Planning Costs and the Theory of Learning by Doing

W. Bentley McLeod

USC Center for Law, Economics & Organization
Research Paper No. C03-16

CENTER FOR LAW, ECONOMICS AND ORGANIZATION RESEARCH PAPER SERIES

Sponsored by the John M. Olin Foundation

University of Southern California Law School
Los Angeles, CA 90089-0071

Planning Costs and the Theory of Learning by Doing

W. Bentley MacLeod*

Department of Economics and The Law School

The University of Southern California

Los Angeles, CA 90089-0253

wmacleod@usc.edu

June 2003

Keywords: decision theory, learning by doing, contingent plans, status quo effects, incomplete contracts, behavioral economics

Abstract

This paper illustrates how the explicit introduction of planning costs into a model of decision making under uncertainty can result is a theory of learning by doing that is empirically implementable. Even when not optimal, it is shown that learning by doing results in convergence to optimal choice under very general conditions. Hence, it may explain why learning by doing (or adaptive learning) is a good first order model of behavior for a wide variety of environments.

*I would like to thank Boyan Jovanovic, Paul Rosenbloom, and Roberto Weber for supplying me with their data. The paper has benefited from the comments of George Akerlof, Colin Camerer, Janet Currie, Dick Day, Ido Erev, Phillip Jehiel, Yaw Nyarko, Bernard Sinclair-Desgagné and Joel Sobel. However, none are to be implicated for the deficiencies of the paper. Finally, I very much appreciated the research assistance provided by Mehdi Farsi and Alexey Ravichev. I am also grateful for the comments during presentations of earlier versions of this work at the American Economics Association Meeting, McArthur Foundation Conference, Hoover Institute, Stanford, Conference on Cognition, Emotion and Rational Choice, Center for Comparative Political Economy, UCLA, the Marshall School of Business, USC, University of British Columbia, Munich University and Harvard University. The financial support of National Science Foundation grant SBR-9709333 and SES-0095606 is gratefully acknowledged. I also benefited from the hospitality of the Center for Economic Studies, Munich while working on this paper.

1 Introduction

In the movie *Groundhog Day,* Phil, played by Bill Murray, must constantly repeat one day in his life in the town of Punxsutawney, Pennsylvania until he finally gets it right. Given that the problems faced by Phil do not change, one wonders why he simply cannot think them through and work out in advance how he should behave. The purpose of this paper is to show that when an individual faces a problem that is sufficiently complex (in the sense that the number of possible scenarios one faces is very large) then Phil's strategy may be optimal - learning by doing is an optimal decision procedure. The goal of the paper is to highlight how complexity costs as measured by the number of potential states can provide insights into a number of features of observed behavior, including learning by doing.

It is well understood that in practice the set of possible states describing most decision problems is much too large to be contemplated. As Savage (1972) points out, "the look before you leap principal is preposterous if taken to extremes".[1] At the time Savage wrote these words there was not a widely accepted theory of decision making in the face of uncertainty, the goal of the book in which these words appear. More recently many people would argue, and Simon (1978) has done so explicitly, that modern economics is often guilty of taking the "look before you leap principal" to preposterous extremes. The problem is that there is no widely accepted way to incorporate decision costs into a simple, tractable model.[2]

Moreover, the claim that complexity as measured by the number of possible states is controversial. For example, contracts, like decisions, are contingent plans of action. Yet Hart and Holmström (1987) in their discussion of the work by Dye (1985), argue that one cannot easily measure complexity based upon the number of potential contingencies, a point of view recently reiterated by Tirole (1999). To highlight to role of these complexity costs, it is assumed that the evaluation of future contingencies is the *only* relevant cost. Surprisingly, this results in a theory that is consistent with a number of the features of observed behavior and decision making.

The basic model supposes that each period the individual faces the following sequence of decisions and choices. At the beginning of the period she spends time making a contingent plan in anticipation of an event that will require an immediate response. This period corresponds to "decision making". A random event then occurs requiring a response that corresponds to a "choice". By choice we mean that the individual is aware of her choice, but does not have sufficient time to explore all the implications of the choice, and hence it may not necessarily be optimal. The difference here is based upon Alan Newell (1990)'s distinction between decision making in the "rational band" - where contingent planning is possible - and choices in what

[1] Page 16.

[2] See Conlisk (1996) for a review of this literature.

he calls the "cognitive band" - the individual is aware of the choice made, but contingent planning is not possible in the time allotted. Choices are linked to decisions by supposing that if an individual has made a contingent plan for an event that subsequently occurs, or has been previously experienced (and hence learned by doing), then the choice is optimal and consistent with her preferences.[3] In the absence of a plan or previous experience with an event, it is assumed there is a chance the individual chooses an action that is inconsistent with her true preferences - in other words the individual makes a mistake.

Given that planning for all possible contingencies is costly, a trade-off arises. An individual can either make a complete contingent plan, or save upon planning costs by using an incomplete plan and hope that she has included in her plan events that are likely to occur. If an unexpected event occurs, then there is a chance she will make a mistake. However, should that event occur again in the future then she is assumed to learn from her experience and make an optimal choice at that time. I follow Rothschild (1974), and model uncertainty by assuming individuals have beliefs over the possible probability distributions determining the likelihood of different events occurring. With experience individuals not only learn how to respond to particular events, they also learn which events are more likely. It is shown that as uncertainty increases, then the amount of planning that is optimal decreases, and in the limit it may be optimal to engage in no planning.

It is then shown that this model can explain the behavioral anomaly known as probability matching. That is when given the choice between two random rewards individuals tend to allocate time in proportion to the probability of reward, rather than making the optimal choice that consists of allocating all time to the choice with the highest probability of success.

A limitation of the approach is that complexity is measured by the number of possible contingencies. The question then is how to think about this problem when the state space is infinite. It is well known that one technique that individuals use for dealing with complexity is to group states together using some similarity measure. The way similarity is measured is a very deep and open question.[4] We show that a very simple and naive similarity measure is sufficient to ensure that the learning by doing algorithm converges to an optimal decision for a large class of decision problems. For this class of problems it is not possible to prove the existence of a optimal similarity measure, however it provides an additional illustration of the power of learning by doing as a general algorithm for decision making.

Finally, it is shown that the model can be used to derive an empirically implementable model of learning

[3] The difference here is based upon Alan Newell (1990)'s distinction between decision making in the "rational band" - where contingent planning is possible - and choices in what he calls the "cognitive band" - the individual is aware of the choice made, but contingent planning is not possible in the time alloted.

[4] See Churchland and Sejnowski (1993) for a good discussion of the issues.

that provides a theoretical foundation for a class of learning curves, first suggested by Newell and Rosenbloom (1981). The empirical potential of the model is illustrated using data on corporate culture taken from Weber and Camerer (2001), and it is compared to a more standard Bayesian learning model studied by Jovanovic and Nyarko (1995).

The agenda of the paper is as follows. The next section introduces the basic model, and provides conditions under which "learning by doing" is an optimal procedure. Section 3 discusses the behavioral implications of the model, while section 4 illustrates an empirical implementation of the model. Section 5 contains a concluding discussion.

2 The Model

Consider an individual making choices at times $t = 0, 1, 2...$ given by $d_t \in D$ in response to an event $\omega^t \in \Omega$, where Ω is a finite set of events. Without loss of generality the decision set is restricted to a binary choice: $D = \{0, 1\}$.[5] In addition, it is assumed that the events are selected by an *i.i.d.* stochastic process, where $\mu(\omega)$ is the probability that ω is chosen in period t. The decision maker does not know this distribution, but updates beliefs over the set of possible prior distributions as events are observed. Suppose that the agent's utility is $U(d|\omega)$ if decision d is chosen when event ω occurs. Throughout, it is assumed that $U(1|\omega) \neq U(0|\omega)$, hence for each event ω there is a unique optimal choice given by $\sigma^*(\omega) = \arg\max_{d \in D} U(d|\omega)$.

It is assumed that the decision is sufficiently complex that there is not sufficient time between the observation of ω and the choice d for individuals to consistently choose the optimal strategy. Before facing the event the individual may prepare a response either through an explicit contingent plan or by training with different possible events before facing a decision. For example, airline pilots train with flight simulators to prepare responses for various possible aircraft failures. Though the pilot may be capable of deducing the appropriate response for a particular failure, he or she may not have sufficient time to carry out such an analysis in the face of an actual failure during flight. Indeed, the point of pilot training is to ensure that the appropriate decision is taken quickly with little apparent thought.

The amount of planning is endogenous, and is the set of events for which the individual has a prepared response in period t, denoted by $\Omega^t \subset \Omega$. This implies that if $\omega^t \in \Omega^t$ occurs then the individual is able to respond optimally with $\sigma^*(\omega^t) \in \{0, 1\}$. If $\omega^t \notin \Omega^t$, then the individual does not have a prepared response, nor does she have sufficient time to determine the appropriate response, and hence she randomizes over D.

[5] If the set D consists of n decisions, then simply identify each d to a binary number and hence one can set $D = \{0, 1\}^k$, where k is the smallest interger such that $k \geq \log_2 n$. One then applies the analysis to a vector of decision functions, where each component takes values in $\{0, 1\}$.

For simplicity suppose that when the optimal response is made the reward is u_g, while if a plan is not in place for an event the expected payoff is $u_b < u_g$.[6] The *choice* of an individual at time t is given by the function:

$$\sigma_t\left(\omega^t\right) = \begin{cases} \sigma_\tau^*\left(\omega^t\right), & \text{if } \omega^t \in \Omega^t. \\ \left\{\frac{1}{2}, \frac{1}{2}\right\} & \text{if} \omega^t \notin \Omega^t. \end{cases}, \tag{1}$$

where $\left\{\frac{1}{2}, \frac{1}{2}\right\}$ denotes the lottery that selects each action in $\{0, 1\}$ with equal probability. The function $\sigma_\tau^*\left(\omega^t\right)$ is the optimal response to ω^t at date τ, where τ is the most recent time at which the agent either constructed a plan for ω^t, or had experienced ω^t. The requirement of using the most recent time τ is not necessary for the current section, but plays an important role in the subsequent sections with non-stationary dynamics.

