Passive Learning – a critique by example*

Godfrey Keller†

First version: December 1998

Revised: February 2000, August 2003, February 2005

This version: December 2005

Abstract

In models of learning by experimentation that exhibit signal dependence, a benchmark using a *passive learner* has been proposed. The use of this benchmark is flawed – first, passive learning does not disentangle the effects of knowing that beliefs, as well as other state variables, might change, and we address this issue directly by introducing a *naïve learner*. Secondly, and more tellingly, passive learning does not do what it is supposed to do, namely help measure the gains from active experimentation; the naïve learner enables us to illustrate this point in the context of a particular example.

Keywords: Experimentation, Learning.

JEL Classification Numbers: D83.

*Thanks for helpful discussions are owed to Donata Hoesch, Thomas Mariotti, and Margaret Stevens; and, for constructive comments, to an anonymous referee. A previous version of this work appeared as Discussion Paper 223, Department of Economics, University of Oxford, with the title "The (in)appropriate benchmark when beliefs are not the only state variable."

†Department of Economics, University of Oxford, Manor Road Building, Oxford OX1 3UQ, UK.

Introduction

Models of learning and experimentation typically exhibit a trade-off between short-term rewards and long-term informational benefits: when the per-period payoff is uncertain, the agent can often incur an opportunity cost in order to resolve (some of) this uncertainty, and consequently improve his payoff in the future.

In the literature on this sort of problem, the agent is commonly said to *experiment* if he deviates from the best short-term action in order to learn. When the only intertemporal link is the agent's belief (about the unknown parameter that determines his per-period payoff), that belief is the natural state variable for the decision problem, and the best short-term action is simply the one that maximizes his current payoff – here, myopia provides a benchmark against which we can measure the actions and payoffs of a rational agent. However, in the most basic model from Lazear [4], for example, there is a further intertemporal link: he considers the two-period problem facing a monopolist with only one unit of a good to sell *across* the two periods – if a sale is made today then there is nothing to sell tomorrow. There being an additional element of the state variable that directly affects payoffs is dubbed *signal dependence* in Datta, Mirman & Schlee [2], and it is unclear what constitutes the appropriate benchmark in this case.

Datta, Mirman & Schlee [1] argue that to use the benchmark from the case in which the monopolist has one unit of the good to sell in *each* period is incorrect and consequently that the derivation in Trefler [5] of Lazear's result (on the direction of experimentation) as a corollary is inappropriate. They propose a benchmark that makes use of a passive learner as opposed to an active experimenter, notions that seem to date back to the article by Freixas [3].[1]

In order to assess this proposal, we analyze the actions and payoffs of five types of seller: a non-learner, a myopic learner, a naïve learner, a passive learner, and a full optimizer. These sellers have an increasing awareness of their economic environment, and we might expect that payoffs increase as the level of sophistication grows – a myopic learner benefits from his experience, a naïve learner takes into account other intertemporal links, a passive learner further takes into account that outcomes generate information, and a full optimizer realizes that he can affect the amount of information generated.

However, payoffs actually fall when we move from a naïve learner to a passive learner in both the cases of signal dependence and signal independence that we study, and, at least in this example, a naïve learner proves to be a better reference point from which

[1] In this context, a passive learner understands that he has only one good and that his belief about the buyer's valuation will change; he takes these into account when setting today's price but he ignores the fact that he can actively affect the information content of the sale/no sale result today, and indeed takes a self-fulfilling action. Note, however, that this consistency requirement of rational expectations does not seem to be made explicitly by Freixas himself.

to measure not only the gains from active experimentation, but also the effects of not being completely short-sighted.

1 The Example

There are two periods, the second period payoff being discounted by δ with $0 < \delta \leq 1$. There is a replicated (non-strategic) buyer, one each period: the buyer in period 2 has the same valuation as the buyer in period 1. There is a single risk-neutral seller whose valuation of the good for sale is normalized to 0, and, for simplicity, his prior belief is that the buyer's valuation v is uniformly distributed on $[0, 1]$.

A seller that learns from the period 1 action and outcome revises his belief about the buyer's valuation using Bayes' rule. If a price p_1 is charged in period 1 and no sale is made then, at the start of period 2, he believes that $v \sim U[0, p_1]$, whereas if a sale is made then he believes that $v \sim U[p_1, 1]$. Thus, in period 2 if he still has a good to sell, the seller wants to

$$\max_{p_2 \in [a,b]} \; [(b - p_2)/(b - a)] \; p_2 \, ,$$

where $[a, b] = [0, 1]$ for a non-learner and $[a, b] = [0, p_1]$ or $[a, b] = [p_1, 1]$ for a learner. Consequently, in period 2 he will charge a price of $p_2^* = \max \{b/2, a\}$, expecting a profit π_2 of either $b^2/4(b - a)$ or a.

