

Uncertainty, Information Acquisition, and Technology Adoption

Canan Ulu

McCombs School of Business, University of Texas at Austin, Austin, Texas 78712,
canan.ulu@mcombs.utexas.edu

James E. Smith

Fuqua School of Business, Duke University, Durham, North Carolina 27708, jes9@duke.edu

Consumers or firms contemplating purchasing a new product or adopting a new technology are often plagued by uncertainty: Will the benefits outweigh the costs? Should we buy now or wait and gather more information? In this paper, we study a dynamic programming model of this technology adoption problem. In each period, the consumer decides whether to adopt the technology, reject it, or wait and gather additional information by observing a signal about the technology's benefit. The technology's actual benefit may be constant or changing stochastically over time. The dynamic programming state variable is a probability distribution that describes the consumer's beliefs about the benefits of the technology. We allow general probability distributions on benefits and general signal processes and assume that the consumer updates her beliefs over time using Bayes' rule. We are interested in structural properties of this model. We show that improving the technology's benefit need not make the consumer better off and that first-order stochastic dominance improvements in the consumer's distribution on benefits need not increase the consumer's value function. Nevertheless, the model possesses a great deal of structure. For example, we obtain monotonic value functions and policies if we order distributions using likelihood-ratio dominance rather than first-order stochastic dominance. We also examine convexity properties and provide many comparative statics results.

Subject classifications: dynamic programming; decision analysis: sequential.

Area of review: Decision Analysis.

History: Received January 2007; revision received August 2007; accepted December 2007. Published online in *Articles in Advance* March 11, 2009.

1. Introduction

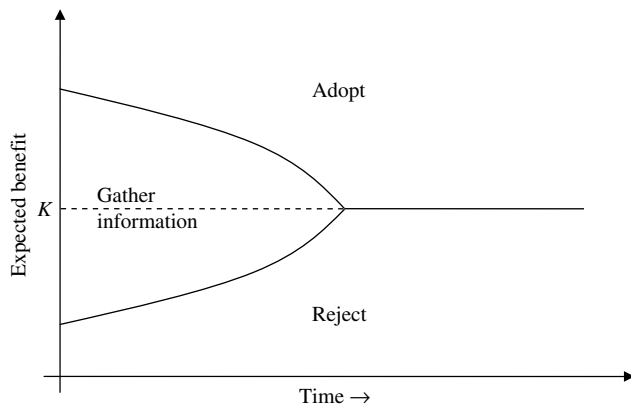
As consumers, we have all grappled with uncertainty when contemplating buying a new product or technology: Should I buy an iPhone now? Will its benefits outweigh the costs? Should I wait and see how others like them and/or wait for a new improved version? Uncertainty also complicates firms' technology adoption decisions, although the stakes may be much larger: Should an electric utility build a coal-fired power plant using the latest technology? Or should it wait until its costs and benefits are better understood and/or the technology improves? In medicine, regulators must decide whether to approve a new drug whose benefits are unknown or wait and request additional clinical trials. Producers of innovative technologies face the opposite side of the consumer's dilemma: The adoption and sales of these products will depend on how consumers' uncertainty about the technology's benefits resolves over time.

In this paper, we study the impact of uncertainty about the benefits of the technology on adoption and information-gathering decisions. More specifically, we formulate a dynamic programming model where, in each period, the consumer adopts or rejects a new technology or waits and gathers additional information about the benefits of technology by observing a signal about the technology's benefit.

The technology's actual benefit may be constant or changing stochastically over time. After observing a signal, the consumer updates her distribution on benefits using Bayes' rule. In this model, the dynamic programming state variable is a probability distribution that describes the consumer's beliefs about the benefits of the technology. We allow general probability distributions on benefits and general signal processes. We are interested in structural properties of the model: When does one distribution on benefits lead to higher values than another? Similarly, when is one signal process better than another? How do changes in the assumptions affect consumers, optimal policies, and the timing of adoption decisions?

Our basic model is similar to the model developed by McCardle (1985). Like us, McCardle considers a dynamic programming model where, in each period, the consumer adopts or rejects a new technology or waits and gathers additional information. However, unlike our model, McCardle's model assumes that the technology is not changing over time, and, rather than allowing general probability distributions and signal processes, it assumes that the consumer's probability distribution on benefits can be summarized by the expected benefit of the technology. In this univariate setting, the consumer's value function is increasing in the expected benefit of the technology,

Figure 1. Policy regions in McCardle’s (1985) model.



and the optimal policy in each period can be characterized by two thresholds: If the expected benefit is above the upper threshold, it is optimal to adopt the technology. If the expected benefit is below the lower threshold, it is optimal to reject. Between the two thresholds, it is optimal to gather additional information. As illustrated in Figure 1, as the consumer gathers additional information over time and the precision of the benefit estimate increases, the upper threshold decreases, the lower threshold increases, and the two thresholds converge to the cost of adoption, represented by K in the figure. The consumer will thus eventually stop gathering information and either adopt or reject the technology. Lippman and McCardle (1987) study the same model and show that the consumer’s value function increases with increases in the uncertainty (in the sense of Rothschild and Stiglitz 1970) about the benefit of the technology.

Allowing general probability distributions and signal processes as well as evolving technologies raises many interesting and sometimes subtle issues. With a probability distribution as a state variable, we can no longer characterize values or adoption policies by simply considering the technology’s expected benefit. In the univariate setting that McCardle considers, an increase in the expected benefit leads to a first-order-stochastic-dominance (FOSD) improvement in the consumer’s distribution on technology benefits. Thus, in the general setting, one might expect a FOSD improvement in the consumer’s probability distribution to lead to an increase in the value function and make adoption more attractive. Similarly, one might expect the result about increases in uncertainty being good for the consumer to generalize. However, despite their intuitive appeal, these conjectures are false, even with a stationary technology.

Although these intuitive results fail to hold, the general model still possesses a great deal of structure and has many nice properties. To simplify the discussion and disentangle the effects of allowing general distributions and information sources and allowing evolving technologies, we first focus on the case where the technology’s benefit is stationary. We begin in §2 by defining the model and describing

some examples. In §3, we study monotonicity properties of this model and show that if the signal-generating process satisfies a monotone likelihood ratio assumption and we use likelihood-ratio (LR) dominance to order distributions, we get natural monotonicity results: A LR improvement in the consumer’s distribution on benefits leads to an increase in the value function and encourages adoption. In §4, we show that, under reasonable conditions, the consumer will eventually decide to adopt or reject the technology.

In §5, we define and study the “derived benefit function” that describes how much benefit the consumer actually receives (net of the information-gathering costs and delays) as a function of the true benefit of the technology. This analysis clarifies why FOSD improvements and increases in uncertainty in the distributions are not necessarily good for the consumer; this derived benefit function need not be increasing or convex. For example, improving a “bad” technology may increase the consumer’s information-gathering costs, increase the probability of adopting this bad technology, and actually make the consumer worse off. In §§6 and 7, we study convexity properties of the value function and optimal policies and the impacts of cheaper and better information, as well as reduced adoption costs.

In §8, we generalize the model to allow the technology to change stochastically over time. Given our earlier results, this generalization is fairly straightforward, and most of the results and intuitions from the stationary case carry over directly to this more general case. However, to guarantee monotonic optimal policies, we must assume that the technology’s expected future improvement is not increasing in the level of the benefit; otherwise, improving the *current* technology may lead the consumer to switch from adopting to waiting. Improving the prospects for *future* technologies naturally makes waiting more attractive.

Technology adoption and the diffusion of innovations have long been studied by researchers in a number of fields. Rogers and Rogers (2003) provide a thorough review of this literature, tracing it back to the 1950s. There are many papers that develop and analyze mathematical models of various aspects of technology adoption and the diffusion of innovations, but relatively few explicitly consider the information-gathering process. As discussed earlier, McCardle (1985) and Lippman and McCardle (1987) consider a model similar to ours, but with a stationary technology and restrictive assumptions on the form of the distributions and signal process. Jensen (1982) earlier studied a model similar to McCardle’s that also assumed that information gathering is costless. Chatterjee and Eliashberg (1990) model the diffusion process using a “micromodeling” approach that explicitly considers the impact of uncertainty on adoption decisions and aggregates these individual models to derive mathematical forms for diffusion curves. Their analysis assumes a stationary benefit and a specific probabilistic model (the normal-normal model discussed in §2.2) and further assumes that consumers are risk averse but myopic: They assume that consumers adopt the technology as soon

as their certainty equivalent is positive without considering the possibility of gathering additional information to reduce the uncertainty about the benefit. More recently, Kornish and Keeney (2008) study the problem of “adopting” a flu vaccine when there is uncertainty about the strain of the virus and thus uncertainty about the effectiveness of the vaccine.

