

ISSN 1471-0498



DEPARTMENT OF ECONOMICS

DISCUSSION PAPER SERIES

**THE (IN)APPROPRIATE BENCHMARK WHEN BELIEFS ARE NOT
THE ONLY STATE VARIABLE**

Godfrey Keller

Number 223

February 2005

Manor Road Building, Oxford OX1 3UQ

THE (IN)APPROPRIATE BENCHMARK WHEN BELIEFS ARE NOT THE ONLY STATE VARIABLE*

Godfrey Keller[†]

First version: December 1998

Revised: February 2000, August 2003

This version: February 2005

Abstract

In models of learning by experimentation, there is a natural benchmark of myopia when the only intertemporal link is the agent's subjective belief (signal independence). An alternative benchmark using a *passive learner* has been proposed when there is a further intertemporal link that directly affects payoffs (signal dependence). The purpose of this note is to suggest that the use of this particular benchmark is flawed for two reasons: first, passive learning does not disentangle the effects of knowing that beliefs might change as well as other state variables, and we offer another benchmark using a *naïve learner* that does, and so necessarily reduces to myopia in the signal independent case; secondly, and maybe more tellingly, passive learning does not do what it is supposed to do, namely help measure the gains from active experimentation, since the payoffs of a passive learner can be markedly lower than those of a naïve learner.

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.

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

Introduction

Many models of learning and experimentation 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* when he deviates from the myopically optimal action. 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 myopically optimal action is the one that just maximizes his current payoff. So myopia provides a benchmark against which we can measure the actions of a rational agent. However, in the most basic model from Lazear (1986), 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; the seller's belief about the buyer's valuation is not the only element of the state variable that directly affects payoffs – if a sale is made today then there is nothing to sell tomorrow. This is dubbed *signal dependence* in Datta, Mirman & Schlee (2002), and it is unclear what constitutes the appropriate benchmark in this case.

Datta, Mirman & Schlee (2000) 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 (1993) of Lazear's result (on the direction of experimentation) as a corollary is inappropriate. The benchmark proposed by Datta *et al.* makes use of a passive learner as opposed to an active experimenter, notions that seem to date back to the article by Freixas (1981). In this context, a passive learner realizes that if he sells his only good today then he will have nothing to sell tomorrow and that if he doesn't sell then his belief about the buyer's valuation might change – he takes these into account when setting today's price; however, 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.¹

The above models tend to focus on period 1 actions when making their comparisons between different sorts of agent, whereas it may be more illuminating to use overall expected payoffs, which we do here. To help clarify matters, we first look at two types of myopic agent – a *non-learner* who stubbornly holds on to his initial prior belief, and a *myopic learner* who revises his belief in accord with the outcome of his action. (These two types take the same action in period 1, but thereafter their actions differ as do their overall expected payoffs.) In order to disentangle experimentation effects from other intertemporal concerns, we then introduce a *naïve learner* who also ignores the fact that his belief might change but understands everything else about his problem, namely the other intertemporal links. (Clearly, if there are no other intertemporal links then a naïve learner does the same as a myopic learner.) Next comes a *passive learner*, and finally a *fully optimizing learner* who understands the problem completely. The five types of seller have an increasing awareness of their economic environment, and we might expect

¹This consistency requirement of rational expectations does not seem to be made explicitly by Freixas himself.

that payoffs increase as the level of sophistication grows.

A myopic learner should do better than a non-learner, since the former uses his experience to refine his belief about the buyer's valuation. A naïve learner should do even better, since he takes into account the intertemporal links other than his belief. A passive learner further takes into account that the period 1 outcome generates information (but without realizing that he can affect the amount), so we might expect his payoff to be an improvement over a naïve learner. Finally, a full optimizer should do best of all.

However, payoffs actually fall when we move from a naïve learner to a passive learner, as we shall see in the main section, in both the cases of signal dependence and signal independence that we study, namely, a simple two-period sales model (“one good across periods”) and the more commonly considered case (“one good each period”).