Events are added to the set Ω^t in two ways. First there is *learning by doing,* if ω^t occurs then the agent evaluates her performance *ex post,* and encodes an optimal response to the event ω^t, which is then added to the set Ω^t. The second method is through the explicit formation of a contingent plan. The individual can expend effort before the realization of ω^t to add additional events to the set Ω^t. This is assumed to be a costly activity, either because acquiring the behavior requires expensive training, or simply because of the cost associated with adding a large number of contingent plans. The goal then is to explicitly model the trade-off between learning by doing and the formation of a contingent plan *ex ante.*

It is assumed that the individual knows that the events in Ω are generated by some unknown stationary distribution $\mu \in \Delta^N$. Since there are a finite number of events, it is possible to allow completely non-parametric beliefs, namely any distribution in Δ^N. A convenient prior belief for this situation is given by the Dirichlet distribution, $f(x|\alpha)$, where $\alpha = \{\alpha_1, ..., \alpha_N\}$. This distribution forms a conjugate family, which means that each period the updated beliefs after observing a draw from this multinomial distribution also has a Dirichlet distribution(see DeGroot (1972)). The parameter α_i represents the weight associated with the event ω_i.[7] If event ω_i is observed in period t, then the posterior belief of the individual using Bayes rule is given by a Dirichlet distribution with parameters $\alpha_j^t = \alpha_j^{t-1}$ if $j \neq i$ and $\alpha_i^t = \alpha_i^{t-1} + 1$, where α_i^{t-1} is the period t belief parameter. The expected value of the probability that event ω_i occurs given a Dirichlet distribution with parameter α is $\frac{\alpha_i}{\sum \alpha_j}$.

It is assumed that initial beliefs are given by $\alpha^0 = b\{\lambda_1, ..., \lambda_n\}$, where $\sum \lambda_i = 1$. The parameter b defines

[6] That is $u_b = (U(0|\omega) + U(1|\omega))/2$.

[7] For $\mu \in \Delta^N$ then:

$$f(\mu|\alpha) = \frac{\Gamma(a_1 + ... + \alpha_N)}{\Gamma(a_1)...\Gamma(\alpha_N)} \mu_1^{\alpha_1 - 1} ... \mu_N^{\alpha_N - 1}.$$

If α_i goes to ∞ while the other parameters remain fixed then this distibution approaches the measure that places probabilty 1 on the $i'th$ event.

the degree of uncertainty regarding beliefs, and determines the speed at which beliefs are undated. When b is very large, then the individual is very certain that the true distribution is close to $\{\lambda_1, ..., \lambda_n\} \in \Delta^N$, and will update beliefs away from this distribution very slowly. In contrast when the agent is very uncertain and b is close to zero, then after one period the individual believes that the probability of an event from Ω^t occurring is close to one. The parameter λ_i defines how the agent's initial beliefs vary over the set of possible events. One special case that will be considered is when the decision maker has prior beliefs given by the principle of insufficient reason , namely $\lambda_i = 1/N$, and initial beliefs are given by $\{b/N, ..., b/N\}$(see Jeffreys (1948)).

The cost of a contingent plan is assumed to be an increasing function of the number of states, and hence planning may be spread over several periods. At the beginning of period t, but before observing ω^t, the individual may decide to make contingent plans for an additional n^t events in $\Omega \backslash \left\{\Omega^{t-1} \cup \{\omega^t\}\right\}$, (it is assumed that the event that occurs in period $t-1$ is always added to Ω^t). The cost of this additional contingent planning is $c(n^t)$, where $c(0) = 0$, $c' > 0, c'' \geq 0$. Let the marginal cost of going from $n^t - 1$ to n^t be given by $mc(n^t) = c(n^t) - c(n^t - 1)$. The individual chooses the amount of planning in period t to maximize her discounted expected payoff given her beliefs. The benefit from adding a plan for event ω arises from a net gain of $u_g - u_b$ that accrues the first time ω^t is observed times the probability that this event occurs. This gain is traded off against the cost of including that state in a contingent plan.

Events that are not in Ω^t have never occurred in the past, and hence the expected probability that an event ω_i not in Ω^t occurs in period t is $\frac{b\lambda_i}{b+t}$. The probability that the first time an event occurs is in period $T > t$, given that it has not occurred before, is:

$$\pi(t, T, b, \lambda_i) = \left(1 - \frac{b\lambda_i}{b+t}\right)\left(1 - \frac{b\lambda_i}{b+t+1}\right)...\left(1 - \frac{b\lambda_i}{b+T-1}\right)\frac{b\lambda_i}{b+T},$$

where as a matter of convention $\pi(t, t, b, \lambda_i) = \frac{b\lambda_i}{b+t}$. Due to the Bayesian updating, for any event that has not been observed, the probability that such an event will occur decreases with time. Since this probability does not depend upon the events that have occurred, and since we have assumed that the distribution is an *i.i.d.* process then the benefit from adding a plan for an event ω that has not been observed is *history independent*. The marginal benefit from adding an event ω_i to the set Ω^t that has not been observed in the past is given by:

$$mb(t, b, \lambda_i, \delta) = (u_g - u_b)\sum_{n=0}^{\infty} \pi(t, t+n, b, \lambda_i)\,\delta^n. \tag{2}$$

The first proposition characterizes the properties of the marginal benefit function.

Proposition 1 *For $t \geq 0$, $b > 0$, $\lambda_i, \delta \in (0,1)$, the marginal benefit function $mb(t, b, \lambda_i, \delta)$ is strictly decreasing in t and strictly increasing in b, λ_i and δ. In addition the marginal benefit function has the* *following limit values:*

$$\lim_{b \to \infty} mb\left(t, b, \lambda_i, \delta\right) = \left(u_g - u_b\right) \frac{\lambda_i}{1 - \delta + \lambda_i \delta}$$

$$\lim_{b \to 0} mb\left(t, b, \lambda_i, \delta\right) = \begin{cases} \left(u_g - u_b\right) \lambda_i & \text{if } t = 0, \\ 0 & \text{if } t > 0. \end{cases}$$

The proof of this proposition is in the appendix. The intuition for the result is as follows. The marginal benefit from planning for a particular event ω_i is given by the marginal value of a good decision $(u_g - u_b)$ times the discounted probability of this event occurring. With time, if the event has not occurred, then one believes it is less likely occur and therefore the benefit from planning falls. The parameter λ_i is the *ex ante* probability of this event occurring and hence the benefit from planning increases if the event is thought to be more likely. In the case of b, an increase in b implies that one beliefs are less sensitive to new information, and hence the fact that event ω_i has not occurred has a smaller downward impact on beliefs, and hence the benefit from planning is greater. Increasing δ implies that the value from the event occurring in the future is higher, and hence the gains from having a plan in place are greater.

The first limiting case corresponds to the case that the individual knows for sure that the probability of event ω_i occurring is λ_i. In this case one can immediately see that marginal benefit from planning is always positive, and constant with time, with an upper bound of $(u_g - u_b)$ that is achieved when $\delta = 1$. The case when b is close to zero corresponds to almost complete uncertainty regarding the true distribution. As a consequence after the first period beliefs assign close to probability one to the set Ω^t, the events that have already occurred. Hence, it immediately follows in this case that for any events not included in a contingent plan in the first period, they will never be included in any future plan.

The fact that the marginal benefit falls with t implies that if one should implement a plan at date t, then the benefit of implementing it earlier is higher. When planning costs are linear this implies that the optimal decision has the following simple structure.

Proposition 2 *Suppose that the marginal costs of planning are constant, $mc(n) = c$ for all n, then given b and δ, there is a belief $\lambda^*(b, \delta)$ such that individuals form a contingent plan at date $t = 0$ for all events ω_i such that $\lambda_i \geq \lambda^*(b, \delta)$. For all remaining events and dates all learning occurs via experience rather than planning.*

Proof. Since $mb(t, b, \lambda_i, \delta) \geq mb(t+1, b, \lambda_i, \delta)$, then $mb(t, b, \lambda_i, \delta) - c > \delta\left(mb(t+1, b, \lambda_i, \delta) - c\right)$, and it is never optimal to delay a planning. Define $\lambda^*(b, \delta)$ by $mb(0, b, \lambda^*(b, \delta), \delta) = c$, and since $mb(t, b, \lambda, \delta)$ is strictly increasing in λ it follows that for all $\lambda_i \geq \lambda^*(b, \delta)$ it is optimal to make a contingent plan for all events such that $\lambda_i \geq \lambda^*(b, \delta)$. ■

Thus the individual makes contingent plans for the most likely events, as long as the marginal costs of planning are not too high. In the extreme case when c is sufficiently high the individual's optimal strategy is to make *no* contingent plans, and improve performance only through learning by doing. Conversely, with sufficiently low marginal costs, complete contingent planning is optimal. Notice that for any positive c, one can bound the fraction of states for which an individual will make a contingent plan.

Corollary 3 *The maximum number of states for which an individual would construct a plan is bounded by* $\bar{N} = int\left(\frac{1}{\lambda^*(b,\delta)}\right)$, *and hence for any* $N > \bar{N}$ *planning is necessarily incomplete.*

Therefore, when the number of states is sufficiently large, all other things being equal, planning is necessarily incomplete.

Consider now the implications of assuming the marginal cost of planning is increasing with the number of states. In this case individuals may spread planning over several periods. To simplify the analysis the principle of insufficient reason is used to construct the prior distribution, and hence $\lambda_i = 1/N$, and N is now a parameter of the model. The extension to a non-uniform prior is straightforward, and the optimal procedure with have the same generic structure, with more likely events included in any plans before less likely events.

Proposition 4 *Suppose that cost of planning is strictly convex and increasing in* n, *then if the number of events* N *is sufficiently large then the optimal level of planning,* $n^*(t, b, N)$, *satisfies:*

$$mc(n^* + 1) \geq mb\,(t, b, 1/N) \geq mc(n^*). \tag{3}$$

Moreover, the optimal amount of planning is decreasing with time, and is increasing with one's certainty regarding the true distribution (greater b*). If the number of states is greater than* $(u_g - u_b)/mc\,(1)$ *then when the individual's beliefs are sufficiently uncertain (*b *sufficiently small) the individual makes no contingent plans, and only learning by doing is optimal.*

This result illustrates the basic trade-off between planning and learning by doing. In particular when the event space is sufficiently complex, and beliefs are sufficiently uncertain, then it is not optimal to make any contingent plans. The next proposition considers the optimal strategy in the case that marginal costs of planning are rising, and the state space is sufficiently small that the individual would in finite time have in place a complete contingent plan. However, due to the rising marginal costs of planning in a particular period the individual may spread planning over several periods.

