[Journal of the Association for Information Systems](https://aisel.aisnet.org/jais)

Volume 10 | Issue 1 Article 1

1-24-2009

Sequential Adoption Theory: A Theory for Understanding Herding Behavior in Early Adoption of Novel Technologies

Eric A. Walden
Carnegie Mellon University, eric.walden@ttu.edu

Glenn J. Browne
University of Virginia, gbrowne@virginia.edu

Recommended Citation

Walden, Eric A. and Browne, Glenn J. (2009) "Sequential Adoption Theory: A Theory for Understanding Herding Behavior in Early Adoption of Novel Technologies," *Journal of the Association for Information Systems*, 10(1), .
DOI: 10.17705/1jais.00181
Available at: https://aisel.aisnet.org/jais/vol10/iss1/1

Journal of the Association for Information Systems
JAIS

Research Article

Sequential Adoption Theory: A Theory for Understanding Herding Behavior in Early Adoption of Novel Technologies*

Eric A. Walden
Rawls College of Business Administration
Texas Tech University
and
Tepper School of Business
Carnegie Mellon University
eric.walden@ttu.edu

Glenn J. Browne
McIntire School of Commerce
University of Virginia
gbrowne@virginia.edu

Abstract

Technology adoption often occurs sequentially, so that later potential adopters can see the decisions (adopt or not adopt) of earlier potential adopters. In this paper we review the literature on observational learning, in which people use information gained by observing the behavior of others to inform their decisions, and note that little prior research has used an observational learning perspective to understand the adoption of information technology. Based on theory and previous literature, we suggest that observational learning is likely to be common in adoption decisions. We develop a model that extends existing observational learning models and use simulation to test the model. The results suggest that following the behavior of other similarly-situated decision makers can be a very useful strategy in adoption situations in which there is a great deal of uncertainty. Implications for research and practice are discussed.

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

* Rob Fichman was the accepting senior editor. Wenjing Duan, Wonseok Oh, and Chris Forman were the reviewers. This was submitted on May 27, 2006 and went through three revisions.

Volume 10, Issue 1, pp. 31-62, January 2009

Sequential Adoption Theory: A Theory for Understanding Herding Behavior in Early Adoption of Novel Technologies

1. Introduction

Sequential decisions under conditions of uncertainty are made in a variety of business and personal domains. People often observe the behaviors of similarly-situated others to aid them in choosing whether to pursue a course of action. One decision of this type that is of particular importance in the information systems (IS) area is technology adoption. In the present research we use theories of observational learning and sequential decision making to develop a new model that extends existing theory and aids in understanding technology adoption decisions.

Observational learning is one of the most ubiquitous and useful means of decision making available to humans. Observational learning occurs when one person observes the behavior of another person and infers something about the usefulness of the behavior based on that observation. A significant body of research has developed in the area of observational learning, including models and empirical findings. Research has shown that people use their observations of others to update their own private beliefs and to take actions (Bandura, 1977), that bandwagon or herding effects occur (sometimes inappropriately) (Abrahamson, 1991, Abrahamson and Rosenkopf, 1997, Oh and Jeon, 2007), that such effects often occur rapidly (Gale and Kariv, 2003), and that such effects are often fleeting and can be reversed fairly quickly and easily (Bernardo and Welch, 2001).

Decisions involving technology are especially prone to observational learning because they are fraught with complexity and uncertainty. Software and other technological components are among the most complex artifacts humans build (Brooks, 1975). Moreover, the impacts of technologies can take years to be realized (Brynjolfsson and Hitt, 1996), so the benefits of adoption decisions are often uncertain. It is precisely this complexity and uncertainty that make observational learning so appealing. Because IT adoption does not usually occur by accident, all of the information an individual has about the adoption decision is consolidated and expressed in his or her behavior. Even if the individual has very little information or is simply jumping on a bandwagon, his behavior still consolidates and expresses what little information he has available. Therefore, 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. Thus, technology decisions, because of their complexity and uncertainty, can be made much easier by observing and utilizing the behavior of others.

There is evidence that observational learning influences technology adoption both in the laboratory (Song and Walden, 2003) and in real world financial markets (Walden and Browne, 2008). Song and Walden (2003) found that subjects' willingness to adopt a technology was significantly related to others' decisions. Walden and Browne (2008) found that stock market reactions to electronic commerce announcements by firm X on day t were strongly predicted by electronic commerce announcements by firm Y on day t-1. In both cases, decision makers showed behavior that suggested they were incorporating the behavior of others into their decision making.

Thus, it is worthwhile to develop a rigorous model of how observational learning impacts technology adoption and to apply this model to develop insights into observational learning issues. In the present research, we develop a theoretical extension of the observational learning model of Bikhchandani et al. (1992). Our extension is particularly useful in technology adoption situations because of (1) the significant uncertainty surrounding emerging technologies, (2) the increasing emphasis on quick technology choices, (3) the path dependency of many technology adoption decisions, and (4) the long lag times between adoption and financial return on a technology. We use the model to develop new insights into situations in which observational learning is a key component of technology adoption decisions. Specifically, we make predictions about the impact of observational learning on the accuracy of adoption decisions, on the relative market share of different technologies, and on the market share of technologies when those technologies are aimed at different sizes of groups.

We address four research questions in this area. First, do adoption decisions converge? That is, if there is little information in the environment, do people simply follow one another and essentially ignore the small amount of information they personally have? Second, we examine the convergence path. How quickly, if at all, do potential adopters start following the decisions of others? Third, we examine what happens when there are decisions that reverse prior strings of identical decisions. Fourth, we examine what happens when decision makers are put into groups of different sizes and allowed to observe the behavior only of those decision makers within the group.

The paper is organized as follows. First, we present background material on technology adoption and important theories explaining imitative behavior. We then develop our model of sequential adoption theory and present our research questions. We then describe our research design, followed by the findings from a simulation study. We conclude with discussion and directions for future research.

2. Background

There are numerous situations in IS research and practice in which members of a community make sequential decisions and can observe the decisions (but not the reasoning) of others. As noted, research has documented observational learning in IT adoption decisions. For example, when asked to evaluate peer-to-peer file sharing technologies, people indicate a higher intention to adopt the technologies if they observe others adopting them even after controlling for network size (Song and Walden, 2003). Research has also shown that observational learning is even more important than professional product reviews for explaining online software downloading (Duan et al., 2009). In the domain of 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 for which others are bidding (Dholakia and Soltysinski, 2004). Purchasing decisions for e-commerce initiatives in financial markets represent another example in which individuals can see bids (and offers) over time, but not the reasons for those bids and offers. There is evidence that the prices 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 (Walden and Browne, 2008).

One of the findings of IT adoption research is that adopters often adopt a technology en masse (Li, 2004). A variety of reasons exist for why many people might make the same decision. First, it might simply be an obviously good decision; for example, adopting the telephone was superior to traveling long distances to communicate a message or sending it via the pony express (at least for many messages). Similarly, computers are both faster and more accurate (and currently less expensive) than human payroll calculators, so many firms use computers to calculate payroll. If a clearly good technology choice presents itself, then we would expect it to be adopted. In such cases, the good choice explanation focuses on the relative value of choices to individuals, and may or may not make reference to the behavior of others. Sometimes this explanation is satisfactory, but other times the behavior of others is particularly salient.

The behavior of others is relevant when acting similarly to others offers some type of benefit. Thus, a second reason that people may make the same decision is that firms may gain social benefits from following the behaviors of other firms. The notion is that a firm that is like other firms is better able to navigate its institutional environment. In such cases, there may be no intrinsic benefit, only social benefit. That is, in the absence of pressure from other firms (e.g., Wal-Mart mandating its suppliers to adopt radio frequency ID tags), the decision is not necessarily “good” from a rational point of view. In fact, IT researchers have found that managers “replicate the selection decisions of other firms even if they believe the copied choices to be inferior or suboptimal” (Tingling and Parent, 2003, p. 114). This explanation focuses on norms and social pressures rather than on the intrinsic benefits of the technology itself.[1]

An important theory for examining situations in which the behavior of others is relevant is network effects theory (Katz and Shapiro, 1985, Katz and Shapiro, 1986). Network effects occur when the value of belonging to a network is a function of the number of others who belong to the same network.

[1] A closely related field of inquiry is heterogeneous diffusion models (Greve 1995; Greve, H. R., D. Strang, and N. B. Tuma 1995). These models provide an empirical method of estimation for social diffusion processes.

This is usually applied to technology adoption by noting that the technology gives access to the network. Good examples include fax machines (Economides and Himmelberg, 1995), computing networks such as BITNET (Gurbaxani, 1990), and ATM networks (Kauffman et al., 2000). In this case, the behavior of others matters because when others join the network by adopting the technology, it increases the value of the technology. This theory focuses on the value of the technology, and the only uncertainty is how others will behave in the future.

Observational learning literature offers another perspective on mass convergence toward a technology. Observational learning suggests that people can augment their own incomplete information by observing the behavior of others. For example, one can learn about a technology by observing others' technology adoption behaviors. It is worth noting that this point of view can supplement the other explanations discussed above. For example, when a person sees others adopting a technology, he might think, "If I adopt that technology then others will like me" or "If I adopt that technology then I can interact with others." However, the observational learning perspective adds the possibility that the person may think, "If others are adopting it, then I should conclude its inherent value is higher than I previously thought." This is similar to mimetic isomorphism (DiMaggio and Powell, 1983) in the strategic management literature.

Information cascade theory (Bikhchandani et al., 1992) was developed to explain the consequences of learning from the behavior of others and is particularly useful in the present research. The underlying notion is that individuals each hold some private information that can be thought of as a signal about the utility of a course of action. The signals are not perfect, so individuals must make their decisions under uncertainty. The decisions, but not the signals, are observable by other decision makers who then use Bayesian updating to revise their beliefs about the appropriate course of action. Information cascade theory has been demonstrated in various laboratory experiments. For example (Anderson, 2001, Anderson and Holt, 1996, Anderson and Holt, 1997), experimenters have presented subjects with an opaque container that either holds two red balls and one green ball or two green balls and one red ball. The subject's task is to decide whether the container from which he is sampling is the two red-one green container or the two green-one red container. 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 (presumably) the decision(s) of the prior subject(s).

Based on the key assumptions of uncertainty about the value of a technology, private information, observability of prior decisions, and rationality in the form of Bayesian updating, several results emerge from these experiments. The first is the tendency of decision makers to "herd." Herding means that all decision makers rapidly converge toward 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 following the lead bird or fish. 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 small piece of new information. This occurs because the entire group's decisions are based on the relatively little information encapsulated 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 fourth subject faces exactly the same information environment as the third. Thus, he must rationally say "red" regardless of his own private information. The cascade is fragile because every person knows that he is deciding based on very little information, and it does not take a great deal of contradictory information to change his mind. The result is that the group as a whole seems flighty—rapidly achieving conformity and then easily reversing its decisions when small amounts of contradictory information are presented—even though each individual is behaving in a fully rational manner.

Information cascade theory provides a complement to the other theories suggested above because the causal process is different. Decision makers follow each other to aggregate information in an uncertain environment. Thus, information cascade theory can explain situations that do not fit the assumptions of the other theories. It can explain situations in which the benefit of a decision does not depend on the number of others making the same decision, or when a particular decision does not lead to legitimacy, or when people have different and limited information. The theory seems particularly suited to the early stages of technology adoption, when there are not enough adopters for network effects to be relevant or future network sizes to be estimated. Further, at this stage there may be too few people making the adoption decision for the technology to be popular enough to precipitate social benefit-based behavior. Of course, in the early stages of adoption, information is poor and decision makers may have very different signals about the utility 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 and deriving benefits from the decisions of others.

In the present research, we enhance the applicability of information cascade theory to IT adoption by developing a more general model of the phenomenon that can be used to develop novel insights. As noted above, there are reasons for exhibiting similar behavior in addition to observational learning. Our model represents an intermediate step between Bikhchandani et al.'s (1992) model and a grand unified model that includes information cascades, network effects, and social benefit-based herding. We accomplish this by expanding the nature of the signal decision makers receive from the environment, so that network effects, social benefit-based herding, and other effects may be included in the signal. Thus, our model illustrates how researchers can expand the nature of the signal to include other factors that might be important to the adoption decision and is a step in the process from reductionism to holism.

