. However, if one assumes for convenience that $u = 2, r = 1, \beta = 2$, there are steady states with $(\eta, p_1, p_2) = (-1.103, 0.196, 0.679)$, $(0,0,0)$, and $(1.103, 0.679, 0.196)$ (x_1, the market share of the first firm, at these three points is 0.224, 0.5 and 0.776 respectively). That is, there is a symmetric outcome, where the market is equally divided. But there also exist steady states, where firm 1 has a high goodwill, and hence high equilibrium price and market share, and a mirror image outcome, where firm 2 is dominant.

Some additional qualitative information can be obtained from drawing a phase diagram. Luckily, it is possible to reduce the original dynamic system in (η, ξ_1, ξ_2) to a two dimensional one in (η, q). From (13), one can obtain

$$\dot{q} - \frac{1 - 2x_1 + 2x_1^2}{x_1(1-x_1)}(\dot{\eta} - \dot{q}) + (u_1+u_2)(\dot{\xi}_1 + \dot{\xi}_2) = 0 \tag{16}$$

It is possible then to solve for $\dot{q}$ though the resulting equation is difficult to work with except for specific examples. For example, for the sample parameter values given above, one can construct the phase diagram in Figure 1. The three equilibrium points identified in the numerical example above are labelled e_1, e_2 and e_3 respectively. From the diagram (and confirmed by numerical analysis) one can see that the steady states e_1 and e_3 are saddlepoints under the dynamics investigated here, and hence approachable under optimal dynamic policies, whereas the symmetric steady state is e_2 is unstable. The next result confirms and generalizes our numerical results: for β large enough there are multiple equilibria and the symmetric outcome is no longer stable.

Proposition 2 *Assume $u_1 = u_2 = u$. Then, there is a symmetric steady state at $(\eta, q) = (0, 0)$. Define*

$$\underline{\beta} = \frac{6(1+r)}{u(3+r)} \geq \frac{2}{u}. \tag{17}$$

Then if $\beta \in (0, \underline{\beta})$, where $\underline{\beta}$ is as given above, then the symmetric steady state is unique, and is a saddlepoint and hence dynamically approachable. However, for $\beta = \underline{\beta}$, there is a bifurcation, and for $\beta > \underline{\beta}$ the symmetric steady state at (0, 0) is dynamically unstable. Furthermore, for $\beta > \underline{\beta}$ there exist two other equilibria, which are saddlepoints.

Proof: In the Appendix. ▮

The above result establishes the possibility of multiple steady states, but there are other good reasons for believing that there should be three possible outcomes. Suppose we look at consumer behavior under this particular learning model under the assumption that each firm adopted a constant price, then we would have

$$\dot{x}_1 = \beta x_1(1 - x_1)(\dot{\eta}) = \beta x_1(1 - x_1)(u(2x_1 - 1) - \eta).$$

But from the choice rule (5), $\beta(\eta - q) = \log x_1 - \log(1 - x_1)$. This gives

$$\dot{x}_1 = \beta x_1(1 - x_1)\left(u(2x_1 - 1) - q + \frac{1}{\beta}(\log(1 - x_1) - \log x_1)\right). \tag{18}$$

This is a perturbed form of the evolutionary replicator dynamic.[11] The replicator dynamics are in effect the limit of these dynamics as β approaches infinity (that is, the replicator dynamics are (18) without the logarithmic terms). Therefore, for β large, this equation will have three equilibria which will be close to those of the replicator dynamics in this context, which are $x_1 = 0$, $x_1 = 1$, and $x_1 = (u + q)/(2u)$. It can be checked that it is the two extreme equilibria that are asymptotically stable. That is, a consumer can, by pure force of habit, become locked into exclusively purchasing one good, and this may not be the one with higher quality. There is a correspondence with the equilibria found above: e_1 represents the consumer becoming locked into the second firm, e_3 being locked into the first firm's product and e_2 is the unstable interior equilibrium.

[11] For more detailed analysis of the connection between different learning models and the replicator dynamics, see Hopkins (2002).