With today’s rapidly evolving technological landscape, understanding technology adoption decisions and the diffusion of innovations remains an important practical problem. Given the inherent uncertainty in such decisions and abundant sources of information, it is particularly important to understand how the dynamic process of uncertainty resolution and environmental factors (e.g., the quality and cost of information) affect these decisions. The results and insights for the model studied in this paper may also provide a foundation for the analysis of other models of technology adoption; we discuss some possible extensions and variations of our model in §9.

We provide proofs for some results in the body of the paper. The other proofs are provided in an online appendix that can be found at <http://or.pubs.informs.org/>.

2. The Stationary Model and Some Examples

We begin by describing the stationary form of the model, and then discuss some specific examples.

2.1. The Model

A consumer is contemplating purchasing a new technology whose benefit is denoted by $\theta \in \Theta \subseteq \Re$; we can think of θ as representing the net present value of the stream of benefits to the consumer provided by the technology. The consumer is uncertain about the benefit of the technology, and her beliefs are described by a probability distribution. For ease of notation, we will assume that the consumer’s probability distribution is continuous and has a density π over Θ . For discrete spaces, we can interpret π as a probability mass function and consider sums instead of integrals; more general probability measures could also be considered. For technical reasons, we assume that θ is π -integrable in that $\int_{\Theta} |\theta| \pi(\theta) d\theta$ is finite.

Time is discrete. In each period, the consumer must choose whether to adopt the technology, reject it, or gather additional information. We will assume that the consumer is risk neutral and makes decisions on an expected value basis. If she decides to reject the technology, she receives nothing and no longer gathers information about the technology. If she decides to adopt the technology, she pays a fixed adoption cost K and receives a net expected benefit of $\int_{\Theta} \theta \pi(\theta) d\theta - K$.

If the consumer chooses to gather additional information, she pays c in that period and observes a signal $x \in X$, drawn with likelihood function $L(x | \theta)$. This signal x could be any kind of message about the technology, including,

for example, a numeric score or categorical rating (e.g., four out of five stars). Having observed signal x , the consumer then updates her prior π using Bayes’ rule and finds a posterior $\Pi(\theta; \pi, x)$ given by

$$\Pi(\theta; \pi, x) = \frac{L(x | \theta) \pi(\theta)}{f(x; \pi)},$$

where $f(x; \pi)$ is the predictive distribution for signals x , $f(x; \pi) = \int_{\Theta} L(x | \theta) \pi(\theta) d\theta$. The consumer then continues into the next period, starting with a new prior distribution that is equal to her posterior distribution from this stage. Our notation will assume that the signal distribution is continuous and $f(x; \pi)$ is a density function on X , but discrete or more general distributions could also be considered. Because our dynamic programming state variable is the distribution itself, we will frequently suppress the domain of the distribution and write the posterior as $\Pi(\pi, x)$ when we want to consider this distribution (on Θ) as a function of the prior π and observed signal x . Similarly, we will write $f(\pi)$ when considering the signal distribution (on X) as a function of the prior π .

The consumer’s optimal value function with k periods remaining, $v_k^*(\pi)$, can be written as a dynamic programming recursion:

$$v_0^*(\pi) = 0, \\ v_k^*(\pi) = \max \left\{ 0, \int_{\Theta} \theta \pi(\theta) d\theta - K, -c + \delta E[v_{k-1}^*(\Pi(\pi, \tilde{x}))] \right\},$$

where δ ($0 \leq \delta \leq 1$) is the discount factor and the expectation of the next period value function is taken over all possible random signals \tilde{x} . Using the signal distribution $f(x; \pi)$, this expectation can be written more explicitly as

$$E[v_{k-1}^*(\Pi(\pi, \tilde{x}))] = \int_x v_{k-1}^*(\Pi(\pi, x)) f(x; \pi) dx.$$

We let $v_{\infty}^*(\pi) = \lim_{k \rightarrow \infty} v_k^*(\pi)$ denote the infinite-horizon limit of the value function. This limit need not exist given our current assumptions. For example, if the consumer is paid for gathering information (i.e., $c < 0$) and there is no discounting ($\delta = 1$), then $v_{\infty}^*(\pi)$ is infinite. However, we will show in §4 below that if $c \geq 0$, then the finite-horizon value functions converge to a well-defined infinite-horizon limit.

2.2. Examples

McCardle (1985) focuses on a *beta-Bernoulli* model that assumes $\theta = Ap^*$ and the consumer’s uncertainty about p^* has a beta distribution $\text{Beta}(\alpha, \beta)$ with density $f(p^*) = (\Gamma(\alpha + \beta) / (\Gamma(\alpha)\Gamma(\beta))) (p^*)^{\alpha-1} (1 - p^*)^{\beta-1}$; this distribution has mean $\bar{p} = \alpha / (\alpha + \beta)$, and $\alpha + \beta$ can be interpreted as a measure of the precision of the mean \bar{p} as an estimate of p^* . The prior for $\theta = Ap^*$ is thus a scaled beta distribution. Signals are generated according to a

Bernoulli process with a probability p^* of a positive signal and $(1 - p^*)$ of a negative signal. After observing a positive signal, the consumer's posterior distribution for p^* becomes $\text{Beta}(\alpha + 1, \beta)$ with mean $(\alpha + 1)/(\alpha + \beta + 1)$. After observing a negative signal, the posterior becomes $\text{Beta}(\alpha, \beta + 1)$ with mean $\alpha/(\alpha + \beta + 1)$. With either signal, the posterior precision is equal to $\alpha + \beta + 1$. Thus, in this model, if we know the prior α and β and the number of periods that have passed, we can fully determine the then-current distribution from its mean \bar{p} ; this allows the model to be formulated as a univariate dynamic program taking \bar{p} as a state variable.¹ The *beta-binomial* model generalizes the beta-Bernoulli model to consider n (rather than one) independent positive or negative reports in each period. After seeing s successes in n trials, the posterior distribution becomes $\text{Beta}(\alpha + s, \beta + n - s)$. If the number of reports at each stage is fixed, the mean of the beta distribution can serve as a univariate state variable.

McCardle (1985) also describes a *normal-normal* model, which is the focus of Lippman and McCardle (1987) and Chatterjee and Eliashberg (1990). In this model, the prior on θ is assumed to be normally distributed with mean m and precision s (or variance $1/s$), and the signals are drawn from a normal distribution with mean θ and precision t . The posterior distribution after seeing signal x is then given by a normal distribution with mean $(sm + tx)/(s + t)$, and the precision $s + t$ is independent of the signal, as in the beta-binomial model. This allows the model to be formulated as a univariate dynamic program with the then-current mean as a state variable.

3. Monotonicity Properties

In this section, we study the impact of changes in the consumer's prior distribution π on the value function and optimal policies. Our first task is to identify a stochastic order that captures the notion of a "more optimistic" prior that allows us to obtain natural monotonicity results. As discussed in the introduction, an FOSD improvement in the prior is not sufficient to ensure an increase in the value function. We must instead use the LR order and assume that the likelihood functions for signals satisfy the monotone-likelihood-ratio (MLR) property. Our interest in the LR order and the MLR property was motivated by their use in economics in the study of agency models and auctions; see, e.g., Milgrom (1981). We begin by introducing the LR order and MLR assumption. We then show that, with the MLR assumption, the value function and policies satisfy natural monotonicity properties in terms of the LR order.

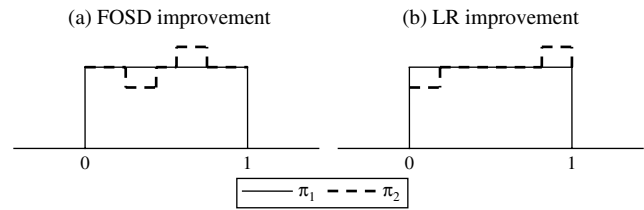
3.1. The Likelihood-Ratio Order

We will first introduce the LR order and then discuss the properties of this order that we will use in our analysis.

DEFINITION 3.1. π_2 LR dominates π_1 ($\pi_2 \succeq_{\text{LR}} \pi_1$) if for all $\theta_2 \geq \theta_1$,

$$\frac{\pi_2(\theta_2)}{\pi_1(\theta_2)} \geq \frac{\pi_2(\theta_1)}{\pi_1(\theta_1)}.$$

Figure 2. FOSD and LR improvements on a uniform $[0, 1]$ distribution.