We consider two cases: one good across the two periods, then one good per period.

1.1 One good across periods

First we consider the signal-dependent case when the state variable *is not* just the agent's belief – the seller has one unit of a good to sell, in *either* period 1 *or* period 2. In period 1, the seller sets a price p_1:

- if the buyer buys, the seller makes a profit of p_1, can revise his belief about the buyer's valuation upwards *but has nothing to sell in period 2*;
- if the buyer does not, the seller makes a profit of 0, can revise his belief about the buyer's valuation downwards and offer the good for sale at a lower price in period 2.

After no sale, a non-learner will choose a price of 1/2 in period 2 and (mistakenly) expect a profit of 1/4, whereas a seller that learns will choose a price of $p_1/2$ in period 2 and expect a profit of $p_1/4$. (Note that when we write $\mathrm{E}[\pi_2]$ below, it is not the seller's expectation in period 1 of his period 2 profit, but what *we*, as modellers, expect his period 2 profit to be, i.e., the profit that he will actually obtain on average.)

Non-learner

A *non-learner* does not think that anything will change from one period to the next; of course, in period 2, he is constrained by the number of goods he has left to sell, but he doesn't use the period 1 outcome to update his belief.

He doesn't expect to learn about v and indeed doesn't learn, but also he doesn't anticipate that he might not have a good to sell in period 2. His initial problem is time-separable and can be solved forwards. It is to

$$\max_{p_1 \in [0,1]} \{p_1.0 + (1 - p_1)\, p_1 + \delta\, [1/4]\}$$

leading to $p_1^* = \frac{1}{2}$. If he doesn't sell, then, since he doesn't learn, he chooses $p_2^* = \frac{1}{2}$ and again fails to sell. Thus his average overall profit when $\delta = 1$ is $\mathrm{E}[\pi_1] + \mathrm{E}[\pi_2] = \frac{1}{4} + 0 = \frac{1}{4}$.

Myopic learner

A *myopic learner* also does not think that anything will change from one period to the next, but, before period 2, he uses the period 1 outcome to update his belief (as well as the number of goods he has left).

He doesn't expect to learn about v but, in fact, does learn; he too doesn't anticipate that he might not have a good to sell in period 2. His *initial* problem is the same as that of a non-learner, namely

$$\max_{p_1 \in [0,1]} \{p_1.0 + (1 - p_1)\, p_1 + \delta\, [1/4]\}$$

again leading to $p_1^* = \frac{1}{2}$. If he doesn't sell, then he cuts his price by choosing $p_2^* = \frac{1}{4}$, and his average overall profit when $\delta = 1$ is $\mathrm{E}[\pi_1] + \mathrm{E}[\pi_2] = \frac{1}{4} + \frac{1}{16} = \frac{5}{16}$.

Note that the optimal p_2 from the perspective of period 1 is $\frac{1}{2} \neq p_2^*$.

Naïve learner

A *naïve learner* understands how the number of goods left to sell might change, but doesn't foresee that his belief will; nevertheless, before period 2, he does update his belief.

He doesn't expect to learn about v but, in fact, he too does learn; however, he does anticipate that he might not have a good to sell in period 2. Consequently, his initial problem must be solved backwards. It is to

$$\max_{p_1 \in [0,1]} \{p_1.0 + (1 - p_1)\, p_1 + \delta\, [p_1(1/4) + (1 - p_1).0]\}$$

leading to $p_1^* = \frac{5}{8}$ when $\delta = 1$. If he doesn't sell, then he too halves his price, chooses

$p_2^* = \frac{5}{16}$, and his average overall profit when $\delta = 1$ is $\mathrm{E}[\pi_1] + \mathrm{E}[\pi_2] = \frac{15}{64} + \frac{25}{256} = \frac{85}{256}$. Again, note that the optimal p_2 from the perspective of period 1 differs from p_2^*.

Passive learner

A *passive learner* understands how the number of goods might change. He further knows that if, at a price of z_1, he doesn't sell in period 1, then he will have a posterior belief that $v \sim U[0, z_1]$. (If he does sell then his posterior belief will be that $v \sim U[z_1, 1]$, but that is not relevant here since he has sold his only good.) He can determine the optimal period 2 action, namely to choose $p_2 = z_1/2$, and calculate the expected period 2 profit conditional on z_1, namely $z_1/4$. Taking z_1 as given, he calculates the optimal period 1 price, which is a function of z_1, then he calculates his posterior belief in the case where he hasn't sold, namely $v \sim U[0, p_1(z_1)]$; p_1^* is the *self-fulfilling* action, that is the price that induces the distribution of the period 2 belief that he used in his period 1 decision.