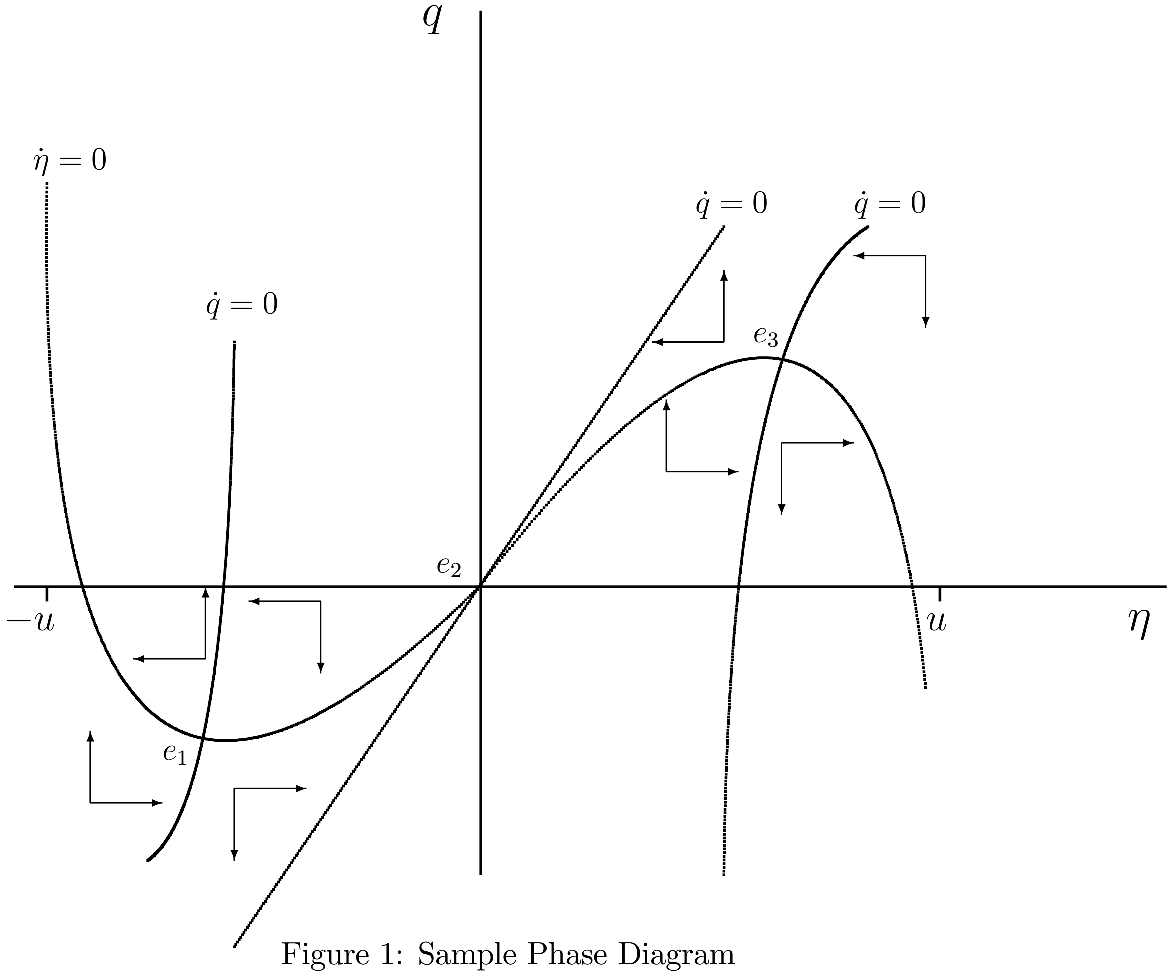

Figure 1: Sample Phase Diagram

Of course, the difference in the full model is that price is not constant, and indeed each firm will choose a dynamic pricing scheme to increase the probability of the consumer becoming locked into its product. Thus, this model implies a considerable first-mover advantage. If initial conditions are such that the consumer has a preference for one particular brand, one would expect convergence to an outcome favorable to that firm. Thus looking at the phase diagram in Figure 1, one can see that if $\eta(0)$ is small but positive, the first firm can choose a low price which will place him on the stable manifold leading to e_3. That is, by choosing a low initial price, the first firm can get naive consumers "hooked". The penalty is that even in the long run, from Proposition 1 the first firm must charge a price below the (myopic) duopoly price.

What happens if the two firms are not symmetric and one holds a quality advantage? By continuity, for u_1 slightly greater than u_2 there must also be multiple steady states close to those we found when u_1 and u_2 were equal. That is, there will still be a steady state where the inferior firm dominates. Of course, we have seen that for low values of the precision parameter β there may be only one steady state. However, the next result shows that for β sufficiently large, there exists a steady state where the inferior firm dominates and has a market share arbitrarily close to one. This is the case no matter

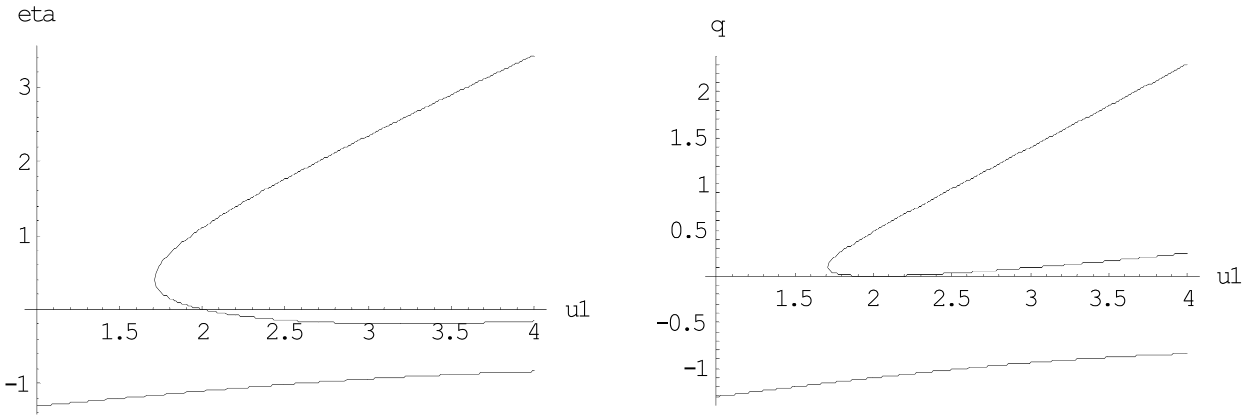

Figure 2: Steady State Values of Relative Goodwill (η) and Relative Price (q)

the degree of the quality advantage of the other firm.

Proposition 3 *Suppose $u_2 > u_1 > 0$: firm 2 has a quality advantage. For any $\epsilon > 0$, there exists a $\beta > 0$ such that there is a steady state outcome where $u_1 - \eta < \epsilon$ and $x_2 < \epsilon$.*

Proof: In the Appendix. ▮

To illustrate this further, Figure 2 presents diagrammatically numerical calculations of how the steady states change with the degree of asymmetry. We fix the parameters β, r, u_2 at 2,1,2 respectively. The quality of the first firm u_1 is varied and appears on the horizontal axis, while steady state values of relative goodwill η (first panel) and relative price q (second panel) are shown on the vertical axis. For $u_1 < 1.72$, there is a unique steady state.[12] For $u_1 > 1.72$, there are three branches each representing a steady state, corresponding to e_1, e_2 and e_3 of Figure 1.