As these two cases are very stylized monopoly pricing models, in the appendix we formulate the abstract two-period model to show how the problem is solved by each of the five types of agent and to illustrate the differences in behaviour between them in a much broader setting; we then generalize it to an arbitrary, but finite, horizon.

1 The Model

There are two periods, 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 his prior belief is that the buyer's valuation v is distributed on $[0, 1]$ according to some CDF $F(\cdot)$.

We consider two cases – either one good across the two periods, or one good per period – and five types of seller: a non-learner, a myopic learner, a naïve learner, a passive learner, and a fully optimizing learner. For any of the learning types, his posterior belief at the start of period 2, $\tilde{F}(\cdot)$, is derived from the prior using Bayes' rule, given the price charged in period 1 and the outcome (0 = no sale, 1 = sale):

$$\begin{aligned} \tilde{F}(v | p, 0) &= \frac{F(v)}{F(p)} \quad \text{if } v < p, & \tilde{F}(v | p, 0) &= 1 & \quad \text{if } v \geq p; \\ \tilde{F}(v | p, 1) &= 0 & \quad \text{if } v < p, & \tilde{F}(v | p, 1) &= \frac{F(v) - F(p)}{1 - F(p)} & \quad \text{if } v \geq p. \end{aligned}$$

The types of seller are progressively more aware of their economic environment.

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

- 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 $\tilde{F}(\cdot \mid z_1, 0)$, and if he does sell then his posterior belief will be $\tilde{F}(\cdot \mid z_1, 1)$; in either case he can determine the optimal period 2 action and calculate the expected period 2 profit conditional on z_1 . Taking z_1 , and thus $\tilde{F}(\cdot \mid z_1, \cdot)$, as given, he calculates the optimal period 1 price, which is a function of z_1 , then he calculates his posterior belief; 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.
- A *fully optimizing learner* correctly foresees how both his belief and the number of goods will change.

For simplicity, we shall focus on the Uniform (Rectangular) distribution. This implies that if a price p is charged and no sale is made then a Bayesian seller believes that the buyer's valuation is uniformly distributed on $[0, p]$, whereas if a sale is made then he believes that it is uniformly distributed on $[p, 1]$. Consequently, we need to consider the more general single-period problem when the seller believes that $v \sim U[a, b]$, for $[a, b] \neq [0, 1]$. Such a seller wants to

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

leading to $p^* = \max\{b/2, a\}$ with an expected profit of either $\pi = b^2/4(b - a)$ or $\pi = a$.

1.1 One good across periods

Here we explore the situation 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, the seller's beliefs in period 2 will be $v \sim U[0, b]$ for some b that depends on the type of seller and on the period 1 choice – for a non-learner $b = 1$, and for a learner $b = p_1$. Thus, in period 2, each type of seller will choose $p_2 = b/2$ leading to $E[\pi_2 \mid b] = b/4$. (Note that when we write $E[\pi_2]$ below, it is what *we* expect the seller's period 2 profit to be, not *his* expectation in period 1 of his period 2 profit.)

Non-learner

In period 1, a non-learner thinks that the situation in period 2 will be unchanged – he doesn't expect to learn about v and indeed doesn't learn; also, he doesn't anticipate that he might not have a good to sell in period 2. Consequently, his initial problem is to

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

leading to

- $p_1^* = \frac{1}{2}$.

If he doesn't sell, then, since he doesn't learn, he still believes that $v \sim U[0, 1]$ leading to

- $p_2^* = \frac{1}{2}$

and no sale again. Thus his average overall profit when $\delta = 1$ is

- $E[\pi_1] + E[\pi_2] = \frac{1}{4} + 0 = \frac{1}{4}$.

Myopic learner

In period 1, a myopic learner also thinks that the situation in period 2 will be unchanged – 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. Consequently, his initial problem is also to

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

again leading to

- $p_1^* = \frac{1}{2}$.

If he doesn't sell, then he updates his belief to $v \sim U[0, 1/2]$ leading to

- $p_2^* = \frac{1}{4}$

and an average overall profit when $\delta = 1$ of

- $E[\pi_1] + E[\pi_2] = \frac{1}{4} + \frac{1}{16} = \frac{5}{16}$.