He doesn't realize in advance that $p_1 = z_1$ – he expects to learn about v but doesn't understand how his period 1 choice affects quite what he will learn; he correctly anticipates that he might not have a good to sell in period 2. His initial problem, which must also be solved backwards, is to

$$\max_{p_1 \in [0,1]} \{p_1.0 + (1 - p_1)\, p_1 + \delta\, [p_1(z_1/4) + (1 - p_1).0]\} \mid z_1 \in [0, 1]$$

leading to $p_1(z_1) = \frac{1}{2}(1 + \frac{1}{4}\delta z_1)$. Imposing the self-fulfilling action $p_1^* = z_1$ leads to $p_1^* = \frac{4}{7}$ when $\delta = 1$. If he doesn't sell, then he chooses $p_2^* = \frac{2}{7}$ and his average overall profit when $\delta = 1$ is $\mathrm{E}[\pi_1] + \mathrm{E}[\pi_2] = \frac{12}{49} + \frac{4}{49} = \frac{16}{49}$.

Fully optimizing learner

A *fully optimizing learner* correctly foresees how both his belief and the number of goods will change. He knows all the implications of his period 1 choice – he expects to learn about v and understands how his period 1 choice affects exactly what he will learn. His initial problem is to

$$\max_{p_1 \in [0,1]} \{p_1.0 + (1 - p_1)\, p_1 + \delta\, [p_1(p_1/4) + (1 - p_1).0]\}$$

leading to $p_1^* = \frac{2}{3}$ when $\delta = 1$. If he doesn't sell, then he chooses $p_2^* = \frac{1}{3}$ with an average overall profit when $\delta = 1$ of $\mathrm{E}[\pi_1] + \mathrm{E}[\pi_2] = \frac{2}{9} + \frac{1}{9} = \frac{1}{3}$.

To summarize:

Type of seller	p_1^*	E[π]	Expected p_2		Actual p_2^*
Non-learner	0.5000	**0.2500**	0.5000	$=$	0.5000
Myopic learner	0.5000	**0.3125**	0.5000	$\neq$	0.2500
Naïve learner	0.6250	**0.3320**	0.5000	$\neq$	0.3125
Passive learner	0.5714	**0.3265**	$z_1/2$	$=$	0.2857
Full optimizer	0.6667	**0.3333**	0.3333	$=$	0.3333

We see that payoffs are *not* monotonically increasing with the sophistication of the seller – there is a drop between a naïve learner and a passive learner. In this example, if we take a non-learner as the reference point, 75% of the total available gains accrue when we simply move to a myopic learner, even though the latter is time-inconsistent; using a myopic learner as the benchmark, 94% of the additional gains are made by moving to a naïve learner, even though he too is time-inconsistent.

The reason that this non-monotonicity has not been noticed before is that the use of a naïve learner is novel to this work – if we remove this type of seller then the non-monotonicity of overall payoffs, and also first period choices, disappears.

1.2 One good each period

Passive learning was introduced specifically for signal-dependent models, in an attempt to help measure the effects of active experimentation. However, in principle, it could also be used with signal-independent models for the same purpose. So, for completeness, we repeat the above analysis for this variant of the example.

The signal-independent case is when the state variable *is* just the agent's belief – the seller has one unit of a good to sell in each period. In period 1, the seller sets a price p_1:

- if the buyer buys, the seller makes a profit of p_1, can revise his belief about the buyer's valuation upwards and offer the good for sale at a possibly higher price in period 2;
- if the buyer does not, the seller makes a profit of 0, can revise his belief about the buyer's valuation downwards and offer the good for sale at a lower price in period 2.

Sale or no sale, a non-learner will choose a price of 1/2 in period 2 and expect a profit of 1/4, whereas the choice of a seller that learns will depend on the period 1 outcome: after no sale, he is in the same position as his counterpart in the previous subsection and so will halve his price in period 2 and expect a profit of $p_1/4$; in period 2 after a sale, if $p_1 < 1/2$ he will increase his price to 1/2 and expect a profit of $1/4(1 - p_1)$, but if $p_1 \geq 1/2$ he will keep his price at p_1 and make a profit of p_1 since he sells for sure. (Again, E[π_2] below denotes *our* expectation, not the seller's.)

Non-learner

His initial problem is as before, and so he chooses a price of $\frac{1}{2}$ in both periods with an average overall profit when $\delta = 1$ of $\mathrm{E}[\pi_1] + \mathrm{E}[\pi_2] = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$.