Proposition 5 *Suppose planning costs are increasing and convex in* n, *then there exists an optimal planning rule* $n^0(r_t, t, b, N)$, *where* r_t *is the number of unexplored states given by* $\Omega \backslash \Omega^{t-1}$. *This rule has the following properties:*

1. $n^0(r_t, t, b, N)$ *is increasing in* r_t.
2. *If* $r_t \leq \tilde{n}(t, b, N)$, *then* $n^0(r_t, t) = r_t$.
3. *If* $r_t > \tilde{n}(t, b, N)$, *then* $\min\{n^*(t, b, N), r_t\} \geq n^0(r_t, t, b, N) \geq \tilde{n}(t, b, N)$, *where* $n^*(t, b, N) = \arg\max_{n \geq 0} n \cdot mb(t, b, N) - c(n)$ *is the optimal one period strategy and* $\tilde{n}(t, b, N)$ *satisfies:*

$$mb_t - mc(\tilde{n}) \geq \max\{p_{t+1}\delta(mb_{t+1} - c(1)), 0\} \geq mb_t - mc(\tilde{n} + 1),$$

where $p_t = (N + t/b)^{-1}$ *is the probability that an event that has not been observed occurs in period* t, *and* $mb_t = mb(t, b, 1/N)$.

Hence, an optimal strategy exists regardless of the number of events, and that the amount of planning is increasing with the number of unexplored states. In other words the pattern of planning is the same as in the previous case, except that now the individual may choose to explore less than $n^*(t, b, N)$ events due to the option value of waiting a period to finish making her contingent plan.

2.1 Uncertain Choice

It has been assumed that given the observed event there is a unique optimal choice that does not vary over time. Consider now the situation in which there is a chance that the optimal strategy varies each period. The introduction of uncertainty regarding choice is important for many decisions, for example, when deciding to whether or not to sell an asset one's decision is sensitive to whether or not one expects the price to rise or fall in the next period. The model can be extended to deal with such cases by letting each $\omega = \{\omega_0, \omega_1\}$, with the interpretation that if ω_0 occurs, then choosing action 0 is optimal, while if ω_1 occurs choosing 1 is optimal.

It is assumed that in period t the individual observes the event $\omega^t = \{\omega_0^t, \omega_1^t\} \in \Omega$ before making a decision, but not ω_0^t or ω_1^t. After the decision is made the individual learns which would have been the optimal choice. Further, suppose that the probabilities evolve according to the following rule:

$$\Pr\{\omega^t = \omega_0^t | \omega^0, \omega^1, ..., \omega^{t-1}\} = \begin{cases} a_{\omega_0} & \text{if } \omega_0 \text{ occurred last time event } \omega^t \text{ occurred,} \\ (1 - a_{\omega_1}) & \text{if } \omega_1 \text{ occurred last time event } \omega^t \text{ occurred,} \end{cases} \tag{4}$$

Notice that if $a_{\omega_0} = a_{\omega_1} = 1$ then this is exactly the process studied above in which the optimal strategy does not change over time. When $a_{\omega_0}, a_{\omega_1} > 1/2$ then the process exhibits persistence, that is if ω_0 occurred previously, then it is more likely to occur again when ω^t occurs.

Finally, one needs to consider the gains from planning in the context of this model. Suppose that the parameters $a_{\omega_0}, a_{\omega_1}$ are not known, but in period t if the agent plans for event ω^t, then she will know the optimal choice for that period. Due to the introduction of uncertainty, the gains from planning are reduced, and one has the following proposition.

Proposition 6 *Suppose the events follow the Markov chain given by 4 and the agent believes* $a_{\omega_0}, a_{\omega_1} \in (1/2, 1]$. *Then if* $c > mb(0, b, 1/N)$ *(as defined by 2) learning by doing is an optimal Bayesian decision procedure.*

Proof. If event ω_0 occurs, when ω is observed, then the next time ω occurs it is optimal to choose $d = 0$ because the probability of 0 being optimal is $a_{\omega_0} > 1/2$. However, the marginal benefit from planning is less than $mb(0, b, N)$ because $a_{\omega_0} \leq 1$, and hence the result follows immediately from proposition 2. ■

This proposition captures the basic normative feature of the learning by doing algorithm. In situations for which the optimal choice in the past for a given event remains optimal in the future, then learning by doing is optimal in a complex environment. Moreover, behavior is *adaptive,* that is if in a given period the optimal choice changes, then the individual's behavior also changes to this choice in the future. It highlights one of the desirable features of learning by doing, namely that behavior can evolve to conform to the optimal choice in a non-stationary environment, as long as the environment does not change too quickly, as captured by the assumption $a_i > 1/2$.

3 Behavioral Implications

3.1 Probability Matching

It has been shown that when the environment is sufficiently complex and stable, in the sense of proposition 6, then the learning by doing procedure is optimal. This does not imply that it is optimal for all complex environments. In fact, it is well known that when planning is costly, a globally optimal procedure does not exist. In brief, one can use resources to determine an optimal rule, however the determination of how much resources to allocate to such a process is itself a costly optimization exercise, and so one faces an infinite regress of optimization exercises.[8]

[8] See Day and Pingle (1991) and Lipman (1991) on this issue.

However, if the learning by doing algorithm is generically a good procedure that is widely adopted by individuals, then this will have implications for observed behavior if it is used in environments for which it is in fact not optimal. An example of such an environment is the following two armed bandit problem. Each period an individual must select which of two lights, L or R, will turn on, and receives a fixed reward R if she guesses correctly.

If the decision maker believes that this is in fact a bandit problem for which the probability of L is fixed over time, then the optimal strategy is to choose the arm which she believes is most likely to occur, and then to update beliefs over time. In the context of the model of section 2.2 this corresponds to setting $a_L =$ probability of L. With these parameters $\Pr(L|R) = \Pr(L|L) = a_L$, while $\Pr(R|R) = \Pr(R|L) = 1 - a_R$. If the decision maker does not know a_L, then she can use the updating procedure for multinomial random variables described above to form beliefs a_L^t each period, and choose L if and only if $a_L^t \geq 1/2$. This would result in a behavior in which the decision maker stays with one arm until there are sufficient draws on the other arm to cause her to change her choice. Over time, as beliefs settle down, the individuals will choose one arm only, regardless of the pattern of lights.

What is in fact observed is the well known phenomena of "probability matching" (see Estes (1976) for a literature review), that was widely studied in the 1960's by psychologists, and considered evidence that people do not make rational decisions, a point that has been much emphasized by Richard Herrnstein (1997). This behavior is described by individuals choosing L with approximately the probability that L occurs, rather than sticking to a single choice. From the point of view of Bayesian decision making, such behavior is only irrational if individuals know the true underlying model, which in general was not the case. What is interesting is that the learning by doing algorithm (which is optimal when the process generating the lights is the appropriate Markov chain), predicts precisely this behavior. This is because under the learning by doing algorithm the individual selects the side that was optimal the previous period. Hence the frequency with which L is chosen is exactly equal to its frequency of occurrence.

This result illustrates the difficulty of testing the "rationality" of behavior using a fixed environment. In particular, the fact that a general "optimization algorithm" does not exist, implies that for *any* decision making procedure it is possible to find problems for which the observed choices are "irrational". Hence, individuals should only be judged irrational if they are performing poorly for the decision problems that they are facing on average in practice. If the learning by doing algorithm is on average a good decision procedure, then the fact that individuals engage in probability matching is not necessarily evidence that they are irrational or on average are making poor decisions.

3.2 Similarity Judgements and Optimality in the Long Run

The learning by doing algorithm supposes that individuals improve performance by acquiring experience with events that occur frequently. This raises the issue that in practice an individual is unlikely to experience *exactly* the same event again in the future. The purpose of this section is to illustrate that the model is robust to allowing individuals to use a rule that associates the same action to events that are "similar".

There is a voluminous literature in psychology that models individual choice as resulting from the exercise of similarity judgements, which, as Tversky and Kahneman (1981) have shown, implies that in many situation individuals make decisions that are inconsistent with rational choice theory.[9] For example, Tversky and Kahneman (1981) show that a subject's response to the question of whether a vaccine should be given to a population depends upon whether or not the risks are presented in terms of mortality rates or survival rates. These, and many other similar results, illustrate that when responding to a question the typical individual does not in fact explore all the implications of the data before making a decision. Hence in the context of Newell (1990)'s decision typography, these are choices made in the "cognitive" time frame.

In economics there is a literature, beginning with Luce (1956), that takes similarity judgements as given, and then asks how one may formally model such behavior.[10] The question that is not explored in economics is the converse, namely to what extent do similarity judgements lead to rational choice. This question is natural in the context of the learning by doing algorithm when it is extended to deal with more complex event spaces. Suppose that the set of possible events, Ω, is now a convex subset of $\Re^d$, and that $\{\omega^t\}_{t=1}^{\infty}$ is an *i.i.d.* process represented by a measure μ on $\Re^d$ that is absolutely continuous with respect to Lebesgue measure (in other words the probability that $\omega^t = \omega^{t'}$ for $t \neq t'$ is zero). Let the space of choices be given by $Z = \{0,1\}^n$, for some $n \geq 1$, and that the individual's utility function, $U(z|\omega^t)$, bounded and Borel measurable on $Z \times \Omega$. This ensures the existence of a Borel measurable optimal choice rule $\sigma^*(\omega^t) \in \arg\max_{z \in Z} U(z|\omega^t)$, but in general the optimal choice may not be continuous function of ω^t.

Under these assumptions it is the case that the probability of an event occurring again is zero ($\Pr\left\{\omega^t = \omega^{t'}\right\} = 0$ for all $t \neq t'$). Therefore, under the learning by doing algorithm as stated in 1, performance can never improve with experience. Thus, a *necessary* condition for learning based upon experience is that the individual must *extrapolate* from past experience to decide on how to behave today. An example of such a decision procedure is the nearest neighbor rule. When the individual observes ω^t, she then finds the event that occurred

[9] See Churchland and Sejnowski (1993) for a review of the cognitive science literature.

[10] Rubinstein (1988) and Leland (1998) are papers that build upon the work of Luce (1956) that introduce concepts of similarity into preferences. Gilboa and Schmeidler (1995) present a model in which such judgements are built up from past experiences or cases, while Sonsino (1997), Jehiel (2001) and Samuelson (2000) explore the implications of such judgements for games.

in the past that is closest to ω^t, say $\omega^{t'}$, and then chooses the optimal decision that is associated with the event $\omega^{t'}$, given by $\sigma^*\left(\omega^{t'}\right)$, which was previously encoded in the individual's memory. As in the previous section, it is assumed that after a decision is made the individual has sufficient time to contemplate her response, and to associate with the experienced event ω^t the optimal response $\sigma^*(\omega^t)$. Hence, the decision z^t at time t is the optimal response to the most similar previous event $\omega^{t'}$, $z^t = \sigma^*\left(\omega^{t'}\right)$.