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

Naïve learner

A naïve learner knows that if he sells in period 1 then he has nothing to sell in period 2, but he thinks that his belief in period 2 will be unchanged – 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 is to

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

leading to

- $p_1^* = \frac{5}{8}$

when $\delta = 1$. If he doesn't sell, then he updates his belief to $v \sim U[0, 5/8]$ leading to

- $p_2^* = \frac{5}{16}$

and an average overall profit when $\delta = 1$ of

- $E[\pi_1] + E[\pi_2] = \frac{15}{64} + \frac{25}{256} = \frac{85}{256}$.

Again, note that the optimal p_2 from the perspective of period 1 differs from p_2^* .

Passive learner

A passive learner knows that if he sells in period 1 then he has nothing to sell in period 2, and further knows that if he doesn't sell at a period 1 price of $z_1 \in [0, 1]$ then he will

revise his belief to $v \sim U[0, z_1]$, choose $p_2 = z_1/2$ and expect $\pi_2 = z_1/4$, but 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. Consequently, his initial problem is to

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

leading to $p_1^* = \frac{1}{2}(1 + \frac{1}{4}\delta z_1)$. 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 updates his belief to $v \sim U[0, 4/7]$ leading to

- $p_2^* = \frac{2}{7}$

and an average overall profit when $\delta = 1$ of

- $E[\pi_1] + E[\pi_2] = \frac{12}{49} + \frac{4}{49} = \frac{16}{49}$.

Fully optimizing learner

A full optimizer 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; he also correctly anticipates that he might not have a good to sell in period 2. Consequently, his initial problem is to

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

leading to

- $p_1^* = \frac{2}{3}$

when $\delta = 1$. If he doesn't sell, then he updates his belief to $v \sim U[0, 2/3]$ leading to

- $p_2^* = \frac{1}{3}$

and an average overall profit when $\delta = 1$ of

- $E[\pi_1] + E[\pi_2] = \frac{2}{9} + \frac{1}{9} = \frac{1}{3}$.

To summarize:

Type of seller	p_1^*	$E[\pi]$	Expected p_2	Actual p_2^*
<i>Non-learner</i>	0.5000	0.2500	0.5000	= 0.5000
<i>Myopic learner</i>	0.5000	0.3125	0.5000	≠ 0.2500
<i>Naïve learner</i>	0.6250	0.3320	0.5000	≠ 0.3125
<i>Passive learner</i>	0.5714	0.3265	$z_1/2$	= 0.2857
<i>Full optimizer</i>	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.

1.2 One good each period

Now we explore the situation when the state variable *is* just the agent's belief – the seller has one unit of a good to sell 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.

The seller's beliefs in period 2 will be $v \sim U[0, b]$ after no sale, or $v \sim U[a, 1]$ after a sale, for some a, b that depend on the type of seller and on the period 1 choice – for a non-learner $a = 0, b = 1$, and for a learner $a = b = p_1$. Thus, in period 2 after no sale, each type of seller will choose $p_2 = b/2$ leading to $E[\pi_2 | b] = b/4$; in period 2 after a sale, if $a < 1/2$ he will choose $p_2 = 1/2$ leading to $E[\pi_2 | a] = 1/4(1 - a)$, and if $a \geq 1/2$ he will choose $p_2 = a$ leading to $E[\pi_2 | a] = a$ since he sells for sure. (Again, $E[\pi_2]$ below denotes *our* expectation, not the seller's.)

Non-learner

In period 1, a non-learner thinks that the situation in period 2 will be unchanged. Consequently, his initial problem is to

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

leading to

- $p_1^* = \frac{1}{2}$.

Whether or not he sells, he still believes that $v \sim U[0, 1]$ since he doesn't learn, leading to

- $p_2^* = \frac{1}{2}$

and the same outcome again. Thus his average overall profit when $\delta = 1$ is

- $E[\pi_1] + E[\pi_2] = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$.