In addition, we are able to investigate important IT-related questions. In particular, IT adoption is often of interest from the point of view of the seller of IT, and the questions the seller might ask concern how to influence the system rather than how the system will behave. To this end, we investigate the finite steps in the adoption path rather than the adoption equilibrium that occurs in the limit. Because prior work in this area has been from an economics perspective, it has focused on the limiting equilibrium, minimizing the consideration of the path adopters might follow to reach the equilibrium. For IT adoption in particular, the path adopters follow is a serious consideration for two reasons. First, there is often a finite number of adopters, so it is not clear that asymptotic results are generalizable. Second, technology changes very rapidly, so it is not clear there is time to reach a stable equilibrium.

Bikhchandani et al.'s (1992) model, with its focus on binary signals, provides an excellent example of the problem and its limits but is too coarse to give a good illustration of the finite sample properties. In particular, their model says that if two people adopt a product or technology, then everyone else will also adopt. In contrast, our model allows us to understand the change in probability if two people adopt as well as the rate of convergence of probability (Bikhchandani et al.'s (1992) model only allows probabilities of 0 percent, 50 percent, or 100 percent, which is useful for theoretical illustration but less useful for real world applications). Similarly, Smith and Sørensen (2000) highlight this point when they state about their own article, "This paper is unified by two natural questions: (i) What are the robust long-run outcomes of learning in a sequential entry model with observed actions? (ii) Do we in fact settle on any one?" (p. 372). In the present research, on the other hand, we are very concerned about the *short-run* outcomes.

Finally, we ask a question that Bikhchandani et al. (1992) could have answered, but did not ask. We ask what happens if an IT vendor splits adopters into different subgroups and how such groupings impact the vendor's risk (i.e., the variance of adoption). Other models have only been interested in what happens in the limit, and so have never considered how splitting a finite number of adopters into smaller subgroups might change the dynamics of the system.

3. Model

Our goal is to produce a rigorous mathematical model of information cascades that can be applied by IT researchers. Information cascade theory is not very amenable to direct application, and making minor changes is not a trivial task. We believe this has retarded the growth of this very important aspect of adoption. Thus, we aim to create a model that is both deployable and easily adjustable.

Our model makes various simplifying assumptions, which we detail below. The model is valid to the extent that the assumptions either mirror reality or do not make a significant difference in this context. Moreover, these assumptions give researchers a starting point for adjusting the model to fit different circumstances. The assumptions are as follows:

Assumption #1: Decision makers must choose between adoption of two technologies called A and B. Some examples might include PC or Macintosh, open source or proprietary software, or peer-to-peer or mainframe computers. The model might also represent a choice between adoption and the status quo (i.e., no adoption). The limitation of this assumption is that it only considers two choices, even though some situations will, of course, contain many choices.

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 difference between A and B. If A is the 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 from which distribution the signal came and, hence, which technology is the better choice.
By "signal" we mean the overall perception a decision maker has about the merits of a technology. We characterize it as a number for the sake of simplicity, but it is, in fact, a pattern of neural activation based on the information available to the decision maker and the decision maker's specific knowledge.

For technology adoption, there is usually no shortage of information; in fact, the opposite is true. Given the vast amounts of information available today, whether from the Internet or myriad other sources, it is safe to assume that different decision makers receive different information. For example, receiving information about an SAP product from the SAP company is probably different from receiving it from an SAP consultant, which is different from receiving it from the representative of a company with a competing product. However, even if two decision makers receive the *same* information about the technologies in question, they still may interpret it differently because of their knowledge bases. In other words, two people sitting in the same room receiving the same information from the same person at the same time about the same technology may have two very different perceptions about the technology. For instance, if one decision maker knows the speaker works for the company and the other thinks the speaker is an independent researcher, then the two will reasonably have different perceptions of the information they receive.

Representing the precise nature of a signal, either mathematically or psychologically, is essentially impossible due to its complexity. For the sake of simplicity, and following standard research practice, we assume it can be compressed into a single number. This number is a measure of the relative values of the two choices. We can think of this value as the answer to the question, "How confident are you that A is better?" A signal such as "A is much better than B" is an extreme private signal that is an endorsement for A (although choosing A still may be wrong). A signal such as "A is a little better than B" is a small endorsement for A, and choosing A is more likely to be wrong than with the first signal.

Thus, we define a signal as an overall evaluation of the relative merits of a technology based on all the information the decision maker has accessed and all the knowledge he possesses. It may include demos, sales pitches, cost benefit analysis, prior experience with similar technologies, magazine articles, or any other sources of information that decision makers use. A signal reflects the personal interpretation of all the information by the decision maker.

Assumption #3: The values of the technologies do not vary across decision makers. Either A is better or B is better for all individuals. In some situations this assumption clearly will not hold, but 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 adopters. 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 capabilities. Thus, a technology that works well within one adopter's constraints and capabilities probably works well within other adopters' constraints and capabilities. For example, people frequently need to search for information online or type memos or visualize numbers, and thus benefit from search engines, word processors, and spreadsheets. More specifically, people may be familiar with a certain menu layout in a word processor and, thus, may benefit more from using WordPerfect than Word (or vice versa).

To the degree that people's relative values for technology are correlated, the model we present is valid. 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 (Yaniv, 2004). Such a discount value for imperfect correlation simply makes decision makers' responses to others' decisions less dramatic.

We note that in cases in which each decision maker's valuation for the two technologies is uncorrelated, there is no information to be gained from observing the behavior of others. Thus, the model does not address those situations.

Assumption #4: The perceived costs and benefits of making a correct or incorrect decision do not change during the adoption period, and they are identical for all decision makers. In other words, decision makers are homogeneous. This is a simplifying assumption (made to keep a variable called k constant) and a limitation of this work. The decision a person makes should depend not only on the probability of being correct, but on the benefit of being correct and the cost of being wrong. Thus, decision makers must evaluate the cost of choosing B when A is actually correct and vice versa, the benefit of choosing A when A is correct, and the benefit of choosing B when B is correct. We assume these values are constant across decisions.

Assumption #5: The distributions from which decision makers receive private signals do not change. In most cases, over time, more information about the relative merits of technologies becomes available to all decision makers, which may change their private evaluations. For example, decision makers' perspectives on electronic commerce technology have changed dramatically over the last decade, and it is unreasonable to believe that current decision makers would receive private signals about electronic commerce from the same distributions as did decision makers in the late 1990s. However, over a short period of time, distributions can remain relatively static. Certainly, those who have adopted will begin receiving information relatively rapidly, but they will have no incentive to share the information about the quality of the technology with competitors. In fact, many companies have employees sign non-compete clauses to keep competitors from gaining access to information about how a technology works. Again, we could relax this assumption 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 decision makers who decide before and some who decide after her. 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 those decisions were made. A decision maker following the status quo may have evaluated the alternative and decided against it, or she may not yet have performed an evaluation. One possible way to alleviate the status quo problem would be to discount the decisions of those people 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 possibility later, when we derive results.

Another issue relevant to the status quo is that following the status quo is not irreversible. If someone chooses at a decision point not to adopt, she may later choose to adopt. In fact, this issue is not limited to the adopt vs. status quo decision; adopters can choose A today and change to B at a later date. This assumption could be relaxed, a possibility we consider later in directions for future research.

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 modeling assumption. In most cases, decision makers will likely only observe a few other individuals. We analyze this in more detail in the section on distributions of potential adopters.

Assumption #8: Decisions are observable, but reasons for decisions (private information) are not. This means that decision makers can see the technology adopted by prior decision makers. For the case of competing technologies, this is a reasonable assumption, but as discussed above, for cases of a new technology vs. the status quo, it may be questionable. We also note that the Internet and similar technologies probably allow decision makers to know more about the reasons for adoption now than in the past.

There are two issues that arise from these assumptions. First, many of the assumptions could be relaxed and made more realistic by discounting prior decision makers' behavior. We do not include this complication in our analysis, so our presentation can be thought of as an upper bound on information cascade effects.

The second issue is that not all these assumptions are relevant in all adoption environments. Thus, it is worthwhile to consider when they hold more strongly. In general, we contend that the assumptions hold for technologies with considerable uncertainty in the short run. We call this early adoption of novel technologies. 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 enterprise resource planning (ERP) systems. It may take many months (or even years) from the time of the adoption decision (i.e., writing the first check) to having a system implemented and operating in the organization. During that time, technologies may not change and other adopters may not gain (or look for) any additional information upon which to base their decisions.

The short-run time frame and novel technologies also suggest small numbers of adopters. The group of technophile or cutting-edge early adopters is often limited and they make their decisions relatively quickly. Early adoption is particularly important because it sets the stage for subsequent technology decisions, especially for technologies that have network effects. Even when technologies do not have network effects, they are often subject to considerable path dependency, so that early decisions shape how a technology evolves. Thus, while the assumptions we make may not apply to all adoption situations, they apply to a very important subset of such decisions.

The short-run, small-number-of-adopters focus of the present research differentiates it from past information cascade work (Bikhchandani et al., 1992). Other work is usually concerned with asymptotic results (Smith and Sørensen, 2000). However, the short-run adjustment is arguably most appropriate for IT adoption decisions.

4. Theory Formalization

Given the background and assumptions discussed 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 relative merits of the two technologies, which comes from a NORMAL(μ_A, σ^2) distribution if A is the better choice and a NORMAL(μ_B, σ^2) distribution if B is the better choice. (Here, μ_A and μ_B represent the difference in values between A and B, so a more descriptive subscript might be $\mu_{(A-B > 0)}$ and $\mu_{(A-B < 0)}$. However, in the interest of readability, we will keep μ_A and μ_B.) The private signal is a single realization from the distribution, and the decision maker's problem is to determine from which distribution the realization came.

Let lower case letters denote the decision maker's choice, 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 information suggests that A is better than B with probability $p_{A,}$ and B is better with probability 1-p_A.

Given the prior, the decision maker then receives a private signal an based on this private signal must choose between A and B. Thus, the potential adopter's decision criterion can be represented as a threshold that depends on the probabilities of A being better and B being better and the relative costs of the decisions, which we discuss below. If an observed private signal exceeds this threshold, then the observer will conclude that technology A should be adopted (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 A is the better technology is $Prob(a|\mu_A)$,[2] while the probability of incorrectly concluding that B is the better technology when A is the better technology is $Prob(b|\mu_A) = 1 - Prob(a|\mu_A)$. The probability of incorrectly deciding that A is better when it is not is $Prob(a|\mu_B)$, and the probability of correctly deciding that B is better when it actually is better is $Prob(b|\mu_B) = 1 - Prob(a|\mu_A < \mu_B)$. These possibilities are shown in Table 1.

Table 1: Possible outcomes of decision task

		Potential adopter's identification	
		A better (*a*)	B better (*b*)
Reality	A better (μ_A)	$Prob(a\|\mu_A)$	$Prob(b\|\mu_A)$
	B better (μ_B)	$Prob(a\|\mu_B)$	$Prob(b\|\mu_B)$

Given that both distributions have the same variance, a measure of the ability of a potential adopter to differentiate between the merits of A and B is $d' = \mu_A/\sigma - \mu_B/\sigma$, which is simply the standardized difference between the two means. These assumptions are illustrated in Figure 1.

