Informing Science: the International Journal of an Emerging Transdiscipline Volume 11, 2008

Resonance within the Client-to-Client System: Criticality, Cascades, and Tipping Points

T. Grandon Gill
University of South Florida, Tampa, Florida, USA

ggill@coba.usf.edu

Abstract

Resonance refers to the ability of a rigorous and relevant message to inform and impact a client. In multi-client systems, a critical part of achieving resonance is client-to-client informing. The paper examines this process, developing three alternative models. The *Criticality* model depends upon the message's ability to motivate clients to resend it; it is most applicable to simple sticky messages, such as rumors and urban myths. The *Information Cascade* model is client-motivated and depends heavily on the recipient's perception of what other clients are doing; it is most applicable to choice situations where decisions are visible and informing subsequent to a choice is self-regulated by each client. The *Tipping Point* model is based upon classical diffusion models with heterogeneous individuals (mavens, persuaders, and connectors) introduced into the community; it is most applicable to complex informing situations where continued sender-client involvement is useful throughout the informing process. The behavior of each of these models is studied and spreadsheet-based simulations are also presented. The conclusions characterize each model according to its domain of applicability and also consider how the emerging filed of network theory is impacting our understanding of client-to-client processes.

Keywords: informing systems, diffusion, information cascades, tipping points, criticality, critical mass, resonance.

Introduction

Within an informing system, resonance refers to the ability of a communication to make its way from sender to client once it has already met the standards of quality (e.g., rigor) and usefulness (e.g., relevance). When the term was introduced (Gill & Bhattacherjee, 2007), it was further proposed that resonance had two distinct aspects. The first was the ability of the message to inform a single client. The second was the ability of the message to produce subsequent client-to-client informing activities.

Because the theory behind achieving single client resonance is addressed elsewhere (Gill, 2008), the present paper will focus strictly on achieving resonance between clients within the client-to-client component of an informing system. This can be a particularly important process within systems that involve the transfer of complex information. The diffusion of innovation literature (e.g., Rogers, 2003) finds, for example, that client-to-client processes dominate all but the earliest stages of knowledge transfer. Moreover, several different models, mathematical and empirical in origin, have been proposed for a variety

Editor: Eli Cohen

of information tasks, but relatively few have been employed within the informing sciences. Thus, the overriding objective of the present paper is to introduce some of these models and consider their respective domains of applicability.

We begin by assuming that a message—of acceptable rigor and relevance—has been transmitted by a sender to a very small number of clients within a client community who have subsequently absorbed the message into their own mental models. Our interest is then to understand what subsequently happens to that message within that client community. In the present paper, we consider three models that make fundamentally different assumptions about the nature and motivation for communications:

1. *Criticality models*: This model is based on the concept of a critical system, most commonly used in the context of nuclear engineering. The simplest of the three models, it could be described as client-sender motivated communications, since it is applicable only when one client who possesses the information is strongly motivated to inform other clients about it.

2. *Information Cascade models*: Introduced originally in economic theory, this model is normally presented in terms of a client's making a choice between two options for which information about prior client adoptions is available. Although often applied to products (e.g., VCR formats, movies), it can also been applied to pure informing situations, such as the enrollment decision made between alternative classes or the choice of a research topic. It can be characterized as client-recipient motivated informing, since it is the potential recipient who actively decides which option to pursue.

3. *Tipping Point models*: Building upon assumptions presented in Gladwell's (2000) widely read book *The Tipping Point*, this model is typical of general diffusion models (e.g., Rogers, 2003) that examine how innovations—including ideas—migrate through communities. It could be characterized as a social-task model, since informing is motivated by both task performance-related criteria and by social criteria.

As each of the three models is presented, relevant literature is reviewed. For each model, key parameters that impact informing are identified and a simulation based upon the basic concepts of the particular model is developed. Interesting areas of behavior highlighted by the simulation are then presented and discussed. Finally, areas where each simulation could benefit from refinement are noted.

At the end of the paper, some general conclusions regarding the application of the models are presented. Central to these conclusions is the assertion that the client-to-client informing phenomenon is very important. It is no exaggeration to say that it is the principal mechanism by which many innovative ideas make their way through a client community. Thus, if the informing sciences are to evolve and prosper as a transdiscipline, it is critical that we include better understanding these processes in our research agenda.

Criticality Models

Terms such as criticality and critical mass are often applied to informing contexts. In this section we examine these models and consider the types of informing situations to which they might be applicable. We begin, however, by examining the origins of the criticality concept in nuclear engineering.

An Introduction to Criticality

The term criticality is used in many contexts, particularly in the context of complex systems (e.g., Bak, 1996). When applied to a process, the term is generally used to suggest the process is self-sustaining—meaning that it will continue without external inputs. If we are interested in studying client-to-client communications and how they can be sustained after an initial message is passed into the community by the sender, the applicability of the concept therefore seems obvious. Before using criticality in the informing context, however, it makes sense to see how the term is applied in its original context, nuclear engineering.

In nuclear engineering, the term criticality is most commonly applied to nuclear reactors. Nuclear reactors work through nuclear fission—utilizing the binding energy that is released when the atoms of certain heavy elements (most commonly uranium or plutonium) are split into smaller atoms (such as lead). Although fission of individual atoms occasionally occurs spontaneously, this does not happen often enough for the process to generate useful energy. You can, however, encourage fission to occur by hitting these heavy element atoms with neutrons travelling at an appropriate speed. Furthermore, it turns out that when an atom experiences fission, a certain number of unattached neutrons are also produced (typically 2 or 3 from a uranium atom's fission). Thus, if these neutrons can be utilized to produce subsequent fissions, it becomes possible for a chain reaction—also known as criticality—to occur.

Naturally, there are a great many engineering challenges that need to be addressed in order to establish and control a nuclear reaction. For example, in a typical uranium reactor, there are 6 key factors associated criticality (DOE, 1993):

1. Neutrons produced by fission are travelling far too fast to produce fission in Uranium-235, the most common fuel source. They can, however, occasionally produce fissions in Uranium-238 particles that are also in the core. These are known as fast fusions and add slightly to the number of neutrons in the reactor core.

2. Neutrons travelling at high speeds may leak from the reactor core, making them unavailable as a source of subsequent fissions.

3. While the neutrons are slowing down, they may be absorbed by fuel and non-fuel elements through a process called resonance absorption, which does not produce any fissions.

4. Once neutrons have reached a suitably slow speed, they may also escape from the core.

5. Slow (thermal) neutrons may be absorbed by non-fuel elements, such as structural materials used to build the reactor and control rods (which contain materials, such as boron, whose specific purpose is to absorb neutrons, thereby allowing the rate of fission to be controlled).

6. Slow neutrons may be absorbed by suitable U-235 fuel elements, producing fission and, as a side-effect, more neutrons that then can be used to produce subsequent reactions.

Multiplying these factors together, we get a six factor formula for reactor criticality:

k_{eff} =

(1) Percentage increase in neutrons created by fast fission *
(2) Likelihood that fast neutrons won't leak from the core *
(3) Likelihood that neutrons won't be captured by resonance absorption *
(4) Likelihood that slow (thermal) neutrons won't escape from the core *

(5) Likelihood that thermal neutrons won't be absorbed by non-fuel elements *
(6) Average number of neutrons produced by thermal fission

The value k_{eff}, also known as the criticality factor or effective multiplier, determines how many neutrons will be produced in each succeeding generation. Where the value is greater than 1, you have exponential growth (super criticality) that can be used to increase the power output of the reactor. Where it is less than 1, you have neutron production levels that decline with each generation—the total effect being similar to that of a multiplier in economic theory. Where the value is exactly 1, your reactor is critical.

The values of the 6 factors leading to criticality change over the life of a reactor. For example, as fuel is depleted, the degree to which (4) or (5) occurs will naturally rise for a given reactor configuration. Thus, reactors are designed with adjustable components (e.g., control rods) that can be partially removed from the core to reduce absorption by non-fuel elements (5) and make it possible to maintain criticality with less fuel.

The term "critical mass" is also frequently used in informing contexts. The idea behind this concept is that for a given core configuration (also sometimes referred to as buckling), if you do not have a certain amount of fuel, it will be impossible to overcome neutron leakage (items 2 and 4) through the surface area of the core. Because volume grows more rapidly than surface area with added mass, the greater the mass, the lower the ratio of escaping neutrons to those that remain inside the core.

With these concepts in mind, we now turn to how criticality can be applied to informing contexts.

Criticality and Pair-Wise Rumor Models

Certain types of client-to-client informing processes have traditionally been modeled in a manner that closely parallels the nuclear criticality model. The most widely known of these models involves the transmission of rumors. The basic model, which is derived from epidemiology, is as follows:

1. There are three groups: susceptibles, infectives, and removed cases (Lefevre & Picard, 1994). *Susceptibles* have never heard the rumor before. *Infectives* are individuals who are actively spreading the rumor. *Removed cases* are those who have heard the rumor, but are no longer spreading it.

2. Every time a susceptible meets an infective, the susceptible becomes an infective. This is how the rumor spreads.

3. Every time an infective meets a removed case, he or she also becomes a removed case. This is viewed as the infective losing interest in transmitting the rumor.

4. There are two variations of the model regarding what happens when two infectives meet. In the first, initially proposed by Daley and Kendall (1965), both become removed cases. In the second version, initially proposed by Maki and Thompson (1973), only one of the two becomes a removed case (Lefevre & Picard, 1994). A third possible variation is also possible: when two infectives meet it reinforces their enthusiasm, causing both to continue spreading the rumor.

The basic models for rumor propagation are simple enough that a variety of stochastic and closed form solutions have been proposed (e.g., Daley & Kendall, 1965; Dietz, 1967; Dunstan, 1982; Lefevre & Picard, 1994; Pittel, 1990). For our purposes, however, it is sufficient to observe that the rumor spreading process, as proposed in these models, fits the notion of criticality quite well.