Myopic learner

In period 1, a myopic learner also thinks that the situation in period 2 will be unchanged. Consequently, his initial problem is also to

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

again leading to

- $p_1^* = \frac{1}{2}$.

If he doesn't sell, then he updates his belief to $v \sim U[0, 1/2]$, and if he does sell, then he updates his belief to $v \sim U[1/2, 1]$, leading to

- $p_2^* = \frac{1}{4}$ or $p_2^* = \frac{1}{2}$

and an average overall profit when $\delta = 1$ of

- $E[\pi_1] + E[\pi_2] = \frac{1}{4} + \frac{5}{16} = \frac{9}{16}$.

Note that the optimal p_2 from the perspective of period 1 is $\frac{1}{2} \neq p_2^*$ when he doesn't sell.

Naïve learner

A naïve learner thinks that his belief in period 2 will be unchanged, and as this is the only intertemporal link, his choices 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 $E[\pi_2 \mid z_1]$ in either case, but he doesn't realize in advance that $p_1 = z_1$. Consequently, his initial problem is either to

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

or to

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

In both cases, the 'present' objective is a concave quadratic centred on $1/2$ and the 'future' objective is a linear decreasing function, so the solution is that to the 'either' problem, given by $p_1^* = \frac{1}{2}(1 - \frac{1}{4}\delta[1/(1 - z_1) - z_1])$. Imposing the self-fulfilling action $p_1^* = z_1$ leads to

- $p_1^* = \frac{11 - \sqrt{37}}{14} \simeq 0.3512$

when $\delta = 1$. If he doesn't sell, then he updates his belief to $v \sim U[0, p_1^*]$, and if he does sell, then he updates his belief to $v \sim U[p_1^*, 1]$, leading to

- $p_2^* = p_1^*/2$ or $p_2^* = 1/2$

and an average overall profit when $\delta = 1$ of

- $E[\pi_1] + E[\pi_2] = (1 - p_1^*) p_1^* + (p_1^{*2}/4 + 1/4) \simeq 0.5087$.

Fully optimizing learner

A full optimizer knows all the implications of his period 1 choice. After a sale in period 1, the functional form of his expected period 2 profit, conditional on p_1 , depends on whether or not $p_1 < 1/2$. Consequently, his initial problem is either to

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

or to

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

In both cases, the 'present' objective is a concave quadratic centred on $1/2$ and the 'future' objective is strictly increasing at least on $[0, 2/3)$, so the solution is that to the 'or' problem,² given by

²For a full optimizer, the assumption that $p_1 < 1/2$ leads to a contradiction, whereas for a passive learner, it is the assumption that $p_1 \geq 1/2$ that does so. The reason for this is as follows. With regard to period 2 profit, a passive learner prefers the state where a sale was made in period 1 ($E[\pi_2 \mid z_1, 0] < E[\pi_2 \mid z_1, 1]$), thus he increases the probability of that state by lowering p_1 . A full optimizer has the same incentive to lower p_1 ($E[\pi_2 \mid p_1, 0] < E[\pi_2 \mid p_1, 1]$), but realizes that raising p_1 means a higher relative payoff after a sale ($dE[\pi_2 \mid p_1, 1]/dp_1 > 0$); this trade-off leads to his choice of p_1 being higher than that of a passive learner.

- $p_1^* = \frac{4}{7}$

when $\delta = 1$. If he doesn't sell, then he updates his belief to $v \sim U[0, 4/7]$, and if he does sell, then he updates his belief to $v \sim U[4/7, 1]$, leading to

- $p_2^* = \frac{2}{7}$ or $p_2^* = \frac{4}{7}$

and an average overall profit when $\delta = 1$ of

- $E[\pi_1] + E[\pi_2] = \frac{12}{49} + \frac{16}{49} = \frac{4}{7}$.