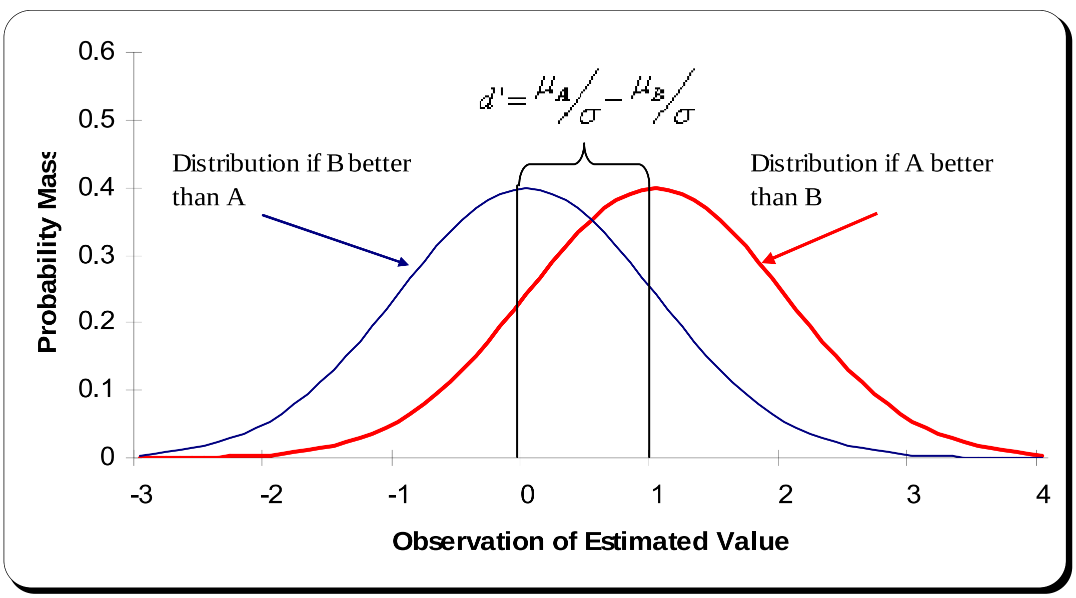


Figure 1: Distributions of private signals concerning an information technology

The potential adopter's problem is to select a threshold beyond which he will conclude that the private signal came from the μ_A distribution.[3] 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

[2] The notation μ_A means that the observation came from the distribution with μ_A as a mean. It does not refer to the mean of the distribution; rather, it refers to the distribution from which the signal came. The term should literally be read as "NORMAL(μ_A,σ^2)." Thus, $Prob(a|\mu_A)$ is shorthand for Prob(a|(The signal comes from a NORMAL(μ_A,σ^2) distribution)). The notation μ_B means that the observation came from the distribution with μ_B as a mean.
[3] Because the labels *A* and *B* are arbitrary, we will let *A* be the label of the better technology for this analysis. The results are symmetrical if *B* is better. This can be seen simply by re-labeling all the *A*s as *B*s and vice versa.

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 (*Prob*($a|\mu_A$)) also increases the probability of falsely concluding that A is better when it is not (*Prob*($a|\mu_B$)).

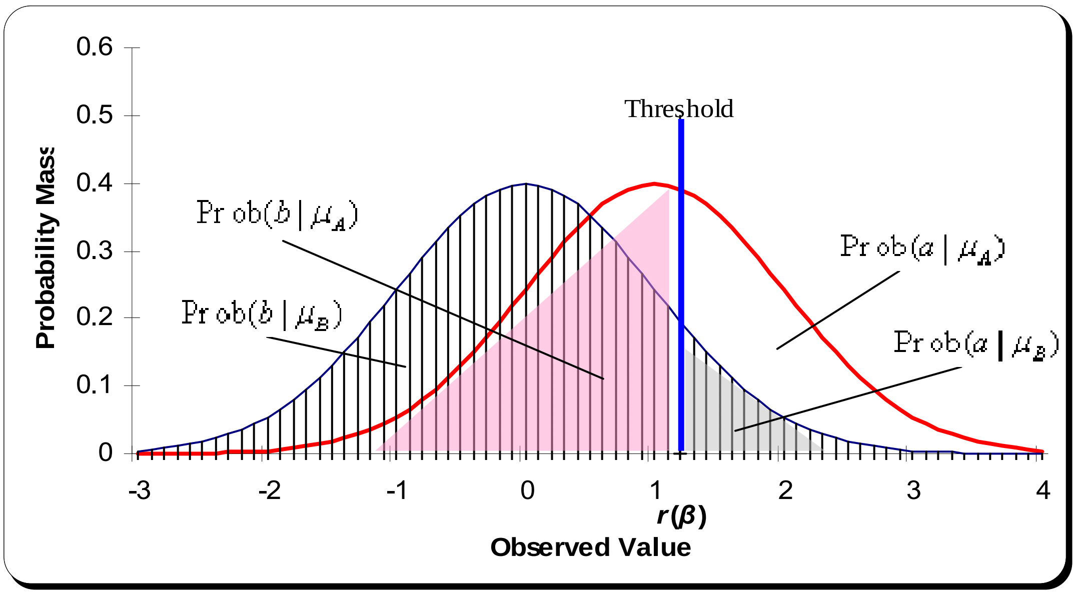


Figure 2: Regions for different decision outcomes

A threshold is chosen to be the location where the ratio of the heights of the probability density functions of distributions A and B is greater than some value. This is expressed mathematically as choosing *a* if

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

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 β in Figure 2 is set to two. Thus, at the threshold represented by the vertical line $r(\beta)$, the height of the μ_A distribution is twice the height of the μ_B 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 β (see Green and Swets, 1966) in this case is

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

The variable *k* represents the relative benefits of the outcomes shown in Tale 1. Though *k* is mathematically simple, it is conceptually complex and requires explanation. The variable *k* represents the relative differences between the values of the technologies in different states of the world. The numerator is the difference between the value of adopting A and the value of adopting B given that the unknown state of the world is *B is better*. The decision maker does not know whether B is better, but she can form some estimate of the value of each technology in the case that B is better. The denominator is the difference between the value of adopting A and the value of adopting B given that the unknown state of the world is *A is better*. The implication is that as *k* increases, a decision maker will require a greater private observation to convince her to adopt the technology.

There is a variety of ways that *k* can be made large (or small). Large or small values of *k* do not require huge differences in the technologies. Rather, extreme values of *k* require that in one state of the world the technologies are fairly similar in value and in one state they are fairly different. In fact, it is easy for *k* to be extreme when the technologies are very similar in each possible state of the world. For example, if in one state of the world the technologies are so similar as to only produce a $1 difference in value, and in the other state they are even more similar, differing in value by a mere penny, then *k*=100 (or 1/100, depending on which state has a dollar's difference). Thus, large and small values of *k* do not suggest any dramatic technical differences; rather, they just suggest how relatively costly it is to make a mistake.

As a practical matter, it is absurd to apply this model if the differences are a dollar and a penny. We would expect that in real situations there is some meaningful difference in the relative values in different states of the world. In situations in which the cost of choosing incorrectly is high (i.e., benefits are low), and the benefit of choosing correctly is high, we might see some high values of *k*. That is, high-risk, high-return projects represent a class of extreme *k* decisions that would be important, whereas choices differing by dollars and pennies would be a class of extreme *k* projects that would be of little interest. However, we must emphasize that the value of *k* gives us no information about the relative merits of each technology. Instead, it tells us the relative merits of choosing a technology in different states of the world.

Consider the following example. Recently, a company called SCO filed a patent infringement suit against IBM. SCO alleged that users of IBM Linux should pay licensing fees to SCO, and SCO sent licensing contracts to Linux users. For the sake of the example, assume that if SCO's claims had been upheld in court Linux users would have had to pay $1,000; if the claims were not upheld Linux users would have had to pay $0. On the other hand, users could purchase Microsoft Windows for, say, $500, regardless of the outcome of the lawsuit. If Linux is technology A and Windows is technology B, and we assume the gross value (i.e., before paying the licensing fee) of an operating system license is $1,000, then benefit($a|\mu_A$) = benefit of Linux given that IBM wins the lawsuit = $1,000-$0 = $1,000. Similarly, benefit($b|\mu_A$) = benefit of Windows given that IBM wins the lawsuit = $1,000-$500, so the denominator is $1,000-$500 = $500. On the other hand, if IBM loses and B is the correct choice, then benefit($b|\mu_B$) = benefit of Windows, given that IBM *loses* the lawsuit = $1,000-$500 = $500, and benefit($a|\mu_B$) = $1,000-$1,000 = $0, so the numerator is $500. This yields a *k* of 1.

Of course, there is no reason SCO should limit the licensing fee to $1,000. If they instead charged $2,000, then Linux is still better if IBM wins and still worse if IBM loses, but benefit($a|\mu_B$) = $1,000-$2,000 = -$1,000, so the numerator is $1,500. This yields a *k* of 3. On the other hand, if SCO charged a fee of $600, Linux is still better if IBM wins and still worse if IBM loses, but benefit($a|\mu_B$) = $1,000-$600 = $400, so the numerator is $100. This yields a *k* of 0.2.

The source of uncertainty in this example is whether IBM will win, which depends in part on facts and opinions about the strength of SCO's case. Ignoring any technical distinctions between the two operating systems, a private signal would be the opinion of a firm's legal counsel, and the public observation would be whether other firms chose Linux or Windows.

Of course, our description simplifies the case, and we chose costs for ease of addition, but this is a real situation that is not well explained by traditional theories such as network externality theory or social benefit-based herding. It is reasonable to assume that firms consulted their legal counsels and that they attended to the actions of other firms.

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 his 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 and Rosenkopf, 1997, Fichman, 2000).

As noted above, in uncertain environments with sequential choices, potential adopters can increase their own information by considering the observed choices of prior potential adopters (Bikhchandani et al., 1992, Bikhchandani et al., 1998, Li, 2004, Walden and Browne, 2002)).[4] To formalize this idea, assume that potential adopters can perfectly identify prior potential adopters' decisions but cannot identify the private information that led to those decisions (as discussed in the assumptions section above). Nor can they observe the benefits accruing to other IT adopters, because those benefits take too long to become apparent (Brynjolfsson and Hitt, 1998, Brynjolfsson and Yang, 1997). Thus, potential adopters can make use of prior information if they condition their own estimates of the probabilities of A and B on the prior potential adopters' choices. Therefore, $\Pr(\mu_B)$ becomes $\Pr(\mu_B|$prior potential adopters' IT adoption decisions) and $\Pr(\mu_A)$ becomes $\Pr(\mu_A|$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.

Consider a situation in which each potential adopter faces the same costs and benefits. Denote potential adopter t's prior beliefs to be $\Pr(\mu_A)_t$ and $\Pr(\mu_B)_t$, which will depend on the sequence of adoption decisions that occurred before time t. Potential adopter t+1 will have prior beliefs denoted $\Pr(\mu_A)_{t+1} = \Pr(\mu_A|D_t, D_{t-1}, D_{t-2}, \ldots D_1)$ and $\Pr(\mu)_{Bt+1} = \Pr(\mu_B|D_t, D_{t-1}, D_{t-2}, \ldots D_1)$, where D_t is the t^{th} potential adopter's observable decision. By Bayes' theorem, it can be shown that

$$\Pr(\mu_A)_{t+1} = \Pr(\mu_A \mid D_t, D_{t-1}, D_{t-2} ..., D_1) = \frac{\Pr(\mu_A)\Pr(D_t, D_{t-1}, D_{t-2} ..., D_1 \mid \mu_A)}{\Pr(D_t, D_{t-1}, D_{t-2} ..., D_1)}, \quad (3)$$

and

$$\Pr(\mu_B)_{t+1} = \Pr(\mu_B \mid D_t, D_{t-1}, D_{t-2} ..., D_1) = \frac{\Pr(\mu_B)\Pr(D_t, D_{t-1}, D_{t-2} ..., D_1 \mid \mu_B)}{\Pr(D_t, D_{t-1}, D_{t-2} ..., D_1)}. \quad (4)$$

Substituting these two results into (2) yields

$$\begin{aligned}\beta_{t+1} &= \frac{\Pr(\mu_B)_{t+1}}{\Pr(\mu_A)_{t+1}} k = \frac{\dfrac{\Pr(\mu_B)\Pr(D_t, D_{t-1}, D_{t-2} ..., D_1 \mid \mu_B)}{\Pr(D_t, D_{t-1}, D_{t-2} ..., D_1)}}{\dfrac{\Pr(\mu_A)\Pr(D_t, D_{t-1}, D_{t-2} ..., D_1 \mid \mu_A)}{\Pr(D_t, D_{t-1}, D_{t-2} ..., D_1)}} k \\ &= \frac{\Pr(D_t, D_{t-1}, D_{t-2} ..., D_1 \mid \mu_B)}{\Pr(D_t, D_{t-1}, D_{t-2} ..., D_1 \mid \mu_A)} \left[\frac{\Pr(\mu_B)}{\Pr(\mu_A)} k \right],\end{aligned} \quad (5)$$

where k is the constant determined by the costs and benefits of each outcome (assumed to be the same for each potential adopter), and $\Pr(\mu_B)$ and $\Pr(\mu_A)$ 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)). Define A_t as the set of all prior decisions so that $A_t = \{D_{t-1}, D_{t-2}, \ldots D_1\}$ for all t>1. Then (5) can be rewritten as

[4] 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.