This illustrates there are two possible regimes in the asymmetric case. For $u_1 < 2$, the second firm (quality fixed at $u_2 = 2$) is the high quality firm. For $u_1 < 1.72$, there is only one steady state. Here, both η and q are negative, and the second firm, the higher quality one, dominates. But at $u_1 = 1.72$ two further branches of steady states appear corresponding to e_2 and e_3. As u_1 increases above 2, the quality advantage to firm 1 grows. The distance between e_1 and e_2 narrows and the distance between e_2 and e_3 increases. Remember that if the initial value of η is intermediate between that at e_1 (e_3)

[12] As we have seen, there are multiple steady states only when β is high. Low values of u_1 and u_2, given the logit choice rule, are like low values of β: low incentives like a low value of the precision parameter mean a low probability of a best response.

and e_2 we would expect the system to converge to e_1 (e_3). That is, the relative size of the basin of attraction of the higher quality firm is growing with its quality advantage.

This illustrates that even adaptive consumers do respond to quality differences. First, if the parameter constellation is such that there is only one steady state, then in that steady state the higher quality firm dominates. Second, when there are multiple steady states, the relative size of the basin of attraction of the steady state where the higher quality firm dominates is larger than the steady state where the lower quality firm dominates. However, it remains true that dominance by a low quality firm is a possibility if the initial value of goodwill η is sufficiently close to the appropriate steady state.

5 Comparison with Belief-Based Learning

In this section, I compare the conclusions of the previous section with a similar analysis when learning is based on beliefs, rather than familiarity. I find that, in contrast, asymptotically beliefs are correct, and so in the long run, firms' pricing decisions are the same as in a static model. However, in the short run, firms may make "introductory offers", that is, a low price to induce consumers to try a product for which initially they have a low opinion. These results on dynamic pricing are similar to those found by Shapiro (1983) and, more recently, Bergemann and Välimäki (2004) for the monopoly case. The point made here is that results with belief-based learning are similar to those with more traditional models, but are quite different from those with learning with familiarity we have just seen.

We consider belief based learning, but adaptive and model free, in that agents have no beliefs about the payoff generating process. In particular, Sarin and Vahid (1999) propose an adaptive learning model which is particularly applicable when an agent has partial information in the sense of Rustichini (1999). Or, in the present context, it should be appropriate for the case of experience goods. Fudenberg and Levine (1998, Chapter 4) propose a similar model. The essence of both is that the agent keeps track of the average realized payoff to the different actions available, and chooses the action with the highest average with high probability. The rule was given in Section 2 as rule (3).

The difference between the updating rule of Sarin and Vahid (3) and the rule (2) in the previous section is that now the goodwill toward the good not chosen does not decay. This may seem a slight difference, but it is crucial. In the model of the previous section, the goodwill toward the good not chosen deteriorated. Hence, the consumer became continuously more convinced that she had chosen correctly. Here, because one's estimate of the quality of the good not chosen does not change, this prevents one becoming locked into the other good through pure force of habit.

Moving to expected motion and continuous time, one obtains

$$\dot{\theta}_1 = x_1(u_1 - \theta_1), \quad \dot{\theta}_2 = (1 - x_1)(u_2 - \theta_2). \tag{19}$$

Notice one important thing. In the case of experience goods/partial information, the consumer can only update her estimate of the quality of good i when she chooses good i. The speed of learning of the quality of a good is therefore proportional to the probability of choosing it. For example, $\dot{\theta}_1$ is proportional to x_1. Therefore, a rise in the price of a firm by lowering the probability of purchase will slow the consumer's learning about that firm's quality.

The Hamiltonian for each firm is now

$$H_i = p_i x_i + \mu_i x_i(u_i - \theta_i) + \nu_i x_j(u_j - \theta_j) \tag{20}$$

This gives us first order conditions of

$$\frac{\partial H_i}{\partial p_i} = x_i + p_i \frac{\partial x_i}{\partial p_i} + \mu_i \frac{\partial x_i}{\partial p_i}(u_i - \theta_i) - \nu_i \frac{\partial x_i}{\partial p_i}(u_j - \theta_j) = 0 \tag{21}$$

for $i = 1, 2$ and $j = 3 - i$. From this, it is possible to obtain

$$p_i = \frac{1}{\beta x_j} - \mu_i(u_i - \theta_i) + \nu_i(u_j - \theta_j) \tag{22}$$

where μ_i and ν_i are the respective solutions to

$$\dot{\mu}_i = \mu_i(x_i + r) - x_i, \quad \dot{\nu}_i = \nu_i(1 - x_i + r) + x_i. \tag{23}$$

In the steady state one can calculate from (19) that $\theta_i = u_i$ for $i = 1, 2$ and from (22) one can see that the price solves $p_i = 1/(\beta(1 - x_i))$, just as in (9). Hence, we have the following result.

Proposition 4 *In the model of experience goods with the Sarin and Vahid learning model, asymptotically the consumer has a correct perception of the quality of the two goods, $\theta_i = u_i$, and the firms charge the myopic duopoly prices $\hat{p}(u^*)$.*

Proof: In the Appendix. ▮

That is, there is complete learning. In the limit, the consumer knows the true values of u_1 and u_2. While this result follows directly from the specification of the learning rule, it remains important for two reasons. First, it highlights the source of the problems with the reinforcement rule. As this learning rule performs well under partial information, partial information in itself cannot be the reason that causes lock-in. Rather, it is how one adjusts assessments of actions not chosen. Second, as most research on learning in economics has concentrated on reinforcement learning, this type of result while simple is still novel.[13] This concentration of researchers on reinforcement learning in turn suggests that writing down a learning rule that is well-behaved in a situation of partial information is not as simple as it might seem.

