SEQUENTIAL ADOPTION THEORY:
A Theory For Understanding Herding in Technology Adoption Decisions

Eric A. Walden
Rawls College of Business Administration
Texas Tech University
ewalden@ba.ttu.edu

Glenn J. Browne
Rawls College of Business Administration
Texas Tech University

ABSTRACT
Technology adoption often occurs sequentially, so that later potential adopters can see the decisions (adopt or not adopt) of earlier potential adopters. Here, we construct a model of the technology adoption decision where 1) adoption occurs sequentially, 2) each decision maker has some private information about the fitness of a technology, and 3) decision makers can see the decisions, but not the information of earlier decision makers. This characterization of some neglected aspects of technology adoption decisions allows us to build a model that illustrates technology adoption when decision makers supplement their own incomplete information with information deduced from the behavior of others.

Keywords: Technology adoption decisions, Sequential decision making, Signal detection theory, Imitative behavior, Simulation

I. INTRODUCTION
Observational learning is one of the most ubiquitous and useful means of decision making available to human beings. Observational learning occurs when one person observes the behavior of an other person and infers something about the expediency of the behavior based on that observation. In other words, if someone else is doing it, maybe it is a good idea. This is a powerful decision making tool because it requires little mental effort and it is often accurate. Thus,
human beings, and a wide variety of animals, imitate the behavior of others <ref BHW>.

One decision of particular interest to information systems (IS) researchers is technology adoption. Research has shown that observational learning influences technology adoption both in the laboratory <song walden>, in real world financial markets <B&W>. It also influences bidding behavior in online auctions <ref>. Hence, it seems useful to develop a rigorous model of how observational learning might proceed in technology adoption, and to apply this model to develop insights into observational learning issues. We offer this work as an explanation of how adopters can learn from the adoption behavior of others.

WHY OBSERVATIONAL LEARNING?

Monkeys <ref> and birds <ref> learn about the expediency of behaviors from observing one another, and human beings preferences for mates depends on the behavior of others vis-à-vis potential mates <ref>, but why would we think that technology decisions would be managed in a similar manner? Technology decisions are fraught with complexity and uncertainty. Software is the most complex artifacts humans build <ref man month>, and is usually just one component of a technology. Moreover, the impacts of technologies can take years to be realized <ref bryn>, so the benefits of adoption decisions are often uncertain.

It is precisely this complexity and uncertainty that makes observational learning so appealing. All of the information an individual has about the expediency of adoption is consolidated and expressed in his behavior. Thus, learners can save a great deal of cognitive effort by inferring that if an individual adopts a technology then his personal information must have suggested that the technology was worth adopting. Of course, this is a gross simplification, but that is the beauty of it—it is a simplification. Using observational learning takes a complex multidimensional and uncertain decision making environment and simplifies it. Using simplifications is most important for cognitively demanding decisions, because it yields the most bang for the buck. If a decision is cognitively easy, then one gains very little by trying to simplify it more. However, simplifying a difficult decision reduces cognitive effort immensely.

Thus, technology decisions, by virtue of their complexity and uncertainty, can be made much easier by using the behavior of others. This reduces complexity by reducing the range of consideration to a single factor—others’ behavior—rather than the multitude of factors that may be relevant. Observational learning mitigates uncertainty because it actually aggregates information from more than one decision maker. Moreover, the loss of certainty from ignoring other factors is less important in uncertain environments than in certain environments. In other words, if a decision maker’s best guess is not very certain anyway, then there is not much to loose by using the second best guess instead. In conclusion, there are real cognitive savings from simplification of technology adoption decisions, which make observational learning attractive.
II. THEORY

There are a variety of situations of interest in IS research and practice, in which members of a community make sequential decisions and can observe the decisions, but not the reasons, of others. For example, in online auctions individuals place bids in time order and can observe the bids of others, but not the reasons for the bids. Research shows that bidders make use of observational learning in this situation, so that in two identical auctions (same product, same seller, and same time) people prefer to bid for the product that others are bidding for <ref>. Financial markets are another example, where individuals can see bids (and offers) over time, but not the reasons for those bids and offers. Research demonstrates that the price investors are willing to pay for firms pursuing electronic commerce initiatives depend heavily on the willingness of others to pay for prior electronic commerce initiatives <ref>.

Adoption decisions are often characterized at binary decisions, with choices being adopt or do not adopt (or a choice between two options, like PC vs. Mac). Again, research has documented observational learning in these sort of IT adoption decisions. When asked to evaluate peer-to-peer file sharing technologies, subjects indicate a higher intention to adopt the technologies if they observe others adopting them <ref>. <Glenn, do we have some more here?>

Information cascade theory <ref>, is a general body of research has been developed to explain the consequences of learning from the behavior of others. The underlying notion is that individuals each hold some private information, which can be thought of as a signal about the expediency of a course of action. The signals are not perfect, so that individuals are make decisions under uncertainty. Given this signal, a decision maker makes a decision. The decision, but not the signal is observable by other decision makers who then use Bayesian updating to revise their belief about the course of action.

For example, in laboratory experiments <ref> subjects are faced with an opaque contain that either holds two red and one green balls <is the plural correct here?> or two green and one red balls. Each subject is allowed to privately draw and view a single ball, representing his private information, from the container. Clearly, the color of the ball is a signal about the total content of the container. The subject then replaces the ball and calls out his decision about the contents of the container, so that other subjects can hear it. The next subject repeats the procedure and makes a decision based upon both the color of the ball he observes and the decision of the prior subject.

Based on the key assumptions of uncertainty, private information, observability of prior decisions, and of course rationality in the form of Bayesian updating, several results emerge. The first is the tendency of decision makers to “herd”. Herding means that all decision makers rapidly converge on the same decision simply because they saw others make that decision. One might imagine a flock of birds or a school of fish that all turn right at the same time because the lead fowl or fish did. This herding is called an information cascade because the information contained in the decision of the first decision maker propagates to other decision makers who observe him.
A second key result is that information cascades are fragile. This means that it is relatively easy to change the emergent behavior of the group by introducing just a little bit of new information. This occurs because the entire group’s decisions are based on relatively little information constrained in the behavior of the first few decision makers. In the container game described above, if the first two decision makers indicate they think they are dealing with the container that contains two red balls then the third decision maker should rationally concur, even if he draws a green ball. If two subjects see red and one sees green then the correct answer is most likely red. In this case, there is no information in the decision of the third subject, and the forth subject faces exactly the same information environment as the third and thus, must rationally say red regardless of his own private information. The cascade is fragile because everyone knows that they are deciding based on very little information, and it does not take a great deal of contradictory information to change their minds.

The interesting thing is that the group as a whole seems flighty—rapidly achieving conformity and then easily reversing their decisions when small amounts of contradictory information are presented—even though each individual is behaving in a fully rational manner.

There are a variety of other theories that suggest similar behavior, particularly rapid convergence to the same outcome. Network externality theory <ref> predicts that as more people make a particular adoption decision that decision becomes more appealing and hence others are more likely to make it. Institutional isomorphism <ref> suggests that making decisions similar to others, gives an organization legitimacy, which in turn gives it access to resources. Hence, organizations make similar decisions to signal their legitimacy. Individuals may also all make the same decision when it is the correct one and they all have good information (or when it is the wrong one and they all have the same bad information).