$$\beta_{t+1} = \frac{\Pr(D_t \mid \mu_B, A_t)\Pr(D_{t-1} \mid \mu_B, A_{t-1})\ldots\Pr(D_2 \mid \mu_B, A_2)\Pr(D_1 \mid \mu_B)}{\Pr(D_t \mid \mu_A, A_t)\Pr(D_{t-1} \mid \mu_A, A_{t-1})\ldots\Pr(D_2 \mid \mu_A, A_2)\Pr(D_1 \mid \mu_A)}\left[\frac{\Pr(\mu_B)}{\Pr(\mu_A)}k\right]$$
$$= \left(\frac{\Pr(D_t \mid \mu_B, A_t)}{\Pr(D_t \mid \mu_A, A_t)}\right)\left(\frac{\Pr(D_{t-1} \mid \mu_B, A_{t-1})}{\Pr(D_{t-1} \mid \mu_A, A_{t-1})}\right)\ldots\left(\frac{\Pr(D_2 \mid \mu_B, A_2)}{\Pr(D_2 \mid \mu_A, A_2)}\right)\left(\frac{\Pr(D_1 \mid \mu_B)}{\Pr(D_1 \mid \mu_A)}\right)\left[\frac{\Pr(\mu_B)}{\Pr(\mu_A)}k\right]. \quad (6)$$

This equation is interesting because the portion in brackets is β_1. Notice further that β_2 is the term in brackets multiplied by the last term in parentheses, and β_3 is the term in brackets multiplied by the last two terms in parentheses. This can be expressed more generally as

$$\beta_{t+1} = \left(\frac{\Pr(D_t \mid \mu_B, A_t)}{\Pr(D_t \mid \mu_A, A_t)}\right)\beta_t . \quad (7)$$

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 β is a monotonic function and thus can be inverted. Inverting the β 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

$$\Pr(b \mid \mu_B)_t = \int_{-\infty}^{r(\beta_t)} \phi(\mu_B)dr > \int_{-\infty}^{r(\beta_t)} \phi(\mu_A)dr = \Pr(b \mid \mu_A)_t \quad (8)$$

and

$$\Pr(a \mid \mu_B)_t = \int_{r(\beta_t)}^{\infty} \phi(\mu_B)dr < \int_{r(\beta_t)}^{\infty} \phi(\mu_A)dr = \Pr(a \mid \mu_A)_t . \quad (9)$$

Combining this with (5) shows

$$\frac{\Pr(a \mid \mu_B)_t}{\Pr(a \mid \mu_A)_t} < 1 \Rightarrow \beta_{t+1} < \beta_t \quad (10)$$

and

$$\frac{\Pr(b \mid \mu_B)_t}{\Pr(b \mid \mu_A)_t} > 1 \Rightarrow \beta_{t+1} > \beta_t . \quad (11)$$

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 *β* function at any decision time *t* can be written as

$$\beta_t = \left(\prod_{i=1}^{t-1} \left(\frac{\Pr(a \mid \mu_B, A_i)_t}{\Pr(a \mid \mu_A, A_i)_t} \right) \right) \left[\frac{\Pr(\mu_B)}{\Pr(\mu_A)} k \right]. \tag{12}$$

After the first three decisions, this equals

$$\beta_3 = \left(\frac{\displaystyle\int_{r\left[\left(\frac{\int_{r(\beta_1^*)}^{\infty}\phi(\mu_B)dr}{\int_{r(\beta_1^*)}^{\infty}\phi(\mu_A)dr}\right)[\beta_1^*]\right]}^{\infty} \phi(\mu_B)dr}{\displaystyle\int_{r\left[\left(\frac{\int_{r(\beta_1^*)}^{\infty}\phi(\mu_B)dr}{\int_{r(\beta_1^*)}^{\infty}\phi(\mu_A)dr}\right)[\beta_1^*]\right]}^{\infty} \phi(\mu_A)dr} \right) \left(\frac{\displaystyle\int_{r(\beta_1^*)}^{\infty} \phi(\mu_B)dr}{\displaystyle\int_{r(\beta_1^*)}^{\infty} \phi(\mu_A)dr} \right) \left[\beta_1^*\right]. \tag{13}$$

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 correct or incorrect relative to the real state of the world. The probability of correctly adopting A when A is actually better can be represented as

$$\Pr(a \mid \mu_A)_{t+x} = \int_{r(\beta_{t+x})}^{\infty} \phi(\mu_A)dr, \tag{14}$$

where

$$r = \frac{\ln(\beta)2\sigma^2 - \bar{\mu}_B^2 + \bar{\mu}_A^2}{2(\bar{\mu}_A + \bar{\mu}_B)}.^{5} \tag{15}$$

This means that

$$r(\beta_{t+x}) = \frac{\ln\left(\left(\prod_{0}^{x-1} \left(\frac{\Pr(D_{t+x} \mid \mu_B, A_{t+x})}{\Pr(D_{t+x} \mid \mu_A, A_{t+x})} \right) \right) \left[\frac{\Pr(\mu_B)}{\Pr(\mu_A)} k \right] \right) 2\sigma^2 - \bar{\mu}_B^2 + \bar{\mu}_A^2}{2(\bar{\mu}_A + \bar{\mu}_B)}. \tag{16}$$

As discussed above, in (5), it is straightforward to show how *β* 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.

[5] Recall that in an earlier footnote we defined the notion μ_A to mean the text "NORMAL(μ_A,σ^2)." To avoid confusion, we note that here and in the next equation the bar over $\bar{\mu}_A$ refers to the mean of the distribution, not the fact that the observation came from the distribution. In other words, the random variable μ_A has mean $\bar{\mu}_A$.

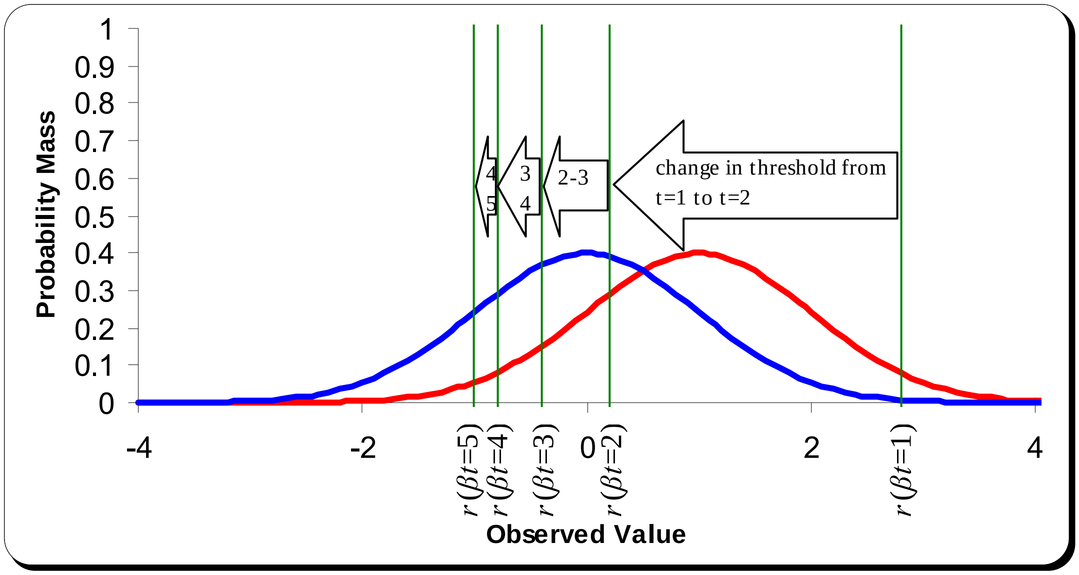


Figure 3: Changes in threshold for consecutive adoption decisions

The figure is graphed assuming $\mu_A = 1$, $\mu_B = 0$, $\sigma^2 = 1$, $k = 10$, $Pr(\mu_A) = Pr(\mu_B) = ½$, 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 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 were true was high.

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 μ_A distribution exceeding 2.8 is .036. However, the probability of an observation from the μ_B distribution exceeding 2.8 is only .003. Thus, the probability of an observation coming from the μ_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 μ_B distribution conditional on the 2.8 threshold. Therefore, 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 cumulative density functions 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 *A-is-better* distribution conditional on the observation being greater than the threshold is .79, but the probability that an observation came from the *B-is-better* 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.

It is worth noting 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 behavior (Abrahamson, 1991, Fichman, 2000, Kauffman and Li, 2003). 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 *B-is-better* and *A-is-better* distributions. Thus, regardless of the actual distribution, subsequent observed decisions lead to increased probability of making the same decision.

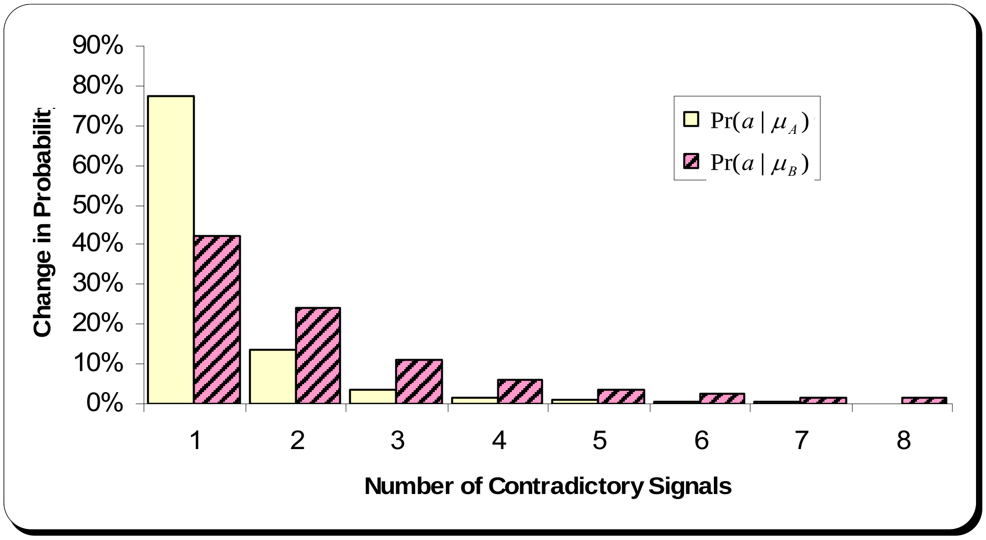


Figure 4: Marginal impacts of successive identical adoption decisions

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 important 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 a large number of prior decisions into their own decision (Miller, 1956). 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 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 utilize 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?*

[6] It is worth noting the relevance of the present discussion to 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.

Innovations often diffuse through social networks based on strong communication channels (Abrahamson and Rosenkopf, 1997, Fichman, 2000, Rogers, 1995, Zmud, 1983). 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:

> ***Research question 4****: What is the effect of groups of potential adopters on the number of correct decisions?*

5. Research Design

Simulation is often used to examine models of technology diffusion (Abrahamson, 1991, Abrahamson and Rosenkopf, 1997, Oh and Jeon, 2007) 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. As Abrahamson and Rosenkopf observe, "Traditional rate-oriented models of innovation diffusion do little to explain the occurrence, extent, and persistence of bandwagons" (Abrahamson and Rosenkopf, 1993). To resolve this issue, these authors used a complicated simulation that incorporated additional model parameters. The advantage of this approach is a more informative model, but the drawback is the difficulty in solving the model. We follow these authors in using simulation to solve our model. This technique requires that we specify the model and then have a computer solve it many times under a variety of different assumptions. We then report the cumulative results.

The simulation incorporates both implications and sensitivity. For example, different technologies will surely have different costs and benefits and thus different values of k. To establish how the model changes with these different levels of k, we simulate a wide range of ks in Figure 7 (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 *A-is-better* distribution. The *A-is-better* and the *B-is-better* distributions 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 Green and Swets, 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 $\Pr(\mu_A)$ and $\Pr(\mu_B)$ were both 0.5, indicating that, in the absence of any other information, they were equally likely. We varied k, the relative cost and benefit, thereby varying β. We note the value of k below each graph.

6. Data and Results

6.1. Convergence

We begin by addressing our first two research questions, which concern whether the decisions of sequential potential adopters converge and the appearance of the convergence path. The answers are not intuitive from the equations and so bear testing. To answer the first question, we ran a string of 1,000 repetitions of 100 decisions. The results of this simulation are shown in Figure 5.

The figure shows two important characteristics. First, there is not convergence to the correct decision after 100 decisions. Specifically, 96.4 percent 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 percent 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 percent of the potential adopters reversed in the incorrect direction.