Specifically, let us assume that we break the space of clients into randomly assigned pairs. For every pair with at least one infective, at least one rumor transmission occurs. Considering only the cases where there are not two infectives, the multiplier for rumor production in the subsequent generation will be:

$$k_{eff} = 1 + p_s - p_{rc}$$

Where p_s is the probability of the partner being a susceptible (given that one infective is already known) and p_{rc} is the probability of a removed case, with the 1 representing the infective continuing to spread disseminate the rumor in the next period, and the other two terms representing the chance of gaining a new infective (p_s) and the chance of losing the existing infective (p_{rc}). This multiplier still applies if we assume that two infectives meeting does not change the state of either (two rumors in this period, two rumors in the next). If the assumption is that both of the infectives change state to a removed case, then the formula becomes:

$$k_{eff} = 1 + p_s - p_{rc} - p_i$$

where p_i is the probability that the second member of the pair is an infective. If, however, only one of the infectives transitions to a removed case, the expression becomes:

$$k_{eff} = 1 + p_s - p_{rc} - (1/2)p_i$$

For example, if everyone were an infective under this assumption, then p_i would be 1 and the pairing of infectives would lead to half as many rumor transmissions in the next generation.

In a large population where a single sender initiates a rumor, $p_s >> p_i$ and p_{rc}. In such situations, super criticality is going to be experienced for a number of generations of rumor spreading—leading to exponential growth in the population familiar with the rumor (p_i and p_{rc}). Eventually, however, p_i and p_{rc} will become large enough so that k_{eff} is driven below one and the rate of propagation subsides.

Even the early developers of the rumor transmission model recognized that some assumptions might need to be changed. For example, the basic Daley and Kendall model can be extended by including a probability that that a susceptible becomes an infective upon a pairing instead of making it a certainty. Similarly, one can assume that it will take multiple encounters with removed cases or other infectives before an infective ceases to spread the rumor (Dietz, 1967). In addition, a wide range of extensions can be made when assumptions of homogeneity in the client population are dropped and social networks are included, as well as the possibility that rumors may be spontaneously forgotten (Nekovee, Morenob, Bianconic, & Marsilic, 2007).

Criticality and the "One Shot" Rumor Simulation Model

When specifically employed in the informing systems context, some of the assumptions of the typical pair-wise model are unduly limiting. For example, when rumors are transmitted by email, it is quite possible—indeed, likely—that multiple clients will be recipients. Similarly, the assumption that the infective client will immediately discern the reaction of the recipient is also questionable in an electronic format. Fortunately, it is relatively easy to construct a new model that addresses these issues. Consider the following set of assumptions:

1. Upon encountering a rumor transmission, the recipient may chose to ignore the transmission with probability p_{IGNORE}. We'd expect this probability to vary according to the recipient's previous experiences (if any) with the sending client.

2. Upon deciding not to ignore the transmission, the individual may decide to commit it to memory with probability p_{COMMIT}. We would expect this to depend on the nature of the message. In particular, messages that are sticky (Gladwell, 2000; Heath & Heath, 2007)

are much more likely to be internalized.

3. Upon committing a transmission to memory, the individual may decide to disseminate it, with probability $p_{DISSEMINATE}$. This would probably depend on both message stickiness and would also vary by recipient.

4. Upon deciding to disseminate the rumor, the individual transmits it to N individuals. This value would undoubtedly vary by recipient.

With these assumptions, it is no longer necessary to make the infective/removed case distinction, since dissemination by an individual automatically stops after N transmissions. Assuming P_i is the informed population at time i, our approximate criticality formula now becomes:

$$k_{eff} = (1 - P_i)*(1 - p_{IGNORE})* p_{COMMIT} * p_{DISSEMINATE} *N$$

If we know the precise number of informed individuals in two periods, a more precise version of the formula involves computing the probability that a particular client will not be hit. Assuming C_0 is the original population, the less approximate criticality formula is as follows:

k_{eff} = Number of Clients Informed in Period i/Number of Clients Informed in Period i - 1

$k_{eff} = (C_0 - P_i)* (1\text{- Probability of non-absorption}^{\text{Expected Number of Messages/Client}})/ (P_i - P_{i-1})$

Where:

Probability of absorption $\equiv 1 - (1 - p_{IGNORE})* p_{COMMIT}$

Expected Number of Messages/Client $\equiv (P_i-P_{i-1})* p_{DISSEMINATE} * N / C_0$

By incorporating p_{IGNORE}, p_{COMMIT} and $p_{DISSEMINATE}$ into this model, we explicitly allow for the possibility that the nature of the rumor/message being spread can have a significant impact on its dissemination. To demonstrate this, a simulation was developed (see Appendix A) that allows

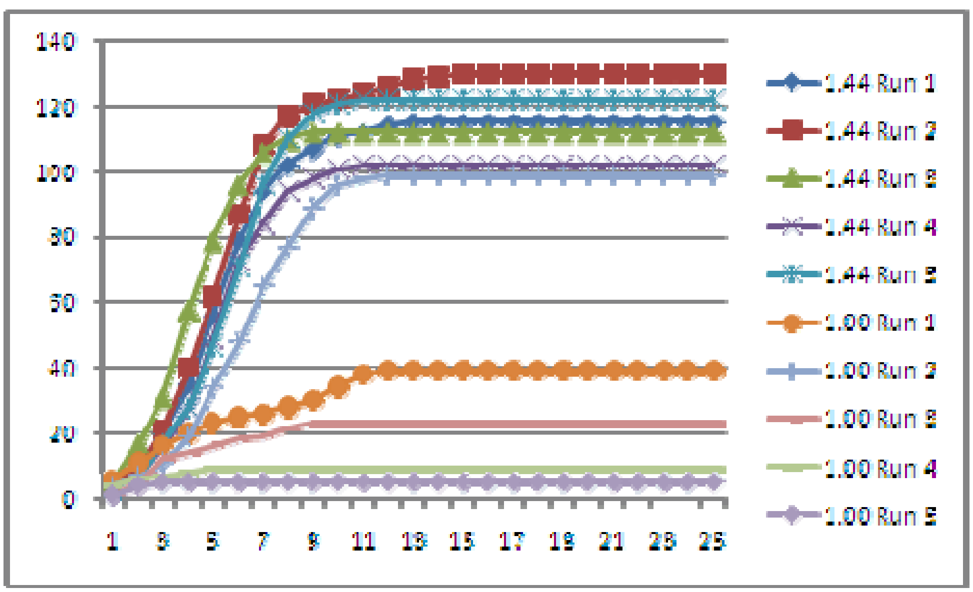


Figure 1: Example Criticality Runs. The run results illustrate how different criticality factors lead to substantially different levels of information penetration (complete penetration would exist at 150). It also shows that randomness can exert a substantial influence on penetration levels (i.e., note the differences in final penetration for the five runs in conducted for each criticality factor).

testing for parameter impact on the dissemination process. Figure 1, for example, plots period (X-axis) and number of clients informed (Y-axis) for a series of simulation runs. Specifically, the results of 5 runs with an estimated initial criticality factor of 1.44 and 5 runs with an estimated initial criticality factor of 1.00 are displayed. The figure illustrates three important characteristics of the criticality informing model: 1) that higher stickiness generally leads to dissemination that is both greater and faster, 2) that the random nature of rumor-based informing can lead to considerable variation in penetration even for the same parameters, and 3) even high initial stickiness factors do not necessarily produce 100% informing throughout the system (which would be a value of 150 for this particular simulation).

The reason for the lack of compete dissemination has to do with the continuous decline in the value of the criticality factor as clients become informed. In Figure 2, both estimated and exact criticality are plotted over the course of a 1.44 initial estimated criticality run, along with the fraction of clients who have been informed (ranging from 0 to 1). As more clients are informed, the number of available sites that a message can reach declines, similar to the original rumor model. Also, the series labeled Criticality—which is based on the probabilistic formula—drops to 0 as soon as the messages stop, since it cannot be computed if no prior messages are sent. That is why it is useful to have the estimated formula available as well.

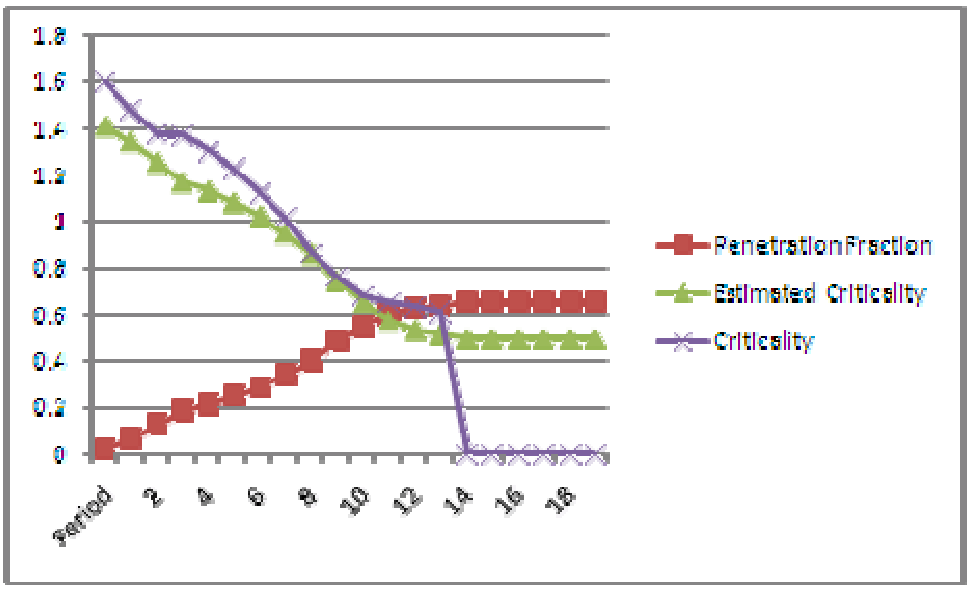


Figure 2: Criticality Declines as Clients are Informed. In such a model, we would nearly always expect some individuals in the community to remain uniformed.

Criticality Conclusions

Criticality-based informing models all share two features in common: 1) if k_{eff} is super critical, they exhibit an exponential pattern of growth until a significant fraction of the client population has been informed, 2) they are assumed to be sender-motivated. In these models, recipients do not demand to be told a rumor. Instead they encounter the rumor by chance.