Since in the limit neither firm can influence the consumer's beliefs, the price asymptotically approaches its myopic level. However, pricing away from the steady state will not necessarily be myopic. Indeed, it is possible to characterize this difference more precisely. As a convenient simplification, assume that the consumer initially has correct beliefs about second brand, but still has some learning to do about the first firm's product. That is, assume $\theta_2(0)$ is equal to u_2 but that $\theta_1(0)$ is not equal to u_1. Then, one can show that, away from the steady state, the optimal price for both firms is higher (lower) than the myopic price $\hat{p}$ if θ_1 is greater (lower) than u_1. Clearly, when the consumer is initially pessimistic about the first brand believing it worse that it really is, the first firm has an incentive to induce the consumer to try its product as this will improve the consumer's opinion. However, when consumers are optimistic, the firm has an incentive to raise prices to slow consumer learning.[14] A higher price will decrease frequency of purchase and hence reduce the speed at which θ_1 falls to u_1.

Proposition 5 *Assume $\theta_2(0) = u_2$ but that $\theta_1(0) \neq u_1$. Then, let $p^*(\eta)$ solve (22). If $\theta_1(0) > u_1$ ($\theta_1(0) < u_1$), the price of the first firm is set higher (lower) than the myopic level, that is, $p_1^*(\eta) > \hat{p}_1(\eta)$ ($p_1^*(\eta) < \hat{p}_1(\eta)$), for all finite time. If $\theta_1(0) > u_1$ ($\theta_1(0) < u_1$), the price of the second firm is set higher (lower) than the myopic level, that is, $p_2^*(\eta) < \hat{p}_2(\eta)$ ($p_2^*(\eta) > \hat{p}_2(\eta)$), for all finite time.*

Proof: In the Appendix. ∎

Note that the above proposition reveals that the firm whose quality is known will also respond to the consumer's uncertainty about its rival. If the consumer's opinion about the other product is initially lower than the true value, that is $\theta_1(0) < u_1$, the second firm has an incentive to charge a price lower than in a static duopoly for the same level of goodwill. First, this represents a competitive response to the lower price of firm 1, which is trying to build up custom. But there is a second motive: it is to slow learning about the quality of the rival product.

6 Heterogeneity and Open Loop Equilibrium

There are well known limitations to open loop equilibrium as a solution concept for dynamic games. As it involves the firms choosing their pricing policies once and for all

[13] This type of result seems well known in artificial intelligence, however. See Sutton and Barto (1998).

[14] This effect is over and above the incentive to charge a high price because θ_1 is high. This latter effect is included in $\hat{p}$.

at the start of the game, it does not allow them to revise their choices in the light of experience. Here, a literal reading of the formal model presented here would support the use of open loop equilibrium as the evolution of the representative consumer's goodwill is deterministic and so both firms can make accurate forecasts of how demand will evolve given initial conditions and their choice of pricing policy. However, there are many problems with using a representative agent or consumer (Kirman (1992)). Indeed, the deterministic equations employed here are only approximations of the true point of interest: the behaviour of a large number of heterogeneous consumers whose own experience is history dependent and driven by random shocks. Therefore, it is important to ask exactly how good an approximation this is.

In a static setting, logit demand functions are used precisely in order to model the aggregate demand of a large population (see, for example, Caplin and Nalebuff (1991)). In the current notation, suppose a large population chose brand 1 if and only if $\eta - q > 0$ and goodwill η was distributed in the population according to a logistic distribution, then aggregate demand would be given by the logit demand (5). However, the problem in the current dynamic setting is that individual goodwill will evolve stochastically according to which good is purchased and the realisations of payoff shocks. Thus even if the initial distribution of goodwill is logistic, it is unlikely to remain so. Since modelling an endogenous distribution is extremely challenging, some kind of approximation is called for. Furthermore, I would argue that averaging over this changing population is not nonsensical.

This argument is stronger in the belief-based learning case. In this case, asymptotically beliefs will be correct. Or, more precisely, one can adapt the results here, by use of stochastic approximation theory, to show that in the truly stochastic case beliefs will be nearly correct with high probability. Exactly how dispersed beliefs will be depends on the model's parameters, particularly β and δ. It remains possible for an individual to have a series of realisations that would take her beliefs far from the correct level. However, if the population is large, then the law of large numbers would ensure that a large proportion of the population at any time would have beliefs close to being correct (there are no aggregate shocks in this model only individual).[15] Beliefs for most of the population will also be close to the average belief, and so the average consumer will be a good approximation for the population distribution.

In the reinforcement learning case, we have seen that when the the precision parameter β is sufficiently high, consumers will tend to become locked into choosing one brand only. With a population of consumers, there would be a critical level of relative goodwill such that for consumers with goodwill greater (lower) than this, they would be expected to be attracted toward always purchasing brand one (two). Even this would not be problematic for a deterministic approximation based on a consumer with average goodwill, provided all consumers have initial goodwill close to the average level. In this

[15] If one wanted to model the case where payoff shocks are correlated (e.g. firms sometimes produce a bad batch of the product and/or there are taste shocks that are driven by fads that transmit across individuals), then indeed one should look at the Markov perfect equilibrium of a truly stochastic model.

case, (nearly) all consumers will become locked in to the same brand as the average consumer, and the average consumer provides a good approximation.

However, suppose goodwill was initially relatively dispersed so that there were significant numbers on both sides of the critical level. Then a substantial proportion of consumers would be attracted to a different brand than the average consumer. The deterministic model predicts dominance for the initially advantaged firm, yet many consumers would actually be faithful to the other brand. That is, in this case the current model does not fit well. What would do better? Tracking a diverging population is technically difficult. Simulation, such as in Kirman and Vriend (2001), is one way forward. Another might be to have two representative consumers, each one representing a proportion of the population. Preliminary work in this direction indicates the possibility of even more equilibria than in Section 4. The additional equilibria capture the possibility of each firm having a mass of loyal consumers. However, equilibria with the market dominated by one firm still exist. Detailed exploration of these possibilities is left to further research.