For our purposes, there are three key properties of the LR order. The first key property is that LR dominance implies FOSD dominance. Recall the definition of the FOSD order: $\pi_2 \succeq_{\text{FOSD}} \pi_1$ if and only if for all increasing functions $\phi(\theta)$, $\int_{\theta} \phi(\theta)\pi_2(\theta) d\theta \geq \int_{\theta} \phi(\theta)\pi_1(\theta) d\theta$. Whitt (1979) showed that LR dominance implies FOSD dominance and, even stronger, that LR dominance is equivalent to requiring FOSD dominance for all conditional distributions of the form $\pi(\theta | \theta \in A)$ where $A \subseteq \Theta$; that is, $\pi_2 \succeq_{\text{LR}} \pi_1$ holds if and only if $\pi_2(\theta | \theta \in A) \succeq_{\text{FOSD}} \pi_1(\theta | \theta \in A)$ for all A with positive probability under both π_1 and π_2 .

Although LR dominance implies FOSD dominance, the converse is not true. Figures 2(a) and 2(b) show two FOSD improvements of a uniform distribution π_1 on $[0, 1]$. In Figure 2(a), we form π_2 by shifting mass from one interval to another higher interval. In this case, $\pi_2(\theta)/\pi_1(\theta)$ is non-monotonic and π_2 does not LR dominate π_1 . In Figure 2(b), we form π_2 by shifting mass from a left-tail interval to a right-tail interval. In this case, $\pi_2(\theta)/\pi_1(\theta)$ is monotonically increasing and $\pi_2 \succeq_{\text{LR}} \pi_1$. Thus, LR improvements, like FOSD improvements, can be constructed by transferring mass from lower to higher values of θ . However, LR improvements require the transfer of mass from a left-tail interval to a right-tail interval to ensure that $\pi_2(\theta)/\pi_1(\theta)$ is increasing.

The second key property of the LR order is that it survives Bayesian updating: If two priors are LR ordered, the corresponding posterior distributions are also LR ordered. This result is stated formally in the next proposition.

PROPOSITION 3.2. *Given any signal x , the posteriors are LR ordered if and only if the priors are LR ordered: $\pi_2 \succeq_{\text{LR}} \pi_1 \Leftrightarrow \Pi(\pi_2, x) \succeq_{\text{LR}} \Pi(\pi_1, x)$ for all $x \in X$.*

PROOF. From Definition 3.1, $\Pi(\pi_2, x) \succeq_{\text{LR}} \Pi(\pi_1, x)$ requires that for all $\theta_2 \geq \theta_1$,

$$\frac{\Pi(\theta_2; \pi_2, x)}{\Pi(\theta_2; \pi_1, x)} \geq \frac{\Pi(\theta_1; \pi_2, x)}{\Pi(\theta_1; \pi_1, x)}.$$

Writing out the posteriors using Bayes' rule and canceling common terms, we see that this is equivalent to $\pi_2 \succeq_{\text{LR}} \pi_1$. \square

This survival property is important for the recursive dynamic programming structure because we will need LR dominance at one stage to imply LR dominance in the next

stage. FOSD dominance is not preserved in this way: It is not difficult to construct examples where the priors are FOSD ordered and the posteriors are not ordered for some signals.

The final important property of the LR order is that, when we have totally ordered signals (e.g., four-star ratings are more favorable than three-star ratings) and “monotonic” likelihood functions, we get natural monotonic relationships among priors, signals, and posteriors. We formalize this monotonicity assumption as follows.

DEFINITION 3.3. The signal process has the MLR property if the signal space X is totally ordered and $L(x | \theta_2) \succeq_{LR} L(x | \theta_1)$ for all $\theta_2 \geq \theta_1$. In other words,

$$\frac{L(x_2 | \theta_2)}{L(x_2 | \theta_1)} \geq \frac{L(x_1 | \theta_2)}{L(x_1 | \theta_1)} \quad \text{for all } x_2 \geq x_1 \text{ and } \theta_2 \geq \theta_1.$$

The requirement that the signals be totally ordered is natural but does rule out some cases that are potentially of interest. For example, if we had a beta-binomial model with uncertainty about the number of reports observed, we may not be able to order all signals: Observing four out of five reports to be positive is not necessarily more favorable than observing seven out of ten reports to be positive. Although the monotonicity results of §§3.2 and 8 require this MLR assumption, most of the other results in the paper do not.

If the signal process satisfies the MLR property, then LR improvements in the prior lead to LR improvements in the signal distribution and more favorable signals lead to LR improvements in the posterior.

PROPOSITION 3.4. *If the signal process satisfies the MLR property, then*

- (i) $\pi_2 \succeq_{LR} \pi_1 \Rightarrow f(\pi_2) \succeq_{LR} f(\pi_1)$,
- (ii) *for any prior* π , $x_2 \geq x_1 \Leftrightarrow \Pi(\pi, x_2) \succeq_{LR} \Pi(\pi, x_1)$.

PROOF. These results are well known (see, e.g., Karlin 1968 and Whitt 1979, Theorem 4), but simple direct proofs are provided in Online Appendix A.1. \square

The LR ordering is quite natural in both the beta-binomial and normal-normal examples introduced in §2.2. In both models, an improvement in the expected benefit of the technology (\bar{p} or m) in any period leads to an LR improvement in the underlying distribution and both signal processes satisfy the MLR assumption.

3.2. Monotonicity of the Value Function and Policies

We now show that if the signal process satisfies the MLR property, the optimal value function and optimal policies are both monotonic in that LR improvements in the prior distribution π lead to higher values and move policies away from rejection and toward adoption.

To formalize the effect of changes in the prior on the value function, we say a function u defined on distributions on Θ is LR increasing if $\pi_2 \succeq_{LR} \pi_1$ implies $u(\pi_2) \geq u(\pi_1)$;

u is LR decreasing if $-u$ is LR increasing. A key step in proving that the value function is LR increasing is to show that Bayesian updating preserves the LR-increasing property; we show this in a lemma and then state the monotonicity result.

LEMMA 3.5. *Suppose that the signal process satisfies the MLR property and $\pi_2 \succeq_{LR} \pi_1$; let \tilde{x}_2 and \tilde{x}_1 denote the random signals corresponding to π_2 and π_1 . Then, for any LR-increasing function u ,*

$$E[u(\Pi(\pi_2, \tilde{x}_2))] \geq E[u(\Pi(\pi_1, \tilde{x}_1))].$$

PROOF. First, recall that Proposition 3.4 implies that the signal distributions $f(\pi_2)$ and $f(\pi_1)$ satisfy $f(\pi_2) \succeq_{LR} f(\pi_1)$. Then,

$$\begin{aligned} E[u(\Pi(\pi_2, \tilde{x}_2))] &= \int_x u(\Pi(\pi_2, x))f(x; \pi_2) dx \\ &\geq \int_x u(\Pi(\pi_1, x))f(x; \pi_2) dx \\ &\geq \int_x u(\Pi(\pi_1, x))f(x; \pi_1) dx \\ &= E[u(\Pi(\pi_1, \tilde{x}_1))]. \end{aligned}$$

The first inequality follows because u is LR increasing and, for each signal x , $\pi_2 \succeq_{LR} \pi_1$ implies $\Pi(\pi_2, x) \succeq_{LR} \Pi(\pi_1, x)$ by Proposition 3.2. The second inequality follows because the LR order on the signal distributions implies FOSD dominance and $u(\Pi(\pi_1, x))$ is an increasing function of x ; the fact that $u(\Pi(\pi_1, x))$ is an increasing function of x follows from Proposition 3.4(ii) because u is LR increasing. \square

PROPOSITION 3.6. *If the signal process satisfies the MLR property, then, for all k , the value function $v_k^*(\pi)$ is LR increasing. In other words, $\pi_2 \succeq_{LR} \pi_1$ implies $v_k^*(\pi_2) \geq v_k^*(\pi_1)$.*

PROOF. We show this by induction. The terminal value function, $v_0^*(\pi) = 0$, is trivially LR increasing. Now suppose that $v_{k-1}^*(\pi)$ is LR increasing. By the previous lemma, the value if the consumer waits, $-c + \delta E[v_{k-1}^*(\Pi(\pi, \tilde{x}))]$, is LR increasing. The rewards if the consumer adopts ($\int_{\theta} \theta \pi(\theta) d\theta - K$) or rejects (0) are also LR increasing. Then $v_k^*(\pi)$, as the maximum of three LR increasing functions, is also LR increasing. \square

We next consider how the optimal policy responds to changes in the consumer’s distribution of benefits. First, recall that, in the univariate setting, McCardle (1985) showed that the policy in each period can be characterized by adoption and rejection thresholds. Intuitively, if LR dominance correctly captures the notion of a “more optimistic” distribution of benefits, we might expect optimal policies to have analogous structures in the more general setting. We will show that this is indeed the case. We state and prove this threshold result after first establishing a lemma that places a bound on the rate of increase of the value function.