Information cascade theory provides a complement to these other theories because the causal process is different. The idea is that decision makers follow each other in order to aggregate information in an uncertain environment. Thus, it can explain situations that do not fit the assumptions of the other theories. It can explain situations where the benefit of a decision does not depend on the number of others making the same decision, or where a particular decision does not lead to legitimacy, or where people have different and limited information. This may characterize the early stages of adoption, where there are not enough adopters for network effects to be relevant or future network sizes to be estimated. In the early stages of adoption, there may be too few people making the adoption decision for it to be legitimate. Of course, in the early stages of adoption information is poor and decision makers may have very different signals about the expediency of a course of action. Information cascade theory can also complement other theories in later stages of adoption and in situations in which the assumptions of the other theories hold. Nothing forbids decision makers from both inferring information from the decisions of other and deriving benefit from the decisions of others.
What information cascade theory provides is an additional perspective that may be more or less appropriate depending on the adoption environment in question. Where the true characteristics of the environment match the assumptions of the model—uncertainty, private information and observable behavior—it is useful to understand what the theory has to say. Thus, in this work we explore what happens in environments where these assumptions hold.

MODEL

Our goal is to produce a rigorous model of information cascades, and by rigorous we mean mathematical. Mathematical models have the advantage of precision, and the disadvantage that reality is often not precise. We have to make a number of simplifying assumptions to achieve precision, which we discuss in detail below. The important thing to keep in mind is that our theoretical contribution is an in-then statement. We claim that if the assumptions hold then the results hold. As reality deviates from the assumptions the model becomes less useful, and some deviations change the outcomes greatly, while others modify it slightly. Thus, it is useful to discuss the assumptions in some detail.

Assumption #1: Decision makers must choose between adoption of two technologies called A and B. Some examples might include PC or MAC, open source or proprietary, or peer-to-peer or mainframe. The model might also represent a choice between adoption and the status quo, which means no adoption. The limitation of this is that it only considers two choices, when there may be many choices in reality.

Assumption #2: Decision makers have some private information about the relative merits of A and B. Specifically, decision makers receive a signal, which is a single observation from a random normal distribution representing the value difference between A and B. If A is a better choice then the random distribution will have a greater mean than if B is the better choice. The decision maker’s problem is to decide which distribution the signal came from and hence which technology is the better choice.

Assumption #3: The values of the technologies do not vary across decision makers. This means that we invoke technological determinism. Either A is better or B is better for all individuals. In some situations this clearly will not hold. In many situations it will hold to some degree. The benefits of a technology to a specific adopter are often correlated with the benefits of the technology to other adopter. Technologies perform some function and adopters usually adopt them with that function in mind. Moreover, people looking for some functionality are probably facing similar problems with similar constraints and abilities. Thus a technology that works well within one adopter’s constraints and abilities probably works well within an other adopter’s constraints and abilities. For example, people frequently need to search for information online or type memos or visualize numbers and hence benefit from the search engines, word processors, and spreadsheets. More specifically, people may be familiar with a certain menu layout in a word processor and hence may benefit more from using Word than WordPerfect.
To the degree that people’s relative values for technology are correlated, the model we present holds. We assume perfect correlation, but relaxing the assumption does not change the implications of the model. Imperfect correlation could be modeled as a discount factor on the information of others, which has grounding in psychological research <ref>. Such a discount value for imperfect correlation simply makes decision maker’s responses to others decisions less dramatic. In the parlance of the model, which we discuss below, it would move the decision threshold less.

We must caution that in cases where each decision maker’s valuation for the two technologies is uncorrelated there is no information to be gained from observing the behavior of others, and hence, the model does not address those situations.

Assumption #4: There are no changes in the valuation of the technologies during the adoption period. What this means the distribution from which decision makers’ private information comes does not change. Inherent in this is the notion that the decision maker’s private information is based on reality. We do not distinguish between the decision makers’ private information and the true nature of the technology.

The assumption that technology does not change over the adoption period can be fulfilled if technology value never changes or if changes to value are only realized after the adoption period is over. Thus, we do not rule out things like network effects <ref>, we just require them to be realized after the adoption has occurred. We leave it to future research to examine the situation in which the distribution from which decision makers receive private signals (i.e. the technology value) changes over time. Our guess is that this would be a special case of the correlation issue discussed above and decision makers would discount prior decision maker’s behavior appropriately for the rate of change of the technology value, and the overall effect would be a smaller change in the decision threshold than we propose here.

Assumption #5: There are no global sources of information during the adoption period. Again this means that the distributions from which decision makers receive private signals do not change. Again, we could relax this by discounting the weight that decision makers place on the behavior of prior decision makers.

Assumption #6: Decision makers make choices sequentially. This means that each decision maker has some others who go before and some who go after. This also means that the time of decision making is known. For the choice between two technologies this is not an unreasonable assumption. However, when the choice is between a technology and the status quo, then for those who follow the status quo, it is not necessarily clear when the decision to follow the status quo was made. A decision maker following the status quo may have evaluated the alternative and decided against it, or he may not yet have performed an evaluation. One possible fix for the status quo problem would be to discount the decisions of those following the status quo. In this case, the impact of a status quo decision would be less than the impact of an adoption decision. We will discuss the implications of this later, when we derive results.
Assumption #7: Decision makers know the sequence of prior decisions and the initial conditions. Thus, they are fully informed and rational. Though they face uncertainty, they face no ambiguity. This is a standard economic assumption, and though the concept of bounded rationality has existed for half a century, no one has yet found an acceptable way to model it.

Assumption #8: Decisions are observable, but reasons for decisions (private information) are not. This means that decision makers can see what technology those that went before adopted. For the case of competing technologies, this is a reasonable assumption, but as discussed above, for cases of a new technology versus the status quo, it may be questionable.

There are two things that present themselves in these assumptions. First, many of the assumptions could be relaxed and made more realistic by discounting other’s behavior. We do not include this complication in our analysis, so what we present can be thought of as an upper bound on information cascade effects.

The second issue is that not all these assumptions hold in all adoption environments. Thus, it is worthwhile to consider when they do hold. In general, they hold in the short-run and in small groups. In particular, the assumptions that require the distribution from which decision makers receive signals to remain constant are easily satisfied in the short-run. One can imagine, for example, firms implementing ERP systems. It may take some years from the adoption decision (i.e. writing the first check) to having a system up and running. During that time, technologies may not change and adopters may not gain any additional information upon which to base their decisions.

The assumptions of sequential adoptions and observable behavior are quite reasonable in small groups, particularly where the small groups are cognizant of one another because they are competitors or participants in the same industry or friends for personal adoption.

In general, the assumptions make sense in the short-run for small, close groups. In particular, this seems to capture the environment during early stages of adoption. The group of technophile early adopters is often small and they make their decisions relatively quickly. Early adoption is particularly important because it sets the stage for what happens next, particularly for technologies that have network effects. Thus, while the assumptions may not apply to all adoption situations, they apply to a very important subset of adoption decisions.