7 Empirical Implications

An interesting question is whether the differing theoretical predictions of the learning models described above can be subject to empirical testing. Of course, there is now a large literature on testing learning models with data from experiments.[16] However, the work of Chintagunta and Rao (CR) (1996) and Ho and Chong (2002) suggests the fascinating possibility of using field data instead. In this section, we formulate a testable hypothesis and consider whether CR's existing empirical work helps to distinguish between different models of learning.

CR analyze scanner panel data on the purchases of yogurt over a two year period. An immediate difference between such marketing data sets and experimental data is that while choices, that is which brand was purchased, may be recorded, payoffs in this context are fundamentally unobservable. We have no way of knowing the level of satisfaction derived from the consumption of a good, or, in the present notation, we can neither observe u_i or $\tilde{u}_i$. To circumvent this problem, CR normalize the utility from each purchase to one. It is still possible, using the observed choices, to construct values for each $\theta_i(t)$ by using the following formula based on the reinforcement learning rule (2),

$$\theta_i(t+1) = (1-\delta)\theta_i(t) + \delta I_i(t). \tag{24}$$

Here, $I_i(t)$ is the indicator function, taking value 1 if the consumer purchased brand i

[16] Some works of many in this field are Erev and Roth (1998), Camerer and Ho (1999) for games, and Erev and Barron (2001) for decision problems.

at time t and zero otherwise. CR then employ logit regressions of the form

$$\Pr[\text{choose good } i \text{ at time } t] = \frac{\exp(\alpha_i + \beta_i \theta_i(t) + \gamma_i p_i(t))}{\sum_{j=1}^{2} \exp(\alpha_j + \beta_j \theta_j(t) + \gamma_j p_j(t))} \tag{25}$$

This specification allows for different sensitivities β_1, β_2 to the stock of goodwill for different brands. Thus different quality levels for the two brands will enter through different estimates for each β_i rather than different values for $\tilde{u}_i$.

One important conclusion from our earlier analysis of consumer learning using the reinforcement learning rule is that outcomes will be history dependent. A consumer will become locked into the brand which she purchases most frequently, possibly simply because of an initial preference. Thus, if this model is an accurate description of actual consumer behavior, a logit regression should find positive and significant estimated coefficients on the measures of goodwill θ_i that depend on past purchases.

We could take a similar approach using the Sarin-Vahid belief-based model. However, one crucial difference becomes apparent. By Proposition 4, in the steady state, under the Sarin-Vahid model, each θ_i is at its "correct" value and independent of the consumer's purchase history. Differing levels of quality should instead appear in differing estimates of α_1 and α_2. We can summarize this argument in the following hypothesis.

Hypothesis 1 *Consider logit regressions of the form (25) with θ constructed from the procedure (24). If the reinforcement learning model correctly describes consumer behavior, a regression including the θ_i variables that reflect recent purchase history should outperform a regression with them omitted. However, if learning is belief-based, the reverse should be true.*

The interesting thing is that CR effectively tested this hypothesis in their original paper by running an alternative regression in which the θ variables were omitted. This regression performed significantly worse in terms of log likelihood than the regression with the θ_i included. When the θ_i were included, the coefficients β_i were significant and positive. Ho and Chong (1999) also perform logit regressions on consumer data but using a somewhat more complex reinforcement learning model.[17] They also find effects from recent purchases. Thus, it seems that actual consumer behavior gives stronger support for the reinforcement learning model than for the model of Sarin and Vahid.

Some caution, however, should be exercised in making this assessment. A crucial assumption in making Hypothesis 1 above, was that the empirical data reflects steady state behavior, an assumption also made by CR in their analysis. As we have seen in Section 5, out of the steady state, in the Sarin and Vahid model learning is affected by

[17] The published version (Ho and Chong (2003)) does not use logit, but the conclusions are similar. They also find an effect similar to hypothetical reinforcement, even though the products in question are experience goods. It seems that seeing a product in a store may reinforce one's memory of it, even if it is not purchased at that time.

the choices made. It is only asymptotically that one's quality estimates θ_i are independent of choices. Thus, the fact that the model including purchase history outperforms the model with it omitted, may only reflect non-equilibrium behavior, not a failure of the Sarin-Vahid model. This problem of discontinuity, that near the steady state the propensities θ have a significant effect but at the steady state their influence disappears, is highlighted by Blume et al. (2002). They argue that it is therefore relatively difficult to separate different learning models econometrically once play is at or close to equilibrium, as opposed to when learning is still active.

In this spirit, we can offer a new test between the two forms of learning, that does not depend on equilibrium having been reached. Although there have been several attempts to distinguish between belief-based and reinforcement learning they have concentrated on whether agents do or do not use information about foregone payoffs, that is, they test between rules (4) and (2). For example, Camerer and Ho (1999) estimate an model that nests those two rules. In a similar way, we could could nest rules (2) and (3) by replacing the empirical formulation (24) with

$$\begin{aligned} \theta_i(t+1) &= (1-\delta_1)\theta_i(t) + \delta_1 I_i(t) \\ \theta_j(t+1) &= (1-\delta_2)\theta_j(t), \text{ for all } j \neq i \end{aligned} \tag{26}$$

That is, if $\delta_1 = \delta_2$ we have the reinforcement learning model, and if $\delta_2 = 0$ and $\delta_1 > 0$ we have the belief-based model.