For the second of these reasons, criticality models are most likely to be exhibited in situations where the message involved is both sticky and simple (with simplicity often being a characteristic associated with stickiness; Heath & Heath, 2007). If a message does not quickly resonate with an individual recipient, the recipient is unlikely to put forth the effort required to become a subsequent sender. The same can be said for messages that are overly complex. As a consequence, we postulate that other mechanisms will dominate in situations where more substantial informing needs to take place.

A particularly significant limitation of criticality models relates to the type of community to which they are applicable. Specifically, the assumption underlying both the pair-wise and one shot models is that the probability of any two individuals coming together is essentially random. This assumption is plausible in small communities, where everyone has linkages to everyone else, or in randomly connected networks, where the probability that a connection exists between two nodes is random. The assumption would not be valid, however, for small world networks (e.g., Watts, 2003) or for networks where node connectivity levels are determined by a power law (e.g., Barbasi, 2002). Networks with these properties lead to the formation of tightly knit communities loosely connected with other communities—producing the small world effect—and to the development of hub nodes that exhibit vastly higher connectivity that the typical node. In the real world, however, many phenomena (e.g., web sites, membership on corporate boards, behavior of cell proteins, degrees of separation from actor Kevin Bacon; Barabasi, 2002; Watts, 2003) exhibit these behaviors. Moreover, the same epidemiology models upon which the criticality models are based have been shown to suffer from the same shortcoming. The propagation of a disease such as AIDS, along with that of many computer viruses, can be better modeled when hub nodes of vastly higher connectivity are incorporated (Barabasi, 2002). Thus, the criticality model is likely to remain useful only for modeling localized spread of sticky communications.

Information Cascade Models

Economists use the term *information cascade* to refer to choice situations where decision-makers ignore their own private perceptions of alternatives in favor of information regarding what other decision-makers have chosen. For example, if your decision to purchase a particular model of car is heavily influenced by the observation that you are suddenly seeing a lot of that particular model car out on the road, then—if others are similar to you in their decision-making process—more and more people will make the same decision and the popularity of that particular model will explode. While the concept is quite general in terms of the choices it can refer to—e.g., consumer products, films—it is sometimes applied to informing situations as well, most notably the development of paradigms.

Information Cascades in Economic Theory

The theory of information cascades (also referred to as informational cascades) was originally developed to explain rapid changes in behavior within a group, as are often observed in fashion, financial markets, science, and medicine (Bikhchandani, Hirshleifer, & Welch, 1992). In addition, it can be used to explain consistencies in behavior that are not necessarily rational, such as tendencies of employers to discount applicants with a long stretch of unemployment, perceived as evidence of rejection by previous employers (Kubler & Weizsacker, 2003) or for the tendency for a paper rejected by one journal to be rejected by subsequent journals (Bikhchandani et al., 1992). Prior to cascade theory, explanations of such behavior included (Bikhchandani et al., 1992, p. 993):

(1) sanctions on deviants,
(2) positive payoff externalities [meaning that consistency of adoption leads to higher overall payoffs, as might occur when everyone adopts the same communication standard],
(3) conformity preference, and
(4) communication [such as occurs when prior adopters extol the benefits of their choice].

Although these explanations certainly help account for rapid adoptions of a particular behavior, they are not particularly useful in explaining why behaviors might suddenly shift.

The basic concepts behind an information cascade are straightforward. In the simplest case, assume that that multiple individual clients are each faced with making a choice between two alternatives. In order to make that choice, two sources of information are available:

1. Private information, including information from direct observation by the individual client of the alternatives.
2. Public information regarding what previous clients have adopted each alternative.

If the clients involved use public information in preference to private information, then very quickly clients will all begin to choose the same alternative, leading to the cascading or herding phenomenon. One of the two original formulations of the model also extended the analysis to include choices between many options (Banerjee, 1992).

Experimental studies have found that information cascade behaviors are relatively easy to produce (e.g., Anderson & Holt, 1997). They also demonstrate that even when subjects are aware of the information cascade phenomenon, they will often fail to recognize that a cascade is taking place and therefore tend to be overconfident in judgment regarding the suitability of a choice (Grebe, Schmid, & Stiehler, 2008).

The information cascade model has also been extended to incorporate some heterogeneity. For example, experts can play an important role in information cascades. In a cascade situation, where all participants make a decision based upon what other participants have chosen, the state-of-the-world has little influence on behavior. Including just a few experts—who base decisions on the state-of-the-world rather than on the behavior of other agents—can lead to changes in the system when exogenous events occur (Bowden & McDonald, 2008). Similarly, information cascades can be influenced by opinion leaders, who are not necessarily experts, which means that cascade phenomena can sometimes be inhibited by aggregating information so that details—such as who voted for a particular proposition in a meeting—are hidden (Arya, Glover, & Mittendorf, 2006).

A number of criticisms and limitations have been raised with respect to the pure information cascade model. For example, in some experimental simulations, temporary cascades are more common than extended cascades (Goeree, Palfrey, Rogers, & McKelvey, 2007), a phenomenon not predicted by the basic model. When allowed to pay for private information, experimental subjects have been found to pay more than an optimal amount in early stages of play, preventing cascade formation in some cases (Kubler & Weizsacker, 2004). Improvements in decision making with payoffs, an effect not predicted by the model, have also been observed (Anderson, 2001). In field settings, cascades do not always appear as expected. For example, motion picture revenues are better modeled with an extension that allows individuals to report quality to each other (De Vaney & Lee, 2001). Social groups and collective knowledge may also play a larger role in decision consistency than is credited by cascade models (Shiller, 1995). As was also the case for criticality models, a pure threshold model does not account for the underlying topology of the network through which information passes. Thus, structures such as small world networks (e.g., Watts, 2003) and modular scale-free networks (e.g., Barabasi, 2002) could exhibit substantially different behaviors.

Since information cascades are an informing phenomenon, they have the potential to impact any multi-client informing system. When the choice to be made by a client is whether or not to become informed on a particular topic, to attend a particular channel, or to adopt a particular mental model, the relationship between information cascades and informing becomes even more direct. It is relatively easy to come up with examples of these types of choices, such as:

- Deciding what classes to attend
- Choosing what threads to read in an online discussion group
- Choosing what videos to watch on a site such as YouTube
- Deciding what topics to research
- Deciding what books to read or purchase
- Deciding what publications or web sites to subscribe to
- Deciding what technological standards to employ in an informing system
- Deciding what sender to attend to in a multi-sender system (see Gill & Bhattacherjee, 2007)

In all these situations, knowledge of prior client choices is likely to be relevant. Where the informing system involves technology, information useful in assessing past adoptions at low cost is frequently available—such as the view counts on YouTube, post and view counts on social networking or discussion sites, or sales rankings on sites such as Amazon.com. In these cases, the popularity of each information source becomes a form of selectivity bias that then acts on your information acquisition behavior.

Cascades and Externalities

From a practical standpoint, one of the greatest limitations of information cascade research is the decision to treat cascades independently of positive or negative payoff externalities. For example, where positive reinforcement is present—such as was the case for the consumer decision regarding what VCR standard (VHS or Betamax) to purchase (Arthur, 1988)—the greater the number of individuals choosing one option, the more desirable that option becomes. On the other hand, the attractiveness of entry into a particular new industry is likely to decline with the number of competitors who have already entered the industry (Porter, 1980).

Virtually all information cascade research involves scenarios in which the number of agents adopting a particular choice does not impact the desirability of that choice. Such an assumption makes great sense from the perspective of trying to understand the information cascade phenomenon in isolation. Unfortunately, it is also likely to be valid mainly for choices made in laboratory conditions—such as trying to identify which bin a colored ball comes from (e.g, Anderson & Holt, 1997)—rather than for choices in the field. For example, the desirability of seeing a particular motion picture (e.g., De Vany & Lee, 2001) or TV show is likely to rise with the percentage of individuals in your social circle who have seen it, since it grows more likely to become a topic of conversation. On the other hand, the objective desirability of a particular restaurant (e.g., Banerjee, 1992) may decline with its popularity, as long waits and crowded table placement may ensue.

Going hand-in-hand with the question of externalities is the option of postponing a decision. If, for example, you are trying to decide whether or not to see a particular motion picture in the theaters, you may decide to wait a week before choosing. Such delays can be particularly important when combined with self-reinforcing phenomena. For example, it might be perfectly rational for an individual who wanted a VCR to postpone the purchase until seeing which format achieved dominance. Indeed, in such situations there might be a collection of agents literally waiting for clear evidence of a cascade developing, at which point they join it.

To illustrate how externalities might impact a client-driven informing decision, consider the question of choosing a topic to research. In making this decision, it is reasonable to assume that you start with two sources of information. First, you would have a base of private information, presumably far from perfect in quality, about the desirability of the topic. Second, you would have information regarding how many individuals have already researched the topic—as can be rapidly determined by performing a few keyword searches.

How to appropriately weigh these two information sources is a complex problem—and one that is likely to vary considerably across individuals. To begin with, you might come to a particular topic with a passion or with a novel idea that you want to develop. Spurred by passion, it is not clear that the presence or absence of a large body of prior choice history (i.e., existing research) would make much difference to you. On the other hand, you might approach the decision from a purely opportunistic standpoint with career benefits in mind. In this case, you would want to know whether or not the research topic is likely to produce publications beneficial to your career—a question upon which prior considerations of the topic will certainly have some bearing. Unfortunately, the nature of the relationship between past research and future prospects is far from straightforward. For example:

- If the base of prior research is very small, there is a strong likelihood that it will be hard to publish similar research and, if published, your research may not receive many subsequent citations.

- If the base of prior research is very large, it will take considerable time to develop expertise on the topic and the field may already be crowded; once again, this could limit the degree to which your research is cited.

Thus, from the opportunistic perspective, the "ideal" topic might be one with moderate size body of research that appears to be growing. The same reasoning might apply to deciding what threads to read on a large discussion site.