Myopic learner

He also has the same initial problem, again choosing $p_1^* = \frac{1}{2}$. Here, he chooses $p_2^* = \frac{1}{4}$ or $p_2^* = \frac{1}{2}$ depending on the outcome in period 1, with an average overall profit when $\delta = 1$ of $\mathrm{E}[\pi_1] + \mathrm{E}[\pi_2] = \frac{1}{4} + \frac{5}{16} = \frac{9}{16}$. Note that he is again time-inconsistent after no sale in period 1.

Naïve learner

As beliefs are the only intertemporal link, the choices of a naïve learner mimic those of a myopic learner, as does his average overall profit.

Passive learner

A passive learner knows that with a period 1 price of $z_1 \in [0, 1]$ he will revise his belief to either $v \sim U[0, z_1]$ or $v \sim U[z_1, 1]$ depending on the outcome (no sale/sale), and can calculate his expected period 2 profit, conditional on z_1, in each case. Consequently, his initial problem is either to

$$\max_{p_1 \in [0,1]} \{p_1.0 + (1 - p_1)\, p_1 + \delta\, [p_1(z_1/4) + (1 - p_1)/4(1 - z_1)]\} \mid z_1 \in [0, 1/2)$$

or to

$$\max_{p_1 \in [0,1]} \{p_1.0 + (1 - p_1)\, p_1 + \delta\, [p_1(z_1/4) + (1 - p_1)\, z_1]\} \mid z_1 \in [1/2, 1]\,.$$

The solution is that to the 'either' problem, given by $p_1(z_1) = \frac{1}{2}(1 - \frac{1}{4}\delta[1/(1 - z_1) - z_1])$, and imposing the self-fulfilling action $p_1^* = z_1$ leads to $p_1^* = (11 - \sqrt{37})/14 \simeq 0.3512$ when $\delta = 1$. He chooses $p_2^* = p_1^*/2$ or $p_2^* = 1/2$ depending on the outcome, leading to an average overall profit when $\delta = 1$ of $\mathrm{E}[\pi_1] + \mathrm{E}[\pi_2] = (1 - p_1^*)\, p_1^* + (p_1^{*2}/4 + 1/4) \simeq 0.5087$.

Fully optimizing learner

After a sale in period 1, the functional form of his expected period 2 profit depends on whether or not $p_1 < 1/2$. So, his initial problem is either to

$$\max_{p_1 \in [0,1/2)} \{p_1.0 + (1 - p_1)\, p_1 + \delta\, [p_1(p_1/4) + (1 - p_1)/4(1 - p_1)]\}$$

or to

$$\max_{p_1 \in [1/2,1]} \{p_1.0 + (1-p_1)\,p_1 + \delta\,[p_1(p_1/4) + (1-p_1)\,p_1]\}\ .$$

The solution is that to the 'or' problem, given by $p_1^* = \frac{4}{7}$ when $\delta = 1$. He chooses $p_2^* = \frac{2}{7}$ or $p_2^* = \frac{4}{7}$ with an average overall profit when $\delta = 1$ of $\mathrm{E}[\pi_1] + \mathrm{E}[\pi_2] = \frac{12}{49} + \frac{16}{49} = \frac{4}{7}$.

To summarize:

Type of seller	p_1^*	$\mathrm{E}[\pi]$	Expected p_2		Actual p_2^*
Non-learner	0.5000	**0.5000**	0.5000	$=$	0.5000
Myopic learner *& Naïve learner*	0.5000	**0.5625**	0.5000	$\neq$	0.2500 or 0.5000
Passive learner	0.3512	**0.5087**	$z_1/2$ or 0.5000	$=$	0.1756 or 0.5000
Full optimizer	0.5714	**0.5714**	0.2857 or 0.5714	$=$	0.2857 or 0.5714

Again, we see that payoffs are *not* monotonically increasing with the sophistication of the seller – there is a very steep drop between a myopic or naïve learner and a passive learner. Here, if we take a non-learner as the reference point, 88% of the total available gains accrue when we simply move to a myopic or naïve learner; using a myopic learner as the benchmark, additional gains are made only by moving to a full optimizer. Indeed, a passive learner does only slightly better than a non-learner.

References

1. Datta, M., Mirman, L., Schlee, E.: Learning with Noiseless Information & Payoff-Relevant Signals. Economic Theory **16**, 63–75 (2000)
2. Datta, M., Mirman, L., Schlee, E.: Optimal Experimentation in Signal-Dependent Decision Problems. International Economic Review **43**, 577–607 (2002)
3. Freixas, X.: Optimal Growth with Experimentation. Journal of Economic Theory **24**, 296–309 (1981)
4. Lazear, E.: Retail Pricing & Clearance Sales. American Economic Review **76**, 14–32 (1986)
5. Trefler, D.: The Ignorant Monopolist: Optimal Learning with Endogenous Information. International Economic Review **34**, 565–581 (1993)