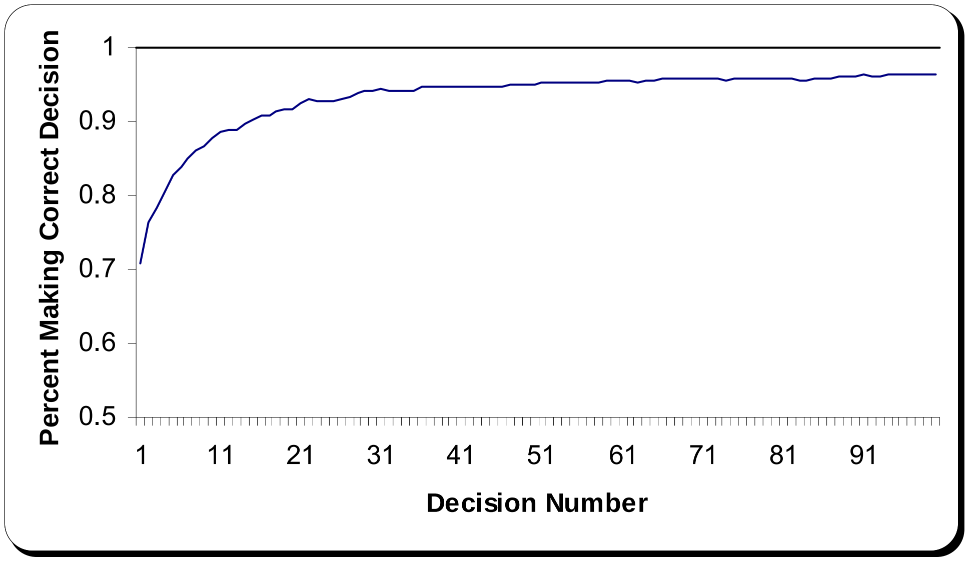


Figure 5: Convergence of Decisions, 1,000 samples of 100 decisions, *k* = 1, *d'*=1

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 and the discriminability of signals. The variable *k* captures these costs and the variable *d'* captures discriminability. Varying *d'* moves the signal distributions either closer together or farther apart. The units of *d'* are standard deviations, so *d'* = 1 means the distributions are one standard deviation apart. This is a measure of the uncertainty of the signal. As *d'* approaches zero, the ability to discriminate between distributions approaches zero, and as *d'* approaches infinity, the ability to discriminate approaches infinity. The effects of different levels of *d'* on the rate of convergence are reported in Figure 6. As discrimination improves, decision makers more quickly converge on the correct decision because both their own private signals and the inferred signals of others are more reliable.

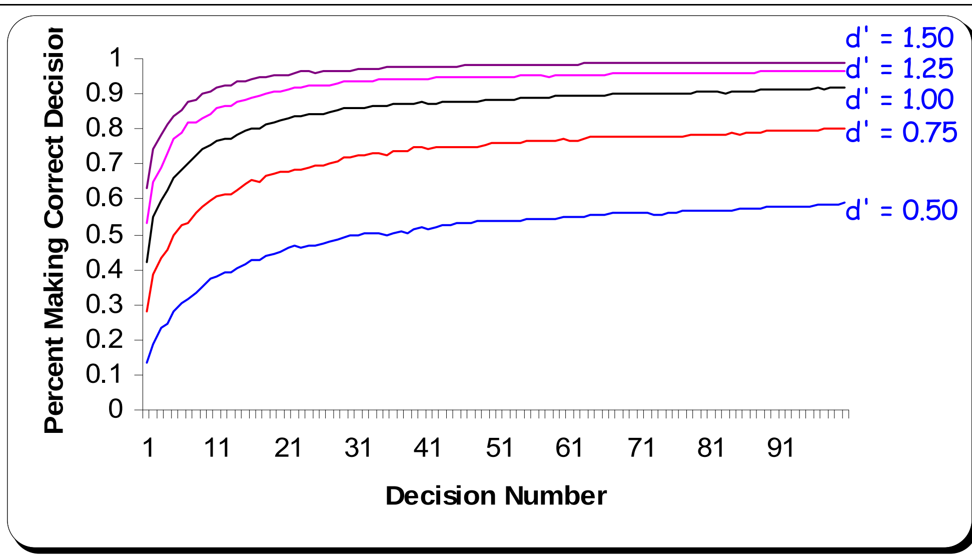


Figure 6: Effect of Discrimination on Convergence in 1,000 samples of 100, *k* = 2

Varying *k* produces different rates of convergence, as shown in Figure 7. Note that the observations are actually coming from the *A-is-better* distribution, so the correct decision is *a*.

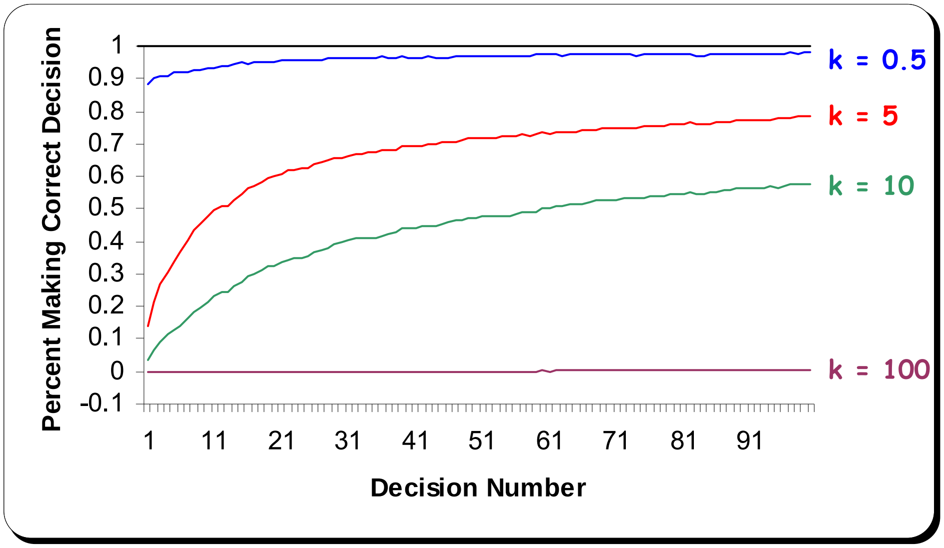


Figure 7: Effect of Costs on Convergence in 1,000 samples of 100, *d'* = 1

Notice that when *k* is very high, it is not clear that there is any convergence to the correct decision. Large *k* implies that the difference in benefit between the two technologies when B is the correct choice is large relative to the difference in benefit when A is the correct choice. This means that if B is the correct choice, then B is much better than A, but if A is the correct choice there is not much difference. Returning to the Linux licensing example, high *k* indicates that if IBM wins, there is not much difference in cost between Linux and Windows, but if IBM loses the price of Linux will be much greater than the price of Windows. In this situation, even if SCO's case is weak, potentially adopting firms may not be willing to take the risk of adopting Linux. Note further that with increases in *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 percent likely to make the correct decision and 21.5 percent likely to make the wrong one. Even if *k* is small, we still do not achieve convergence after 100 decisions.

If the net benefit of choosing A when A is better is greater than the net benefit of choosing B when B is better,[7] 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, if B is the better technology, then this situation 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). 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 in at least some areas of the popular press seem to have become more negative (Carr, 2003). If this is the case, we may experience a period of underadoption of technically efficient IT.

We include k=100 to show the limiting behavior. This level of k is probably rare, but is fairly easy to imagine. One good example is the recent patent case concerning Research In Motion (RIM) Co.'s Blackberry (Krazit and Broache 2006). At some point in the past, RIM faced a decision about what technology to use to connect wireless devices to wired networks to deliver email messages from the wireless Blackberry. They chose a technology (A) and later claimed to have developed a workaround

[7] This can occur for a variety of reasons. Ceteris paribus, if the gross benefit of choosing A when A is better increases, then *k* decreases. Ceteris paribus, if the gross benefit of choosing B when A is better is very small or negative, then *k* decreases. Also, if the benefit of choosing A and the benefit of choosing B are very close when B is better, then *k* decreases. Put another way, if choosing B when A is correct is a significant mistake, but choosing A when B is correct is a minor mistake, then *k* is small.

(technology B) (Hamblin, 2006). The *k* they faced can be decomposed based on a state of the world in which technology A may or may not have already been patented. We assume that if technology A had not been patented, it is slightly more efficient and the better choice; thus, the benefit of a|A is some number and the benefit of b|A is a similar but smaller number because we assume that A is somewhat more efficient. On the other hand, if A had been patented and B had not been, then B is the better choice by far. This means the benefit of b|B is much greater than the benefit of a|B. In this particular instance, technology A had been patented and RIM was forced to make a settlement payment of $612.5 million to the owner of technology A (Krazit and Broache, 2006). We can easily imagine that from a technical perspective A may have only been a million dollars better, which would give a *k* of 611.5 for this particular technology. This situation also occurred when Microsoft settled with Eolas for $521 million for using patented technology in Internet Explorer Plug-ins, and when Microsoft settled with Intertrust for $440 million for choosing the wrong technology for digital rights management. This patent issue is present any time a firm makes choices about a technology to include in a product. The point is that in these cases *k* is large and there are very few observable differences between technologies. The private signal represents the quality of the patent search performed prior to implementation.

It is useful to pause here and discuss how these results are different from Bikhchandani et al. (1992). In the former paper decision makers could converge to either the correct or incorrect decision. However, in our formulation of the model decision makers seem to be converging to the correct decision almost all the time. This is because each new decision lets new information enter the system, as shown in Figure 4. Thus, even false starts are overcome with time, as anticipated by Bikhchandani, et al. (1998). Even when many people choose the wrong distribution, eventually someone receives a signal high enough to overcome it and choose the correct distribution. However, if the costs (i.e., *k*) are high enough, the probability of someone receiving a high enough signal is extraordinarily low. For example, at k=100, the probability of receiving a signal extreme enough to make someone choose the correct distribution is less than one in two million.

Based on the results of the first simulation, a natural question to ask is whether 100 decisions are enough to achieve convergence to the correct decision. Perhaps more repetitions are necessary. More to the point, it is not clear whether the threshold moves faster than the probability needed to overcome it. To test this question, we ran simulations of 1,000,000 decisions, as shown in Figure 8. However, 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 from the number at 1,000,000, suggesting that even after 100,000 observations potential adopters may reverse if they receive a private signal that is sufficiently extreme.

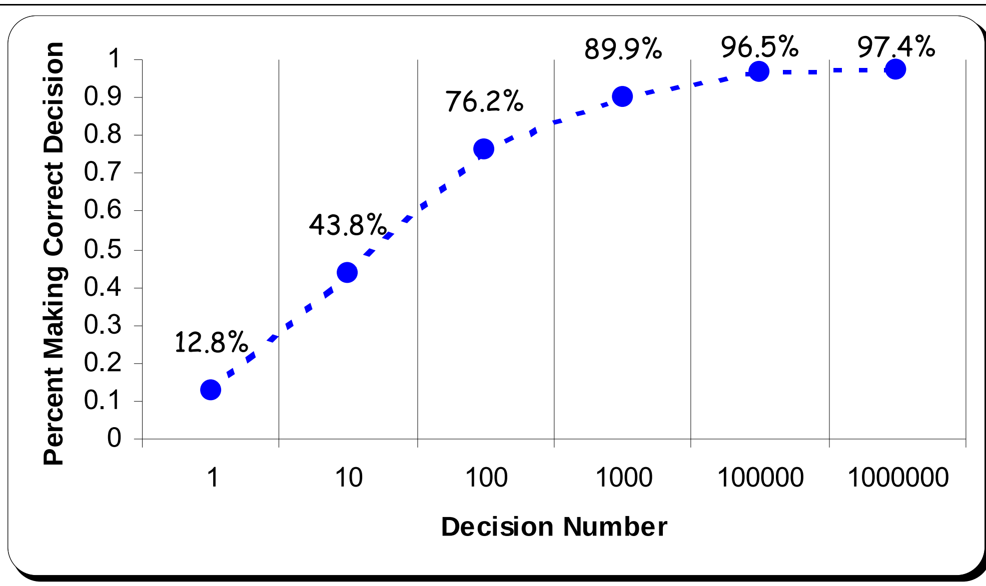


Figure 8: Convergence of Decisions, 1,000 samples of 1,000,000 decisions, *k* = 5, *d'*=1

We have kept k constant across all potential adopters in a particular simulation. This implies that the costs and benefits for every potential adopter are similar. It is worth noting, though, that this might not always be the case. For example, in the SCO vs. RIM case discussed above, after SCO filed its case, Hewlett-Packard (HP) indemnified all its Linux users against SCO licensing litigation. In other words, HP said it would pay the licensing fee if SCO won the case. This would clearly change k by making it smaller. Similarly, when SCO filed the lawsuit, it increased k. Therefore, we explore the impacts of different ks in Figure 9. We specify $k = 5$ for the first 10 decision makers and then $k = 100$ for the remainder of the decision makers.