More formally, given a history of events and corresponding optimal responses at time t :

$$H^t = \left\{\left(\omega^1, z^1\right), \left(\omega^2, z^2\right), \ldots, \left(\omega^{t-1}, z^{t-1}\right)\right\},$$

and given a new event ω^t, let $\Omega(\omega^t, k, H^t)$ be the k closest events, where closeness is measured by the Euclidean distance$\|\omega' - \omega\|$. Then a generalization of the nearest neighbor rule is given by the *$k-$nearest neighbor rule.*

Definition 7 *The $k-$*nearest neighbor rule *is defined for odd k by:*

$$\sigma_i\left(\omega^t|H^t\right) = \left\{\begin{array}{l} 1, \ \textit{if} \ \sum_{\omega^\tau \in \Omega(\omega^t, k, H^t)} \sigma_i^*(\omega^\tau)/k > 1/2, \\ 0, \ \textit{otherwise.} \end{array}\right\}. \tag{5}$$

This rule requires the agent to use the average best response for the k events most similar (in terms of Euclidean distance) to event ω^t.Under the optimal rule the expected utility each period of the individual would be $U^* = E\{U(\sigma^*(\omega)|\omega)\} = \int_{\omega\in\Omega} U(\sigma^*(\omega)|\omega)\,d\mu$, while the expected utility in period t would be a function of the history and given by $U(H^t) = E\{U(\sigma(\omega|H^t)|\omega)|H^t\}$. The performance of the behavioral rule can be evaluated by looking at performance for a typical history, which is found by taking the expected value over all possible histories, and involves extending μ to a measure on Ω^∞ in a natural way using the rule $\sigma(\omega^t|H^t)$. Letting $U^t = E\{U(H^t)\}$, one has the following proposition (proved in the appendix).

Proposition 8 *If the optimal decision rule $\sigma^*(\omega^t)$ is Borel measurable, then under the $k-$nearest neighbor rule for any odd $k = 1, 3, 4...$ with a Euclidean similarity measure expected performance approaches the optimum:*

$$\lim_{t\to\infty} U^t = U^*$$

This somewhat surprising result demonstrates the power of using even a very crude similarity measure in ensuring that learning by doing behavior converges to the optimum. It is based upon some rather deep results in mathematics that demonstrates that measurable functions can be approximated by continuous functions. Also the convergence result does not depend upon the individual's prior beliefs, and hence it illustrates the robustness of the learning by doing procedure.

Given that the starting point of the model is the idea that individuals do not have time to determine an optimal response to ω^t, then one may wonder if the similarity computation itself may also be too difficult to carry out in the time allowed. As it happens, such procedures are very fast, and as Churchland and Sejnowski (1993) observe, the human brain is optimized to carry out such procedures very quickly.

A remarkable feature of the learning by doing procedure augmented with a similarity judgement is that no knowledge of the function $\sigma^* (\omega^t)$ is required. The procedure works with a wide variety of similarity judgements, and illustrates that the learning by doing algorithm is a robust, non-parametric procedure that results in behavior that converges utility maximization in the long run. In particular, the model is consistent with the observation that in certain domains, where the individual has a great deal of experience, performance may be close to optimal, while in domains where experience is limited then behavior may be inefficient.

The fact that the model predicts that behavior converges to the optimum in the long run is not inconsistent with the recent work in behavioral economics incorporating elements of irrational behavior into economic models. This is because the model makes no prediction regarding behavior in the short run, the case considered by behavioral models. Moreover, one can choose rules $\sigma^* (\omega^t)$ in such a way to make convergence arbitrarily slow. Hence there are situations where the utility maximization model would perform very poorly, and observed behavior may be better predicted using models based upon decision making shortcuts. (see for example the collection of papers in Kahneman, Slovic, and Tversky (1982) and Payne, Bettman, and Johnson (1993)).

4 Learning Dynamics

The purpose of this section is to derive the learning curve implied by this model, and to compare it to a standard one parameter Bayesian learning model due to Jovanovic and Nyarko (1995) using an interesting data set from Weber and Camerer (2001). Only the simplest possible variant of the model is considered, and it is assumed that the parameters b and $c(\cdot)$ are such that planning is never optimal, and hence the individual engages only in learning by doing.[11] Suppose that there are N possible events and the true probability of N occurring is $1/N$. This is consistent with some planning occurring as long as all planning occurs in the first period, and for all the remaining events the individual uses the principle of insufficient reason to assign unobserved events equal probabilities. At time t let $m^t \in \{0, 1, .., N\}$ be the number of events that have been experienced, which, given the equal probability hypothesis, is a sufficient statistic for the state of an

[11] Alternatively, one may suppose the current model is the reduced form after all planning has occured.

individual's knowledge. In period t if an event from these m^t events occurs then the individual earns u_g, otherwise the expected payoff is u_b. Hence the expected payoff in period t is:

$$V(m_t) = u_g \cdot m^t/N + u_b \left(1 - m^t/N\right).$$

Whenever a new event is experienced this is added to the repertoire Ω^t, and m^t is increased by 1. The evolution of the state over time is a Markov chain with transition probability function:

$$P\left[m^{t+1}|m^t\right] = \begin{cases} m^t/N, & \text{if } m^{t+1} = m^t, \\ 1 - m^t/N, & \text{if } m^{t+1} = m^t + 1, \\ 0, & \text{in all other cases.} \end{cases} \tag{6}$$

In the data one observes neither N nor the number of events experienced. Moreover, the data reports the average performance for a number of individuals, rather than actual performance after t trials. Thus the expected payoff is used as the performance measure in the model, x_m^t is the probability that an individual has experienced m events in period t, and $x^t = (x_0^t, x_1^t, ..., x_N^t)$ is the vector of probabilities. The initial state is assumed to be $x_0 = (1, 0, .., 0)$, which assigns probability 1 to having no experience. Let P denote the Markov transition matrix, where P_{ij} is the probability of going from state i to state j, then the probability distribution of experience at date t is given by:

$$x^t = x^0 P^t,$$

where P^t is P to the power t. Accordingly, the expected payoff and variance of performance is given by:

$$\begin{aligned} \hat{V}^t &= x^t \hat{U} = x^0 P^t \hat{U}, \\ \hat{S}^t &= \sum_{m=0}^{N} x_m^t \left(V(m) - \hat{V}^t\right)^2, \end{aligned}$$

where $\hat{U} = \begin{bmatrix} V(0) \\ \vdots \\ V(N) \end{bmatrix}$.

The reduced form learning curve can be written as:

$$V_{LD}^t (u_b, \delta, N) = u_b + d \cdot X_t (N)$$

where

$$X_t(N) = x_0 P^t \begin{bmatrix} 0 \\ 1/N \\ 2/N \\ \vdots \\ (N-1)/N \\ 1 \end{bmatrix},$$

and u_b is the lowest possible performance, and d is the difference, $u_g - u_b$, between the lowest and highest performance level. This formula describes a learning curve with some of the basic features of observed learning curves. In particular it exhibits the "power law of learning", namely learning initially proceeds quickly and then slows down (Snoddy (1926)). The speed of this effect is determined by the complexity parameter N, with u_b determining the starting point, and d determining the maximum gain in performance that is possible.

In this model the matrix P is upper diagonal, thus the eigenvalues of the matrix are the diagonal elements and are given by $\lambda_i = i/N$, $i = \{0, 1, .., N\}$. And hence we may write the formula for mean performance in terms of powers of the eigenvalues, or:[12]

$$\hat{V}^t = \sum_{i=0}^{N} z_i e^{t \ln \lambda_i}. \tag{7}$$

This is precisely the functional form that Newell and Rosenbloom (1981) suggests provides the best fit to data they explore. This implies that *a priori* this model will provide a good fit to a wide variety of learning curves.

For purposes of comparison consider the following learning curve due to Jovanovic and Nyarko (1995), based upon a simple Bayesian decision problem. In their model it is assumed that the decision maker each period chooses $z_t \in \Re$ to yield performance:

$$q_t = A\left[1 - (y_t - z_t)^2\right].$$

The parameter y_t is not known at the time z_t is chosen, and is assumed to be normally distributed, satisfying $y_t = \theta + w_t$, where θ is a time invariant constant and w is a $N\left(0, \sigma_w^2\right)$ random variable, assumed to be serially uncorrelated with time. Though the variance of w_t is known, the agent does not know θ, but is assumed to have a normally distributed unbiased estimate $\hat{\theta}$, with variance σ_θ^2.

[12] More precisely we can rewrite $P = VLM$, where L is a diagonal matrix with the eigenvalues along the diagonal. Then $z_i = a_i b_i$, where a_i and b_i are the i'th coordinate of the row vector $x^0 V$ and the column vector $M\hat{U}$ respectively.

At the end of each period the individual observers the realization of y_t, and the optimal strategy each period is to set $z_t = E\left(y_t|\hat{\theta}, y_0, y_1, ..., y_{t-1}\right)$. Thus the *expected* performance at date t is given by:

$$V_{PL}^t\left(A, \sigma_w^2, \sigma_\theta^2\right) = A\left[1 - \sigma_t^2 - \sigma_w^2\right],$$

where σ_t^2 is the posterior variance of θ given by:[13]

$$\sigma_t^2 = \frac{\sigma_w^2 \sigma_\theta^2}{\sigma_w^2 + t \cdot \sigma_\theta^2}.$$

This model generates a three parameter learning curve, where A controls the range of learning (between 0 and a maximum of $A\left[1 - \sigma_w^2\right]$), while σ_θ^2 and σ_w^2 jointly determine the speed of learning. Throughout, let index LD refer to the learning by doing model, while PL refers to this Parametric Learning model.

4.1 Learning a Corporate Culture

Though the LD and PL models highlight different aspects of the learning process, they both share the feature of having the same basic properties of a learning curve - performance increases quickly at the beginning, and then slows down with experience. Hence, to empirically distinguish between the models one must depend upon differences in their shapes. This section presents the results from estimating the two learning models using data from an experiment performed by Weber and Camerer (2001). They report results from an experimental study of the effect of mergers on team performance that seem to be appropriate for the learning by doing model introduced in this paper.