An important consideration in threshold models can also be the degree to which fitness depends on how many other individuals choose the same peak—the previously introduced idea of payoff externalities. One possibility is that peaks might be occupancy independent, meaning that a peak's fitness does not depend on its popularity. Your choice of salad dressing at a restaurant might meet this criterion, provided they don't run out. As previously noted, this is the most common assumption of information cascade models. There are, however, three other categories of possible peak that might be routinely encountered, which will be termed cascading, inverted-U, and competitive. These categories are defined as follows:

- *Cascading*: the fitness of an option grows with each adoption. The VCR example fell into this category, as do many types of standards-driven behaviors. This type of peak should amplify any initial tendency towards information cascades.

- *Inverted-U*: Fitness peak is somewhere between 0 and universal occupancy. The research topic example fits this profile.

- *Competitive*: Desirability of an option declines as additional clients choose the option. For example, the first person to patent an invention will typically receive substantially greater rewards than the second person coming up with the same idea. Because such peaks inherently resist the cascade phenomenon, to the extent cascades can occur at all for such peaks it would only be in cases where the visibility of individual choices lags the choice itself, allowing many individuals to seek the same peak without realizing its popularity. Thus, such cascades would tend to happen very quickly, or not at all.

A Cascade Simulation

To examine how a variety factors not included in traditional economic models of information cascades impact client decision-making, it is useful to develop a simulation. The model, described in Appendix B, incorporates the following parameters:

1. *Optimal Occupancy Percent*: Used to simulate the type of peak (e.g., cascading is 100%, competitive is 0% and inverted-U is somewhere in the middle).
2. *Percent Deviation Cost*: Allows the cost of deviating from optimal occupancy to be adjusted. By setting the parameter to 0, occupancy independence is simulated.
3. *Private Observation Error:* Allows client errors in assessing the true value of the option, prior to occupancy effects, to be simulated. 0 assumes that the underlying value of the option to the client can be assessed with perfect accuracy.
4. *Taste Difference Factor:* Allows variations in option desirability across clients to be tuned.
5. *Behavior Constant:* Allows differences in client-perceptions of the value of adoptions by other clients (i.e., observed behaviors) to be tuned.
6. *Observation Weight:* Percent of weight given to private observations compared with occupancy (prior adoption) data. Setting the value to 0 means that choice is entirely driven by occupancy differences (independent of their effect on option quality). 100% means that the decision is based entirely on perceived option quality (including any fitness impacts of occupancy).
7. *Threshold Factor*: The difference in value between the two options that must be achieved before a decision is made.
8. *Urgency Factor*: Adjusts the threshold to make decisions more likely (+) or less likely (-) as time passes during the simulation.
9. *Allow Switch*: Allows clients to switch between options in the periods after the initial decision has been made.

As further discussed in Appendix B, the model employs the following logic:

A. Two options are assumed. Initially, each client (150 in all) is assigned a random value between 0 and 1 that reflects the client's underlying preference (taste) for the option. The level of preference can be adjusted with *a taste difference factor*. The randomness element (0.5 is no preference) means that, on average, the two options should be perceived as being equally fit—although fitness will vary by individual.

B. During each period, individual difference in private preference for each option is assessed. This preference is based upon:
   a. The taste factor discussed in A.
   b. A random error (times *private observation error*) that changes each period
   c. The incremental differences in option values based on previous period occupancy (computed using the absolute value of *occupancy percent – optimal occupancy* times the *percent deviation cost*).

C. For each period, a behavior-based preference is also assessed, taking the difference in prior period occupancies for the options and multiplying them by the *behavior constant*.

D. Private and behavior-based preferences are then summed, weighted as follows:

   observation weight * (private preference) +
   (1- observation weight) * (public preference)

E. If the difference between the two preferences exceeds the individual's decision threshold, the option is chosen and included in the total for that period's occupancy—which will impact other agent choices in the next period. The threshold is determined by a uniformly distributed random threshold variable for each individual between 0 and 1 times the *threshold factor* parameter. If a positive *urgency factor* value is set, the factor is used to reduce the threshold by an additional increment each period.

F. If *allow switch* is FALSE, once a preference is chosen, it cannot be reversed. If not, preferences may switch any time the threshold value for an individual is reached.

As is typical with simulation models, the wealth of parameters available makes a complete exploration of model behavior combinatorially prohibitive. Also, because many values are established for tuning purposes (e.g., *taste difference factor*, *behavior factor*, *threshold factor*), specific numeric results are not going to be of particular interest. What is of interest, however, is some of the qualitative behaviors that are exhibited by the model under different initial conditions. Some relevant highlights of these, at different levels of *optimal occupancy percentage*, are now explored.

Occupancy independent

The occupancy independent case makes the same basic assumption as most economic models of information cascades but adds the notion of a threshold for decision-making, as opposed to forcing a decision for each period. The impact of the threshold is that a certain percentage of clients may remain undecided throughout the 25 period simulation run. By reducing the threshold factor, or by adding an urgency factor, the percentage making a decision can be increased.

For high values of the *observation weight* (e.g., preference is determined almost entirely by private values), choices are generally equally split between the two options (reflecting the underlying taste distribution). Where individuals are allowed to move back and forth (*allow switch* is TRUE), as *observation weight* declines this distribution remains relatively stable until a relatively sharp inflection point is reached after which 3 distinct distributions are appear: all option 1, all option 2, and 50-50 options 1 and 2. As *observation weight* further declines, the 50-50 case becomes less common, but does not vanish entirely. Thus, the impact of the threshold is to enable both information cascades and non-cascading situations to be possible for the same set of parameters. Figure 3 illustrates these different stability points for 3 runs created using the same initial parameter values. Two lines are plotted over time for each of the 3 runs, one showing Option A

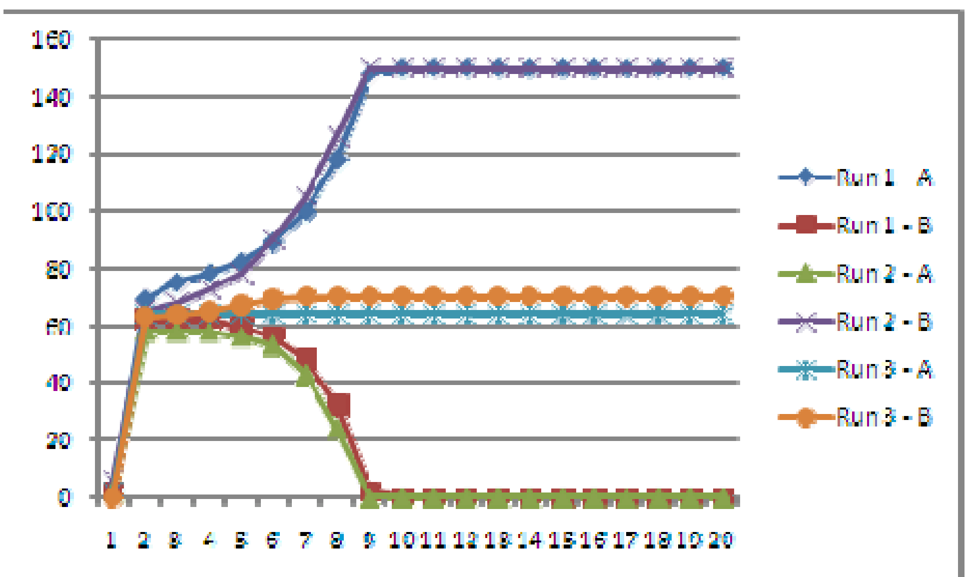


Figure 3: Cascade-dominated simulation where switching is allowed. This illustrates how information cascades can be quite unpredictable, with three distinct outcomes arising from the same parameter selections. In the figure, Options A and B necessarily add to 150 or less (less signfiying some individuals remained undecided). Run 1 signifies a cascade to Option A, Run 2 signifies a cascade to Option B, run 3 signifies a 50-50 outcome.

choices, one showing Option B choices. The sharp spike from period 1 to period 2 represents primarily taste-driven selections. The subsequent separation into 3 paths represents the two cases where a single option dominates (top and bottom) and the case where the options are close enough so that initial taste differences continue to dominate.

Where switching after a choice is not allowed, the situation is somewhat similar except that cascading does not occur until lower values of *observation weight* are reached, and instead of 100%-0% and 0%-100% distributions, we see X% - (100-X%) with X moving away farther away from 50% the smaller that the *observation weight* gets. This reflects the fact that in a threshold model, some low threshold individuals will always make a choice before a consensus (i.e., information cascade) is reached. The difference between allowing switching and not allowing switching can have a dramatic impact on volatility of outcomes. For example, Figure 4 and Figure 5 both plot Option A selections for the same parameter values, differing only with respect to allowing switching, across 10 runs. (An urgency value was set so no non-choice values remained at the end, meaning that the value of Option B was always 150 – Option A, which is why it was not plotted).

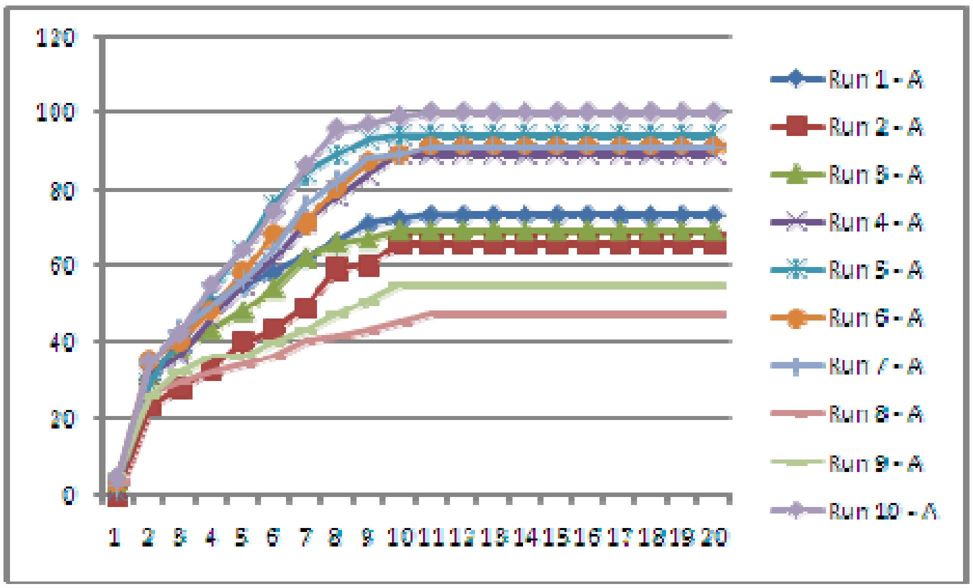


Figure 4: Example runs with no switching allowed. This figure illustrates how eliminating switching leads reduces the impact of cascades, with alternatives instead tending towards the 50-50 case (where Option A is 75).