One of the focuses of this research which differentiates it from other information cascade work is that it concerns itself with the short run, small population dynamics. Other work is usually concerned with asymptotic results. However, the short run adjustment is what is appropriate for IT adoption decisions, so that is what we focus on.

THEORY FORMALIZATION
Given the setup above we can now formalize the model. Assume that decision makers are faced with two technologies: A and B. They receive a private signal about the merits of the two technologies, which comes from \( \text{NORMAL}(\mu_A, \sigma^2) \) if A is the better choice and \( \text{NORMAL}(\mu_B, \sigma^2) \) if B is the better choice. Here \( \mu_A \) and \( \mu_B \) represent the difference in values between A and B, so a
more descriptive, but more difficult to read subscript might be $\mu_{A-B > 0}$ and $\mu_{A-B < 0}$. However, in the interest of readability we will keep $\mu_A$ and $\mu_B$. This means that $\mu_A > \mu_B$. The private signal is a single realization from the distribution, and the decision maker’s problem is to determine which distribution the realization came from.

Let lower case letters denote the decision maker’s decision, so that $a$ indicates that the decision maker chooses technology A and $b$ indicates that the decision maker chooses technology B. A decision maker would like to make choice a when $\mu_A > \mu_B$ and choice b otherwise. The first decision maker’s prior suggests that A is better than B with probability $p_A$, and thus, B is better with probability $1 - p_A$.

Given the prior, the decision maker then receives a private signal and based on this private signal must make a choice between A and B. Thus, the potential adopter’s decision criterion can be represented as a threshold. If an observed private signal exceeds this threshold then the observer will conclude that technology A is the better technology (recall that the distributions represent the value of A minus the value of B). For a given decision criterion, the probability of correctly concluding that an A is the better technology is $\text{Prob}(a|\mu_A)$, while the probability of incorrectly concluding that A is the better technology, when B is the better technology is $\text{Prob}(b|\mu_A) = 1 - \text{Prob}(a|\mu_A)$. However, the lower the threshold for accepting that A is better when it is not is $\text{Prob}(a|\mu_B)$, and the probability of correctly deciding that B is better when it actually is better is $\text{Prob}(b|\mu_B) = 1 - \text{Prob}(a|\mu_A < \mu_B)$. These possibilities are shown in Table 1.

<table>
<thead>
<tr>
<th>Reality</th>
<th>A better ($\mu_A$)</th>
<th>B better ($\mu_B$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A better</td>
<td>$\text{Prob}(a</td>
<td>\mu_A)$</td>
</tr>
<tr>
<td>B better</td>
<td>$\text{Prob}(a</td>
<td>\mu_B)$</td>
</tr>
</tbody>
</table>

A measure of the ability of a potential adopter to differentiate between the merits of A and B is $d' = \frac{\mu_A}{\sigma} - \frac{\mu_B}{\sigma}$, which is simply the standardized difference between the two means. These assumptions are illustrated in Figure 1.

---

1 The notation $\mu_A$ means that the observation came from the distribution with $\mu_A$ as a mean.
The potential adopter’s problem is to select a threshold beyond which he will conclude that the private signal came from the $\mu_A$ distribution. In other words, the potential adopter chooses the minimum observed value of a private signal that will lead him to believe that A is the better technology. Any observed values above that threshold will lead him to decide that A is better and any observations below the threshold will lead him to decide that B is better.

The four possible outcomes of a decision task are graphed in Figure 2. It can be seen that the probabilities of all possible outcomes are determined by the choice of threshold. Further, it is apparent that increasing the probability of concluding that A is better when it actually is better ($\text{Prob}(a|\mu_A)$) also increases the probability of falsely concluding that A is better, when it is not ($\text{Prob}(a|\mu_B)$).
A threshold is chosen to be the location where the ratio of the probability of a correctly identifying A as better to the probability of incorrectly identifying A as better is greater than some value. This is expressed mathematically as choosing $a$ if

$$\frac{\Pr(a \mid \mu_A)}{\Pr(a \mid \mu_B)} \geq \beta. \quad (1)$$

The probabilities define a likelihood ratio, which is the ratio of the heights of the distributions at a point in Figure 2. For example, the value of $\beta$ in Figure 2 is set to two. Thus, at the threshold represented by the vertical line $r(\beta)$, the height of the $\mu_A$ distribution is twice the height of the $\mu_A$ distribution.

The optimal threshold is chosen by balancing the costs and benefits of each outcome along with the prior probability of each distribution. The optimal value of $\beta$ (see Green, 1966 #344) in this case is

$$\beta = \frac{\Pr(\mu_B)(\text{Benefit}(b \mid \mu_B) - \text{Benefit}(a \mid \mu_B))}{\Pr(\mu_A)(\text{Benefit}(a \mid \mu_A) - \text{Benefit}(b \mid \mu_A))} = \frac{\Pr(\mu_B)k}{\Pr(\mu_A)}. \quad (2)$$

The variable $k$ represents the relative benefits of the outcomes shown in Table 1. The numerator is simply the net benefit adopting B if B is better, while the denominator is the net benefit of adopting A if A is better. As $k$ increases, a decision maker will require a greater private observation to convince him to adopt the technology. Commonly, a high value of $k$ will be associated with a technology that is very expensive or risky to deploy, such as enterprise resource planning (ERP) packages.\(^2\)

Thus far, we have established the decision task of the potential adopter. The task is relatively straightforward. The potential adopter sets an acceptance threshold based on the relative benefits of adoption and rejection, forms an opinion of the technology, and adopts the technology if her opinion is higher than the threshold and rejects it otherwise. The problem is that potential adopters know they may be wrong and would like to incorporate better information into their decisions. One way they can do this is by considering the behavior of others {Abrahamson, 1997 #431; Fichman, 2000 #454}.

In uncertain environments with sequential choices, potential adopters can increase their own information by considering the observed choices of prior potential adopters {Bikhchandani, 1992 #252; Bikhchandani, 1998 #339; Li, 2004 #354; Walden, 2002 #355}).\(^3\) To formalize this idea, assume that potential adopters can perfectly identify prior potential adopters’ decisions but cannot...

\(^2\) For example, Unisource Worldwide Inc. recently wrote off $168 million because of a failed ERP implementation, and such costs tend to be very high for other firms as well Nicolaou, A. I. (2004) “ERP Systems Implementation: Drivers of Post-Implementation Success.” The 2004 IFIP International Conference on Decision Support Systems (DSS2004), Prato, Tuscany, Italy, 2004..