In their experiment individuals are divided into two group (firms), in which one person (the manager) is exposed to a sequence of pictures (the events), who then has to communicate this information to the second person (the worker), who executes a task as a function of the picture shown. After several periods of learning, firms were merged by having one manager fired, with the remaining manager required to transmit the information to the remaining two workers.

The loss of performance when merger occurs has two interpretations, depending upon the learning model used. The PL model is fitted to the data by supposing the difference between z_t and y_t represents the average error in responses in period t. When a merger occurs, this error is assumed to increase on average, with performance steadily increasing at the same rate. Thus there are 4 parameters in the model, A, variance of the noise parameter, σ_w^2, the prior variance of θ at the beginning of the experiment and just after the merger, given by $\sigma_{\theta_1}^2$ and $\sigma_{\theta_2}^2$ respectively. The learning by doing model supposes that individuals have coded the optimal response for a fraction of the possible states. It is assumed that after merger the number of such

[13] See equations 7 and 8 of Jovanovic and Nyarko (1995).

states decreases, and then increases with the new manager. This decrease is estimated by q, the probability that the memory for an event is lost, and hence the learning by doing model has 4 parameters, the initial payoff, u_b, the maximum increase in performance possible, d, the number of events, N and the probability of a forgetting an event after the merger event, q.

Following Jovanovic and Nyarko (1995), the parameters are estimated by minimizing the sum of the squared errors:

$$\min_{\beta} \sum_{t=1}^{T} \left(v_t - V_M^t(\beta)\right)^2,$$

where v_t is observed performance in period t, β is the set of parameters, and $M \in \{LD, PL\}$ denotes the model. Let $\hat{\beta}_M$ be the predicted parameter values for model M, $v = [v_1, ..., v_T]^T$ the vector of data, $\hat{v}_M = \left[V_M^1\left(\hat{\beta}_M\right), ..., V_M^T\left(\hat{\beta}_M\right)\right]^T$ the vector of predicted values for model M, and $\hat{J}_M = \left[\nabla V_M^1\left(\hat{\beta}_M\right), ..., \nabla V_M^T\left(\hat{\beta}_M\right)\right]^T$ the Jacobian matrix at the estimated parameter values. There is no Jacobian for N since it is a discrete variable, and hence is found by systematically searching over the possible values of N, with (u_b, d, q) estimated at each step to minimize squared error.

In terms of testing we compare the learning by doing model against the parametric learning model using a non-nested hypothesis test, called the P test, due to Davidson and MacKinnon (1981). The purpose of this test is to see if the alternative hypothesis is to be preferred, and proceeds as follows. Suppose that the null hypothesis is that LD is correct, then one can write the data as a convex combination of the estimated values from the two models:

$$v = (1-\alpha)\,\hat{v}_{LD} + \alpha\hat{v}_{PL} + residuals.$$

Where v, $\hat{v}_{LD}$ and $\hat{v}_{PL}$ are vectors with actual performance and estimated performance for each model, each period. The parameter α is estimated using OLS (ordinary least squares), and determines the optimal combination of estimates from LD and PL that fit the data. Davidson and MacKinnon (1981) show that if LD correct, then the estimated α should be zero, using the standard errors from the OLS estimate. If not then LD is rejected. The regression that is run to compute the test is:

$$v - \hat{v}_{LD} = \hat{J}_{LD} b + \alpha\left(\hat{v}_{PL} - \hat{v}_{LD}\right) + residuals,$$

where b and α are both estimated using OLS. This is called a Gauss-Newton artificial regression. The reason that it is artificial is that if $\hat{\beta}_M$ is estimated precisely then $(v - \hat{v}_{LD})\,\hat{J}_{LD} = (0,0)$, and hence the estimated values of b should be zero. If not, then the optimization routine needs to have a tighter convergence requirement. Once the regression has been run, one can use the estimated $t-statistics$ for α to test the hypothesis $\alpha = 0$. It should be emphasized that rejecting LD, does not imply accepting PL. One has to run a separate regression for model PL. If both models are rejected when tested against each other then a linear combination of the models provides a better fit, and neither model fits the data better at the exclusion of the other.

The data consists of two series, one with 25 rounds, with merger occurring at round 15, while in the other series there are 30 rounds, with merger occurring in round 20. The results from estimating these models are reported in table 1. These are graphed in figures 1 and 2. Both learning curves do a good job of fitting the data, though the learning by doing model provides a better fit.

Parametric Learning Model

Data Set	**A**	$\sigma^2_{\theta_1}$	$\sigma^2_{\theta_2}$	σ^2_w	R^2
Short	899.944	0.753	0.343	0.091	0.956
Long	856.812	0.735	0.439	0.140	0.910

Learning by Doing Model

	u_b	d	N	q	R^2
Short	-22.241	213.853	5.000	0.241	0.961
Long	-26.477	237.753	6.000	0.423	0.942

Table 1: Estimates for Corporate Culture Data

Learning by Doing Model		**Parametric Learning Model**	
a	t-statistic	a	t-statistic
.070	0.33	.963	3.42

Table 2: P-Test Results for Corporate Culture Data

This is particularly evident when the P-test is performed. As one can see in table 2, the LD model is accepted when tested against the PL model, and the PL model is rejected when tested against the LD model. The results highlight the fact that learning in this case is well modelled by a procedure in which individuals learn with experience the optimal responses to a large number of discrete events. It also illustrates a general problem with inference with learning curves, namely they all fit the data rather well (an R^2 for the PLM of greater than .9 is usually taken as evidence of a good fit). This would have also been the case that if one used a standard log-linear model, as is typical in the literature (see Alchian (1963)). This model was also tests, and found that though it provides a good fit, both the PLM and the LD models soundly reject the log-linear specification.

These results in and of themselves do not prove that this learning by doing model is correct. However, it does illustrate that one can construct an empirically testable model based upon the costs of contingent planning alone, something that the current literature would lead one to believe is impossible.

5 Discussion

The optimality of learning by doing is analogous to Simon (1955)'s result for satisficing. He observes that individuals involved in costly search seem to use the procedure of satisficing, that is they continue to search until the payoff reaches some threshold, called their reservation value, at which point they stop search. As Simon (1955) demonstrates in the appendix to his paper, if the cutoff at which search stops is chosen appropriately, then the satisficing procedure is an optimal search algorithm.[14]

Moreover, viewing the satisficing procedure as an optimal decision generates some additional predictions regarding how observable characteristics of the decision problem affect an individual's reservation value. Even though individuals may use crude methods for determining their reservation values, if the procedure is understood as an approximation to an optimal procedure, then the comparative static predictions from the optimizing model are likely to remain valid for satisficing individuals.[15]

Similarly, there is a long tradition in economics, beginning with Cournot (1974), that explores models based upon the observation that individuals are adaptive, and learn from experience. Rational expectations models, and more generally game theoretic models, have made the important point that individuals are also strategic decision makers. Namely they build simple models of the world, including models of other decision makers, that they use to guide their own decision making. However, in focussing upon the role of beliefs, these models create new puzzles, namely that in many situations individuals simply do not act with as much foresight as the theory predicts.

For example, Weisbuch, Kirman, and Herreiner (2000) demonstrate that one can explain much of the behavior in the Marseille fish market using simple adaptive model in which individuals adjust their behavior

[14] See Lippman and McCall (1981) for a review of the optimal search literature. See also Sobel (2000) for a discussion of learning for a class of bandit problems, and their implications for social learning.

[15] See Devine and Kiefer (1991) for a review of the empirical evidence in labor economics and Camerer (1995) for a review of the results from the experimental literature. MacLeod and Pingle (2003) provide some additional experimental evidence regarding the factors that affect an individuals reservation value.

in response to trading experience. Madrian and Shea (2002) provide strong evidence that individuals do not systematically search over possible savings plans when they are hired at a new job. Madrian and Shea (2002) explicitly argue that their results suggest that non-economic factors need to be considered to understand these savings behaviors.

The results of this paper suggest that this may not necessarily be the case, and in fact the focus upon decision costs may provide insights into how one should design enrollment into retirement savings plans. In everyday life individuals face a number of competing uses for their time, and hence they rationally attend to only those events that appear important or pressing. The results of this paper suggest that the algorithm of modifying behavior for a contingency only when that contingency occurs can be optimal.

The problem for retirement savings is obvious. The consequence of not saving enough will not be salient for an individual for many years, and hence for this particular problem learning through experience does not result in optimal behavior, nor will behavior converge to the optimum. If one can repeat one's life over and over again, like Phil in the film *Groundhog Day*, individuals would eventually learn to save optimally. Hence, for those events with consequences far into the future it may be socially optimal to require people to pay attention to the decision, or possibly put into place a default this is known to be an average optimal for the target population.

5.1 On Parsimony

The model and results in this paper can be (and have been) criticized because of their simplicity, or what one might more generously call their parsimony. There are two kinds of criticism to consider. The first is that human behavior is much more complex than is captured by the model. More damming still is the observation that there is a large amount of research on human decision making that is able to explain a wide variety of behaviors not captured with this model. The second criticism is that complexity costs are not the whole story, but that actual problems deal with a number of different sources of complexity, such a computational complexity, measurement error, and strategic interactions.

In terms of the complexity of human behavior, one of the striking conclusions that comes from the recent research is the large amount of variation in observed behavior. For example, Payne, Bettman, and Johnson (1993) and Gigerenzer and Goldstein (1996) provide reviews of the psychological research on human decision making and show that there is no single model of behavior with the power and simplicity equal to the simple utility maximization model. Payne, Bettman, and Johnson (1993) show that decision costs are one of the most important inputs into understanding observed behavior. Since their goal is to explain decision making of individuals in a very specific environments, their models must include a wide variety of decision making strategies.

Hence, the question for these researchers is how can we best understand human behavior in a particular context. By identifying which algorithms an individual is using, one can construct a good predictive model of *this* individual's behavior. This research has found that the mix of algorithms used in day to day decision making varies from person to person. In contrast, economists are concerned with policies that affect millions of individuals, and hence the goal is to understand average behavior of a population, and to construct models with implications that are broadly applicable.