While cascading phenomena obvious exerted some impact on the "no switching allowed" run for Figure 4, taste clearly played dominant role, with the 50% mark (75 Option A selections) being roughly the midpoint of the runs. In Figure 5, using the same parameter values with switching, however, we see a clear bifurcation pattern in outcomes resulting from cascades.

Competitive

The competitive case assumes that the more occupied a peak becomes, the less attractive it will be. For such a fitness space, there is no clear overall fitness optimum, since any distribution will tend to produce the same overall fitness summed across individuals. While we would not expect such a fitness profile to be particularly prone to information cascades, we do—from time to time—see examples of such behavior in everyday life. For example, when you enter an unfamiliar environment for a relatively routine activity (such as going into a motor vehicle department to

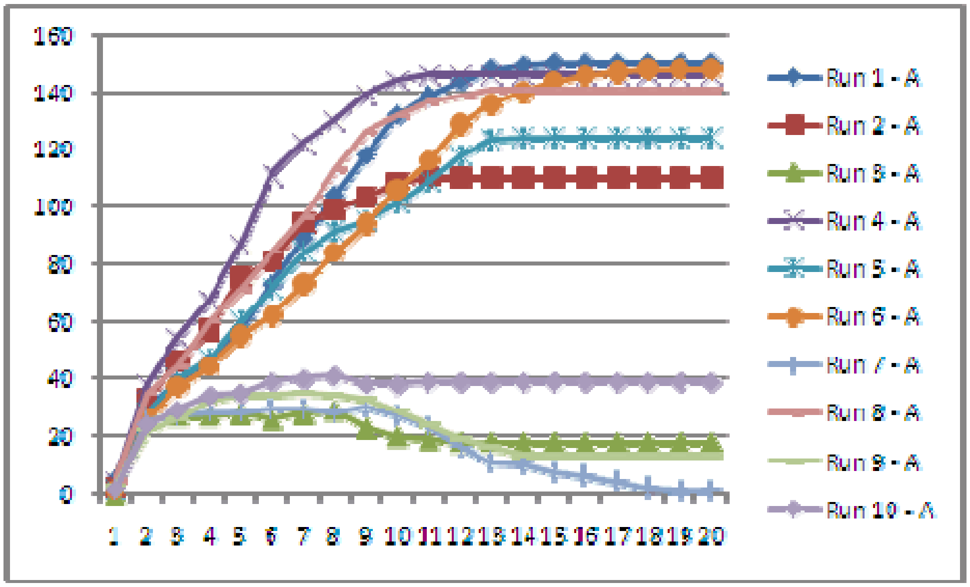


Figure 5: Example cascade runs (using same parameters as Figure 4) with switching allowed. In these runs, we find that cascading phenomena becoming much more pronounced, as individuals switch to popular peaks after initially choosing less popular ones. For this set of parameters, the 50-50 case became unstable, with either all Option A (close to 150) or all Option B (close to 0) becoming the principal stable states.

renew a driver's license) it might not be entirely irrational to stand in the *longest* line on the mistaken assumption that the other lines might be for a different—and less common—activity.

The competitive case exhibits qualitative behaviors very similar to the occupancy independent case, particularly where switching is not allowed. Where switching is allowed, on the other hand, as the observation weight declines, we begin to see period-to-period oscillations. In the most extreme case, this involves 0%-100% to 100%-0% switches each period—a predictable outcome since as soon as you discover that you are on an over-occupied peak it makes sense to move off of it. To a great extent, these huge swings are a consequence of the discrete model being used and would not occur in a continuous model. Nonetheless, it would be reasonable to expect that some oscillations could take place in the competitive case if not sufficiently damped by high switching costs.

Cascading

The cascading case, so named because of the likelihood that cascading behaviors will occur, assumes that the more occupied a peak becomes, the more attractive it will be. For this model, 100% occupancy for a single peak is clearly the most desirable outcome.

In the cascading case, where switching is allowed, even with very high observation weights (e.g., 95%) we see the emergence of three outcomes 100%-0%, 0%-100% and roughly 50%-50%. As observation weight drops when switching is not allowed, we see a similar pattern, but there will always be some individuals who have chosen the non-dominating option (the Betamax owners, in the VCR example). When threshold is reduced or urgency is increased, the number of individuals on the non-dominant option grows. This corresponds to the situation where a decision is forced prior to full information on other adoptions being available.

Inverted-U

The inverted-U case (simulated as an inverted-V, with a peak at the optimum percentage) has a number of interesting variations. The most interesting of these is probably the 50-50 case in this simulation, since a clear economic maximum occurs where clients divide themselves equally across the two peaks. Where observation weight is high, this distribution tends to be the outcome.

As observation weight declines and switching is allowed, we reach a fairly sharp inflection point where 4 distinct stable outcomes emerge (0%-100%, 100%-0%, 50%-50%, 60%-40%, 40%-60%). The last two of these—which are approximate—represent situations where small cascades cause an overshoot of the desired 50-50 split. The first two represent larger cascades. As observation weight declines further, the full cascade outcomes become completely dominant. Where switching is not permitted, we see the same outcome as previously reported: as observation weight declines, outcomes skew towards one peak or the other.

For the inverted-U case where the optimum occupancy is 25%, total fitness summed across all clients is constant for a plateau of occupancy between 25% and 75%. Once all clients have decided, however, if either is below 25% (and, correspondingly, above 75%), total fitness is suboptimal. For high observation weights, nearly all outcomes are 50-50, although a few non-converging outcomes were noted, signifying oscillations. Thus, this scenario appears to be somewhat less stable than the others. As observation weight is reduced, a sharp inflection point occurs after which the familiar stable outcomes (100%-0%, 0%-100%, and 50%-50%) emerge. For the inverted-U case where the optimum occupancy is 75%—which has the same profile in terms of total fitness across all clients—these outcomes are observed almost immediately, at very high levels of observation weight. Indeed, the 75% optimal occupancy runs were nearly identical in behavior to the cascading runs.

Cascades Conclusions

Just as the criticality model is useful for thinking about multi-client resonance when client-senders are motivated by sending, the cascade model is useful for thinking about situations where client-recipients are trying to choose between alternate informing sources. To be realistic, however, the economic model of information cascades needs to be extended in a variety of ways, such as:

- Allowing the postponement of choices that don't meet a desired threshold
- Allowing the "occupancy" of a choice to impact its desirability
- Allowing for choices to be specified as either being reversible or irreversible.

The basic model used to examine the impact of these factors shows that regions generally exist where cascades are uncommon, where they may or may not occur (with 50-50 distributions being the most common outcome if they do not, in our choice between two equivalent options model), and regions where they nearly always occur—often to the detriment of the clients involved.

The cascades model also makes a number of assumptions that are worth considering. First, the aggregate of decisions made by other individuals should be visible to all individuals considering a decision. This would imply the model would function effectively only in a small community setting (e.g., a group or gathering) or in a more dispersed network where some other means of determining aggregate behavior is broadcast (e.g., polls prior to an election, sales rankings). It also means that the nature of the choice must be simple enough to be tallied easily. Choice of a VCR standard would fit that requirement; philosophies of life probably would not.

Another aspect of the cascades model is its client focus. Unlike the criticality model, where the sending client derived utility from passing the information to other client recipients, the cascades model assumes that the client-recipient is completely driving the informing process. It does not address the issue of why other clients within the system are willing to make their choices known to undecided clients. This suggests that it would not be a good model for decisions with a competitive peak or where substantial informing effort on the part of the already decided clients was necessary once an undecided client chooses to adopt a particular choice. Thus, it seems most applicable to situations where the client-recipient, after having made the choice of an option, then ends up taking most of the responsibility for subsequent action—as might be the case in choosing between self-regulated learning options. Choosing a topic to research and choosing a type of VCR represent good examples of this type of situation. Many situations, however, require commitment on the part of both a client-sender and a client-recipient in the event effective informing is to take place within the multi-client community. For these situations we need a different model, one based on the principles of diffusion through a social network.

Tipping Point Model

Malcolm Gladwell's (2000) book *The Tipping Point* popularized a great many concepts that had long been evolving in the innovation diffusion literature (e.g., Rogers, 2003). We begin by summarizing some of the key conclusions of that literature, then examining some of the specific ideas presented by Gladwell in his synthesis. Finally, we examine the behavior of a computer simulation model designed according to the tipping point principles, the details of which are presented in Appendix C.

Diffusion and Tipping Point Research

The diffusion of innovations literature is huge in size (as of 2003, an estimated 5200 publications on the subject; Rogers, 2003, p. xvii) and quite informative in its conclusions. The seminal book in the field, *Diffusion of Innovations*, was written by Everett Rogers, a researcher whose pioneering efforts in early studies of diffusion—conducted in the 1950s after patterns started to become apparent in the adoption of farming technologies during the 1930s and 1940s—is now in its fifth edition (Rogers, 2003) and continues to be an important force in the field. Some of the key findings from this research stream, as summarized by Rogers, are as follows:

- Certain characteristics tend to make some innovations easier to diffuse than others. Examples of these are simplicity, compatibility with previous models or ideas, relative advantage compared to previous ideas, trialability (the ability to try out the innovation prior to adopting it), and observability (Rogers, 2003, p. 222). Ideas without these characteristics take much longer to diffuse.

- Diffusion does not occur immediately but, instead, through a gradual process of adoption within the client community. Two forces that are particularly important for this process are mass media (i.e., any communication where a single sender provides information to multiple clients concurrently) and interpersonal communications within the client network. In general, mass media communications are more important in the earlier stages of communications, while interpersonal communications dominate later stages (Mahajan, Muller, & Bass, 1991, cited in Rogers, 2003).