To summarize:

Type of seller	p_1^*	$E[\pi]$	Expected p_2	Actual p_2^*
<i>Non-learner</i>	0.5000	0.5000	0.5000	= 0.5000
<i>Myopic learner & Naïve learner</i>	0.5000	0.5625	0.5000	≠ 0.2500 or 0.5000
<i>Passive learner</i>	0.3512	0.5087	$z_1/2$ or 0.5000	= 0.1756 or 0.5000
<i>Full optimizer</i>	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.

Concluding remarks

Naïve learning provides a more useful benchmark than does passive learning because it can be used to isolate the effect of learning by experimentation from other intertemporal considerations in the cases of both signal dependence and signal independence – indeed, it is identical to the standard (myopic) benchmark in the latter case. Moreover, since passive learning can do worse than even myopic learning, it is not a good reference for measuring the gains from active learning.

Appendix

We formulate the problem in a (finite-horizon) dynamic programming framework.

Let θ denote the unknown parameter that affects the agent's payoff, and let $S = \langle \mu, s \rangle$ denote the state vector, where μ is the CDF representing the agent's belief about θ , and s represents the remaining state variables.

Let $X(s)$ denote the actions available to the agent, $x \in X(s)$ the agent's choice, y the outcome, and π the payoff. The outcome is a deterministic function $y = y(s, x; \theta)$ of the state and the action, given the unknown parameter, and the payoff π is a deterministic function $\pi = \pi(s, x, y; \theta)$ of the state, the action and the outcome, given the unknown parameter.

The state transition rule comprises $\mu \mapsto \Lambda_\mu(\mu, s, x, y; \theta)$, revising the belief of a rational, learning, agent using Bayes' rule, and $s \mapsto \Lambda_s(s, x, y; \theta)$, updating everything else for all types of agent – they will all face the changed environment.

The two-period problem

Consider a two-period set-up with initial state $S_1 = \langle \mu_1, s_1 \rangle$. The agent's problem in period 1 is to choose $x_1 \in X(s_1)$ to maximize

$$E[\pi(x_1) + \delta V_2(S_2) \mid S_1]$$

where $0 < \delta \leq 1$ is the discount factor, and

$$V_2(S_2) = \max_{x_2 \in X(s_2)} E[\pi(x_2) \mid S_2]$$

is the agent's maximum expected period 2 payoff from the perspective of period 1. (We have suppressed the dependence of the payoff on variables other than the agent's choice.) The agent's actual problem in period 2 is then to choose $x_2 \in X(s_2)$ to maximize $E[\pi(x_2) \mid S_2]$ where $s_2 = \Lambda_s(s_1, x_1)$, but μ_2 depends on what, if anything, the agent learns from the period 1 outcome. (Again, we have suppressed the dependence of the state transition rule on variables other than the current state and the agent's choice.) We consider five types of agent: a non-learner, a myopic learner, a naïve learner, a passive learner, and a fully optimizing learner.

Clearly, it is not possible to fully characterize optimal actions and their associated expected overall payoffs without making the various parameters and functions used in the model more explicit. However, we can describe precisely how the problem is solved by each of the five types of agent.

Non-learner

In period 1, a non-learner thinks that $S_2 = \langle \mu_1, s_1 \rangle$; in period 2, his belief is unchanged, but he is constrained by the fact that the remaining state variables might have.

Formally, he uses $V_2(\langle \mu_1, s_1 \rangle)$ in period 1 (so his problem is time-separable), but then in period 2 he chooses $x_2 \in X(s_2)$ to maximize $E[\pi(x_2) \mid \langle \mu_1, \Lambda_s(s_1, x_1) \rangle]$.

Myopic learner

In period 1, a myopic learner also thinks that $S_2 = \langle \mu_1, s_1 \rangle$, but, before period 2, he updates his belief (as well as the other state variables).

He also uses $V_2(\langle \mu_1, s_1 \rangle)$ in period 1, but then in period 2 he chooses $x_2 \in X(s_2)$ to maximize $E[\pi(x_2) \mid \langle \Lambda_\mu(\mu_1, s_1, x_1), \Lambda_s(s_1, x_1) \rangle]$. Consequently, his action in the first period is the same as that of a non-learner (his problem is also time-separable), but thereafter their actions differ.