\(^3\) It is worth noting that the present work can be distinguished from research that has investigated sequential decisions made by the same individual. Such decisions have been studied in a wide variety of contexts (e.g., Busemeyer 1982, Mussi 2002, Puterman 1994, Seale and Rapaport 1997, Shanteau 1970, Sullivan et al. 1995). In the present research, we are concerned with each of several decision makers who face the same decision.
identify the private information that led to those decisions. Nor can they observe the benefits accruing to other IT adopters because those benefits take too long to become apparent {Brynjolfsson, 1997 #215; Brynjolfsson, 1998 #154}. Potential adopters thus make use of prior information if they condition their own estimates of the probability merits of A and B on the prior potential adopters’ choices. Therefore, Pr(µ_B) becomes Pr(µ_B|prior decision makers IT adoption decisions) and Pr(µ_A) becomes Pr(µ_A|prior potential adopters’ IT adoption decisions). In other words, if one potential adopter sees a prior potential adopter adopt, then he infers that the prior adopter must have had a sufficiently high opinion of the IT to make that choice.

For the sake of clarity, it is valuable to pause here and consider what the prior means. We follow Abrahamson (1991) and assume that the values of the technologies are determined before decisions are made. Thus, the prior probability of the true value is one, while the prior probability of the non-true value is zero. However, the true value is precisely what is unknown to the potential adopter. Therefore, the potential adopter must form a belief about the prior probability. It is to this belief that we refer when we discuss the prior probability of the value of a technology.

Consider a situation in which each potential adopter faces the same costs and benefits. Denote potential adopter t’s prior beliefs to be Pr(µ_A_t) and Pr(µ_B_t), which will depend on the sequence of adoption decisions that occurred before time t. Potential adopter t+1 will have prior beliefs denoted Pr(µ_A_{t+1}) = Pr(µ_A_t|D_t, D_{t-1}, D_{t-2}, ... D_1) and Pr(µ_B_{t+1}) = Pr(µ_B_t|D_t, D_{t-1}, D_{t-2}, ... D_1), where D_t is the tth potential adopter’s observable decision. By Bayes’ theorem, it can be shown that

\[ Pr(\mu_{A_{t+1}}) = \frac{Pr(\mu_{A_{t}}|D_{t}, D_{t-1}, D_{t-2}, ..., D_{1})}{Pr(D_{t}, D_{t-1}, D_{t-2}, ..., D_{1})} \]  

(3)

and

\[ Pr(\mu_{B_{t+1}}) = \frac{Pr(\mu_{B_{t}}|D_{t}, D_{t-1}, D_{t-2}, ..., D_{1})}{Pr(D_{t}, D_{t-1}, D_{t-2}, ..., D_{1})} \]  

(4)

Substituting these two results into (2) yields

\[ \beta_{t+1} = \frac{Pr(\mu_{B_{t+1}})}{Pr(\mu_{A_{t+1}})} k = \frac{Pr(D_{t}, D_{t-1}, D_{t-2}, ..., D_{1})}{Pr(\mu_{B_{t}}) Pr(D_{t}, D_{t-1}, D_{t-2}, ..., D_{1})} \frac{Pr(\mu_{A_{t}})|D_{t}, D_{t-1}, D_{t-2}, ..., D_{1}}{Pr(\mu_{A_{t}})|D_{t}, D_{t-1}, D_{t-2}, ..., D_{1}} k \]

(5)

where k is the constant determined by the costs and benefits of each outcome (assumed to be the same for each potential adopter), and Pr(µ_B_t) and Pr(µ_A_t) are the prior probabilities assumed by the first potential adopter.

It is important to note that the probability of an observed action, D_t, given a particular distribution, is dependent on all prior decisions (such situations have been referred to as “history-dependent” {Mussi, 2002 #388}). Define A_t as the set of all prior decisions so that A_t = {D_{t-1}, D_{t-2}, ... D_1} for all t>1. Then (5) can be rewritten as
\begin{equation}
\beta_{t+1} = \frac{\Pr(D_1 | \mu_A, A_t) \Pr(D_{t-1} | \mu_A, A_{t-1}) \cdots \Pr(D_1 | \mu_A, A_1)}{\Pr(D_1 | \mu_A, A_t) \Pr(D_{t-1} | \mu_A, A_{t-1}) \cdots \Pr(D_1 | \mu_A, A_1)} \frac{\Pr(\mu_B)}{\Pr(\mu_A)}_k \\
= \left( \frac{\Pr(D_1 | \mu_B, A_t)}{\Pr(D_1 | \mu_A, A_t)} \right) \left( \frac{\Pr(D_{t-1} | \mu_B, A_{t-1})}{\Pr(D_{t-1} | \mu_A, A_{t-1})} \right) \cdots \left( \frac{\Pr(D_1 | \mu_B)}{\Pr(D_1 | \mu_A)} \right) \frac{\Pr(\mu_B)}{\Pr(\mu_A)}_k 
\end{equation}

(6)

This equation is interesting because the portion in brackets is \( \beta_1 \). Notice further that \( \beta_2 \) is the term in brackets multiplied by the last term in parentheses, and \( \beta_3 \) is the term in brackets multiplied by the last two terms in parentheses. This can be expressed more generally as,

\begin{equation}
\beta_{t+1} = \left( \frac{\Pr(D_1 | \mu_B, A_t)}{\Pr(D_1 | \mu_A, A_t)} \right) \beta_t.
\end{equation}

(7)

This means that the decision threshold for any given potential adopter depends on the prior potential adopter’s observed decision, and is, in fact, the prior potential adopter’s threshold multiplied by some factor dependent upon the prior decision. We can also unambiguously show the direction of the change from (5).

Note that the assumption that each distribution is normal with the same variance implies that \( \beta \) is a monotonic function and thus can be inverted. Inverting the \( \beta \) function gives us the observed value that corresponds to each level of the likelihood ratio.

Given the value of \( r \) and the decision, it can be seen that the relevant probabilities are

\begin{equation}
\Pr(b_t | \mu_B) = \int_{r(B_t)}^{r(B_t)} \phi(\mu_B) dr > \int_{-\infty}^{r(B_t)} \phi(\mu_A) dr = \Pr(b_t | \mu_A),
\end{equation}

(8)

and

\begin{equation}
\Pr(a_t | \mu_B) = \int_{-\infty}^{r(B_t)} \phi(\mu_B) dr < \int_{r(B_t)}^{\infty} \phi(\mu_A) dr = \Pr(a_t | \mu_A).
\end{equation}

(9)

Combining this with (5) shows

\begin{equation}
\frac{\Pr(a_t | \mu_B)}{\Pr(a_t | \mu_A)} < 1 \Rightarrow \beta_{t+1} < \beta_t,
\end{equation}

(10)

and

\begin{equation}
\frac{\Pr(b_t | \mu_B)}{\Pr(b_t | \mu_A)} > 1 \Rightarrow \beta_{t+1} > \beta_t.
\end{equation}

(11)

Thus, if potential adopter \( t \) chooses \( b \), then potential adopter \( t+1 \) has a more lax decision threshold than potential adopter \( t \), meaning that potential adopter \( t+1 \) is more likely to choose \( b \) than potential adopter \( t \). Conversely, if potential adopter \( t \) chooses \( a \), then potential adopter \( t+1 \) has a more strict decision threshold than potential adopter \( t \), meaning that potential adopter \( t+1 \) is more likely to choose \( a \) than potential adopter \( t \).