This is precisely the point of the recent research by Erev and Roth (1998) in the context of explaining behavior in games. Stahl (1996) and Camerer and Ho (1999) have shown that for a given game, a detailed model of behavior and belief formation can explain much of the variation in the data, and can certainly do this better than the simple adaptive learning model of Erev and Roth (1998). However, what is surprising about the results of Erev and Roth (1998) is that a single adaptive learning model provides a good fit of the data across a variety of games. For the economist this result is important because in most cases one simply cannot observed the detailed structure of the decision problem a person is facing, but would never the less like to be able to make some predictions regarding how behavior adjusts to observable features of the economic environment.

Like the optimal search foundations for the model of satisficing behavior, the model of this paper provides some additional insights into the conditions under which learning by doing is an optimal algorithm. Environments for which the individual has little experience and are complex in the sense that it is not possible to evaluate all the alternatives, then performance is expected is increase with experience at a rate that depends upon the complexity of the problem.

The extension of the model to similarity measures also makes the point that when an individual faces a new problem, then she is likely to use past experience to guide her choice, even when this might not be optimal. This is a rather strong causal implication of the model. A standard decision or game theoretic approach implies that an individual's behavior is explained by the optimal response to the current decision problem. In contrast, the learning by doing model predicts that for new problems, behavior is explained by observed *past* behavior for similar problems, but with experience behavior will converge to the optimal response for the current problem.

This implication is consistent with the results on status quo effects, as discussed in Samuelson and Zeckhauser (1988) and Madrian and Shea (2002). It also distinguishes this model from standard Bayesian learning models such as Jovanovic and Nyarko (1995). In those models, both the behavior and speed of learning is explained by features of the problem at hand, and would be independent of unrelated previous experience. In particular, status quo effects are not a predicted outcome of standard Bayesian learning models.

These results follow from a very simple premise, namely the existence of planning costs. Given the ubiquitous nature of planning costs the model predicts that learning by doing is in many cases an optimal strategy, and, given the convergence result from section 3.2, it is never a poor long run strategy. This may help explain why learning by doing is a good first order model of behavior that has proven to be useful in a wide variety of contexts.

5.2 Implications for Organization and Contract Theory

The results also have some implications for the theory of organizations and contracts that may be worth developing in future research. The planning cost for each contingency can be interpreted as the cost of acquiring information from a competitive market. If the number of possible contingencies is very large, then an individual may choose not to buy information, but rather acquire the skill via learning by doing. Since individual learning experiences are likely to be different, this would imply that equally able individuals would have different observed performances, even in a competitive market.

Rosen (1972)'s paper on learning in firms assumes that knowledge cannot be acquired on the market, but is gathered through experience with the production process. The results here provide a formal justification for that hypothesis. Moreover, the analysis of Rosen (1972) suggests that knowledge acquired in this manner can be viewed as an asset. Since it is not economic to buy the knowledge directly, one can acquire the expertise only by buying the firm that embeds the knowledge of the production process.

Another example is a form contract. As Dye (1985) has shown, planning costs imply that optimal contracts are incomplete. With experience different unexpected contingencies occur that find themselves included in future contracts, and hence over time contract become more complete. An example of this are the American Institute of Architects form contracts. They began drafting contracts in 1888, and have been publishing form contracts for use in the construction industry since 1911. These contracts have got longer and more detailed over time in response to learning about contingencies that give rise to disputes (see Sweet (1999)). If complete contracts were possible and inexpensive, individuals would not need to depend upon these form contracts.[16]

In these examples knowledge accumulation is continuous, and would be complete only if the environment were very simple, or if the same situation were to occur repeatedly. In practice, the environment is constantly

[16] See Korobkin (2003) for a discussion of the complexity of form contracts and the legal implications of these complexity costs.

changing, and hence learning by doing is likely to be an ongoing process. Life is unfortunately not like the movies. The character Phil in the movie *Groundhog Day* gets the chance to keep learning by playing out the same day until finally he gets it right and finds a meaningful relationship and happiness. The rest of us must simply bumble along, and may only achieve such understanding in the very long run, if ever!

References

Alchian, A. (October, 1963). Reliability of progress curves in airframe production. *Econometrica 31*(4), 679 – 693.

Camerer, C. (1995). Individual decision making. In J. H. Kagel and A. H. Roth (Eds.), *The Handbook of Experimental Economics*, pp. 587–683. Princeton University Press: Princeton University Press.

Camerer, C. and T.-H. Ho (July, 1999). Experience-weighted attraction learning in normal form games. *Econometrica 67*(4), 827–74.

Churchland, P. S. and T. J. Sejnowski (1993). *The Computational Brain*. Cambridge, MA: MIT Press.

Conlisk, J. (June, 1996). Why bounded rationality? *Journal of Economic Literature 34*(2), 669–700.

Cournot, A. A. (1974). *Recherches sur les principes mathâematiques de la thâeorie des richesses*. [Paris]: Calmann-Lâevy. Prâef. de Henri Guitton. 19 cm.

Davidson, R. and J. G. MacKinnon (1981). Several tests for model specification in the presence of alternative hypotheses. *Econometrica 49*, 781–93.

Day, R. H. and M. A. Pingle (1991). Economizing economizing. In R. Frantz, H. Singh, and J. Gerber (Eds.), *Behavioral Decision Making: Handbook of Behavioral Economics*, pp. 509–22. Greenwich CT: JAI Press.

DeGroot, M. H. (1972). *Optimal Statistical Decisions*. New York, NY: McGraw-Hill Book C.

Devine, T. J. and N. M. Kiefer (1991). *Empirical Labor Economics*. Oxford, U.K.: Oxford University Press.

Devroye, L., L. Györfi, and G. Lugosi (1996). *A Probabilistic Theory of Pattern Recognition*. New York, NY: Springer-Verlag.

Dye, R. A. (February, 1985). Costly contract contingencies. *International Economic Review 26*(1), 233–250.

Erev, I. and A. E. Roth (September, 1998). Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *American Economic Review 88*(4), 848–81.

Estes, W. K. (1976). The cognitive side of probability learning. *Pschological Review 83*, 37–64.

Gigerenzer, G. and D. G. Goldstein (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review 103*(4), 650–669.

Gilboa, I. and D. Schmeidler (August, 1995). Case-based decision theory. *Quarterly Journal of Economics 110*(3), 605–640.

Hart, O. D. and B. Holmström (1987). The theory of contracts. In T. Bewley (Ed.), *Advances in Economic Theory: Fifth World Congress.*, pp. 71–155. Cambridge, U.K.: Cambridge University Press.

Herrnstein, R. J. (1997). *The Matching Law*. Cambridge, MA: Harvard University Press.

Jeffreys, H. (1948). *Theory of Probability (Second Edition)*. Oxford, UK: Clarendon Press.

Jehiel, P. (March, 2001). Analogy-based expectation equilibria. Technical report.

Jovanovic, B. and Y. Nyarko (1995). A bayesian learning model fitted to a variety of empirical learning curves. In *Brookings Papers on Economic Activity*, pp. 247–305.

Kahneman, D., P. Slovic, and A. Tversky (Eds.) (1982). *Judgement under Uncertainty: Heuristics and Biases*. Cambridge, UK: Cambridge University Press.

Korobkin, R. (2003). Bounded rationality and unconscionability: A behavioral approach to policing form contracts. *University of Chicago Law Review*. forthcoming.

Leland, J. W. (May, 1998). Similarity judgements in choice under uncertainty. *Management Science 44*(5), 659–672.

Lipman, B. L. (July, 1991). How to decide how to decide how to...: Modeling limited rationality. *Econometrica 59*(4), 1105–25.

Lippman, S. A. and J. J. McCall (1981). The economics of uncertainty: Selected topics and probabilistic methods. In K. J. Arrow and M. D. Intriligator (Eds.), *Handbook of Mathematical Economics*, Volume 1 of *Handbooks in Economics*, pp. 211–284. Amsterdam: North-Holland Publishing Company.

Luce, D. R. (April, 1956). Semiorders and a theory of utility discrimination. *Econometrica 24*(2), 178–191.

MacLeod, W. B. and M. Pingle (2003, May). Aspiration uncertainty: Its impact on decision performance and process. Technical report, University of Southern California.

Madrian, B. C. and D. F. Shea (2002). The power of suggestion: Inertia in 401(k) participation and savings behavior: Erratum. *Quarterly Journal of Economics 117*(1), 377.

Newell, A. (1990). *Unified Theories of Cognition*. Cambridge, Mass.: Harvard University Press.

Newell, A. and P. S. Rosenbloom (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive Skills and Their Acquisition*, pp. 1–55. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Payne, J. W., J. R. Bettman, and E. J. Johnson (1993). *The Adaptive Decision Maker*. Cambridge, U.K.: Cambridge University Press.

Rosen, S. (August, 1972). Learning by experience as a joint production. *Quarterly Journal of Economics 86*, 366 – 82.

Rothschild, M. (1974). Searching for the lowest price when the distribution of prices is unknown. *Journal of Political Economy 82*, 689–711.

Rubinstein, A. (1988). Similarity and Decision-Making under risk (is there a utility theory resoltion to the allais paradox?). *|jet| 46*, 145–153.

Samuelson, L. (September, 2000). Analogies, adapatation and anomalies. Technical report, University of Wisconsin.

Samuelson, W. and R. Zeckhauser (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty 1*, 7–59.

Savage, L. J. (1972). *The Foundations of Statistics*. New York, N.Y.: Dover Publications.

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics 69*, 99–118.

Simon, H. A. (1978). Rationality as a process and as a product of thought. *American Economic Review 68*(2), 1–16.

Snoddy, G. S. (1926). Learning and stability. *Journal of Applied Psychology 10*(10), 1–36.

Sobel, J. (2000). Economists' models of learning. *Journal of Economic Theory 94*, 241–261.

Sonsino, D. (1997). Learning to learn, pattern recognition, and nash equilibrium. *Games and Economic Behavior 18*, 286–331.

Stahl, D. (1996). Boundedly rational rule learning in a guessing game. *Games and Economic Behavior 16*, 303–330.

Sweet, J. (1999). *Legal Aspects of Architecture, Engineering, and the Construction Process.* West Publishing Company.

Tirole, J. (July, 1999). Incomplete contracts: Where do we stand? *Econometrica 67*(4), 741–782.

Tversky, A. and D. Kahneman (January, 1981). The framing of decisions and the psychology of choice. *Science 211*(4481), 453–458.

Weber, R. A. and C. F. Camerer (2001). Cultural conflict and merger failure: An experimental approach. Technical report.

Weisbuch, G., A. Kirman, and D. Herreiner (2000, April). Market organisation and trading relationships. *Economic Journal 110*(463), 411–36.