Naïve learner

In period 1, a naïve learner thinks that $S_2 = \langle \mu_1, \Lambda_s(s_1, x_1) \rangle$ – he doesn't foresee that his belief will change, but understands that the remaining state variables might – but, before period 2, he also updates his belief.

He uses $V_2(\langle \mu_1, \Lambda_s(s_1, x_1) \rangle)$ in period 1, and later chooses $x_2 \in X(s_2)$ to maximize $E[\pi(x_2) \mid \langle \Lambda_\mu(\mu_1, s_1, x_1), \Lambda_s(s_1, x_1) \rangle]$ like other types of learner.

Clearly, if there are no state variables other than the belief μ , then the actions of a naïve learner mimic those of a myopic learner. Also, a naïve learner is time-inconsistent as is a myopic learner, and like a non-learner unless $s_2 = s_1$.

Passive learner

In period 1, a passive learner thinks that $S_2 = \langle \Lambda_\mu(\mu_1, s_1, z_1), \Lambda_s(s_1, x_1) \rangle$ for some unspecified action z_1 , chooses his first period action $x_1(z_1)$ accordingly, and then imposes the self-fulfilling action $x_1(z_1) = z_1$ before updating all the state variables – he conjectures that his belief will change but doesn't realise that he can directly influence this change; he takes an optimal current action that bears out his conjecture.

He uses $V_2(\langle \Lambda_\mu(\mu_1, s_1, z_1), \Lambda_s(s_1, x_1) \rangle)$ in period 1, then sets $x_1 = z_1$ and chooses $x_2 \in X(s_2)$ to maximize $E[\pi(x_2) \mid \langle \Lambda_\mu(\mu_1, s_1, x_1), \Lambda_s(s_1, x_1) \rangle]$.

Fully optimizing learner

In period 1, a fully optimizing learner thinks that $S_2 = \langle \Lambda_\mu(\mu_1, s_1, x_1), \Lambda_s(s_1, x_1) \rangle$ – he correctly foresees how both his belief and the remaining state variables will change.

He uses $V_2(\langle \Lambda_\mu(\mu_1, s_1, x_1), \Lambda_s(s_1, x_1) \rangle)$ in period 1, then indeed chooses $x_2 \in X(s_2)$ to maximize $E[\pi(x_2) \mid \langle \Lambda_\mu(\mu_1, s_1, x_1), \Lambda_s(s_1, x_1) \rangle]$.

To summarize:

Type of agent	Period 1 beliefs about S_2	Actual S_2
<i>Non-learner</i>	$\langle \mu_1, s_1 \rangle$	$\langle \mu_1, \Lambda_s(s_1, x_1) \rangle$
<i>Myopic learner</i>	$\langle \mu_1, s_1 \rangle$	$\langle \Lambda_\mu(\mu_1, s_1, x_1), \Lambda_s(s_1, x_1) \rangle$
<i>Naïve learner</i>	$\langle \mu_1, \Lambda_s(s_1, x_1) \rangle$	$\langle \Lambda_\mu(\mu_1, s_1, x_1), \Lambda_s(s_1, x_1) \rangle$
<i>Passive learner</i>	$\langle \Lambda_\mu(\mu_1, s_1, z_1), \Lambda_s(s_1, x_1) \rangle$	$\langle \Lambda_\mu(\mu_1, s_1, x_1), \Lambda_s(s_1, x_1) \rangle$
<i>Full optimizer</i>	$\langle \Lambda_\mu(\mu_1, s_1, x_1), \Lambda_s(s_1, x_1) \rangle$	$\langle \Lambda_\mu(\mu_1, s_1, x_1), \Lambda_s(s_1, x_1) \rangle$

The N -period problem

Define

$$V_N(S_N) = \max_{x_N \in X(s_N)} E[\pi(x_N) \mid S_N]$$

and

$$V_n(S_n) = \max_{x_n \in X(s_n)} E[\pi(x_n) + \delta V_{n+1}(S_{n+1}) \mid S_n] \quad \text{for } n = 2, \dots, N-1.$$

Again, the agent's problem in period 1 is to choose $x_1 \in X(s_1)$ to maximize

$$E[\pi(x_1) + \delta V_2(S_2) \mid S_1]$$

given the initial state $S_1 = \langle \mu_1, s_1 \rangle$. He then updates the initial state and solves the problem with a horizon of $N-1$ periods.