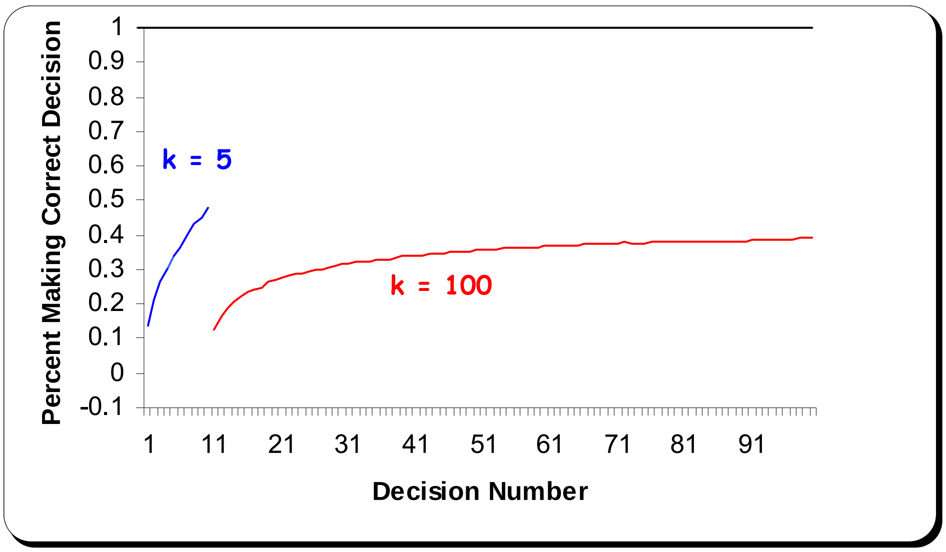


Figure 9: The effect of a change in *k* at the 11th decision in 1,000 samples of 100, *d'*=1

The graph shows that increasing k increases the threshold and thereby reduces the probability of making the correct adoption decision for the 11th 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 11th adopter, whereas it remained constant when all adopters had a k of 100. This occurs because the first ten adopters had a threshold sufficiently low to allow some information into the system. Though the 11th adopter is faced with a very high k, he also has enough information about other adopters' signals to judge the relative probabilities of A and B being the better choice, so he is relatively confident that adopting technology A is a good idea even if his own cost of a mistake is 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 in fact 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 risk 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 $a|\mu_B$). This is precisely the kind of change in k shown above. Comparing k=100 in Figure 9 and Figure 7 shows the benefit to the vendor of mitigating the downside risk.

6.2. Impact of Contrary Decisions

One of the implications of information cascade theory is that because no new information enters the system, decision makers never become particularly confident in their decisions. This property is referred to as "fragility" (Bikhchandani et al., 1992). The concept of fragility is, to some degree, an artifact of the Bikhchandani et al. (1992) model. With binary signals, the magnitude of the signal is fixed, and so it is easy to establish a threshold beyond which the signal cannot have any impact.

However, in our sequential adoption theory model, with continuous signals potential adopters may receive extreme private observations that result in decisions that are contrary to the prevailing cascade. Notice this is subtly different from the Bikhchandani et al. (1992) type of fragility because it is endogenous to the model. Bikhchandani et al. (1992) require some input beyond the model such as a change in preferences or more informed agents. In our model, fragility can result from the luck of the draw based on what information decision makers use and how they interpret that information. 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.

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 10.

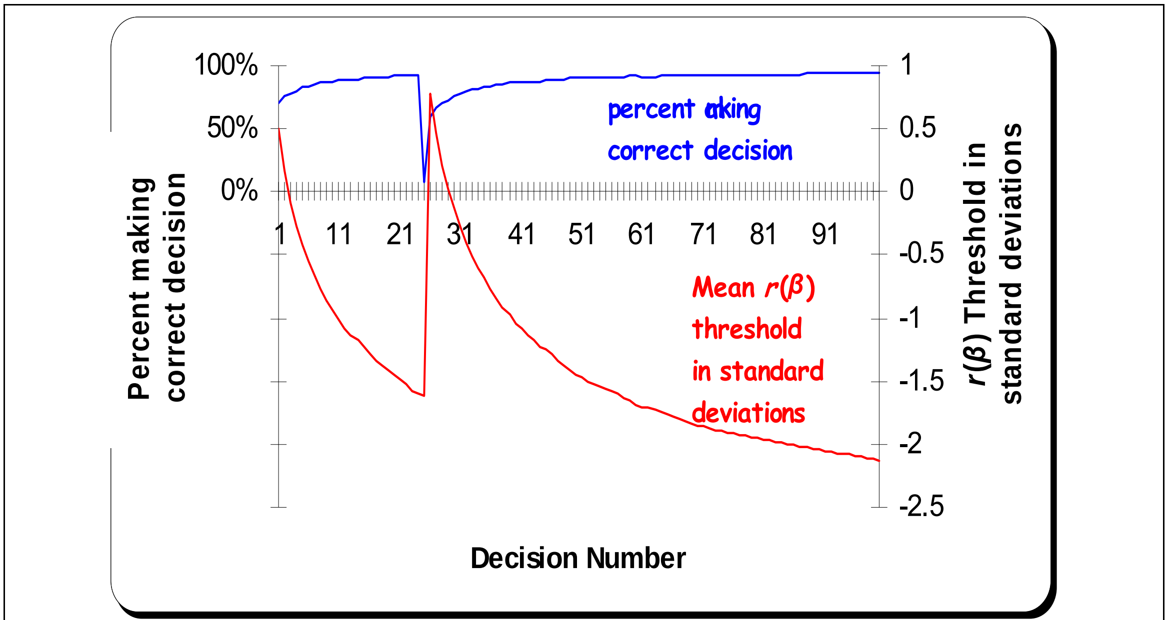


Figure 10: Effects of Contradictory Information in 1,000 samples of 100, *k* = 1, *d'* = 1

The figure shows that the impact of contradictory information is tremendous. The percentage of correct decisions dropped from 93 percent to 59 percent 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 percent 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 12 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.

A common sense explanation may help illustrate our point. At the 24th decision, there are three possible states in which the system can exist: a correct cascade, an incorrect cascade, or no cascade at all. If the 24th decision maker is in a correct cascade and the 25th decision maker chooses

[8] We use the 25th decision because in both the 24th and 25th decisions 92.9% of the individuals make the same choice; thus, by using the 25th decision, we control for extreme observations in the 24th decision.

[9] Note that the probability that the 25th decision maker was correct was 7 percent and the probability that he was incorrect was approximately 93 percent 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.

incorrectly, then the 26th decision maker knows that the 25th decision maker must have had an observation extreme enough to outweigh all of the prior decisions (although he does not know at this point that the 25th decision maker's choice was the opposite of what it should have been), so the 26th decision maker sets his acceptance threshold a little higher than that of the 1st decision maker. However, the 26th decision maker will probably receive a signal that favors the correct answer, so even with the threshold slightly higher, the 26th decision maker has a reasonable chance of making the correct decision. So the correct cascade starts again (although with slightly lower probability than it would have started with in the first place).

On the other hand, if the 24th decision maker is in an incorrect cascade and the 25th chooses correctly, then the 26th decision maker sets his acceptance threshold a little *lower*. This means he is more likely to make the right decision than the 1st decision maker. Thus, contrary information is very damaging to incorrect cascades.[10]

If the 24th person is not in a cascade, and the 25th person makes a decision opposite to that of the 24th, then the 26th decision maker has little information from the string of decisions that have come before. Hence, the 26th decision maker puts a great deal of weight on his private signal, which is likely to be correct, and therefore probably starts a correct cascade.

The point is that extreme observations are very harmful to incorrect cascades, but not very damaging to correct cascades or non-cascades. This explains why we see an upward sloping curve on the probability of being correct. As more decisions are made, extreme observations appear, which are quite likely to reverse incorrect cascades and do little damage to correct cascades. This means that the variance of the signal has a tendency to eventually bring most strings of adoptions toward the correct decision. Moreover, this is an internal property of the model. It does not require an exogenous change. By comparison, in the Bikhchandani et al. (1992) model there is a string of decisions that rapidly converges to some probability less than one and greater than one half. On the other hand, the probability of being correct in our model tends to converge to one (for moderate values of *k* and *d'*). This is an endogenous property of the model, which tends to eliminate incorrect cascades.

6.3. Distribution of Potential Adopters

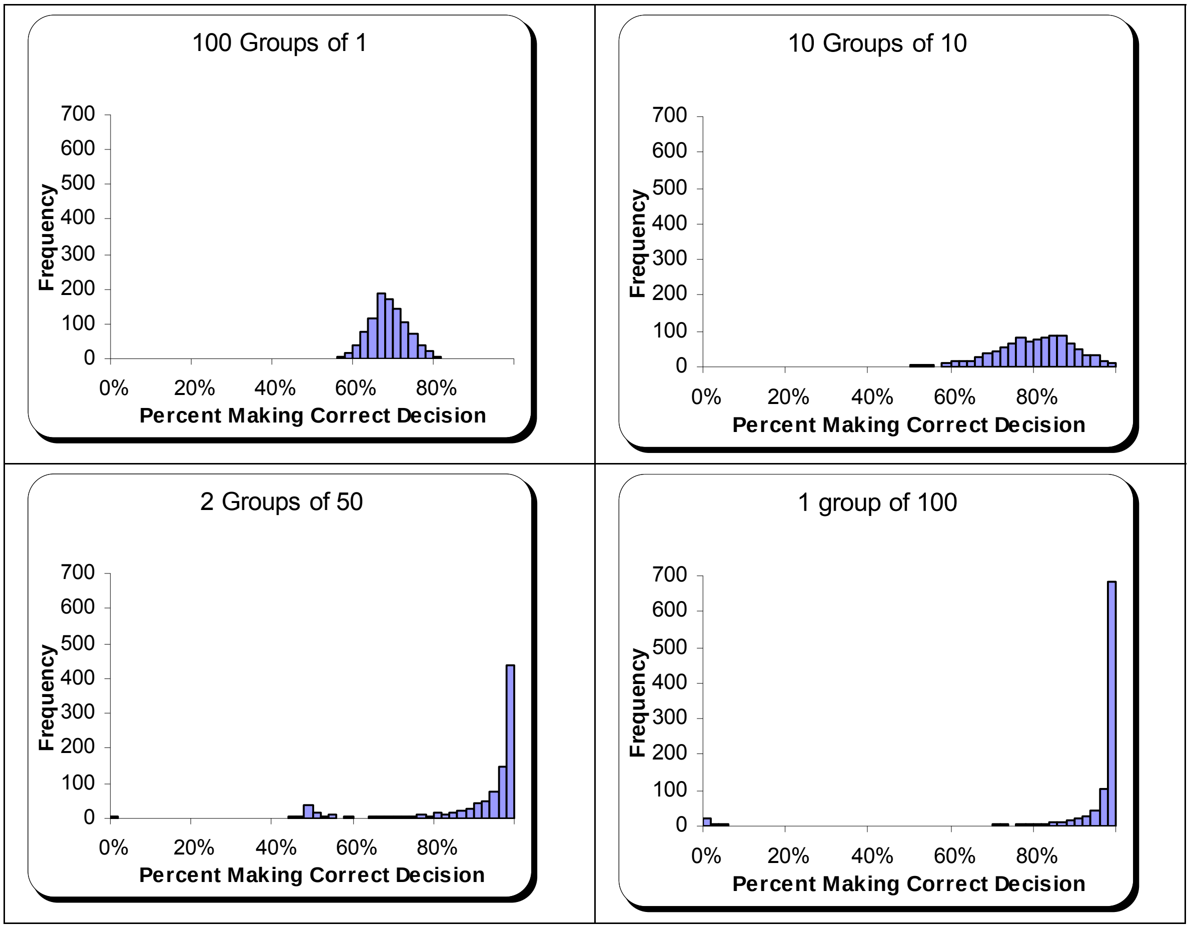
The impact of groups of potential adopters is investigated by our fourth research question. It is particularly useful to consider group decision making from the perspective of the decision motivator. In the case of a novel technology, the decision motivator is the creator of that technology. From this 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 1,000 samples. A group is a set of potential adopters who can observe one another's actions. For example, if the group size is 10, then the second potential adopter can observe the action of the first, and the 10th can observe the actions of the first through the ninth. However, the 11th (the first adopter in a new group of 10) can observe no other potential adopter, but the 12th can observe the 11th and the 20th can observe the 11th through 19th. 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 1,000 simulations, in which adoptions are sequential across groups.