LEMMA 3.7. *Suppose that the signal process satisfies the MLR property. Then, $v_k^*(\pi) - \int_{\theta} \theta \pi(\theta) d\theta$ is LR decreasing.*

PROOF. Note that with Bayesian updating the expected posterior mean is equal to the prior mean $\int_{\theta} \theta \pi(\theta) d\theta = E[\int_{\theta} \theta \Pi(\theta; \pi, \tilde{x}) d\theta]$. Using this, we can write $g_k(\pi) = v_k^*(\pi) - \int_{\theta} \theta \pi(\theta) d\theta$ as a dynamic programming recursion; $g_0(\pi) = -\int_{\theta} \theta \pi(\theta) d\theta$, and, for $k > 0$,

$$g_k(\pi) = \max \left\{ -\int_{\theta} \theta \pi(\theta) d\theta, -K, \right. \\ \left. -c - (1 - \delta) \int_{\theta} \theta \pi(\theta) d\theta + \delta E[g_{k-1}(\Pi(\pi, \tilde{x}))] \right\}.$$

We can then show that $g_k(\pi)$ is LR decreasing using the same recursive argument as in the proof of Proposition 3.6 above: The terminal value function, $g_0(\pi)$, is LR decreasing. Now suppose that $g_{k-1}(\pi)$ is LR decreasing. Lemma 3.5 implies that the continuation value, $E[g_{k-1}(\Pi(\pi, \tilde{x}))]$, is LR decreasing. Thus, the functions associated with each choice in $g_k(\pi)$ are all LR decreasing. Then, $g_k(\pi)$, as the maximum of three LR decreasing functions, is LR decreasing. \square

Because $\int_{\theta} \theta \pi(\theta) d\theta - K$ is the expected value of adopting the technology, we can interpret this lemma as saying that, if we improve the underlying distribution, the value of adopting increases at least as much as the value of waiting or rejecting. This result is critical to establishing monotonicity of the optimal policies.

PROPOSITION 3.8. *Suppose that the signal process satisfies the MLR property and $\pi_2 \succeq_{\text{LR}} \pi_1$:*

- (i) *If it is optimal to adopt with prior π_1 , then it is also optimal to adopt with π_2 .*
- (ii) *Similarly, if it is optimal to reject with π_2 , then it is also optimal to reject with π_1 .*

PROOF. Assume that $\pi_2 \succeq_{\text{LR}} \pi_1$. (i) Let $g_k(\pi) = v_k^*(\pi) - (\int_{\theta} \theta \pi(\theta) d\theta - K)$ be the difference between the value function and the value given by immediate adoption; this differs by a constant from the function studied in the previous lemma. If it is optimal to adopt the technology given distribution π_1 , then $g_k(\pi_1) = 0$. The previous lemma implies that g is LR decreasing. Thus, $g_k(\pi_2) \leq 0$, which implies that it is optimal to adopt given π_2 .

(ii) If it is optimal to reject the technology given distribution π_2 , then $v_k^*(\pi_2) = 0$. v_k^* is LR increasing (by Proposition 3.6), and thus $v_k^*(\pi_1) \leq 0$. Therefore, it is optimal to reject given π_1 . \square

Thus we have established the monotonic structure of the optimal policy and a general version of the two-threshold result: Along any chain of LR-improving distributions, the optimal action moves from rejection toward adoption, perhaps passing through the information-gathering region.

One natural way to construct a chain of LR-improving distributions is by considering the observed signals: Given

the MLR property, by Proposition 3.4(ii), more favorable signals lead to LR-improved posteriors. Combining this result with Proposition 3.8, this implies that in each period there exist threshold signals x_k^a and x_k^r (with $x_k^a \geq x_k^r$) such that it is optimal to adopt in the next period if the observed signal is greater than x_k^a , optimal to reject if the observed signal is less than x_k^r , and optimal to gather additional information if the signal lies between these two thresholds. Note that these signal thresholds, like the posterior distributions, are “path dependent” in that they will depend on the original prior distribution π and the entire history of signals observed. Because increasing any early signal leads to an LR improvement in the posterior distribution, an increase in any observed signal will lead all subsequent signal thresholds to decrease (weakly).

Given the monotonic relationship between signals and posterior means (this follows from Proposition 3.4(ii)), we can also describe these adoption and rejection thresholds in terms of posterior means: At each stage, there exist thresholds θ_k^a and θ_k^r ($\theta_k^a \geq \theta_k^r$) such that it is optimal to adopt in the next period if the posterior mean is greater than θ_k^a , optimal to reject if the posterior mean is less than θ_k^r , and optimal to gather additional information if the posterior mean lies between these two thresholds. These thresholds must also bracket the cost of adoption K ($\theta_k^a \geq K \geq \theta_k^r$) because rejection would be preferred to adoption if the posterior mean were less than K ; similarly, adoption would be preferred to rejection if the posterior mean were greater than K . Thus, the policy regions are analogous to those in McCardle’s univariate model, as illustrated in Figure 1. However, here, unlike the univariate model, the posterior-mean thresholds may depend on the original prior distribution π and the entire history of signals observed.

An LR improvement in the consumer’s prior distribution on benefits not only makes the consumer better off (according to her own expectations), it also makes the producer better off: Even if we hold the set of observed signals constant, by Proposition 3.2(i), an LR improvement in the prior leads to an LR improvement in the posteriors and thus moves the consumer closer to the adoption region. Thus, actions that lead to consumers having an LR-improved prior will increase the probability of adoption and lead to earlier adoptions. More favorable signals are also good for both consumer and producer because they lead to increased values and earlier adoption.

4. Convergence and Decisiveness

The finite-horizon version of the model assumes that the consumer has a limited number of periods to make the decision to adopt or reject the technology. Increasing the number of periods remaining allows the consumer to gather additional information, if desired, and thus makes waiting weakly more attractive and the consumer weakly better off. Formally, it is easy to show that $v_k^*(\pi)$ is nondecreasing in k and, if it is optimal to wait given prior π and

k periods remaining, it is also optimal to wait with prior π and $k + 1$ periods remaining.

However, there is a limit to the benefits of gathering more information: As the number of signals observed increases, the consumer's estimate of the benefit of the technology will almost certainly converge and additional information will have diminishing value.² If information or delay is costly (i.e., $c > 0$ or $\delta < 1$), the cost of gathering information will eventually outweigh its benefit and it will be optimal to stop gathering information. Of course, if information is free and delay costless ($c = 0$ and $\delta = 1$), the consumer might as well gather information forever; however, even in this scenario, additional information has diminishing value and the finite-horizon value functions will converge to an infinite-horizon limit. We summarize these convergence results as follows.

PROPOSITION 4.1. (i) *If information gathering is costly ($c > 0$), then the consumer will almost certainly stop gathering information at some point.*

(ii) *If delay is costly ($\delta < 1$), information is free ($c = 0$), and it is truly optimal to adopt (i.e., the limiting expected value $E[\theta]$ is greater than the adoption cost K), then the consumer will almost certainly adopt the technology.*

(iii) *If the cost of information gathering is nonnegative ($c \geq 0$), then the finite-horizon value functions converge for all π : $\lim_{k \rightarrow \infty} v_k^*(\pi) = v_\infty^*(\pi)$.*

PROOF. See Online Appendix A.2. \square

These convergence results do not assume that the consumer will necessarily learn the true value of θ if she gathered information forever. For example, if the likelihood function assigns the same signal distribution (e.g., “four stars” with probability one) to all technologies with values of θ above some threshold, then the posterior distributions will never distinguish among technologies above this threshold. Nevertheless, the estimated expected values ($E[\theta]$) will converge without any assumptions about the likelihood function.

The producer of a new technology can perhaps take some solace in knowing that, if information is costly to gather, consumers will eventually decide whether to adopt or reject the technology. Moreover, if information is free but delay is costly, the consumer will eventually adopt technologies that are “good for her.” Typically, these adoption decisions (and rejection decisions if information is costly) will be made with some uncertainty about the benefit of the technology remaining and there being some risk of adopting a bad technology (with $\theta < K$) or rejecting a good one. The information gathering costs, the discount factor, and the quality of the information all affect the timing of these decisions and determine how much risk the consumer will take in these adoption and rejection decisions; we discuss these effects further in §7.

5. The Derived Benefit Function

In this section, we study how the consumer's realized benefit (net of information-gathering costs and discounting)

varies with the true benefit of the technology. This characterization will give us some insight into why FOSD improvements and increases in uncertainty in the distribution on benefits do not necessarily lead to increases in the value function.