Next, consider a sequence of decisions in which each decision is the same—\( a \) in this case. Then, the \( \beta \) function at any decision time \( t \) can be written as
\[
\beta_i = \left( \prod_{i=1}^{t-1} \left( \frac{\Pr(a_i | \mu_B, A_i)}{\Pr(a_i | \mu_A, A_i)} \right) \right) \left[ \frac{\Pr(\mu_B)}{\Pr(\mu_A)} \right]_k. \tag{12}
\]

After the first three decisions this equals
\[
\beta_3 = \left( \int_{r(r_{t-1}^c)}^{\infty} \phi(\mu_B) \, dr \right) \left( \int_{r(r_{t-1}^c)}^{\infty} \phi(\mu_A) \, dr \right) \left[ \frac{\Pr(\mu_B)}{\Pr(\mu_A)} \right]_k. \tag{13}
\]

This is a particularly difficult equation to solve because the limits of integration for each successive decision are constrained by the \( r \) function of the product of all prior decisions. In general, this can be expressed as a highly complex recursive equation.

The variable of interest is the probability of adoption, which is actually two probabilities—\( \Pr(a | \mu_A) \) and \( \Pr(a | \mu_B) \). As signal detection theory illustrates, a decision to adopt can be a hit or a false positive. The probability of a hit can be represented as
\[
\Pr(a_{t+x} | \mu_{A_{t+x}}) = \int_{r}^{\infty} \phi(\mu_A) \, dr, \tag{14}
\]
where
\[
r = \frac{\ln(\beta)2\sigma^2 - \mu_B^2 + \mu_A^2}{2(\mu_A + \mu_B)} \tag{15}
\]
This means that
\[
r(\beta_{t+x}) = \frac{\ln\left( \prod_{i=0}^{t-1} \left( \frac{\Pr(D_{t+x} | \mu_B, A_{t+x})}{\Pr(D_{t+x} | \mu_A, A_{t+x})} \right) \right) \left[ \frac{\Pr(\mu_B)}{\Pr(\mu_A)} \right]_k}{2(\mu_A + \mu_B)} \tag{16}
\]

As discussed above, in (5), it is straightforward to show how \( \beta \) changes over time, but it is not obvious how the probability of adoption changes over time because (14) does not have a closed form solution. However, we can graphically examine the question and offer a solution for (14). The graphical examination is shown in Figure 3.

The figure is graphed assuming \( \mu_A = 1, \mu_B = 0, \sigma^2 = 1, k = 10, \Pr(\mu_{A1}) = \Pr(\mu_{B1}) = \frac{1}{2} \), and all potential adopters choose to adopt. It can be seen from the figure that the first adoption decision has a tremendous effect on the threshold

---

\( ^4 \) Note that here and in the next equation the bar over \( \mu_B \) refers to the mean of the distribution not the fact that the observation came from the distribution.

Communications of the Association for Information Systems

Style Guide by Paul Gray and Phillip Ein-Dor (Revised 8-17-99)
set by the second potential adopter. This is due to the fact that \( k \) was set very high, meaning that the cost for adopting if \( B \) was true was high.

![Graph showing changes in threshold for consecutive adoption decisions](image)

Figure 3: Changes in threshold for consecutive adoption decisions

Thus, there is an extreme bias toward choosing not to adopt. The value of \( r \) for the first adopter is 2.8. Any private opinion exceeding 2.8 is highly unlikely, but given that it did occur, and thus the first potential adopter chose to adopt, the second potential adopter received a great deal of information. The probability of an observation from the \( \mu_A \) distribution exceeding 2.8 is .036. However, the probability of an observation from the \( \mu_B \) distribution exceeding 2.8 is only .003. Thus, the probability of an observation coming from the \( \mu_A \) distribution conditional on the knowledge that the observation was greater than 2.8 is 12 times the probability of an observation being from the \( \mu_B \) distribution conditional on the 2.8 threshold. Thus, a decision to adopt by decision maker \( t \) changes the prior belief of the \( t+1 \) potential adopter significantly.

Notice also that as the threshold decreases, the amount of information in a positive adoption decision decreases. This occurs because the relative amounts of probability mass to the right of the threshold are very similar, and thus the ratios of the CDFs are close to one. For example, if the second potential adopter, facing the threshold \( r(\beta_{t-2}) \), chooses to adopt, the updated prior for the third potential adopter does not change much. Specifically, the probability that an observation came from the \( s \) distribution conditional on the observation being greater than the threshold is .79, but the probability that an observation came from the \( n \) distribution conditional on it exceeding the threshold is .42. This ratio of 1.9 is considerably less than the ratio of 12 obtained from the first potential adopter’s adoption choice.

We have only looked at the impact of a series of adoption decisions, but we assume that the effects of a series of rejection decisions will be symmetrical.
and will exhibit the same behavior.\(^5\) That is, decisions not to adopt with a strict \((r \text{ is large})\) criterion will not generate a great deal of information, while decisions not to adopt with a lax \((r \text{ is small})\) criterion will generate a great deal of information. The important point from this analysis is that each successive adoption [rejection] decision makes the criterion for the next adopter more lax [strict], thereby reducing the impact of subsequent decisions.

It is worth noting here that the fact that each subsequent decision has a positive impact on the probability of the next adopter making the same decision is the hallmark of herding \(\{\text{Fichman, 2000 \#454; Kauffman, 2003 \#434; Abrahamson, 1991 \#430}\}. Thus, our theory allows for herding even among rational adopters if information about the relative merits of two technologies is poor.

A graph of the marginal impacts of successive identical decisions based on Figure 3 is presented in Figure 4. The figure illustrates a positive but declining impact of subsequent decisions on the probability of choosing to adopt for both the \(n\) and \(s\) distributions. Thus, regardless of the actual distribution, subsequent observed decisions lead to increased probability of making the same decision.

\[\text{Figure 4: Marginal impacts of successive identical adoption decisions}\]

\[^{5}\text{Under certain circumstances, people make different decisions based on whether a stimulus is framed as a gain or a loss (Kühberger, 1995). This “framing effects” behavior is commonly attributed to risk aversion for gains and risk seeking behavior for losses (Kahneman and Tversky, 1979; Levin, Schneider, and Gaeth, 1998). However, it is not clear that framing effects would apply in the standard way for sequential decisions. For example, would a series of prior positive (or negative) adoption decisions be viewed as a gain frame or loss frame to a current potential adopter? Although asymmetries in adopters’ reactions to positive and negative information are theoretically possible, it is not clear whether and how such reactions would apply in the current context. If they do apply, it seems likely that the risk aversion or risk-seeking behavior would result in slight changes in the placement of thresholds rather than shifts that would alter the basic structure of our model.}\]
The probability of making the same decision increases with the number of identical decisions, but it increases at a decreasing rate. Thus, from a theoretical perspective, it is interesting to ask whether the probability of making the same decision as others converges to some value. In other words, is herding absolute, as herding literature often assumes? Smith and Sørensen (2000) show that under certain conditions the probability of making a particular decision does converge. From an applied perspective, it is more interesting to examine the convergence path. The rate and reliability of the convergence path determine how organizations can make use of this theory in the real world. The behavior at \( n = \infty \) is irrelevant to any real world application and, pragmatically, potential adopters will probably have trouble incorporating more than seven plus or minus two pieces of information into a real decision {Miller, 1956 #437}. Thus, two important research questions are:

**Research question 1**: Do the decisions of sequential potential adopters converge?

**Research question 2**: What is the convergence path?

Herding is fickle if it is based on fashion alone, but rational herding based on information aggregation may not be. When one potential adopter decides against a stream of identical decisions, she changes the threshold in the opposite direction. Thus, if potential adopter \( t \) had a lax decision criterion and still failed to adopt, potential adopter \( t+1 \) would adopt a strict decision criterion. As noted above, with a lax decision criterion a positive adoption decision is not very informative, but a negative adoption decision contains a great deal of information. The magnitude of the change depends on the magnitude of the decision criterion. Based on the fact that subsequent identical decisions quickly move the threshold toward a bias for making the same decision, it can be seen that contrary decisions have more impact than confirmatory decisions.\(^6\) However, the magnitude of contrary decisions is not clear. Thus, it is useful to consider the impact of a contrary decision on the convergence path of sequential decisions. This yields our third research question.

**Research question 3**: What is the effect of contrary decisions on the convergence path?

Innovations often diffuse through social networks based on strong communication channels {Fichman, 2000 #454;Abrahamson, 1997 #431;Rogers, 1995 #360;Zmud, 1983 #453}. Thus, it is important to investigate the effects of our adoption theory if adopters are in closed groups (e.g., industries, geographies, social groups) rather than in a totally open environment in which everyone can see everyone else. Our fourth research question is:

---

\(^6\) It is worth noting the relevance of the present discussion to positivist science and decision-making behavior generally. A single confirmatory observation often makes a great deal of difference in driving a conclusion, but adding additional confirmatory observations makes increasingly less difference to the conclusion. Confirmations quickly lead to asymptotic confidence in a conclusion in many situations. This line of reasoning also explains why science seeks to disconfirm conclusions rather than confirm them.
Research question 4: What is the effect of groups of potential adopters on the number of correct decisions?

III. RESEARCH DESIGN

Simulation is often used to examine herding models of technology diffusion (Abrahamson, 1997 #431; Abrahamson, 1991 #430). This is the case because the complex mathematical modeling has no closed form solutions. Thus, typical derivatives cannot be calculated analytically. Instead, we specify the parameters of the model and solve it many times. We then change the parameters and repeat the process. When we graph the changes in the simulated behavior for different levels of parameters, we offer a graphical representation of the effects of those parameters.

Our purpose with the simulation is to show the implications of the model for various levels of the parameters. Thus, the simulation incorporates both implications and sensitivity. For example, different technologies will surely have different costs and benefits, and hence different values of $k$. To establish how the model changes with these different levels of $k$ we simulate a wide range of $k$s in Figure 6 (discussed below). By plotting several levels of $k$ on the same graph, we show how the model works at different levels of the parameter and develop an understanding of the sensitivity of the model to changes in the parameter.

Throughout, we drew the observed values from the $s$ distribution. Both $s$ and $n$ were normally distributed with variances of one and means of one and zero, respectively. The results are identical with respect to any scaling that preserves the measure $d'$ (see Swets and Green 1966). If $d'$ increases then the convergence will be much quicker because the confidence with which a value is attributed to a particular distribution will be higher. The reverse is true if $d'$ decreases. The priors $Pr(e)$ and $Pr(i)$ were both 0.5, indicating that in the absence of any other information, both were equally likely. We varied $k$, the relative cost and benefit, thereby varying $\beta$. We note the value of $k$ below each graph.

IV. DATA AND RESULTS

CONVERGENCE

We begin by addressing our first two research questions, which concern whether the decisions of sequential potential adopters converge and what the convergence path looks like. The answers are not intuitive from the equations, and so bear testing. To answer the first question, we ran a string of 1000 repetitions of 100 decisions. The results of this simulation are shown in Figure 5.

The figure shows two interesting characteristics. First, there is not convergence to the correct decision after 100 decisions. Specifically, 96.4% of the 100th potential adopters chose the correct distribution. The second item to notice is the jaggedness of the line, which indicates that even at 100 decisions, there are still reversals. For example, 96.3% of the 98th potential adopters made the correct decision, so in one trial “someone” reversed the prevailing decision at
the 99th decision. Note also that even at the extremes, potential adopters may reverse in the incorrect direction. Between the 91st and 92nd decision, 0.3% of the potential adopters reversed in the incorrect direction.

Figure 5: Convergence of Decisions 1000 samples of 100 decisions $k = 1$

This suggests that several forces are at work, making convergence a more difficult issue than previously believed. We should be concerned not with absolute convergence, but with the speed of convergence. This depends on the relative costs and benefits of making correct or incorrect inferences about the merits of A and B. The variable $k$ captures these costs. Varying $k$ produces different rates of convergence, as shown in Figure 6. Note that the observations are actually coming from the A is better distribution, so the correct decision is $a$.

Figure 6: Effect of Costs on Convergence in 1000 samples of 100
Notice that when \( k \) is very high, it is not clear that there is any convergence to the correct decision. Even with moderate values of \( k \), many potential adopters make incorrect decisions even after viewing 99 prior decisions. For example, at \( k = 5 \), the 100th potential adopter is only 78.5% likely to make the correct decision and 21.5% likely to make the wrong one. Even if \( k \) is small, we still do not achieve convergence after 100 decisions.

If technology A is perceived to be highly beneficial, and costs associated with technology B are perceived to be low, then potential adopters will have a low \( k \), which will lead to a bias toward A. This is good if A is, in fact the better technology. However, B is the better technology, then it can lead to many incorrect decisions even after additional information is incorporated by decision makers. This is particularly troubling if B is the status quo, because it leads to a bias toward adoption {Abrahamson, 1991 #430}. If this bias is present, it can help explain why there are so many failed IT implementations. Recently, however, perceptions of the benefits of IT seem to have become more negative {Carr, 2003 #459}. If this is the case we may experience a period of underadoption of technically efficient IT.

If 100 decisions are not enough to achieve convergence to the correct decision, then perhaps more repetitions are necessary. To test this question, we ran simulations of 1,000,000 decisions, as shown in Figure 7.

Even at 1,000,000 decisions, potential adopters make incorrect decisions. At the same time, the number of correct decisions at 100,000 decisions is different than that at 1,000,000, suggesting that even after 100,000 observations, potential adopters may reverse if they receive a private signal that is extreme enough.

![Figure 7: Convergence of Decisions 1000 samples of 1,000,000 decisions k = 5](image)

We have kept \( k \) constant across all potential adopter in a particular simulation. This implies that the costs and benefits for every potential adopter
are similar. This is clearly an unrealistic assumption\textsuperscript{7}. Therefore, we explore the impacts of different \( k \)s in Figure 8. We specify \( k = 5 \) for the first 10 decision makers, and then \( k = 100 \) for the remainder of the decision makers.