The panel indicates that introducing the sequential effects causes the distribution of decisions to become bi-modal as the group size 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 to a technology vendor of introducing a technology choice to a single large group or many small groups depends upon the risk propensity of the vendor.[11]

---

[10] It is important to note that we cannot say whether the probability of receiving a sufficiently extreme value decreases faster than the threshold increases. In other words, the extremeness of an observation needed to reverse the cascade increases with each decision. Given the results of our investigation of *k*, we expect this depends on the value of *k*.

[11] The benefit to a particular event is $b = \Sigma_a$ prob(a)*U(a), where a is a specific outcome and U is a function describing the individual adopter's utility for that outcome. The second derivative of U reveals the adopter's risk attitude. If one is risk averse then $U'' < 0$, so that doubling the outcome a more than doubles the utility. Organizing



Panel 1: Distributions of decisions with different group sizes

When a vendor takes on a project with random payoffs, the utility of those payoffs may not change at the same rate as the monetary value of the payoff. If the utility of a payoff increases more slowly than the monetary value of a payoff, then the vendor is risk averse. Intuitively, doubling a payoff does not always double the utility of a payoff. Therefore, when any decision maker chooses among potential payoffs that have random outcomes, the decision maker must balance the risk she is taking against the expected utility of the outcomes. To put it another way, the decision maker must receive some premium for bearing increased risk. Figure 11 shows that while larger group sizes lead to larger expected numbers of adopters, they also lead to larger standard errors. In the most extreme case, when everyone adopts as a group, a vendor faces the risk of having no adoptions at all in exchange for a high expected value. At the other end of the spectrum, when everyone decides independently, vendors can expect a middling number of adoptions with very high certainty. A firm that is very risk averse would prefer to have a small expected value with high probability, while a firm that is risk seeking (or risk neutral) would prefer a higher expected number of adoptions even if there is a chance of complete failure.

One important consideration with small groups is that they might share their signals with one another, even if they are otherwise competitors. Sharing signals in this way would make each group more likely to choose correctly. Such a situation would skew the results in Panel 1 to the right, and the results would be more skewed the larger the group size.

---

adopters into different sized groups changes prob(*a*) as shown in the panel (which are frequency plots of prob(*a*)). Obviously, changing prob(*a*) changes the overall benefit, but the type of 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.

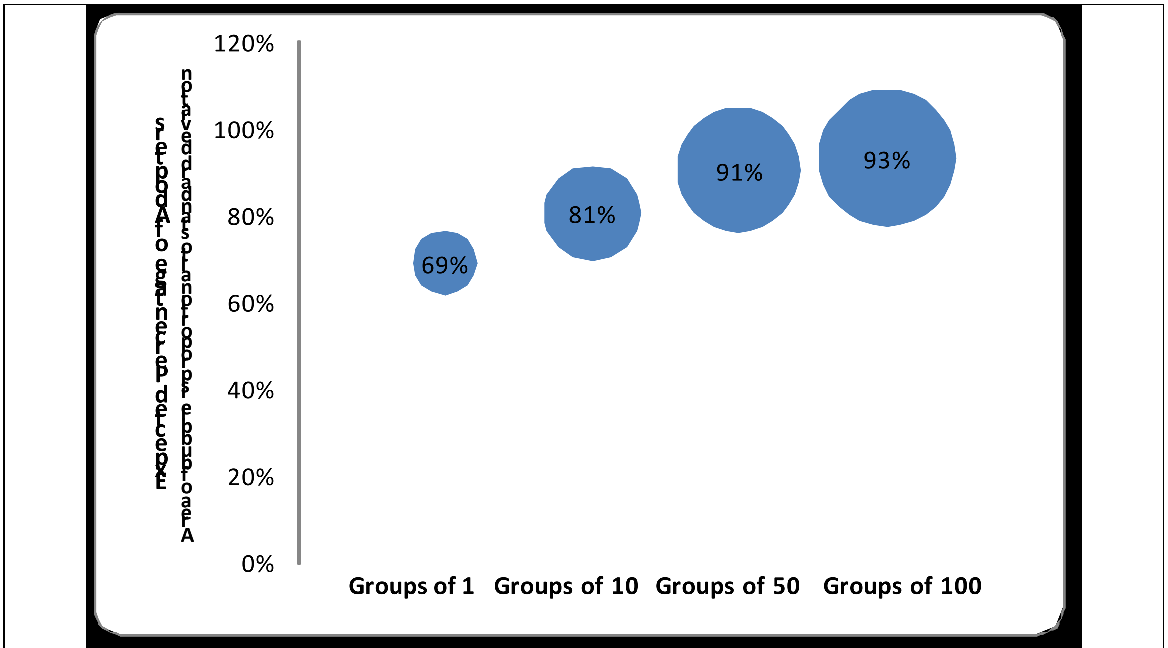


Figure 11: Expected value and standard deviation of adoption by group size

7. Discussion

In this paper we have developed and tested a model of observational learning to explain technology adoption decisions. Previous IS researchers have explained IT adoption in terms of costs and benefits, network externalities, and social benefit-based herding. Our observational learning explanation uses a modified version of the information cascade model of Bikhchandani et al. (1992). While there are many papers that apply the other stated reasons for adoption, there are few (see, e.g., Walden and Browne, 2008) that make use of the information cascades model. We believe this is due in part to the fact that the information cascades model is designed for theoretical elegance but requires non-trivial adjustments to be useful for developing practical explanations. We have made those adjustments and both derived new insights and presented a model that can be used as a building block for developing a more unified view of adoption.

There are several important implications for adoption. Decision makers tend to follow one another (herd), but they tend to do so in the correct direction. The probability of being correct increases as more decisions are made rather than settling to some steady state. Our simulations suggest two reasons for this result. First, we allow for extreme observations in the signals. Whereas others have studied simple binary signals, we allow for signals to come from a normal distribution, and thus decision makers are not limited to simply preferring one option or the other. Rather, they can prefer one option a lot or a little or any amount. Second, when a decision maker reverses the prevailing cascade, it more than resets the prior beliefs of the next decision maker. This is particularly damaging for incorrect cascades. Thus, over time the possibility of extreme signals has a greater tendency to reverse incorrect cascades than to reverse correct cascades. However, the behavior depends on the costs and benefits. When the relative costs and benefits of bad and good choices as embodied in the parameter k become extreme, decision makers do not necessarily converge to a correct cascade. Convergence to a high probability of correct decisions also tends to be fairly slow, and extreme values of k slow down the convergence even more. This means that it may take some time for a decision maker to experience a signal sufficiently extreme to reverse an incorrect cascade.

This result has implications for the behavior of a population of adopters that is divided into groups that can observe their internal members but not external members. If adopters are divided into large groups, then most of the time almost all of them will make the correct decision (however, although rare, occasionally almost all of them will make the incorrect decision). On the other hand, if adopters are divided into small groups, a few decision makers will make the wrong decision, but most will make the correct decision. The mean chance of being correct is higher for large groups, but the variance is also higher.

Because adoption of IT is often quite complex, it is practically and scientifically useful to develop numerous perspectives on the phenomenon. Decision makers in the business world likely consider several different perspectives when making adoption decisions. At some point, most decision makers probably ask questions such as, "Does this technology meet our requirements?," "Can we afford it?," and "Will my superiors and peers see this as a good decision?" To these questions we add a fundamental question from observational learning: "Do other adopters know something we don't know?"

This last question is extraordinarily important because of the role of observational learning in human behavior and decision making. As noted, people generally learn by observing the behaviors of others, making this mechanism one of the most powerful decision tools available. The observational learning perspective is particularly important because the wisdom of considering others' actions is often questioned. Although blindly following others' behavior without due consideration of one's own signal is obviously foolish (Tingling and Parent, 2003), the observational learning perspective suggests that much is to be gained by incorporating the decisions of prior similarly-situated decision makers into one's own decision. In fact, some research shows that the real problem is not that decision makers consider the actions of others, but that decision makers fail to weigh the actions of others heavily enough (Yaniv, 2004).

Our model offers many implications for research and practice, as well as directions for future research. 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 various theories that are relevant to this issue and that could be incorporated into sequential adoption theory.

For example, it is important to ask whether everyone observes the same factors. Currently, we assume that the signal is a function both of the data and the observer, but it can be useful to separate these two concepts. We would expect that the value of others' actions would increase if the others 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.

Another example concerns network effects. It would be straightforward, if perhaps tedious, to incorporate network effects into the model by allowing the benefits and costs of the decision (i.e., the constant k) to depend on network size. Of course, one could include other variables of interest in the relative costs.

In our model we held k (the relative costs and benefits) fixed and inferred the information from the decision. One could also hold the information content constant and infer k or, more specifically, the determinants of the specific elements of k. Of course, one could also infer both k and the information. Moreover, one could apply econometric techniques to estimate the values of the parameters empirically.

This raises the point that k consists of four elements. It is important for adopters to recognize this and incorporate estimates of each of these four components into their decision making. Not only will this improve decision making, but it can also aid decision makers in classifying where problems might emerge. This is not a new result, but we believe it is worth reiterating.

It is also important to understand 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 and Frei, 1998). It is easy to envision this bias as a partial explanation for the dotcom boom of the late 1990s.

Anecdotal evidence in business is also available for the observational learning perspective. 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 advertisement. Several explanations are available that may explain the use of these types of ads. For example, the cost-benefit framework is present. People looking for an inexpensive video game are not likely to buy Oracle E-Business Suite, because it neither does what they want nor fits their budget. Similarly, this ad signals an adoption that may implicate network effects. The ad might also encourage social benefit-based herding, or be viewed as a celebrity endorsement.

However, we believe that the best explanation for the goal of these ads is observational learning. Decision makers in other companies see that 1-800-Flowers has adopted the Oracle product and infer that 1-800-Flowers must have thought that adoption was a good decision. Thus, they incorporate the information contained in that decision into their own decision making. If network effects were the main driver, the ad would have been improved by listing many adopters rather than just one. Similarly, 1-800-Flowers is not large enough or prominent enough to engender wide social benefit-based herding. Being like 1-800-Flowers will not make regulators, customers, and vendors treat a company better, nor will it make others view a company in a more flattering light. (Being like General Electric or Walt Disney might accomplish these things, but not 1-800-Flowers.) We do not completely discount other possible explanations for these ads, as we note above, but we argue that observational learning is a strong potential explanation.

Our results also show that imitation in adoption decisions can be incorrect. Thus, we answer Abrahamson's (1991) call to investigate the diffusion of technically inefficient technologies and the rejection of technically efficient technologies. 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 unanticipated behavior of a pioneering decision maker.

We noted earlier that in most cases decisions to adopt a particular technology or to not adopt (maintain the status quo) are not irreversible (Assumption #6). Relaxing this assumption in our model can lead to numerous additional research questions. A model could be designed in which, rather than specifying discrete decisions, individual decision makers are characterized by hazard functions that specify their probability of adopting in a given period of time. Observational learning would occur by allowing these hazard functions to be functions of other decision makers' observed adoption decisions. Any time one decision maker adopted, the probability of other decision makers adopting over a given increment of time would increase. For increments of time when no one adopted, the probability of other decision makers adopting over a given increment of time would decrease. Such an extension could be very valuable for exploring questions of when adoption occurs and questions about the costs of waiting. Moreover, private signals might become much more important in the beginning stages of adoption, which would probably carry on to future stages. We are also unsure how reversibility would impact the distributions of groups of adopters. This is clearly a limitation of the current work that would be worthwhile to investigate further.

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 and Vinze, 1998, Isenberg, 1986, Lamm, 1988). This group rationalization process can lead to both good decisions (for example, groups agreeing to high levels of charitable giving (Muehleman et al., 1976)) and sometimes to extremely bad decisions (as, for example, in groupthink decisions and mob violence).

As shown in Panel 1, 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.