An adoption/rejection policy for a consumer can be defined by identifying the sets of signals that lead to adoption, rejection, or continuation in each period; these sets will generally depend on earlier signals. For any such policy P (not necessarily an optimal policy), let $v_k(\pi; P)$ denote the value given by following this policy. Clearly, $v_k^*(\pi) = v_k(\pi; P)$ if P is an optimal policy for prior π . As discussed in §3.2, if the signal process satisfies the MLR property, the optimal signal policies are *monotonic* in that they can be characterized in terms of adoption and rejection signal thresholds that are decreasing (weakly) in all observed signals. The following proposition shows that $v_k(\pi; P)$ can be represented as the expected value of a *derived benefit function* $b_k(\theta)$ that describes the net benefit realized by the consumer for a technology with true benefit θ .

PROPOSITION 5.1. *For any k and policy P , there exists a derived benefit function $b_k(\theta)$ such that*

$$v_k(\pi; P) = \int_{\theta} b_k(\theta) \pi(\theta) d\theta.$$

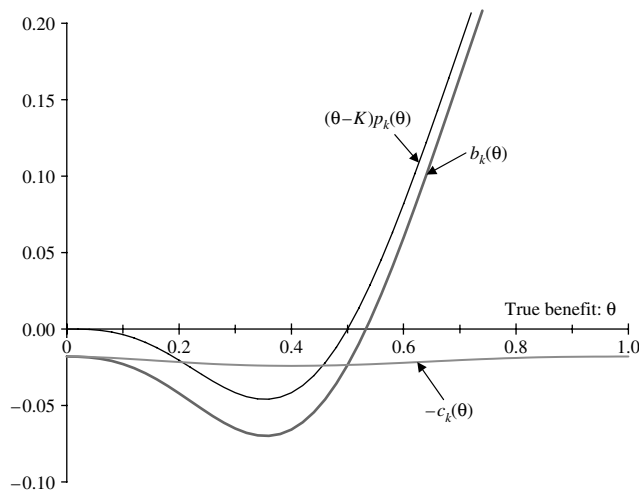
Moreover, $b_k(\theta)$ can be decomposed as $b_k(\theta) = (\theta - K) \cdot p_k(\theta) - c_k(\theta)$, where $p_k(\theta)$ is the discounted probability of adoption and $c_k(\theta)$ is the expected information-gathering costs in present value terms; $p_k(\theta)$ satisfies $0 \leq p_k(\theta) \leq 1$, and $c_k(\theta)$ is nonnegative if $c \geq 0$. If the signal process satisfies the MLR property and P is a monotonic policy, then $p_k(\theta)$ is increasing in θ .

PROOF. See Online Appendix A.3. \square

If the policy P calls for adopting immediately, the discounted probability of adoption is one and the expected information gathering costs are zero. If P calls for immediate rejection, the discounted probability of adoption is zero and the information costs are also zero. Adoptions in later periods are discounted using the discount factor δ so that $(\theta - K)p_k(\theta)$ represents the expected present value of adopting a technology with true benefit θ . The functions $b_k(\theta)$, $p_k(\theta)$, and $c_k(\theta)$ are constructed recursively by considering the probability of adoption and rejection in each period. Given a fixed policy P and the true benefit θ , these adoption and rejection probabilities are determined entirely by the likelihood function $L(x | \theta)$. Thus, $b_k(\theta)$, $p_k(\theta)$, and $c_k(\theta)$ all depend on the policy P but are independent of the prior π .

Figure 3 shows the derived benefit function, decomposed into its two components, for a beta-binomial example. Here we assume that there are $k = 3$ periods to go, and in each period the consumer observes $n = 5$ positive or negative reports. The probability of seeing a positive in each report is θ . The cost of gathering information (c) is 0.018 in

Figure 3. An example derived benefit function $b_k(\theta)$.



each period; the discount factor (δ) is 0.98; the adoption cost (K) is 0.5; and the scaling factor A is 1.0. The policy P assumed here is the optimal policy for a prior π that is uniform on $[0, 1]$. In this case, it is optimal for the consumer to gather information in the first period. If zero or one positives are observed, it is optimal to reject in the next period; if two positives are observed, it is optimal to gather additional information; otherwise it is optimal to adopt.

The form of the derived benefit function in Figure 3 is typical. For values of θ near zero, it is very likely that the consumer will observe zero or one positives and, following the specified policy, will reject in the next period. In these cases, the probability of adoption $p_k(\theta)$ is near zero and the expected information-gathering costs $c_k(\theta)$ are close to the cost of gathering information in the first period. For values of θ near one, the consumer is very likely to observe three or more positives and adopt in the next period. In these cases, the probability of adoption is near one and the expected information-gathering costs are equal to the cost of gathering information in the first period.

As we move from low values of θ to high values, the probability of adopting the technology increases, but this is not necessarily a good thing because adoptions for values of θ that are less than the adoption cost K are not economic. Combining adoption and information costs, the derived benefit function initially decreases as the expected information costs increase and the expected value of adoption, $(\theta - K)p_k(\theta)$, decreases. Thus, in this range, better technologies actually make the consumer worse off! However, for larger values of θ , the derived benefit is again increasing.

This nonmonotonicity in the derived benefit function explains why FOSD improvements in the prior π need not be beneficial to the consumer. If we shift mass from lower values of θ in regions where $b_k(\theta)$ is decreasing (e.g., in the example, shifting mass from the regions near 0.1 to the region near 0.3), the value with this policy will decrease. Of course, with large enough changes in the probability

distribution, the policy that was optimal for the original distribution may no longer be optimal. However, in this case and others, we can generate numerical examples that consider optimal policies and are such that $\pi_2 \succeq_{\text{FOSD}} \pi_1$ and yet $v_k^*(\pi_2) < v_k^*(\pi_1)$. In contrast, as discussed in §3.1, an LR improvement in the prior shifts mass from a left-tail region to a right-tail region. In this case, we can be sure that we are transferring mass from “bad” to “good” values of θ and the consumer will certainly be better off.

The form of the derived benefit function also shows why increases in uncertainty need not be good for the consumer. First, recall that a distribution π_2 is an increase in uncertainty over π_1 if and only if for all convex functions $\phi(\theta)$, $\int_{\theta} \phi(\theta)\pi_2(\theta) d\theta \geq \int_{\theta} \phi(\theta)\pi_1(\theta) d\theta$. Equivalently, π_2 is an increase in uncertainty over π_1 if and only if π_2 is a “mean-preserving spread” of π_1 ; see Rothschild and Stiglitz (1970). The derived benefit functions are generally not convex, and thus increases in uncertainty need not lead to increases in the value function. For example, in Figure 3, the left side of the “valley” of $b_k(\theta)$ is visibly nonconvex. A mean-preserving spread of π that is constructed by moving mass from near $\theta = 0.2$ and splitting it between regions near $\theta = 0.0$ and $\theta = 0.4$ will lead to a decrease in the consumer’s value function.

However, many increases in uncertainty will make the consumer better off. For example, a mean-preserving spread that moves mass from the central valley in Figure 3 to the left and right tails of the distribution will make the consumer better off. In these cases, the increase in uncertainty about θ is good for the consumer because the low or high values of θ are more easily recognized as “bad” or “good” technologies. Because a mean-preserving spread in the prior does not affect the expected rewards associated with adopting or rejecting, if a mean-preserving spread leads to an increase in the value function, it must be because the value of waiting has increased; thus, such an increase in uncertainty will encourage the consumer to move from rejection or adoption toward waiting. Conversely, an increase in uncertainty that makes the consumer worse off must move the consumer from waiting toward adoption or rejection.

Just as FOSD improvements and increases in uncertainty in the priors have ambiguous effects on the consumer’s value function, they also have ambiguous effects on the producer: They may increase or decrease the probability of adoption. Improving the true value of the technology (increasing θ) need not make the consumer better off, but, if the signal process satisfies the MLR property, such improvements are good for the producer (at least in the short term) because they increase the probability of adoption (i.e., $p_k(\theta)$ is increasing in θ). Intuitively, improving the technology makes favorable signals more likely and thereby makes the consumer more likely to adopt and likely to adopt sooner.

6. Convexity

We now study the convexity of the value function and optimal policy regions. These convexity results are useful for

understanding the role of uncertainty in the model and will be important for characterizing the effects of “better information” in the next section. We do not need to assume the MLR property to prove these results. We establish convexity of the value function using the same proof strategy that we used to establish monotonicity in Proposition 3.6; we first show that Bayesian updating preserves convexity in the same way it preserves monotonicity in Lemma 3.5.