Hypothesis 2 *Suppose we jointly estimate the parameters $(\beta, \gamma, \delta_1, \delta_2)$ using the procedure (26) to construct estimates of the goodwill θ. If the reinforcement learning model correctly describes consumer behavior, estimates of δ_2 should be positive and close to those for δ_1. However, if learning is belief-based, estimates of δ_2 should not be significantly different from zero.*

This hypothesis, to my knowledge, has not yet been tested either with experimental data or data on actual consumer choices.

8 Conclusion

This paper explores the consequences of recent advances in adaptive learning theory for the analysis of consumer behavior. The case of experience goods corresponds to partial information in the learning literature. Two different models of learning are compared in this setting. The first, a model of reinforcement learning, may be biased with consumers becoming locked into inferior choices. This leads to the possibility of multiple steady states. When there are multiple states, the stable ones are those which involve dominance by one firm. Under a model of belief-based learning, due to Sarin and Vahid (1999), consumers will learn accurately in the long run and so there is only one long run equilibrium. However, in the short run a seller has an incentive to charge a price different from the myopic maximum to affect the speed at which consumers learn.

Whether these different models can be separated empirically is an interesting question. The availability of consumer scanner data now permits investigation by the examination of individual consumer behavior. In Section 7 of this paper, two simple tests for the identification of different types of learning behavior were suggested. Some of the existing empirical evidence, both from the laboratory and field consumer data, seems to give greater support for the reinforcement learning model that predicts suboptimal behavior even in the long run.

The market outcomes under reinforcement learning are probably best interpreted as an important first mover advantage. Familiarity with an existing brand will make the establishment of an alternative difficult, even if it is higher quality, at least under price competition. It is an open question whether there would be a different conclusion if other forms of competition were included. For example, it has long been asserted that certain forms of advertising convey no information, but only serve to aid familiarity. Thus, the investigation of the effect of advertising when consumers are reinforcement learners seems a natural complement to the current research.

The assumptions and methodology employed in this paper are quite different from those of the strategic approach to dynamic pricing. It would be interesting to analyze the robustness of the two types of model. In particular, both the assumption that all consumers can act as though they understand the intuitive criterion and the present alternative, that all consumers are incapable of any strategic inference, seem extreme. Some heterogeneity amongst consumers would seem more reasonable. For example, how would the current results change if a proportion of consumers were sophisticated rather than adaptive? Or, for example, can one successfully signal high quality when such a signal is simply not understood by a proportion of its intended audience? As a final remark, the existing experimental evidence, for example, Cooper, Garvin and Kagel (1997), as well as supporting heterogeneity, suggests that adaptive learning does better than equilibrium refinements at explaining actual human behavior.

Appendix

Proof of Proposition 1: From the equation (14) one can calculate the equilibrium value of ξ_1 as $x_1/(1+r) > 0$. Second, if at any point ξ_1 were negative, given (14), ξ_1 would clearly diverge to negative infinity. Hence, ξ_1 is always positive on any optimal path. Similarly, one can calculate that ξ_2 is always negative. Comparing (13) with the myopic first order condition (9), it is easy to see that $p_1^* < \hat{p}_1$ and $p_2^* < \hat{p}_2$ if $\xi_1 > 0$ and $\xi_2 < 0$. More formally, differentiating the first order conditions, and evaluating the second order derivatives at the equilibrium point, we have

$$\left[\begin{array}{cc} -\beta x_1 & \beta x_1^2 \\ \beta x_2^2 & -\beta x_2 \end{array}\right]\left[\begin{array}{c} dp_1 \\ dp_2 \end{array}\right]=\left[\begin{array}{r} u_1+u_2)d\xi_1 \\ -(u_1+u_2)d\xi_2 \end{array}\right].$$

It is then easy to verify that $\partial p_1^*/\partial\xi_1 < 0$ and $\partial p_1^*/\partial\xi_2 > 0$, and, equally, $\partial p_2^*/\partial\xi_1 < 0$ and $\partial p_2^*/\partial\xi_2 > 0$. ■

Proof of Proposition 2: From (15) a steady state of the open loop equilibrium with $u_1 = u_2 = u$ will simultaneously satisfy

$$u(2x_1(\eta, q) - 1) = \eta \tag{27}$$

and, from the first order conditions (13) together with the steady state of the costate dynamics (14),

$$\frac{1}{\beta(1-x_1(\eta,q))}-\frac{1}{\beta x_1(\eta,q)}-\frac{2u(2x_1(\eta,q)-1)}{1+r}=q \tag{28}$$

Indeed, it is easily established that $(\eta, q) = (0, 0)$, implying $x_1 = x_2 = 1/2$, is such a state.

We have a dynamical system on (η, q) defined by simultaneous differential equations (11) and (16). The Jacobian at (0,0) can be calculated as

$$J=\left[\begin{array}{cc} \beta u/2-1 & -\beta u/2 \\ -2(2+r-\beta u)/3 & 1+r-2\beta u/3 \end{array}\right].$$

The determinant $|J| = -1+\beta u/2+r(\beta u/6-1)$ which is negative (positive) for $\beta < (>)\underline{\beta}$. Thus, the symmetric steady state (0,0) is a saddlepoint for low β, but as the trace is positive for $\beta = \underline{\beta}$, the equilibrium is a source for $\beta > \underline{\beta}$.