Our model also suggests another possible reason that each Oracle ad targets a different group. A factor that is important when incorporating the decisions of others into a person's own decision is how similar that person is to the others. For example, ERP systems have generally worked well for manufacturing firms but have been less effective for services firms. We have assumed that the correct choice is perfectly correlated across firms in our simulations, but our model allows one to change the correct choice to be imperfectly correlated. In other words, if technology A is the right choice for company 1, then it is possible to model technology A as the correct choice for company 2 only 90 percent (or 80 percent or 73.5 percent) of the time. The important point is that none of the other research perspectives allows us to examine such a question, though it is clearly relevant to actual practice. Thus, we can ask questions with our model that have not yet been posed by IS researchers.

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.

References

Abrahamson, E. (1991) "Managerial Fads and Fashions: The Diffusion and Rejection of Innovations," *Academy of Management Review* (16) 3 (July), pp. 586-612.

Abrahamson, E. and L. Rosenkopf (1993) "Institutional and Competitive Bandwagons: Using Mathematical Modeling as a Tool to Explore Innovation Diffusion," *The Academy of Management Review* (18) 3, pp. 487-517.

Abrahamson, E. and L. Rosenkopf (1997) "Social network effects on the extent of innovation diffusion: A computer simulation," *Organization Science* (8) 3, pp. 289-309.

Anderson, L. R. (2001) "Payoff Effects in Information Cascade Experiments," *Economic Inquiry* (39) 4, pp. 609-615.

Anderson, L. R. and C. A. Holt (1996) "Classroom Games: Information Cascades," *The Journal of Economic Perspectives* (10) 4, pp. 187-193.

Anderson, L. R. and C. A. Holt (1997) "Information Cascades in the Laboratory," *The American Economic Review* (87) 5, pp. 847-862.

Bandura, A. (1977) *Social Learning Theory*. New York: General Learning Press.

Bernardo, A. E. and I. Welch (2001) "On the Evolution of Overconfidence and Entrepreneurs," *Journal of Economics and Management Strategy,* (10) 3, pp. 301-330.

Bikhchandani, S., D. Hirshleifer, and I. Welch (1992) "A theory of fads, fashion, custom and cultural change as informational cascades," *Journal of Political Economy* (100) 5, pp. 992-1026.

Bikhchandani, S., D. Hirshleifer, and I. Welch (1998) "Learning from the Behavior of Others: Conformity, Fads, and Informational Cascades," *The Journal of Economic Perspectives* (12) 3, pp. 151-170.

Brooks, F. P. (1975) *The Mythical Man-Month: Essays on Software Engineering*. Reading, Massachusetts: Addison-Wesley Professional.

Brynjolfsson, E. and L. Hitt (1996) "Paradox Lost? Firm-Level Evidence on the Returns to Information Systems," *Management Science* (42) 2, pp. 541-558.

Brynjolfsson, E. and L. M. Hitt (1998) "Beyond the Productivity Paradox," *Communications of the ACM* (41) 8, pp. 49-55.

Brynjolfsson, E. and S. Yang. (1997) "The Intangible Costs and Benefits of Computer Investments: Evidence from the Financial Markets." *International Conference on Information Systems, Atlanta, GA, 1997*.

Busemeyer, J.R. "Choice Behavior in a Sequential Decision-Making Task." *Organizational Behavior and Human Decision Processes*, 29 (1982), 175-207.

Carr, N. G. (2003) "IT Doesn't Matter," *Harvard Business Review* (81) 5, pp. 41-52.
DiMaggio, P. J., and W. W. Powell (1983) " The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields," *American Sociological Review*, (48) 2 pp. 147-160.
Dholakia, U. M. and K. Soltysinski (2004) "Coveted or Overlooked? The Psychology of Bidding for Comparable Listings in Digital Auctions," *Marketing Letters* (12) 3, pp. 225-237.
Duan, Wenjing; Gu, Bin; and Whinston, Andrew B.. 2009. "Informational Cascades and Software Adoption on the Internet: An Empirical Investigation," MIS Quarterly, (33: 1) pp.23-48.
Economides, N. and C. Himmelberg (1995) “Critical Mass and Network Evolution in Telecommunications,” in G. Brock (Ed.) *Toward a Competitive Telecommunication Industry*, Mahwah, NJ: Lawrence Erlbaum Publishers, pp. 47-63.
El-Shinnawy, M. and A. S. Vinze (1998) "Polarization and Persuasive Argumentation: A Study of Decision Making in Group Setting," *MIS Quarterly* (22) 2, pp. 165-198.
Fichman, R. G. (2000) “The Diffusion and Assimilation of Information Technology Innovations,” in R. W. Zmud (Ed.) *Framing the Domains of IT Management: Projecting the Future…Through the Past*, Cincinnati, OH: Pinnaflex Educational Resources, pp. 105-128.
Gale, D. and S. Kariv (2003) "Bayesian learning in social networks," *Games and Economic Behavior* (45) 2, pp. 329-346.
Green, D. M. and J. A. Swets (1966) *Signal detection theory and psychophysics*. New York, New York: Wiley.
Greve, H. R. (1995) "Jumping Ship: The Diffusion of Strategy Abandonment," *Administrative Science Quarterly* (40) 3, pp. 444-473.
Greve, H. R., D. Strang, and N. B. Tuma (1995) “Specification and Estimation of Heterogeneous Diffusion Models,” in P. V. Marsden (Ed.) *Sociological Methodology*, Oxford: Blackwell, pp. 377-420.
Gurbaxani, V. (1990) "Diffusion in Computing Networks: The Case of BITNET," *Communication of the ACM* (33) 12, pp. 65-75.
Hamblen, M. (2006) "BlackBerry users 'wiggin' out' over RIM/NTP patent battle And IT managers aren't thrilled with the alternatives to RIM's wireless service," Computerworld, accessed May 26 2008, available online at http://www.computerworld.com/action/article.do?command=viewArticleCoverage&articleId=108040&continuingCoverageId=1018&intsrc=article_cc_report_bot.
Hitt, L. M. and F. X. Frei. (1998) “Evaluating Investments in Electronic Distribution: The Case of PC Banking.” *Tenth Annual Workshop on Information Systems and Economics, New York, NY, 1998*.
Isenberg, D. J. (1986) "Group Polarization: A Critical Review and Meta Analysis," *Journal of Personality Social Psychology* (50) 6, pp. 1141-1151.
Katz, M. L. and C. Shapiro (1985) "Network Externalities, Competition, and Compatibility," *American Economic Review* (75) 3, pp. 424-440.
Katz, M. L. and C. Shapiro (1986) "Technology Adoption in the Presence of Network Externalities," *Journal of Political Economy* (94) 4, pp. 822-841.
Kauffman, R. J. and X. Li. (2003) “Payoff Externalities, Informational Cascades And Managerial Incentives: A Theoretical Framework For It Adoption Herding.” *2003 INFORMS Conference on IS and Technology, Atlanta, GA, 2003*.
Kauffman, R. J., J. McAndrews, and Y. M. Wang (2000) "Opening The ‘Black Box’ of Network Externalities in Network Adoption," *Information Systems Research* (11) 1, pp. 61-82.
Krazit, T., and Broache, A. (2006) " BlackBerry saved," CNET news.com, accessed on May 26, 2008, available online at http://news.cnet.com/BlackBerry-saved/2100-1047_3-6045880.html.
Lamm, H. (1988) "A Review of Our Research on Group Polarization: Eleven Experiments on the Effects of Group Discussion on Risk Acceptance, Probability Estimation, and Negotiation Positions," *Psychological Reports* (62) 3, pp. 807-813.
Li, X. (2004) "Informational Cascades in IT Adoption," *Communications of the ACM* (47) 4, pp. 93-97.
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.
Muehleman, J. T., C. Bruker, and C. M. Ingram (1976) "The Generosity Shift," *Journal of Personality and Social Psychology* (34) 3, pp. 344-351.
Mussi, S. (2002) "Sequential Decision-Theoretic Models and Expert Systems," *Expert Systems* (19) 2, pp. 99-108.

Oh, W. and Jeon, S. “Membership Herding and Network Dynamics in the Open-Source Community: The Ising Perspective” Management Science 53 (7), pp. 1086-1101.
Puterman, M.L. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. New York: Wiley & Sons, 1994.
Rogers, E. M. (1995) *Diffusion of Innovations*. New York, NY: Free Press.
Seale, D.A. and A. Rapoport. "Sequential Decision Making With Relative Ranks: An Experimental Investigation of the 'Secretary Problem.'" *Organizational Behavior and Human Decision Processes*, 69, 1997, 221-236.
Shanteau, J.C. "An Additive Model for Sequential Decision Making." *Journal of Experimental Psychology*, 85, 1970, 181-191.
Smith, L. and P. Sørensen (2000) "Pathological Outcomes of Observational Learning," *Econometrica* (68) 2, pp. 371-398.
Song, J. and E. A. Walden. (2003) “Consumer Behavior In The Adoption Of Peer-To-Peer Technologies: An Empirical Examination Of Information Cascades And Network Externalities.” *Ninth Americas Conference on Information Systems, Tampa, Florida, August 4-6, 2003*, pp. 1801-1810.
Sullivan, M.S., D. Burnham, B. Stevens-Wood, and B.C. Sheldon. "Sequential Assessment and Decision-Making in Humans." *Behaviour*, 132, 1995, 571-589.
Swanson, E. B. and N. Ramiller (2004) "Innovating mindfully with information technology," *MIS Quarterly* (28) 4, pp. 553-583.
Tingling, P. and M. Parent (2003) "Mimetic Isomorphism and Technology Evaluation: Does Imitation Transcend Judgment?," *Journal of the Association for Information Systems* (3) 5, pp. 113-143.
Walden, E. A. and G. J. Browne. (2002) “Information Cascades in the Adoption of New Technology.” *Twenty-Third International Conference on Information Systems, Barcelona, Spain, 2002*, pp. 435-443.
Walden, E. A. and G. J. Browne (2008) "Rational fads in investor reactions to electronic commerce announcements: An explanation of the Internet bubble," *Electronic Commerce Research and Applications* (7) 1, pp. 44-54.
Yaniv, I. (2004) "Receiving other people's advice: Influence and benefit," *Organizational Behavior & Human Decision Processes* (93) 1, pp. 1-13.
Zmud, R. W. (1983) "The Effectiveness of External Information Channels in Facilitating Innovation Within Software Groups," *MIS Quarterly* (7) 2, pp. 43-56.

APPENDIX

Below is the pseudo code for the simulation. The actual SAS code is available upon request.

1. Define *k*, priors prob(μ_A)= prob(μ_B)=0.5, σ^2=1, $\overline{\mu}_A$ and $\overline{\mu}_B$ => *d'*.
2. Calculate β = *k**(prob(μ_B)/ prob(μ_A)).
3. Draw a signal from a normal distribution with appropriate mean and variance.
4. Calculate the threshold as $r = \dfrac{\ln(\beta)2\sigma^2 - \overline{\mu}_B^2 + \overline{\mu}_A^2}{2(\overline{\mu}_A + \overline{\mu}_B)}$.
5. If signal > threshold then decision = A otherwise decision = B.
6. Calculate new β = old β *prob(decision| μ_B)/ prob(decision| μ_A)
7. Repeat 3-6 100 times
8. Repeat 1-7 1,000 times

About the Authors

Eric A. Walden is the James C. Wetherbe chair and associate professor of Information Systems and Quantitative Sciences at Texas Tech University and a visiting associate professor at the Tepper School of Business at Carnegie Mellon University. He research interests focus on developing a greater understanding of information systems in the organizational context. This includes issues within and among organizations. Eric combines strong theory from industrial organization in economics with empirical validation within an organizational context. It is his hope that this focus will bridge a gap between pure analytic economics issues and softer, real world organizational issues, making each more accessible to the other and helping to develop real theory unique to the information systems literature. Eric earned his Ph.D. at The University of Minnesota. His prior research has appeared in *Information Systems Research*, *MIS Quarterly*, and *Harvard Business Review*.

Glenn J. Browne is an associate professor at the McIntire School of Commerce at the University of Virginia. His research focuses on cognitive and behavioral issues in systems development, project management, and managerial decision making, including information requirements determination, information search, and stopping rules. His articles have appeared in *MIS Quarterly*, *Management Science*, *Organizational Behavior and Human Decision Processes*, the *Journal of Management Information Systems*, *Decision Sciences*, *Communications of the AIS*, the *International Journal of Electronic Commerce*, and other journals.

Journal of the Association for Information Systems

ISSN: 1536-9323