A *non-learner* uses $S_{n+1} = \langle \mu_1, s_1 \rangle$, for all $n = 1, \dots, N-1$; consequently, he solves the problem forwards, with no concern for the future. He then updates the initial state to $\langle \mu_1, \Lambda_s(s_1, x_1) \rangle$, and proceeds recursively.

A *myopic learner* also uses $S_{n+1} = \langle \mu_1, s_1 \rangle$, for all $n = 1, \dots, N-1$; he too solves the problem forwards, just as the non-learner. He then updates the initial state to $\langle \Lambda_\mu(\mu_1, s_1, x_1), \Lambda_s(s_1, x_1) \rangle$, and proceeds recursively.

A *naïve learner* uses $S_{n+1} = \langle \mu_1, \Lambda_s(s_n, x_n) \rangle$, for all $n = 1, \dots, N-1$; he solves the problem backwards, first calculating the optimal x_N as a function of S_N , then using $V_N(\langle \mu_1, \Lambda_s(s_{N-1}, x_{N-1}) \rangle)$ in the period $N-1$ problem and so on. He then updates the initial state to $\langle \Lambda_\mu(\mu_1, s_1, x_1), \Lambda_s(s_1, x_1) \rangle$, and proceeds recursively.

A *passive learner* uses $S_{n+1} = \langle \Lambda_\mu(\mu_n, s_n, z_n), \Lambda_s(s_n, x_n) \rangle$, for all $n = 1, \dots, N-1$; he also solves the problem backwards, first calculating the optimal x_N as a function of S_N , then using $V_N(\langle \Lambda_\mu(\mu_{N-1}, s_{N-1}, z_{N-1}), \Lambda_s(s_{N-1}, x_{N-1}) \rangle)$ in the period $N-1$ problem; after calculating the optimal x_{N-1} as a function of S_{N-1} and z_{N-1} he imposes $x_{N-1} = z_{N-1}$ before moving on to the period $N-2$ problem; and so on. He then updates the initial state to $\langle \Lambda_\mu(\mu_1, s_1, x_1), \Lambda_s(s_1, x_1) \rangle$, and proceeds recursively.

A *full optimizer* uses $S_{n+1} = \langle \Lambda_\mu(\mu_n, s_n, x_n), \Lambda_s(s_n, x_n) \rangle$, for all $n = 1, \dots, N-1$; he too solves the problem backwards, first calculating the optimal x_N as a function of S_N , then using $V_N(\langle \Lambda_\mu(\mu_{N-1}, s_{N-1}, x_{N-1}), \Lambda_s(s_{N-1}, x_{N-1}) \rangle)$ in the period $N-1$ problem and so on. He then updates the initial state to $\langle \Lambda_\mu(\mu_1, s_1, x_1), \Lambda_s(s_1, x_1) \rangle$, and proceeds recursively.

References

- DATTA, M., MIRMAN, L. and SCHLEE, E. (2000): "Learning with Noiseless Information & Payoff-Relevant Signals," *Economic Theory*, **16**, 63–75.
- DATTA, M., MIRMAN, L. and SCHLEE, E. (2002): "Optimal Experimentation in Signal-Dependent Decision Problems," *International Economic Review*, **43**, 577–607.
- FREIXAS, X. (1981): "Optimal Growth with Experimentation," *Journal of Economic Theory*, **24**, 296–309.
- LAZEAR, E. (1986): "Retail Pricing & Clearance Sales," *American Economic Review*, **76**, 14–32.
- TREFLER, D. (1993): "The Ignorant Monopolist: Optimal Learning with Endogenous Information," *International Economic Review*, **34**, 565–581.