LEMMA 6.1. *Suppose that $u(\pi)$ is convex in π . For any priors π_1 and π_2 and $0 \leq \alpha \leq 1$, we have*

$$E[u(\Pi(\pi_\alpha, \tilde{x}_\alpha))] \leq \alpha E[u(\Pi(\pi_1, \tilde{x}_1))] + (1 - \alpha) E[u(\Pi(\pi_2, \tilde{x}_2))],$$

where $\pi_\alpha = \alpha\pi_1 + (1 - \alpha)\pi_2$ and $\tilde{x}_\alpha, \tilde{x}_1$, and \tilde{x}_2 are the signals corresponding to π_α, π_1 , and π_2 .

PROOF. See Online Appendix A.4. \square

PROPOSITION 6.2. *For all k , the value function $v_k^*(\pi)$ is convex in π : For any distributions π_1 and π_2 and $0 \leq \alpha \leq 1$, we have*

$$v_k^*(\pi_\alpha) \leq \alpha v_k^*(\pi_1) + (1 - \alpha) v_k^*(\pi_2),$$

where $\pi_\alpha = \alpha\pi_1 + (1 - \alpha)\pi_2$.

PROOF. We show this by induction. The terminal value function, $v_0^*(\pi) = 0$, is trivially convex. Now suppose that $v_{k-1}^*(\pi)$ is convex. By the previous lemma, the value if the consumer waits, $-c + \delta E[v_{k-1}^*(\Pi(\pi, \tilde{x}))]$, is convex. The rewards if the consumer adopts ($\int_\theta \theta \pi(\theta) d\theta - K$) or rejects (0) are also convex. Then, $v_k^*(\pi)$, as the maximum of three convex functions, is also convex. \square

Convexity of the value function can be interpreted as an aversion toward uncertainty about the prior π . For example, in the iPhone example, the consumer may be uncertain about the kind of battery used in the device, and this may affect her beliefs about the iPhone’s benefit. Specifically, she may have probabilities α versus $(1 - \alpha)$ chance of having prior π_1 with battery 1 or π_2 with battery 2. A convex value function means that the consumer would prefer to resolve this uncertainty before beginning the information-gathering process (for an expected value of $\alpha v_k^*(\pi_1) + (1 - \alpha) v_k^*(\pi_2)$) rather than begin the information-gathering process with this uncertainty unresolved (for an expected value of $v_k^*(\pi_\alpha)$). Uncertainty about the prior is costly because of its impact on decision making: If the consumer knew whether π_1 or π_2 prevailed, she could make better adoption, rejection, and/or information-gathering decisions.

We next show that the adoption and rejection regions are convex.

PROPOSITION 6.3. *If it is optimal to adopt (reject) with both π_1 and π_2 , then it is also optimal to adopt (reject) with any $\pi_\alpha = \alpha\pi_1 + (1 - \alpha)\pi_2$, where $0 \leq \alpha \leq 1$.*

PROOF. We will prove that the adoption region is convex; the proof for the rejection region is similar. If it is optimal to adopt with both π_1 and π_2 , then the value of adopting at π_α is

$$\begin{aligned} v_k(\text{adopt}, \pi_\alpha) &= \int_\theta \theta \pi_\alpha(\theta) d\theta - K \\ &= \alpha \left(\int_\theta \theta \pi_1(\theta) d\theta - K \right) \\ &\quad + (1 - \alpha) \left(\int_\theta \theta \pi_2(\theta) d\theta - K \right) \\ &= \alpha v_k^*(\pi_1) + (1 - \alpha) v_k^*(\pi_2) \geq v_k^*(\pi_\alpha), \end{aligned}$$

with the inequality following from the convexity of the value function. This implies that it is optimal to adopt with π_α . \square

Thus, if it is optimal to adopt (or reject) at both π_1 and π_2 , uncertainty about which distribution prevails does not affect the optimal strategy. However, the waiting region need not be convex: Although it may be cost effective to gather additional information at both π_1 and π_2 , with uncertainty about which distribution prevails, it may be too expensive to resolve this additional uncertainty, and the consumer may be better off adopting or rejecting based on the current information.

7. Comparative Statics

We now consider the effects of changing various parameters in the model. What happens if we reduce the cost of gathering information? Or make the information “better”? Or reduce the cost of adoption? Or use a different discount rate? We examine how these changes affect the consumer’s value function, the policy regions, and the timing of adoption/rejection decisions.

7.1. Cheaper Information

We first consider the impact of reducing the cost of information gathering.

PROPOSITION 7.1. *Decreasing the cost of information gathering (c) increases the value function, expands the information-gathering region, and delays the adoption or rejection decision.*

PROOF. See Online Appendix A.5. \square

Thus, cheaper information is good for the consumer and leads her to gather more information before she adopts. For example, in the “old days” before the Internet, we might have been willing to make a spontaneous decision to purchase a new gadget upon first seeing it at a store. Now, however, we are very unlikely to buy a gadget that costs more than (say) \$100 without first researching it on the Internet.

Although cheaper information delays the adoption or rejection decision, the effect on the probability of adoption

or rejection is unclear. With expensive information, it may be optimal to either adopt or reject. With cheaper but still costly information, the consumer would invest in information gathering only if it has the possibility of changing her decision. Thus, if the optimal decision with expensive information is to reject immediately, cheaper information can only increase the probability of adoption. Conversely, if the optimal decision with expensive information is to adopt immediately, reducing the cost of information can only decrease the probability of adoption.

7.2. Better Information

What is the effect of better information? Here we define better information in the sense of Blackwell’s “sufficiency” condition (Blackwell 1951). Consider two different signal processes, one producing signal $x \in X$ with likelihood function $L_x(x | \theta)$ and the other producing signal $y \in Y$ with likelihood function $L_y(y | \theta)$. These two processes may, in general, be dependent, and we can write $L_y(y | \theta) = \int_x L_{y|x}(y | x, \theta) L_x(x | \theta) dx$ where the conditional distribution $L_{y|x}(y | x, \theta)$ captures the dependence between signals X and Y . Blackwell’s notion of a signal process X being more informative than (or, in his terms, sufficient for) a signal process Y can be defined as follows.

DEFINITION 7.2. Signal process X is more informative than signal process Y , or $X \succeq_B Y$, if there exists a stochastic transformation $B(y | x)$ such that $L_y(y | \theta) = \int_x B(y | x) \cdot L_x(x | \theta) dx$.

Here the stochastic transformation $B(y | x)$ must satisfy the usual conditions of a conditional probability distribution (for each x , $B(y | x)$ defines a probability distribution on Y). With such a transformation, we can define a signal Y^* with conditional distribution $L_{y^*|x}(y^* | x, \theta) = B(y | x)$ that has the same likelihood function as y ($L_{y^*}(y^* | \theta) = L_y(y | \theta)$) but is conditionally independent of θ given x . Such a signal Y^* is sometimes referred to as a “garbling” of X because it is as if Y^* were generated from X using a stochastic mechanism that is independent of θ . In the beta-binomial model, a signal X consisting of five positive or negative reports is more informative than a signal Y with three reports; we could view Y ’s three reports as a random selection of X ’s five reports. In the normal-normal model, a signal X drawn from a normal distribution with mean θ and precision t_2 is more informative than a signal Y drawn from a normal distribution with mean θ and precision t_1 , whenever $t_2 \geq t_1$; we could view the less precise signal as equal to the more precise signal plus additional noise.

We summarize the effects of having better information in the following proposition. The convexity of the value function established in Proposition 6.2 plays a central role in this proof.

PROPOSITION 7.3. *Increasing the informativeness of the signal process increases the value function and expands the information-gathering region; that is, if it is optimal to wait*

with prior π and signal process Y , then it is also optimal to wait with prior π and signal process X whenever $X \succeq_B Y$.

PROOF. See Online Appendix A.6. \square

Thus, better information is clearly good for the consumer.³ The effects on the producer are ambiguous. With a fixed prior, better information makes waiting more attractive. However, the expected time until adoption or rejection and the probability of eventual adoption may increase or decrease. A consumer who would have otherwise immediately adopted or rejected may be induced by better information to delay and gather more information. On the other hand, a consumer who acquires better information may form precise beliefs more quickly and ultimately decide sooner.

7.3. Reduced Adoption Cost

We next consider the effects of reducing the cost of adoption.

PROPOSITION 7.4. *Decreasing the cost of adoption (K) increases the value function, expands the adoption region, and shrinks the rejection region; the probability of eventual adoption increases, and the expected time until adoption decreases.*

PROOF. See Online Appendix A.7. \square

Clearly, reducing the adoption cost K is good for the consumer. The impact on the producer depends on the interpretation of these costs. If we interpret the adoption cost as the price paid by the consumer to the producer to acquire the technology, then we cannot say whether the benefits associated with reducing these costs (e.g., by offering a rebate) outweigh the direct reduction in revenue. However, if we interpret K as a setup cost paid by the consumer but not paid to the producer (e.g., a cost associated with installing the technology), then a reduction in K is unequivocally good for the producer because it accelerates adoptions and increases the probability of eventual adoption.