- Diffusion processes often have to reach a "critical mass" after which diffusion starts to take off at a very rapid rate (Rogers, 2003, p. 349).

- Individuals within client communities are not homogeneous. Rather, they exhibit different characteristics with respect to their willingness to adopt innovations. These may be modeled in terms of thresholds (Rogers, 2003, p. 355). Idealized categories of adopters are often classified as: innovators, early adopters, early majority, late majority, and laggards (Rogers, 2003). Individuals may also exhibit different degrees of influence on other clients in the community (e.g., opinion leaders; Rogers, 2003, p. 300), awareness of the social nature of the community (e.g., key informants; Rogers, 2003, p. 310), and willingness to venture outside of their community and cumulative past experience (innovators; Rogers, 2003, p. 282).

Because mass media exerts its influence on the most receptive portions of the client community (i.e., innovators), we can expect that interpersonal client-to-client communications will play a much more critical role in idea diffusion as the complexity of the idea grows.

The *Tipping Point* model (Gladwell, 2000) further synthesizes these findings into a series of general principles that guide the flow of information in human systems. The concept of “critical mass” in innovation theory is restated in terms of *tipping points*. As these points are reached, the level of communication of a particular idea within a social system suddenly jumps dramatically. Gladwell organizes his findings into four central themes:

1. *The Law of the Few* (Gladwell, 2000, p. 30): Three types of individuals play a particularly critical role in the diffusion of information within social systems. *Connectors* maintain active communications links with an unusually large number of individuals within and outside of the immediate social network. For example, given a random set of last names from a phonebook, a connector might be able to identify personal connections with 3-10 times as many names as the average individual. *Mavens* act as sinks for information, gathering information from many sources and willingly sharing it with others in the community. *Salesmen*, whom we will refer to as *Persuaders*, are unusually good at convincing other individuals to adopt a particular product or idea.

2. *The Stickiness Factor* (Gladwell, 2000, p. 30): As mentioned previously during the discussion of criticality, certain characteristics of a communication (e.g., simplicity, unexpectedness, concreteness, credibility, emotional impact, story setting; Heath & Heath, 2007) make it particularly likely to be retained by a client.

3. *The Power of Context, Part I* (Gladwell, 2000, p. 133): Small aspects of the decision-making setting can exert a huge influence on overall decision-making.

4. *The Power of Context, Part II* (Gladwell, 2000, p. 169): The effective size of a social community is limited to roughly 150 participants. Beyond this point, there is insufficient cohesion for consistent messages to be shared among all members.

From a research standpoint, what makes findings 1, 2, and 4 particularly attractive is that they are sufficiently specific so that they can form the basis of a computer simulation that can test whether or not “tipping points” actually exist. What finding 3 implies is that such simulations are only likely to be useful in identifying general patterns of behavior; the sensitive dependence upon initial conditions associated with informing contexts means that it will be nearly impossible to predict accurately all the behaviors that will be exhibited in specific situation based purely upon a small set of general situational characteristics.

A Tipping Point Simulation

To construct an example of a tipping point simulation, we combine the various assumptions presented by Gladwell (2000) into a single model. This can be done as follows (see Appendix C for details on the specific implementation):

1. *We assume a community size of 150*. Within each community, we assume that knowledge of who has already been informed is widely known. This is also consistent with the diffusion generalization that the rate of awareness of an innovation diffuses much more rapidly than actual adoptions (Rogers, 2003, p. 214).

2. *We assume a threshold model, similar to the previous information cascade simulation*. For simplicity, we use a uniform distribution for thresholds, similar to that of other proposed threshold examples (e.g., Granovetter, 1978, as quoted in Rogers, 2003, p. 356). Adoption is driven by prior adoptions within the community, a feature shared with information cascade models. Since the tipping point model does not require a choice between alternatives, however, the threshold value can be viewed as combining both stickiness and taste factors, the idea being that as more individuals in the community adopt a particular idea, it becomes easier to be informed about it and it also becomes more socially attractive.

3. *We assume that a certain percentage of individuals within each community fall into the categories of maven, persuader, and connector*. Their influence is modeled as follows:
    a. *Mavens* have a much lower adoption threshold than the typical member of the community. This assumption makes particular sense when what is being diffused is information, since what characterizes a maven is an insatiable thirst for information.
    b. *Persuaders* exert a particularly large influence on the perception of prior adoptions within a single community. For example, when a typical individual within a community adopts the visibility of that adoption is one. When a persuader adopts, on the other hand, the visibility might be 5 or even 10. Persuader thresholds are distributed similarly to other members of the community, reflecting the generalization that opinion leaders are not necessarily more innovative that the average individual, particularly in communities that do not favor change (Rogers, 2003, p. 318).
    c. *Connectors* are assumed to exert visibility across communities, although their influence within a single community is no greater than that of any other individual. For example, if there were 10 150-person communities in the model, a given connector might be visible in 5 of them (or even all 10 of them). Thus, connectors become the mechanism through which innovations diffuse across communities within a broader region.

We also assume that an external agency (e.g., a change agent; Rogers, 2003, p. 365) may, during any period, exert influence leading to the adoption of the innovation by some members of the community. For the purpose of the present model, we normally exert that influence in the initial period, to initiate the client-to-client diffusion process.

For the purposes of performing analysis using the model, the following parameters can all be adjusted:

- *Maven Percent*: The probability that an individual member of the community is a maven.
- *Persuader Percent*: The probability that an individual member of the community is a persuader.
- *Connector Percent*: The probability that an individual member of the community is a connector.
- *Maven Factor*: The factor used to reduce the threshold for any individual designated to be a maven.
- *Persuader Factor*: The factor applied to a persuader adoption.

- *Connector Factor*: The number of communities, on average, in which a given connector is visible.
- *Threshold Adjust*: A multiplicative factor applied to the threshold value of each individual that can be used to make adoption thresholds lower (< 1) or higher (> 1) than the base case.

The results of exploratory simulation runs confirm a number of distinct behaviors relating to diffusion models. These are now discussed.

Sensitive dependence on initial conditions

The simulation of diffusion exhibited considerable volatility with respect to outcomes for many combinations of parameters. Figure 6, for example, shows the results of 12 separate simulation runs, each plotting the average total outcomes across 10 communities (regions) of 150 individuals for the same set of parameters. Although a cluster of three adoptions at close to 100% were observed, other runs showed values as low as 7%. This would seem to confirm the *Power of Context, Part I* (Gladwell, 2000) aspect of the tipping point model.

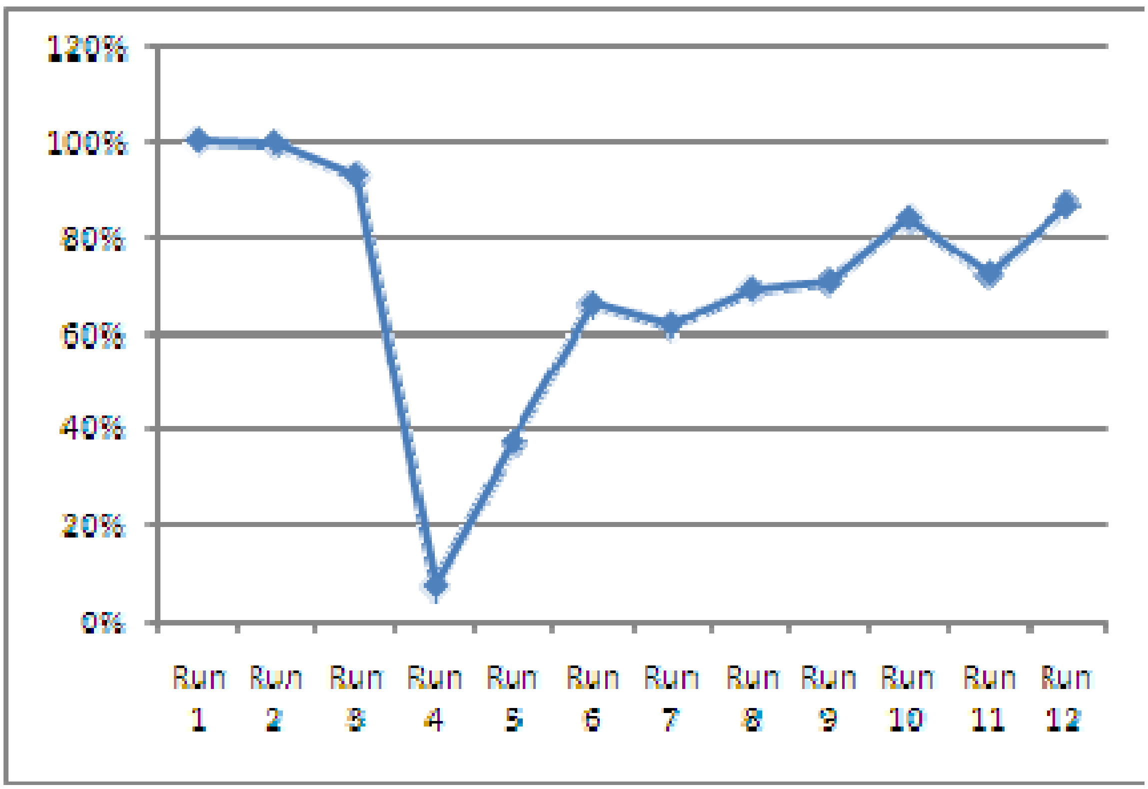


Figure 6: Final diffusion percentage for 12 runs. This illustrates that the diffusion rates for a given set of parameters can vary quite substantially, suggesting sensitive dependence on initial conditions.

Heterogeneity in the community enhances diffusion

An interesting question that comes out of the tipping point model is whether or not the existence of mavens, persuaders, and connectors has greater impact on diffusion than simply lowering the thresholds. The simulation allows us to address this question by assigning an “equivalent threshold” to any combination of mavens, persuaders and connectors. The logic is as follows:

- If everyone were a maven, then that would be precisely the same as dividing the threshold value by the maven weight—since that is how mavens thresholds are calculated—and then treating it as zero mavens. Similarly, if everyone were a persuader, that would be the same as dividing the threshold by the persuader weight. The same applies to connectors.