Figure 8: The effect of a change in \( k \) at the 11th decision in 1000 samples of 100

The graph shows that increasing \( k \) increases the threshold and thereby reduces the probability of making the adoption decision for the 11\textsuperscript{th} and following adopters. However, it does not reduce the probability to near zero as it did when \( k \) was fixed at 100 for all decision makers. Moreover, the probability of adopting continues to increase after the 11\textsuperscript{th} adopter, whereas it remained constant when all adopters had a \( k \) of 100. This occurs because the first ten adopters had a low enough threshold to allow some information into the system. Though the 11\textsuperscript{th} adopter is faced with a very high \( k \), he also has enough information about other adopters' signals to infer a low \( \Pr(i)/\Pr(e) \), so he is relatively confident that adopting the technology is a good idea even if his own costs of mistakes are high.

This helps explain why, for example, firms give away trial copies and why it is so important for vendors to work closely with partners when first introducing a new technology. With a new technology, potential adopters may have a high value of \( k \) because the cost of adopting A if B is better is high (\( a|\mu_B \)). If this is the case, potential adopters are likely not to purchase the technology and no information about it will enter the system. This in turn will cause more potential adopters to pass on the technology. However, if the vendor of a technology can somehow mitigate the downside for a few initial firms (either by giving away trial copies or by working closely with the initial adopters) then it may be able to demonstrate that A is better and overcome future adopters' concerns (particularly \( a|\mu_B \)).

\textsuperscript{7} We thank an anonymous reviewer for suggesting that we address this issue.
Impact of Contrary Decisions

Like fashion-based herding, rational herding can be fragile. This fragility occurs because no new information enters the system (e.g., in the Bikhchandani, Hirshleifer, and Welch (1992) model). However, in our sequential adoption theory model, potential adopters may receive extreme private observations that result in decisions that are contrary to the prevailing cascade. Thus, there may be another type of fragility, and we investigate this fragility in our third research question concerning the impact of contrary decisions on the convergence path.

Recall that if someone behaves according to the prevailing cascade, there is not much information in his decision; however, if he reverses the cascade, there is a tremendous amount of information revealed. It may be the case that a single reversal reveals as much information as all prior confirmatory decisions. For example, many practitioners and researchers have hypothesized a Kodak effect for the adoption of IT outsourcing (Loh, 1992 #38). The notion is that IT outsourcing was unpopular before 1991, but then Kodak decided to outsource its IT to IBM and several other companies. This single event led other companies to change their opinions of the relative merits of outsourcing and in-house development. Thus, it is important to determine the effect of contrary decisions on the behavior of the system. We looked at this effect by introducing a contradictory decision at the 25th decision. Specifically, whatever the 24th potential adopter did, the 25th did the opposite.8 The results of this simulation are displayed in Figure 9.

Figure 9: Effects of Contradictory Information in 1000 samples of 100, k = 1

The figure shows that the impact of contradictory information is tremendous. The percentage of correct decisions dropped from 93% to 59%

---

8 We use the 25th decision because in both the 24th and 25th decisions 92.9% of the individuals made the same choice; thus, by using the 25th decision we control for extreme observations in the 24th decision.
from the 24th to the 26th potential adopter\(^9\). The mean threshold changed from –1.60 to 0.78. This is important, considering that the first potential adopter was 70% likely to be correct and had a threshold of 0.50. This means that the contradictory information of the 25th potential adopter more than reset the information in the string of decisions. This occurred because the 26th potential adopter knew that the 25th potential adopter was aware of all of the prior adoption decisions and had a threshold value larger in absolute value than any prior potential adopter (-1.62 on average). Thus, for the 25th potential adopter to reverse the cascade, he must have received a very extreme private observation. Specifically, an observation of less than –1.62 is twelve times as likely to have come from the B is better distribution as from the A is better distribution. Therefore, the 26th potential adopter is much less likely to follow the first 23 decisions than was the 24th potential adopter.

**Distribution of Potential Adopters**

The impact of groups of potential adopters is investigated by our fourth research question. It is particularly useful to consider group decision making from the perspective of the decision motivator. In the case of a novel technology, the decision motivator is the creator of that technology. From this perspective, the important question concerns the distribution of possible outcomes. We examined this issue by separating 100 potential adopters into groups and repeating the simulations, each with 1000 samples. A group is a set of potential adopters who can observe one another’s actions. For example, if the group size is ten, then the second potential adopter can observe the action of the first, and the tenth can observe the actions of the first through the ninth. However, the eleventh can observe no other potential adopter, but the twelfth can observe the eleventh and the twentieth can observe the eleventh through nineteenth. We are interested in the aggregate behavior of 100 decision makers if they are able to observe certain other potential adopters. A histogram of the decisions is presented in Panel 1. Each histogram shows the outcomes of the group’s or groups’ decisions for 1000 simulations.

---

\(^9\) Note that the probability the 25th decision maker was correct was 7% and the probability of him being wrong was approximately 93% because he was forced to behave in an opposite manner to the 24th decision maker. Thus, the important difference is between the 24th and 26th decision makers.
Panel 1: Distributions of decisions with different group sizes

The panel indicates that introducing the sequential effects causes the distribution of decisions to become bi-modal as the number of potential adopters increases. Increasing the number of potential adopters participating in a sequence increases the mean of the distribution, but it also increases the chance of having all potential adopters make an incorrect decision. Thus, the value of introducing a decision to a single large group or many small groups depends upon the risk propensity of the decision motivator.10

10 The benefit to a particular event is \( b = \sum a \text{ prob}(a) * U(a) \), where \( a \) is a specific outcome and \( U \) is a function describing the individual’s utility for that outcome. The second derivative of \( U \) reveals the person’s risk attitude. If one is risk averse then \( U'' < 0 \), so that doubling the outcome \( a \) more than doubles the utility. Organizing people into different sized groups changes \( \text{prob}(a) \) as shown in the panel (which are frequency plots of \( \text{prob}(a) \)). Obviously, changing \( \text{prob}(a) \) changes the overall benefit, but the nature of the change depends on the nature of the individual’s utility \( U \). If one is risk averse, then the increased probability of very poor outcomes associated with large groups may outweigh the increased probability of very good outcomes, leading one to prefer small groups. However, if one is risk neutral, then the increased expected value of large groups will lead one to prefer large groups. Thus, the optimal group size from the decision motivator’s point of view depends on his or her attitude toward risk.
V. DISCUSSION

When IT adopters are faced with a new IT in an information-poor environment, they use the behavior of other adopters to infer additional information about the adoption decision. This results in several outcomes. First, our results show that potential adopters tend to imitate one another, so that the majority of potential adopters make the same decision. The implication of this is that agents will often jump on bandwagons, exhibiting fad-like behavior, in response to limited information about the relative merits of new IT.

Anecdotal evidence is available for this finding. On the back cover of the July 26, 2004 issue of Forbes magazine an advertisement for Oracle’s E-Business Suite appeared. The ad read, “E*Trade Financial Runs The Oracle E-Business Suite. The Best Companies Run Oracle Applications.” Oracle also has another ad using 1-800-Flowers, and recently SAS has offered a similar ad. This suggests that technology vendors are behaving according to our theoretical framework. These ads simply offer the example of a single adopter of the technology; they do not mention the system’s cost or give any indication of its potential benefits. Nor do they offer any comment on the system’s usefulness or ease of use. If the reasoning for the Oracle ad were network effects, then the ad would have listed the number of adopters. Likewise, if the motivation were legitimacy, the ad should have listed the names of many adopters. However, our theory suggests that the information contained in a single adoption is significant, and that additional adoptions provide very little additional information to the decision maker, so listing only a single company’s behavior should encourage further adoption.