7.4. Reduced Discounting

Finally, we consider the effect of increasing the consumer’s discount factor.

PROPOSITION 7.5. *Increasing the consumer’s discount factor (δ) increases the value function, expands the information-gathering region, and delays the adoption or rejection decision.*

PROOF. The proof is analogous to the proof of Proposition 7.1 in Online Appendix A.5. \square

Thus, a consumer with a higher discount factor is better off and more likely to gather information because the delay induced by information gathering is less costly. As with a decrease in the cost of information, the effect on the adoption probability is unclear.

8. Evolving Technologies

In the previous sections, we assumed that the benefit of the technology, θ , does not change over time. A more general model would allow the technology to evolve over time. For example, the producer may come out with new versions of the technology. We can also think of the change in benefit being driven by the consumer's need for the technology changing in some predictable or unpredictable way. For example, a consumer's desire for an iPhone may be related to her travel plans, which may change because of a change in job status. Note that the changes need not be improvements. For example, the benefit of a new coal-fired power plant may decrease with changes in environmental regulations. A sophisticated consumer needs to consider the potential changes in benefits when making adoption and information-gathering decisions.

Because we formulated the stationary model to allow general probability distributions and signal processes, it is not difficult to generalize the model to accommodate stochastic changes in the technology's benefits. Let $\theta_k \in \Theta$ denote the benefit of the technology in period k . As before, if the consumer decides to reject the technology, she receives nothing. If she decides to adopt the technology, she pays a fixed adoption cost K and receives a net expected benefit of $\int_{\theta_k} \theta_k \pi(\theta_k) d\theta_k - K$, where $\pi(\theta_k)$ is her prior on θ_k . If the consumer chooses to wait, she pays c , and the benefit of the technology changes stochastically according to the conditional probability distribution $g(\theta_{k-1} | \theta_k)$. She then observes a signal x_{k-1} about the new technology with benefit θ_{k-1} , with the signal x_{k-1} drawn from likelihood function $L(x_{k-1} | \theta_{k-1})$. Note that, in this model, waiting not only allows the consumer to gather additional information, but it also allows the possibility of adopting a better technology in the future.

The dynamic programming formulation is the same as the stationary model except the posterior distributions and signal distributions now include the evolution of the technology. Let $\eta(\theta_{k-1}; \pi)$ be the prior on next period's technology benefit: $\eta(\theta_{k-1}; \pi) = \int_{\theta_k} g(\theta_{k-1} | \theta_k) \pi(\theta_k) d\theta_k$. The signal and posterior distributions are then given by $f(x_{k-1}; \pi) = \int_{\theta_{k-1}} L(x_{k-1} | \theta_{k-1}) \eta(\theta_{k-1}; \pi) d\theta_{k-1}$ and

$$\Pi(\theta_{k-1}; \pi, x_{k-1}) = \frac{L(x_{k-1} | \theta_{k-1}) \eta(\theta_{k-1}; \pi)}{f(x_{k-1}; \pi)}.$$

Most of the properties of the earlier model carry over to this more general setting. To establish monotonicity properties analogous to those of §3, we need to assume that the technology transitions, as well as the signal process, satisfy the MLR property: $g(\theta_{k-1} | \theta_k^2) \succeq_{\text{LR}} g(\theta_{k-1} | \theta_k^1)$ for all $\theta_k^2 \geq \theta_k^1$ and for all k . This assumption means that a truly better technology in one period leads to an LR improvement in the distribution on technology benefits for the next period. If the technology transitions and the signal process both satisfy the MLR property, we can use the results of §3.1 to show that LR dominance among the priors for one period

($\pi_2 \succeq_{\text{LR}} \pi_1$) implies that the next-period priors, signal distributions, and posteriors are also LR ordered ($\eta(\pi_2) \succeq_{\text{LR}} \eta(\pi_1)$, $f(\pi_2) \succeq_{\text{LR}} f(\pi_1)$, and $\Pi(\pi_2, x) \succeq_{\text{LR}} \Pi(\pi_1, x)$ for each x). Lemma 3.5 and Proposition 3.6 then go through as before, and we can conclude that the value function $v_k^*(\pi)$ is LR increasing. Thus, if the technology transitions and the signal process both satisfy the MLR property, LR dominance remains an appropriate ordering on priors when the technology is evolving.

Although monotonicity of the value function generalizes with only this MLR assumption, we need an additional assumption to ensure monotonicity of the optimal policies. We can illustrate the issue by considering a special case of the general model where the signal perfectly reveals the true benefit and there are two periods remaining. Taking the discount factor $\delta = 1$ and ignoring the possibility of quitting, the choice between adopting and waiting requires a comparison between $\theta_k - K$ and $E[\theta_{k-1} | \theta_k] - c$; equivalently, the consumer should adopt if $c - K \geq E[\theta_{k-1} - \theta_k | \theta_k]$ and wait otherwise. If $E[\theta_{k-1} - \theta_k | \theta_k]$ is decreasing in θ_k and adoption is preferred for one θ_k , then adoption must also be preferred for all higher values of θ_k . However, if $E[\theta_{k-1} - \theta_k | \theta_k]$ is not decreasing in θ_k , then we may find that adoption is preferred for one θ_k but waiting is preferred for a higher θ_k . The condition required to establish monotonicity in this special case is sufficient for the general case.

PROPOSITION 8.1. *Suppose that the signal process and technology transitions both satisfy the MLR property. If $E[\theta_{k-1} - \theta_k | \theta_k]$ is nonincreasing in θ_k for each k , then*

- (i) $v_k^*(\pi) - \int_{\theta_k} \theta_k \pi(\theta_k) d\theta_k$ is LR decreasing.
- (ii) *The optimal policy satisfies the monotonicity properties of Proposition 3.8.*

PROOF. The proof of (i) is analogous to the proof of Lemma 3.7, taking into account the technology transitions; see Online Appendix A.8 for the proof. Given (i), the proof of the monotonicity of policies in (ii) follows exactly as in the proof of Proposition 3.8. \square

With the exception of the convergence results of §4, the other results and intuitions with stationary technologies continue to hold with evolving technologies without any additional assumptions. Specifically, the existence of the derived benefit function of Proposition 5.1 (now a function of the period- k benefit θ_k) does not require any additional assumptions on the technology transitions except for the last part of this result, which shows that, if the policy is monotonic, the probability of adoption is increasing in θ_k ; this monotonicity result requires the technology transitions and signal processes to both satisfy the MLR property. The results on the convexity of the value function and policies in §6 and the comparative statics results of §7 carry over directly without any additional assumptions on the transitions. The convergence results of §4 would clearly require some additional assumptions to hold: If the technology benefit θ_k is changing over time, there is no reason to expect the estimate of the benefit to converge.

If technologies are potentially changing over time, this raises another interesting comparative statics question: When is one technology transition $g(\theta_{k-1} | \theta_k)$ preferred to another? Or what constitutes an improvement in the future prospects for a technology? Given that we need an LR improvement in the prior to ensure an increase in the value function, we want a condition on the transitions that ensures the prior on next-period’s benefit ($\eta(\theta_{k-1})$) improves in the LR order. We say that a technology transition $g_2(\theta_{k-1} | \theta_k)$ is an LR improvement of $g_1(\theta_{k-1} | \theta_k)$ if both g_2 and g_1 satisfy the MLR property and

$$\frac{g_2(\theta_{k-1}^2 | \theta_k)}{g_1(\theta_{k-1}^2 | \theta_k)} \geq \frac{g_2(\theta_{k-1}^1 | \theta_k)}{g_1(\theta_{k-1}^1 | \theta_k)} \quad \text{for all } \theta_k \text{ and } \theta_{k-1}^2 \geq \theta_{k-1}^1, \quad (1)$$

$$\frac{g_2(\theta_{k-1} | \theta_k^2)}{g_1(\theta_{k-1} | \theta_k^2)} \geq \frac{g_2(\theta_{k-1} | \theta_k^1)}{g_1(\theta_{k-1} | \theta_k^1)} \quad \text{for all } \theta_{k-1} \text{ and } \theta_k^2 \geq \theta_k^1. \quad (2)$$

Here condition (1) requires that, for any current technology benefit θ_k , the conditional distribution for next period’s benefit θ_{k-1} under g_2 to LR dominate the distribution under g_1 . Although (1) seems natural, it is not sufficient in itself because a mixture of LR-dominant distributions need not result in an LR-dominant distribution: that is, $g_2(\theta_{k-1} | \theta_k) \succeq_{\text{LR}} g_1(\theta_{k-1} | \theta_k)$ for all θ_k does not imply that $\eta_2(\theta_{k-1}) = \int_{\theta_k} g_2(\theta_{k-1} | \theta_k) \pi(\theta_k) d\theta_k \succeq_{\text{LR}} \eta_1(\theta_{k-1}) = \int_{\theta_k} g_1(\theta_{k-1} | \theta_k) \pi(\theta_k) d\theta_k$. Condition (2) can be interpreted as requiring the backward transitions from θ_{k-1} to θ_k to satisfy the same LR dominance condition. A result in Karlin (1968) then ensures that $\eta_2(\theta_{k-1}) \succeq_{\text{LR}} \eta_1(\theta_{k-1})$.