- Dividing the threshold weight by a given factor is equivalent to multiplying the weight of each observation by a given factor. This can be used to determine an "equivalent population". For example, if everyone were a persuader and the persuader weight was 5, then 150 persuaders would represent the equivalent population of 750 normal individuals.

Using equivalent populations, we can come up with an "effective population" value of:

Original Population *(Percentage not mavens, persuaders or connectors) +

Original Population *(Maven % * Maven Weight + Persuader % * Persuader Weight + Connector % * Connector Weight)

For a base case of 150 people with 5% values for mavens, persuaders, and connectors, each with associated factors of 5, the equivalent population is:

240 = 150 * (0.85+0.05*5 + 0.05*5 +0.05*5)

Using the adjusted population, we can then compute an adjusted threshold using the formula:

Threshold Value * Original Population / Adjusted Population

For a threshold value of 1.4 (used in the base case runs), the adjusted threshold becomes:

0.875 = 1.4 * 150 / 240

Thus, if having heterogeneity in the simulation does not matter, we would expect the 5% maven, persuader, and connector simulation—each with an associated weight of 5— with a 1.4 threshold factor to produce a distribution of outcomes similar to that of a simulation with no heterogeneous elements (leaving only variation in threshold values) and a 0.875 threshold factor.

Table 1: Comparison of Heterogeneous Base Case and Homogeneous Runs. This illustrates that a heterogeneous mixture of mavens, persuaders and connectors produces a higher level of diffusion than a homogeneous system with the same threshold (Run 3).

	Base Case	Homogeneous Runs 1	Homogeneous Runs 2	Homogeneous Runs 3	Homogeneous Runs 4
Threshold Value	1.4	1.2	1.0	0.875	0.8
Equivalent Threshold	0.875	1.2	1.0	0.875	0.8
Maven Percent	5%	0%	0%	0%	0%
Maven Weight	5	0	0	0	0
Persuader Percent	5%	0%	0%	0%	0%
Persuader Weight	5	0	0	0	0
Connector Percent	5%	0%	0%	0%	0%
Connector Weight	5	0	0	0	0
10 Run Average Diffusion	***0.67***	***0.08***	***0.28***	***0.50***	***0.66***

The Table 1 results show that heterogeneous base case produces a greater level of diffusion (across the sample of 10 runs) than homogeneous case with a comparable equivalent threshold (the Homogeneous Runs 3 column). Indeed, it is not until we get to 0.8 (the equivalent of 6.25% presence for each of the heterogeneous clients) that we reach the same final diffusion. Thus, heterogeneity in the client space appears to aid diffusion beyond the pure impact of the additional observation weight.

Existence of sharp tipping points for a given network

The implied message of the term “tipping point” is that there is a certain threshold that—once crossed—leads to major changes in diffusion. From the perspective of the simulation, there are two ways of looking at this issue. The first is to consider whether or not a given fixed network (i.e., fixed set of threshold values) produces significant changes in diffusion as a result of small changes to parameters. The second is to look for such sharp changes in the averages across multiple runs with different random numbers. We consider the fixed network first.

In virtually every network examined, distinct sharp variations in outcomes resulting from tiny parameter changes could be identified. An example of such a change is presented in Figure 7. In this illustration, one set of values represents penetration in each region for a run with the connector percentage set to 0.054. Only one region (Region 2) has reached full diffusion (value of 150). As the connection percentage is changed to 0.055, however, all of the remaining 9 regions reach full diffusion, shown as the series going across the top.

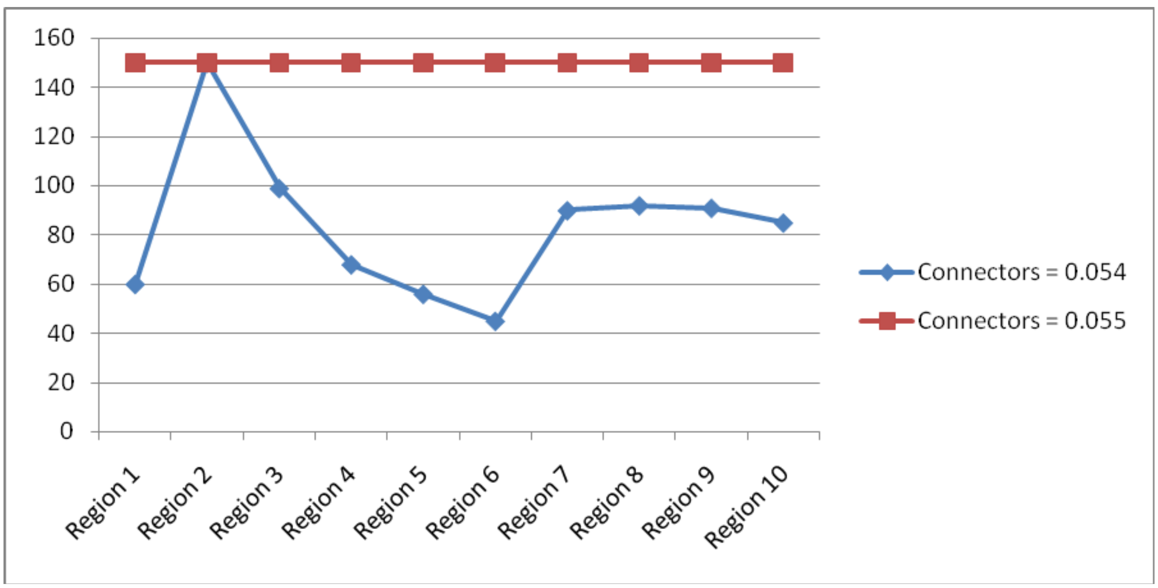


Figure 7: Tipping point for fixed connections on a single 10 region run. This demonstrates how a tiny percentage variation in the number of connectors can dramatically impact adoption levels across regions. Both runs used the same set of random numbers. In this example, the higher percentage resulted in a single connector being added.

Similar inflection points can be observed for nearly every parameter. This behavior should not be particularly surprising. As percentage or weight parameters are adjusted upwards, individual clients accept the innovation earlier. This process is not continuous, however. Instead, it occurs discretely (e.g., in the Figure 7 demonstration, a connection percentage of 0.053 produces a distribution identical to the first series—produced by 0.054—because no elements changed). That means that as one passes over a point that produces one more adoption, the effect cascades through the system. Furthermore, should that element be a persuader or connector, that effect is amplified by whatever weight factor is used. (Mavens are slightly different, as their effect tends to be most pronounced at the early stages of diffusion owing to their low threshold.)

Existence of sharp tipping points for an "average" network

The second type of search for tipping point involves finding inflection points where large changes in diffusion values are experienced, on average, with relatively small changes in parameters. Because of the randomness involved in different runs of the simulation, we would not expect the same sharpness that occurs with varying parameters for a fixed set of threshold characteristics. Nonetheless, we would expect to see fairly substantial swings if small changes to an informing system do produce tipping points.

In Figure 8, 4 different sets of 10 simulation runs are plotted (each ordered by percent diffusion achieved, so make comparisons easier). The base case used a 1.4 threshold factor, 0.05 for maven, persuader, and connector percentages, and 5 for each of the corresponding weights. The three remaining cases each adjusted one of the three percentage values to 0.06 (as shown in the legend). What these results illustrate is how large the impact of these small changes can be (from 0.05 to 0.06 will produce, on average, 1.5 additional individuals of that type, in a given run). This difference is reflected in the averages for the 10 runs, which were 69% for the base case, 82% with mavens at 0.06, 91% with persuaders at 0.06, and 89% with connectors at 0.06 percent.

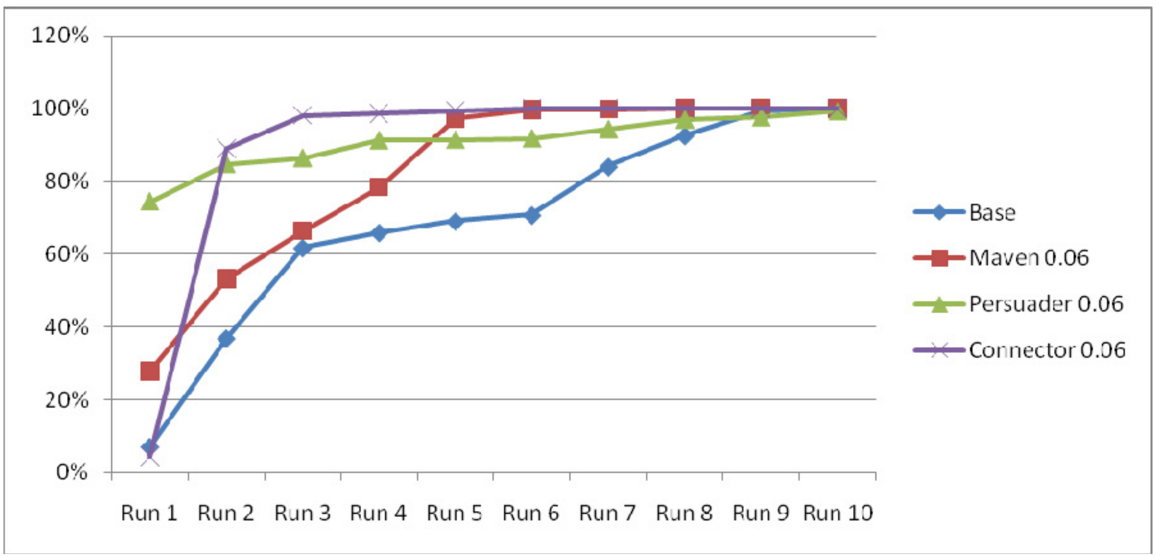


Figure 8: Results of runs with small parameter changes. The 10 runs were ordered based on the level of penetration achieved in the base case. The impact of changing each special type (e.g., maven, persuader, connector) from its base case value of 0.05 to 0.06 is then shown, illustrating how sensitive penetration levels can be to the presence of one or two extra individuals within the community.