A Appendix

A.1 Proofs of Propositions

Proposition For $t \geq 0$, $b > 0$, $\lambda_i, \delta \in (0,1)$ the marginal benefit function $mb\,(t, b, \lambda_i, \delta)$ strictly decreasing in t and strictly increasing in b, λ and δ.

Proof. Without loss of generality set $u_g - u_b = 1$. The existence and continuity of $mb\,(t, b, \lambda_i, \delta)$ follows from the closed form expression for $\pi\,(t, t+n, b, \lambda_i)$, and equation 2. Observe that $\pi\,(t, t+n, b, \lambda_i) < 1$, and decreasing with n. Moreover $mb\,(t, b, \lambda_i, 1) = \Pr\{\text{event } \omega_i \text{ occurs at least once}\} \leq 1$ and $\lim_{b\to\infty} mb\,(t, b, \lambda_i, 1) = 1$. From this it immediately follows that marginal benefit is increasing in δ. This also implies that it is decreasing in t. To characterize the other properties of the marginal benefit, notice that it satisfies the following difference equation:

$$mb\,(t, b, \lambda_i, \delta) = \alpha_t + \delta\,(1 - \alpha_t)\,mb\,(t+1, b, \lambda_i, \delta)\,,$$

where $\alpha_t = \frac{\lambda}{1+t/b}$.

Hence α_t is decreasing in t and is in the interval $(0, 1)$. Thus we can conclude that:

$$mb\,(t+1, b, \lambda_i, \delta) \leq \frac{\alpha_t}{(1-\delta)}.$$

It follows that:

$$\begin{aligned} mb\,(t, b, \lambda_i, \delta) - mb\,(t+1, b, \lambda_i, \delta) &= \alpha_t + (\delta\,(1-\alpha_t) - 1)\,mb\,(t+1, b, \lambda_i, \delta) \\ &\geq \alpha_t \left(1 - \frac{(1 - \delta\,(1-\alpha_t))}{1-\delta}\right) \\ &> 0. \end{aligned}$$

Now observe if one differentiates the difference equation one obtains:

$$\partial mb\left(t,b,\lambda_i,\delta\right)/\partial\lambda_i = \left(1-\delta mb\left(t+1,b,\lambda_i,\delta\right)\right)/\left(1+t/b\right)+\delta\left(1-\alpha_t\right)\partial mb\left(t+1,b,\lambda_i,\delta\right)/\partial\lambda_i.$$

Now for $\delta < 1$, $\delta mb\left(t+1,b,\lambda_i,\delta\right) < 1$, and hence the constant term is positive, and $\delta\left(1-\alpha_t\right) < 1$, from which we conclude that $\partial mb\left(t,b,\lambda_i,\delta\right)/\partial\lambda_i$ exists and is strictly positive. In the case of b one has:

$$\partial mb\left(t,b,\lambda_i,\delta\right)/\partial b = \left(1-\delta mb\left(t+1,b,\lambda_i,\delta\right)\right)\frac{t\lambda_i}{\left(b+t\right)^2}+\delta\left(1-\alpha_t\right)\partial mb\left(t+1,b,\lambda_i,\delta\right)/\partial b,$$

and hence a similar argument demonstrates that $\partial mb\left(t,b,\lambda_i,\delta\right)/\partial b > 0$. ■

Proposition Suppose $\lim_{n\to\infty} mc\left(n\right)=\infty$, then for each b, there is an $N^*\left(b\right)$, such that for $N > N^*\left(b\right)$, the optimal level of planning, $n^*\left(t,b,N\right)$, satisfies:

$$mc(n^*+1) \geq mb\left(t,b,1/N\right) \geq mc(n^*). \tag{8}$$

Moreover $n^*\left(t,b,N\right)$ has the following properties:

1. $n^*\left(t,b,N\right) \geq n^*\left(t+1,b,N\right)$, and there is a T^* such that $n^*\left(t,b,N\right)=0$ for $t \geq T^*$.
2. $n^*\left(t,b,N\right)$ is increasing with b.
3. If $N > \left(u_g-u_b\right)/mc\left(1\right)$, then for b sufficiently close to zero $n^*\left(t,b,N\right)=0$ for all t.

Proof. The hypothesis that $\lim_{n\to\infty} mc\left(n\right)=\infty$, ensures the existence of $n^*\left(t,b,N\right)$. From statement 1 for each N there is a $T\left(b,N\right)$ such that $n^*\left(t,b,N\right)=0$ for $t > T\left(b,N\right)$. Moreover, $T\left(b,N\right)$ is decreasing with N, and hence there exists a smallest $N\left(b\right)$ satisfying:

$$N > \sum_{t=0}^{T(b,N)} n^*\left(t,b,N\right)+T\left(b,N\right).$$

Moreover this expression is satisfied for all $N > N\left(b\right)$. The right hand side specifies the largest number of states that one would have explored after $T\left(b,N\right)$ periods. For $t > T\left(b,N\right)$ only learning by doing is optimal, and hence there is never an option value to delaying the exploration of a state. This ensures that the optimal number of states to be explored each period is $n^*\left(t,b,N\right)$, as given by (8).

Let $c = mc\left(1\right) > 0$ be the marginal cost of planning for a single state. Notice that

$\lim_{t\to\infty}\pi\left(t,t+1,b,1/N\right)=0$, and $\pi\left(t,t+n,b,1/N\right) > \pi\left(t,t+n+1,b,1/N\right)$, hence it follows that $\lim_{t\to\infty} mb\left(t,b,N\right)=0$, and for large enough t, $mb\left(t,b,N\right) < c$, from which statement 1 follows.

For any $T > t$, $\pi\left(t,T+1,b,1/N\right)=\pi\left(t,T,b,1/N\right)\left(\frac{b(N-1)+T-1}{bN+T}\right)$, from which one can show recursively that $\partial\pi\left(t,T,b,N\right)/\partial b > 0$ for $T > t$, and hence the marginal benefit is an increasing function of b from which statement 2 follows

As b approaches zero this corresponds to the agent placing almost all probability mass on Ω^t for $t \geq 1$. Thus the only benefit from planning occurs in period 0, before any observations have been made. In that case the marginal benefit is $(u_g - u_b)/N + \varepsilon(b)$, where $\lim_{b\to 0} \varepsilon(b) = 0$, hence if $N > (u_g - u_b)/mc(1)$ the parameter b can be chosen sufficiently small that no planning takes place, which combined with 1 implies 3. ■

Proposition Suppose planning costs are increasing and convex in n, then there exists an optimal planning rule $n^0(r_t, t, b, N)$, where r_t is the number of states $\Omega \backslash \Omega^{t-1}$ with the following properties:

1. $n^0(r_t, t, b, N)$ is increasing in r_t.
2. If $r_t \leq \tilde{n}(t, b, N)$, then $n^0(r_t, t) = r_t$.
3. If $r_t > \tilde{n}(t, b, N)$, then $\min\{n^*(t, b, N), r_t\} \geq n^0(r_t, t, b, N) \geq \tilde{n}(t, b, N)$, where $n^*(t, b, N) = \arg\max_{n \geq 0} n \cdot mb(t, b, N) - c(n)$ is the optimal one period strategy and $\tilde{n}(t, b, N)$ satisfies:

$$mb_t - mc(\tilde{n}) \geq \max\{p_{t+1}\delta(mb_{t+1} - c(1)), 0\} \geq mb_t - mc(\tilde{n}+1),$$

where $p_t = (N + t/b)^{-1}$ is the probability that an event that has not been observed occurs in period t, and $mb_t = mb(t, b, 1/N)$.

Proof. Let $V_t(r_t)$ be the value obtained from further exploration of states, then from the dynamic programing algorithm the optimal search rule, $n^*(t, b, N)$, and the value function solve:

$$V_t(r_t) = \max_{n \leq r_t} (n_t \cdot mb_t - c(n_t)) + \delta \left\{ \begin{array}{c} p_t(r_t - n_t) V_{t+1}(r_t - n_t - 1) + \\ (1 - p_t(r_t - n_t)) V_{t+1}(r_t - n_t) \end{array} \right\}.$$

The first term is the value of search in the current period, while the second term is the value from delaying search to subsequent periods. Due to the convexity of c the marginal cost of planning increases with n_t, and hence it maybe worthwhile to delay planning for an event until the next period. The benefit from doing this is at least $\delta(1 - p_t)(mb_{t+1} - c(1))$, where $(1 - p_t)$ is the probability the event does not occur in period t (and hence one adds it to the set of known events for period $(t+1)$, and $mb_{t+1} - c(1)$ is the benefit for planning this event in the following period. For any $n_t \leq \tilde{n}(t, b, N)$ it never pays to delay planning, and moreover if $n_t < \tilde{n}(t, b, N)$, then the agent gains by increasing n_t, hence for $r_t \leq \tilde{n}(t, b, N)$, $n^0(r_t, t, b, N) = r_t$.

Since mb_t is decreasing with t and $\lim_{t\to\infty} mb_t = 0$, then $\tilde{n}(t, b, N)$ decreases with t. If $\tilde{n}(0, b, N) < 0$ the individual engages in no planning, and we are done. Suppose not, then there is a first period T such that $\tilde{n}(T-1, b, N) \geq 1$, and $\tilde{n}(T, b, N) < 0$ from period T on, and hence it is optimal to do no further search and $V_t(r_t) = 0$ for all $t \geq T$, and $r_t \geq 0$. This implies that the optimal planning problem can be viewed as a finite horizon problem with a finite state space, thus ensuring the existence of an optimal strategy.

The final step is to characterize $n^0(r_t, t, b, N)$ and show that it is increasing in r_t when $r_t > \tilde{n}(t, b, N)$. Let $\mu_t(r) = V_t(r) - V_t(r-1)$ be the marginal benefit from having more states to explore. It shall be shown by induction that the function is decreasing with r_t from which it will follow that the optimal planning strategy is increasing in r_t. From the above it is the case that $\mu_T(r) = 0$, and hence the function is decreasing in r for $t = T$.