With this notion of improved prospects for a technology, we can conclude that such improvements are good for the consumer and make waiting more attractive.

PROPOSITION 8.2. *Suppose that the signal process and technology transitions both satisfy the MLR property. An LR improvement in the technology transition process increases the value function and expands the waiting region.*

PROOF. See Online Appendix A.9. \square

Thus, improved prospects for future technologies are clearly good for the consumer. The effect on the producer is unclear: Like better information, improved prospects for future technologies may entice a consumer who otherwise would have adopted or rejected the technology to wait instead. Such a change could increase or decrease the probability of ultimately adopting the technology and accelerate or delay the adoption/rejection decision.

9. Conclusions and Possible Extensions

Table 1 summarizes some of the main conclusions of the paper. The rows correspond to changes in assumptions,

Table 1. Summary of key results.

	Impact on value function	Short-term encourages
Better technology θ	+/-	Adoption
More optimistic prior π	+	Adoption
More favorable signal x	+	Adoption
More uncertainty about θ	+/-	—
Reduced adoption cost K	+	Adoption
Cheaper information c	+	Waiting
Better information $L(x \theta)$	+	Waiting
Improved future prospects $g(\theta_{k-1} \theta_k)$	+	Waiting

and the table entries describe the impact of these changes on the consumer’s value function and optimal policy. The impacts of these changes on the consumer’s value function are mostly quite natural. For example, if we define “better information” or “improved future prospects” appropriately, such a change should be good for the consumer. The challenge at times has been to find the appropriate definition of an improvement that captures the essence of the change and ensures the desired natural result. Some of our results are perhaps surprising: For example, we saw that a better technology may not make the consumer better off once we take into account the information-gathering costs and adoption process. This leads us to require a stronger notion of a “more optimistic” prior than we might have expected; specifically, we need to use the LR order rather than the FOSD order.

Although we have shown that this basic model has a great deal of structure, there are some additional potential results that could be of interest. For example, although we have shown that increases in uncertainty about the benefit of the technology do not always make the consumer better off, it may be possible to identify a more refined notion of “increased uncertainty” that is stronger than Rothschild and Stiglitz’s (1970) notion and is unambiguously good for the consumer. This refinement might be analogous to our use of LR dominance instead of FOSD dominance to describe “more optimistic” priors. In addition, in the model with evolving technologies, it would be interesting to characterize how “increased uncertainty” about future technologies affects values and policies.

There are a number of ways our basic model could be extended to address other issues related to technology adoption. First, one can extend the model to consider technologies with multiple attributes and multidimensional signals. For example, one might consider a product’s per-period benefit and its reliability as separate attributes. Multidimensional signals might provide information on these attributes separately: For example, it might be easy to obtain accurate information about the benefit of a technology but harder to obtain information about its reliability. In this setting, we need to consider an aggregate benefit function that combines these attributes into an overall benefit value. The results

of §§4–7 carry over directly to this multivariate setting, but the monotonicity results in §§3 and 8 require the aggregate benefit function to be increasing in each product attribute and use multivariate extensions of the LR order and the MLR property. This extension is discussed in Ulu (2007).

A second possible extension would be to give the consumer a range of different possible information-gathering activities. For example, one might include costless “passive” information gathering as well as more expensive “active” activities. In this more general setting, the basic structure of the model will still hold taking the information gathering region in aggregate. For example, the adoption and rejection regions will be convex and, along any chain of LR-improving prior distributions, we move away from the rejection region and toward the adoption region. However, we suspect that it will be difficult to say much about the choice of information-gathering processes within the information-gathering region. For example, information may be worth more when the prior is near a threshold (in some sense) and less when the priors are farther from the thresholds. Similarly, different kinds of information sources may be more or less valuable depending on specific features of the prior.

A third direction to extend the model would be to allow repeat purchases with evolving technologies. In this extension, adopting a technology would provide some immediate benefits and leave open the possibility of “upgrading” in the future. This form of extension would be particularly interesting if the consumer has a choice about the cost and quality of the new technology to purchase. In this setting, improved prospects for future technologies may encourage consumers to purchase cheaper and lower-quality technologies today and plan to upgrade in the future.

Finally, another direction to extend the model would be to consider richer models of consumer behavior. For example, rather than assuming that the consumers make decisions on an expected-value basis, one might consider risk-averse consumers. Here, we would expect to find that risk-averse consumers are less likely to adopt than their risk-neutral or less risk-averse counterparts. We might go further still and consider nonexpected utility models of consumer behavior. For example, we might consider consumers who have loss-averse or reference-point-dependent utility functions and/or update their beliefs about the benefits of new technologies in a non-Bayesian fashion; Kahneman and Tversky’s (1979) “prospect theory” model provides a rich descriptive model of consumer behavior that might be useful in this context. Alternatively, one might consider a game-theoretic model where producers strategically choose the cost and quality of information to influence consumer behavior and consumers acquire and interpret information knowing that the producer is being strategic in these choices.

We expect that the model formulation and analytic techniques that we have used in this paper will be useful when studying more complex technology adoption models. Moreover, we hope that the results and insights provided here will help clarify our intuitions about what to expect—and

what not to expect—in more complex models of technology adoption and the diffusion of innovations.

10. Electronic Companion

An electronic companion to this paper is available as part of the online version that can be found at <http://or.journal.informs.org/>.

Endnotes

1. Jensen (1982) studied a model with a similar Bernoulli information structure but with only two possible benefit levels. He also assumed that information gathering is costless ($c = 0$).
2. When we say that something will “almost certainly” occur, we mean that this will occur with probability one. For example, there may exist infinite signal sequences such that the estimated benefit does not converge, but the total probability associated with such signal sequences is zero.
3. Lippman and McCardle (1987) incorrectly claim that increasing the precision of the signal in the normal-normal model leads to a decrease in the consumer’s value function; this can be traced to an error in their formula for the variance of the signal distribution.

Acknowledgments

The authors are grateful for helpful comments provided by Laura Kornish, Kevin McCardle, Bob Winkler, and the anonymous associate editor and referees.

References

- Blackwell, D. 1951. The comparison of experiments. *Proc. Second Berkeley Sympos. Math. Statist. Probab.*, University of California Press, 93–102.
- Chatterjee, R., J. Eliashberg. 1990. The innovation diffusion process in a heterogeneous population: A micromodeling approach. *Management Sci.* **36**(9) 1057–1079.
- Jensen, R. 1982. Adoption and diffusion of an innovation of uncertain profitability. *J. Econom. Theory* **27**(1) 182–193.
- Kahneman, D., A. Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* **47**(2) 263–292.
- Karlin, S. 1968. *Total Positivity*, Vol. 1. Stanford University Press, Stanford, CA.
- Kornish, L. J., R. L. Keeney. 2008. Repeated commit-or-defer decisions with a deadline: The influenza vaccine composition. *Oper. Res.* **56**(3) 527–541.
- Lippman, S. A., K. F. McCardle. 1987. Does cheaper, faster, or better imply sooner in the timing of innovation decisions? *Management Sci.* **33**(8) 1058–1064.
- McCardle, K. F. 1985. Information acquisition and the adoption of new technology. *Management Sci.* **31**(11) 1372–1389.
- Milgrom, P. R. 1981. Good news and bad news: Representation theorems and applications. *Bell J. Econom.* **12**(2) 380–391.
- Rogers, E. M., E. Rogers. 2003. *Diffusion of Innovations*, 5th ed. Free Press, New York.
- Rothschild, M., J. E. Stiglitz. 1970. Increasing risk I: A definition. *J. Econom. Theory* **2**(3) 225–243.
- Ulu, C. 2007. Dynamic programming models of technology adoption and use. Ph.D. thesis, Duke University, Durham, NC.
- Whitt, W. 1979. A note on the influence of the sample on the posterior distribution. *J. Amer. Statist. Assoc.* **74**(366) 424–426.