Second, our results show that imitation in adoption decisions can be incorrect. Thus, we answer Abrahamson’s (1991) call to investigate the diffusion of technically inefficient technology and the rejection of technically efficient technology. However, both incorrect and correct fads can be reversed by a sufficiently extreme private signal. Thus, the beginning and end of a fad are preceded by unexpected behavior of a pioneering decision maker.

Placement of the adoption threshold also has important implications from the technology vendor’s point of view. For example, our simulation showed that if a technology is particularly risky because the benefit of non-adoption is much larger than the benefit of adoption if the technology turns out to be bad, and/or the benefits of adoption and non-adoption are very similar if the technology turns out to be good, then early adopters’ thresholds will be too high for them to adopt and very little information will enter the marketplace. Further, later adopters will face the same situation and not adopt. A technology vendor can solve this problem by subsidizing the benefits of adoption if the technology turns out to be bad, usually by reducing the costs of acquisition and/or implementation. This reduces the threshold and encourages early adoption, which increases the amount of information in the marketplace and facilitates later adoption by additional firms. Thus, vendors may need to subsidize early adopters to start a sequential adoption process {Economides, 1996 #349}.

Another implication of our analyses is the rationality of fads. Although fads have a negative connotation, many fads are good (e.g., food companies’
fortification of foods with vitamins) and people who follow fads often fare better. The cumulative knowledge of many prior adopters is typically greater than the knowledge of any one potential adopter (Surowiecki, 2004 #438), and this cumulative knowledge thus should not be ignored. However, it is important that potential adopters understand that each prior potential adopter’s decision was based only partially on her own knowledge. Further, the private observation of each decision maker must be weighed carefully against the collective wisdom of the previous adopters. Finally, as noted above, the relative costs and benefits of all the consequences listed in Table 1 should be considered. However, the bottom line is that if many other organizations are adopting a particular technology, then that technology deserves careful consideration.

Our results also show that stable equilibria do not seem to exist for reasonable numbers of potential adopters. Instead, potential adopters tend to fall quickly into a semi-stable equilibrium with occasional extreme events reversing that equilibrium. This suggests that we should not think of the presence or absence of fads, but of the degree of fad.

We also found that extreme private signals more than reset the information framework. An observation extreme enough to make a rational potential adopter reverse a fad must contain more information than all previous decisions. Thereafter, subsequent potential adopters are more cautious (this is the reverse of the Kodak effect). This is clearly a rational reaction. If people are jumping on a technology bandwagon, and then someone chooses not to adopt, it must be because he has more and/or better information than the previous potential adopters. Thus, later potential adopters adjust their thresholds appropriately to account for this contrary decision.

Another important finding is that the introduction of groups of potential adopters makes the distribution of decisions bi-modal. These results (shown in Panel 1) are consistent with findings in behavioral decision making research. Groups often make better decisions than individuals due to increased discussion and differing points of view. This is consistent with the higher means of correct decisions by groups in Panel 1. However, groups also are more likely to make extreme decisions, both good and bad, due to phenomena such as group polarization, in which the dominant view in the group is adopted and then rationalized and justified by the group members {El-Shinnawy, 1998 #461;Isenberg, 1986 #462;Lamm, 1988 #460}. This group rationalization process can lead to both good decisions (for example, groups agreeing to high levels of charitable giving {Muehleman, 1976 #463}) and sometimes to extremely bad decisions (as, for example, in groupthink decisions and mob violence).

When there are many small social groups that are essentially independent, it is very unlikely that all the groups will make the same correct (or incorrect) decision. Instead, some groups will adopt a particular IT and some groups will not, subject largely to random events that gave the first adopters in that group a strong positive (or strong negative) signal. This could explain the resilience of technologies that clearly do not perform as well as rivals. Absent illegal behavior or overwhelming incentives, local preferences and random events generally prevent a single technology from dominating all others when numerous
independent groups are operating in the environment. This may also explain why Oracle ran its ads using several different companies. E*Trade is a financial services firm, while 1-800-Flowers is a retail firm. Thus, each ad targets a different group.

Many directions for future research are suggested by our theory and findings. For example, we have assumed that potential adopters can observe some characteristics of the IT that hint at its nature (adopters’ private observations). However, we have not investigated those characteristics. Although this allows the model to be generalized to a number of useful theories, it is also important to consider what characteristics potential adopters evaluate in a technology. Is ease of use or complexity or network size important, or are there other factors? Are these indicators different across different technologies? Are some better indicators than others? There are a variety of theories that are relevant to this issue and that could be incorporated into sequential adoption theory.

Relatedly, it is interesting to ask whether everyone observes the same factors. Currently, we assume that the opinion is a function both of the data and the observer, but it can be useful to separate these two concepts. Intuitively, we would expect that the value of others’ actions would increase if they had observed different characteristics because it would allow for better discrimination. However, this depends on whether and how well different characteristics predict the same outcome.

It would also be interesting to see how potential adopters assess the benefits for the different outcomes in Table 1. There may be systematic biases that potential adopters introduce into their judgments of benefits. IT adopters in particular may place more negative value on failing to adopt a good technology than is warranted {Hitt, 1998 #210}. It is easy to envision this bias as a partial explanation for the dotcom boom and bust.

In sum, sequential adoption theory offers an important explanation of the behavior of sequential, similarly-situated potential adopters using a rigorous mathematical foundation based on established theory. Our theory offers both researchers and practitioners a valuable tool for understanding sequential adoption of technologies and other artifacts.

VI. REFERENCES


Communications of the Association for Information Systems

Style Guide by Paul Gray and Phillip Ein-Dor (Revised 8-17-99)


Miller, G. A. (1956) "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information," The Psychological Review (63) 2, pp. 81-97.


VII. ABOUT THE AUTHORS

Glenn J. Browne is an Associate Professor at the Rawls College of Business Administration at Texas Tech University and Director of the Performance Economics Institute Research Center and the Institute for Internet Buyer Behavior. He received his Ph.D. from the University of Minnesota. His research focuses on information requirements determination for systems development, consumer behavior issues in e-commerce, strategies for enterprise regeneration, and behavioral decision-making processes. His articles have appeared in Management Science, Organizational Behavior and Human Decision Processes, the Journal of Management Information Systems, and other journals.
Eric A. Walden received his Ph.D. at the University of Minnesota and is currently an Assistant Professor at the Rawls College of Business Administration at Texas Tech University. His research interests include electronic commerce, IT outsourcing, interorganizational relationships, and decision-making processes for technology adoption from a rational perspective. Prior research has appeared in *MIS Quarterly, Information Systems Research, The Harvard Business Review, The International Journal of Electronic Commerce*, and other journals.