Let $\lambda_t(n) = mb_t - mc(n)$ be the net marginal benefit in period t from increasing planning from $n-1$ to n. The first order conditions for the optimal amount of planning satisfy:

$$\tilde{\mu}_t\left(r_t - n^0\right) \geq \lambda_t\left(n^0\right) \geq \tilde{\mu}_t\left(r_t - n^0 + 1\right), \tag{9}$$

where

$$\tilde{\mu}_t(r) = \delta\left\{p_t(r-1)\,\mu_{t+1}(r) + (1 - p_t r)\,\mu_{t+1}(r)\right\}$$

is the increase in the expected value from increasing the stock of unexplored states from $r-1$ to r. From the induction hypothesis that $\mu_t(r)$ is decreasing in r it follows that $\tilde{\mu}_t(r)$ is also decreasing in r. Since $\lambda_t(n)$ is also decreasing with n, then it follows from 9 that $n^0(r, t, b, N)$ is increasing in r.

Finally, using this result one can show that $V_t(r) - V_t(r-1) = \lambda_t\left(n^0(r, t, b, N)\right)$ (there are two cases, depending upon whether or not n^0 increases when going from $r-1$ to r), which implies that $\mu_t(r)$ is decreasing in r, and hence the induction hypothesis is satisfied and we are done. ■

Proposition Under the k-nearest neighbor rule with a Euclidean similarity measure expected performance approaches the optimum:

$$\lim_{t \to \infty} U^t = U^*$$

Proof. Since utility is bounded, then $M = \max_{z,z',\omega} |U(z|\omega) - U(z'|\omega)| < \infty$. Then we have:

$$\begin{aligned}
U^* - U^t &= E\left\{E\left\{U(\sigma^*(\omega)|\omega) - U\left(\sigma\left(\omega|H^t\right)|\omega\right)|H^t\right\}\right\} \\
&\leq E\left\{M \cdot \Pr\left\{\sigma^*(\omega) \neq \sigma\left(\omega|H^t\right)|H^T\right\}\right\} \\
&\leq E\left\{M \cdot \sum_{i=1}^{n} \Pr\left\{\sigma_i^*(\omega) \neq \sigma_i\left(\omega|H^t\right)|H^T\right\}\right\} \\
&\leq M \cdot \sum_{i=1}^{n}\sum_{i=1}^{n} E\left\{\Pr\left\{\sigma_i^*(\omega) \neq \sigma_i\left(\omega|H^t\right)|H^T\right\}\right\}
\end{aligned}$$

For a given coordinate i, let $X_t = \omega^t$ and $Y_t = \sigma_i^*\left(\omega^t\right) \in \{0, 1\}$, then the problem can be viewed as one of prediction, namely given X_t, can one predict Y_t. This is formally the problem studied in the statistical pattern recognition literature (Devroye, Györfi, and Lugosi (1996)). One measure of the asymptotic performance of the rule is the expected probability of error, or for the $k-$nearest neighbor rule:

$$L_{knn} = \lim_{t \to \infty} E \left\{ \Pr \left\{ \sigma_i^* (\omega) \neq \sigma_i \left(\omega | H^t \right) | H^T \right\} \right\},$$

which is the limiting value of the last expression in the previous inequality. The best performance occurs if one knows $\sigma_i^* (\omega^t)$, and is given by:

$$L^* = E \left\{ \min \left\{ \eta \left(X_t \right), 1 - \eta \left(X_t \right) \right\} \right\},$$

where $\eta (\omega^t) = \Pr \{ \sigma_i^* (\omega^t) = 1 \}$. Notice that since $\sigma_i^* (\omega^t)$ is a deterministic function, then it is the case that $L^* = 0$. From theorem 5.4 of Devroye, Györfi, and Lugosi (1996) one has that $L^* \leq L_{knn} \leq 2L^*$, and hence one concludes that:

$$\lim_{t \to \infty} M \cdot \sum_{i=1}^{n} E \left\{ \sum_{i=1}^{n} \Pr \left\{ \sigma_i^* (\omega) \neq \sigma_i \left(\omega | H^t \right) | H^T \right\} \right\} = 0,$$

and we are done. ■

A.2 Figures

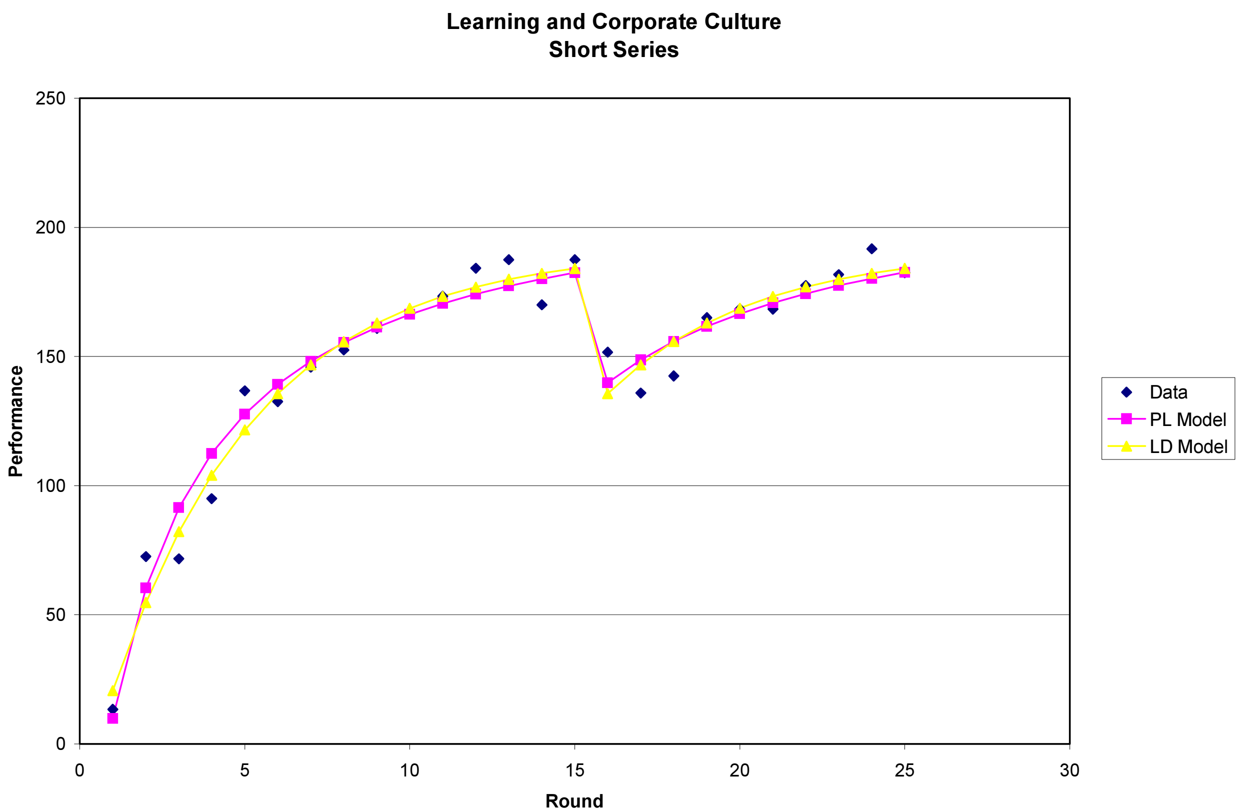

Figure 1: Learning with Corporate Culture - Short Series

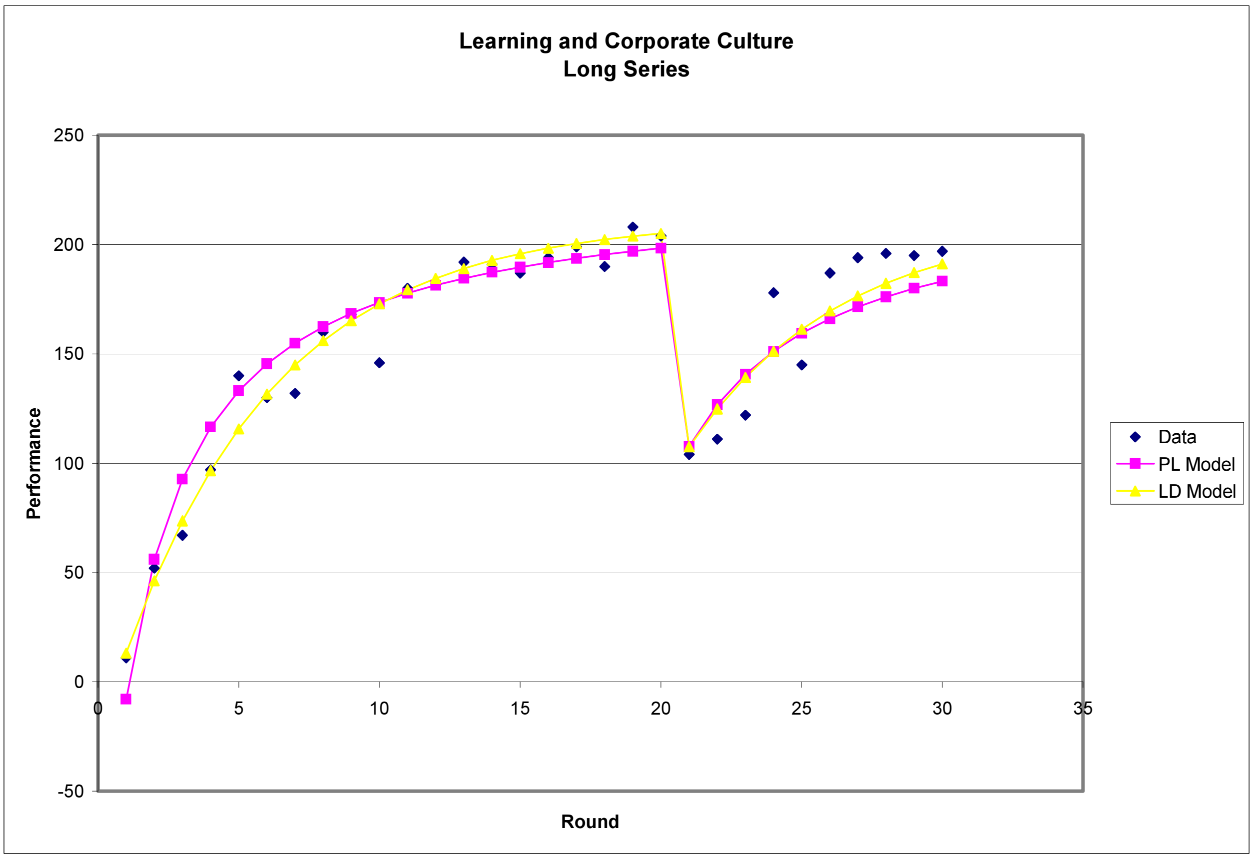

Figure 2: Learning with Corporate Culture Data - Long Series