Thus we have a bifurcation at $\underline{\beta}$ (and at $\underline{\beta}$ alone, as this is the only value for which J is singular). We now show that it is a pitchfork bifurcation, so that as the symmetric steady state changes from a saddlepoint to a source, two new saddlepoint steady states are created. We verify this by application of Theorem 10.1a in Tu (1994), which provides conditions that the bifurcation is not degenerate.[18] If the conditions are met, the theorem states that two curves of equilibria, in the space (η, q, β) intersect at $(0, 0, \underline{\beta})$, so that there are multiple equilibria for $\beta > \underline{\beta}$. The conditions are that $Jy_1 = 0$ and $y_2 \cdot J = 0$ for non-zero vectors y_1, y_2, that is there are non-degenerate left and right eigenvectors corresponding to the zero eigenvalue of J evaluated at (0,0), but also that $y_2 \cdot J_\beta y_1 \neq 0$, where J_β is the derivative of J with respect to β. We have

$$y_2\cdot J_\beta y_1=\left[\begin{array}{c} -(1-r)/3 \\ 1 \end{array}\right]\left[\begin{array}{cc} u/2 & -u/2 \\ 2u/3 & -2u/3 \end{array}\right]\left[\begin{array}{c} 3(1+r)/(2r) \\ 1 \end{array}\right]=\frac{u(3+r)^2}{12r}\neq 0.$$

[18] An example of a “degenerate” bifurcation is a linear system, where there is always a single equilibrium point even when a change in parameters change the sign of one of the eigenvalues.

There must be two additional equilibria and they must be saddlepoints by verification that the bifurcation is of the supercritical pitchfork type.[19] To do this, I apply results from Kuznetsov (1995, Chapter 7). The dynamical system we consider is entirely symmetric, in that $(\dot{\eta}(\eta, q), \dot{q}(\eta, q)) = -(\dot{\eta}(-\eta, -q), \dot{q}(-\eta, -q))$. That is, if we write $y = (\eta, q)$ and $\dot{y} = f(y, \beta)$, then $Rf(y, \beta) = -f(Ry, \beta)$, where $R = -I$ and I is the identity matrix. Therefore, in the terminology of Kuznetsov, the dynamical system is $\mathbf{Z}_2$-equivariant. Kuznetsov defines the set X^- such that $Ry = -y$ for $y \in X^-$, thus here $X^- = \mathbb{R}^2$. Now, by Theorem 7.7 of Kuznetsov (1995), for a $\mathbf{Z}_2$-equivariant system, when the eigenvector corresponding to the zero eigenvalue is in the set X^-, a bifurcation will be of the pitchfork type. Now, as here X^- is whole space, the eigenvector is in X^-, and the bifurcation is a pitchfork. Furthermore, as we know from above that the negative eigenvalue of J moves to positive as β increases, the bifurcation is supercritical. That is, we move from one to three equilibria, with the new equilibria having the same stability properties that the original equilibrium possessed for $\beta < \underline{\beta}$. That is, they are saddlepoints as claimed. ∎

Proof of Proposition 3: The first step is to show that in any steady state where $\eta > 0$, then $q < \eta$. The myopic duopoly equilibrium given by (9) defines implicitly a myopic level of relative price $\hat{q}(\eta)$. We have, clearly, $\hat{q}(0) = 0$ and, by the implicit function theorem, $\partial\hat{q}(\eta)/\partial\eta = ((1 - x_1)^2 + x_1^2)/((1 - x_1)^2 + x_1) < 1$. So, $\hat{q}(\eta) < \eta$ for $\eta > 0$. Comparison of the myopic conditions (9) with the steady state of the dynamic equilibrium (15) shows that the dynamically optimal level q^* will be less than $\hat{q}$ for $\eta > \hat{q} > 0$ as this implies $x_1 < 1/2$. So, $q^* < \eta$ in any steady state with $\eta > 0$. Fix q at its equilibrium level $q^*(\eta)$ and substitute into the second equation in the steady state conditions (15). Since, $q^*(\eta) < \eta$, we have by the properties of the logit choice function (5), $\lim_{\beta\to\infty} x_1(\eta, q^*(\eta)) = 1$, and consequently $\lim_{\beta\to\infty} \eta^*(q^*(\eta)) = u_1$. The result follows. ∎

Proof of Proposition 4: Inspection of the system of differential equations in (19) reveals that for $x_1 \in (0, 1)$, there is only one fixed point which is $(\theta_1, \theta_2) = (u_1, u_2)$. It is easy to prove (e.g. $V = (\theta_1 - u_1)^2 + (\theta_2 - u_2)^2$ is a suitable Liapunov function) that this fixed point is a global attractor, again for $x_1 \in (0, 1)$. But, given the functional form (5), it is always true that $x_1 \in (0, 1)$. ∎

Proof of Proposition 5: If $\theta_1(0) > u_1$ it is easy to demonstrate that θ_1 converges asymptotically to u_1 but that $\theta_1(t) > u_1$ for all finite t. One can then compare the myopic first order conditions (9) with (22) and see that $p_1^* > \hat{p}_1$ if $\mu_1 > 0$. Again, examining (23), it is clear that μ_1 must always be positive to be able to attain its steady state value which is positive. For p_2^*, it is clear that ν_2 must always be negative to be able to attain its steady state value which is negative. More formally, differentiating the first order conditions, and evaluating the second order derivatives at the equilibrium

[19] A pitchfork bifurcation occurs when, as the bifurcation parameter (here β) passes the critical level, there is a change from 1 to 3 equilibria. This is supercritical if, at the same time, the original equilibrium changes from stable to unstable.

point, we have
$$\begin{bmatrix} -\beta x_1 & \beta x_1^2 \\ \beta x_2^2 & -\beta x_2 \end{bmatrix} \begin{bmatrix} dp_1 \\ dp_2 \end{bmatrix} = \begin{bmatrix} (u_1 - \theta_1) d\mu_1 \\ -(u_1 - \theta_1) d\nu_2 \end{bmatrix}.$$
So, we have $\partial p_1^* / \partial \mu_1 > 0$ when $\theta_1 > u_1$. ∎

References

Arthur, W.B. (1993). On Designing Economic Agents that Behave like Human Agents. Journal of Evolutionary Economics 3, 1-22.

Bagwell, K., Riordan, M. (1991). High and declining prices signal product quality. American Economic Review 81, 224-239.

Benaïm, M. (1998). Recursive algorithms, urn processes and chaining number of chain recurrent sets. Ergodic Theory and Dynamical Systems 18, 53-87.

Benaïm, M. (1999). Dynamics of Stochastic Algorithms. In: J. Azéma et al. (Eds), Séminaire de Probabilités XXXIII, Springer-Verlag: Berlin.

Bergemann, D., Välimäki, J. (1996). Learning and strategic pricing, Econometrica 64, 1125-1149.

Bergemann, D., Välimäki, J. (2004). Monopoly pricing of experience goods, working paper.

Blume, A., Dejong, D.V., Neumann, G.R., Savin, N.E. (2002). Learning and communication in sender-receive games: an econometric investigation, Journal of Applied Econometrics 17, 225-247.

Börgers, T., Sarin, R. (2000). Naive reinforcement learning with endogenous aspirations, International Economic Review 41, 921-950.

Börgers, T., Sarin, R., Morales, A. (2004). Expedient and Monotone Learning Rules, Econometrica 72, 383-405.

Camerer, C., Ho, T-H. (1999). Experience-weighted attraction learning in normal form games, Econometrica 67, 827-874.

Caplin, A., Nalebuff, B. (1991). Aggregation and imperfect competition: on the existence of equilibrium, Econometrica 59, 25-59.

Cellini, R., Lambertini, L. (1998). A dynamic model of differentiated oligopoly with capital accumulation, Journal of Economic Theory 83, 145-155.

Chintagunta, P.K., Rao, P.V. (1996). Pricing strategies in a dynamic duopoly: a differential game model, Management Science 42, 1501-14.

Cooper, D., Garvin S., Kagel, J. (1997). Adaptive Learning versus Equilibrium Refinements in an Entry Limit Pricing Game, Economic Journal 107, 553-575.

Ellison, G., Fudenberg, D. (1995). Word-of-Mouth Communication and Social Learning, Quarterly Journal of Economics 110, 93-125.

Erdem T. et al. (1999). Brand equity, consumer learning and choice, Marketing Letters 10, 301-318.

Erev, I., Barron, G. (2001). On adaptation, maximization and reinforcement learning among cognitive strategies, working paper, Columbia University.

Erev, I., Haruvy, E. (2001). Variable pricing: a customer learning perspective, working paper.

Erev, I., Roth, A.E. (1998). Predicting how people play games: reinforcement learning in experimental games with unique, mixed strategy equilibria, American Economic Review 88, 848-881.

Fudenberg, D., Levine, D. (1998). The Theory of Learning in Games. MIT Press: Cambridge, MA.

Gabaix, X., Laibson, D. (2003). Some industrial organization with boundedly rational consumers, working paper.

Harrington, J.E., Chang, M-H (2003). Co-Evolution of firms and consumers and the implications for market dominance, Journal of Economic Dynamics and Control.

Ho, T-H., Chong J-K. (1999). A Parsimonious Model of SKU Choice: Familiarity-based Reinforcement and Response Sensitivity, working paper, Wharton School.

Ho, T-H., Chong J-K. (2003) A Parsimonious Model of Stock-Keeping Unit Choice, Journal of Marketing Research 40, 351-365

Hopkins, E. (2002). Two competing models of how people learn in games, Econometrica 70, 2141-2166.

Hopkins, E., Posch, M. (2005). Attainability of Boundary Points under Reinforcement Learning, Games and Economic Behavior 53, 110-125.

Hopkins, E., Seymour, R. (2002). The stability of price dispersion under seller and consumer learning, International Economic Review 43, 1157-1190.

Kamien, M.I., Schwartz, N.L. (2000). Dynamic Optimization. Second Edition. North Holland: Amsterdam.

Kirman, A.P. (1992). Whom or what does the representative individual represent, Journal of Economic Perspectives 6(2), 117-36.

Kirman, A.P., Vriend, N.J. (2001). Evolving market structure: an ACE model of price dispersion and loyalty, Journal of Economic Dynamics and Control 25, 459-502.

Kuznetsov, Y.A. (1995). Elements of Applied Bifurcation Theory. Springer-Verlag: New York.

Milgrom, P., Roberts, J. (1986). Prices and advertising signals of product quality, Journal of Political Economy 94, 796-821.

Rabin, M., Schrag, J.L. (1999). First impressions matter: a model of confirmatory bias, Quarterly Journal of Economics 114, 37-82.

Rustichini, A. (1999). Optimal properties of stimulus-response learning models, Games and Economic Behavior 29, 244-73.

Sarin, R., Vahid, F. (1999). Payoff assessments without probabilities: a simple dynamic model of choice, Games and Economic Behavior 28, 294-309.

Schmalensee, R. (1978). A model of advertising and product quality, Journal of Political Economy 86, 485-503.

Seetharaman, P.B., Chintagunta, P.K. (1998). A model of inertia and variety-seeking with marketing variables, International Journal of Research in Marketing 15, 1-17.

Shapiro, C. (1983). Optimal pricing of experience goods, Bell Journal of Economics 14, 497-507.

Smallwood D.E., Conlisk, J. (1979). Product quality in markets where consumers are imperfectly informed, Quarterly Journal of Economics 93, 1-23.

Sutton, J. (1991) Sunk Costs and Market Structure. MIT Press: Cambridge, MA.

Sutton, R.S., Barto, A.G. (1998) Reinforcement Learning: An Introduction. MIT Press: Cambridge, MA.

Tu, P.N.V (1994). Dynamical Systems. Second Edition. Springer Verlag: Berlin.

Weisbuch G., Kirman A., Herreiner D. (2000). Market Organisation and Trading Relationships, The Economic Journal 110, 411-436.