Another aspect of Figure 8 worth remarking upon is the degree to which it suggests that the notion of a "critical mass", often mentioned in the diffusion literature (e.g., Rogers, 2003, p. 343), is not purely an analogy. In each of the non-base cases, very high levels of diffusion were achieved frequently. Where the parameters are raised just a bit more, virtually all runs produce 100% diffusion. Thus, the tipping point simulation exhibits that same take-off behavior associated with criticality that is observed for other diffusion models.

Tipping Point Conclusions

The tipping point model seems the best fit for those informing situations where a fairly complex body of knowledge needs to be transferred among clients and where both client-recipients and client-senders need to be involved in the informing process as it proceeds. Because it is a threshold model, its behavior can easily resemble that of the information cascade model in those cases where the tipping point's distinctive elements—its mavens, persuaders, and connectors—are set to very low values. The results of the simulation suggest, however, that ignoring the heterogeneity that has been observed in client communities could lead to serious misjudgments regarding the likelihood that client-to-client informing is going to diffuse across the community.

The high degree of sensitivity to heterogeneous individuals within the community exhibited by the model also suggests that the efficiency of informing may be improved by specifically identifying these unique individuals (mavens, persuaders, and connectors) and paying special attention to informing them. This is consistent with Gladwell's notion of a "maven trap" (e.g., a toll free number printed on an Ivory soap wrapper; Gladwell, 2000, p. 276) as a tool for jump-starting diffusion. Similarly, identifying and informing persuaders and connectors as diffusion proceeds could pay substantial dividends both in terms of the speed at which an idea travels and in its ultimate level of penetration.

Although the tipping point model does not capture network topology, it is qualitatively closer to the prevailing network models than either the criticality or cascades model. In the small world model (e.g, Watts, 2003), for example, closely linked sub-communities are linked by infrequent links between communities, as shown in Figure 9. This is similar to the topology presented in the tipping point model. These might be viewed in terms of connectors. The small world model, however, does not appear to have analogs to mavens or persuaders.

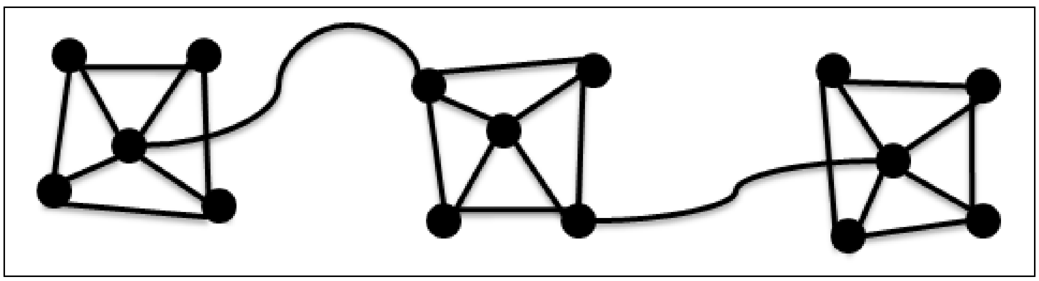


Figure 9: Small world model (Watt, 2003), showing tightly connected clusters tied together by a small number of cross cluster links. Such networks can arise, for example, when individuals link to affiliated (e.g., closely linked) networks but may have more than one affiliation. For example, a faculty member could be linked to other faculty in his or her college (geographic affiliation) while also being tightly linked to the members of his or her discipline (which would not be geographically constrained in the same way).

Similarly, the scale free network model (e.g., Barabasi, 2002) proposes that the connection densities of nodes will be governed by a power law, rather than by a more typical normal distribution. Such a network naturally evolves under circumstances where new nodes gravitate towards connections with existing nodes that are highly connected ("the rich get richer", Barabasi, 2002, p. 79). As networks grow large, this will naturally lead to the emergence of hub nodes that play a particular influential role in enabling communications across the system, as shown in Figure 10. The hub node, therefore, plays a similar role to the connector nodes of tipping point model.

Barabasi's (2002) scale free model also introduces the notion that node fitness, in addition to the number of connections, might influence the likelihood of further connections. This would suggest, as a possibility, that mavens might tend to acquire more connections than the typical individual—a proposition that could be tested as a hypothesis.

While the correspondence between these network models and the proposed tipping point model is far from exact, we would expect that some qualitative similarities in behavior would be observed—particularly since the tipping point model was largely empirical in its origins, based upon observed behaviors. Ultimately, the mathematical study of networks should help us to distinguish the emergent properties common to nearly all networks from those properties that are relatively unique to human multi-client systems.

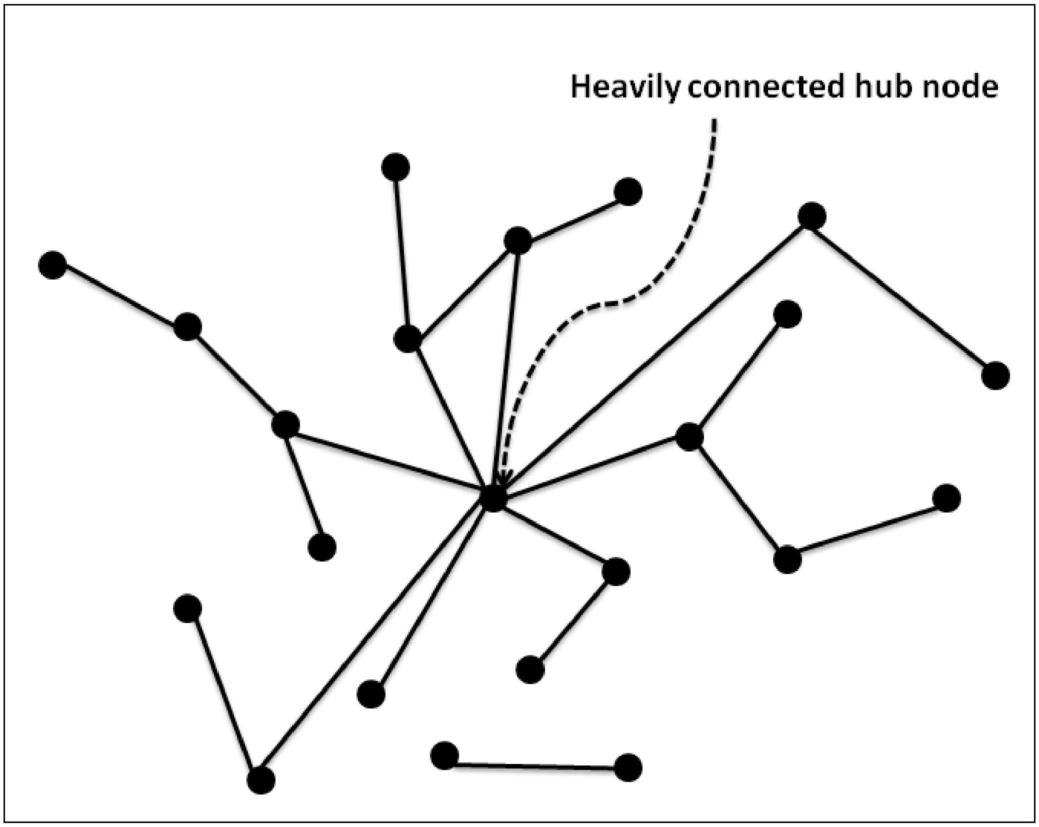


Figure 10: Representation of scale-free network with emergent heavily connected hub node. In a scale free model, the logarithm of the likelihood of any given number of connections for a particular node declines proportionately to the logarithm of the number of connections.

Conclusions and Directions for Future Research

This paper has been written with three distinct objectives in mind:

1. *To highlight the importance of client-to-client communication processes within an informing system.* Where an informing system consists of an information technology involved in the performance of a relatively routine task, such processes can probably be safely ignored. Where the informing is of a non-routine nature, the message is complex, and the channels are not well established or are flexible in nature, however, such client-to-client informing is likely to dominate the overall informing process. If we do not consider these types of complex informing within the informing sciences, then we are ignoring one of the most important areas to which we may be able to contribute.

2. *To introduce a partial taxonomy of models of client-to-client informing*. The sender-driven (criticality), client-driven (information cascade), and socially-constructed (e.g., tipping point) models presented here offer alternative ways of looking at the client-to-client informing process. In presenting them, the goal has been to identify situations where each might be appropriate; there is no basis for asserting that one is inherently better than another across all situations.

3. *To demonstrate the use of simulation in modeling client-to-client informing processes*. As a general rule, economists tend to prefer models that can be mathematically formalized.

> The price to be paid for such formalization is the need to assume away the aspects of each model that lead to intractability. By using simulations, particularly for the information cascade and tipping point models, it becomes possible to supplement whatever intuitions we might have based on reasoning about these forms of informing with objective results. Naturally, using simulations comes with its own price tag: the presence of many parameters for which values and underlying statistical distributions are unknown. We should not, however, allow the presence of such unknown parameters to deter us from using simulation for two reasons. First, simulations may still help us understand characteristic behaviors even where they are not terribly useful for making specific predictions; it is for this purpose that they have been used here. Second, the fact that we have no clear idea regarding what value a parameter might have (e.g., the percentage of connectors within a typical population) does not make it unimportant. Sometimes, trying to put together a reasonable simulation helps us better understand what we need to find out if our knowledge is to advance.

Beyond these general goals, a certain number of specific observations may be made about the models presented. First, all exhibit the typical s-curve of diffusion processes, signifying gradual early adoption, followed by rapid diffusion, followed by a tailing off as maximum penetration is reached. Second, all exhibit sensitivity to certain key parameters that can dramatically impact the ultimate level of penetration that may be reached. Third, all have a sufficiently large number of parameters so that applying the models to real world situations is likely to be difficult; qualitative insights into typical system behaviors are the most likely benefit of the models for the foreseeable future. Finally, all the models have specific domains of applicability. The criticality model, for example, incorporates assumptions that are most appropriate for small, tightly connected communities. The information cascades model is applicable to similar communities, where each individual's decisions are visible to the remaining individuals, but also to settings where a score sheet of prior decisions (e.g., polls, sales data) is maintained. Perhaps most importantly, the applicability

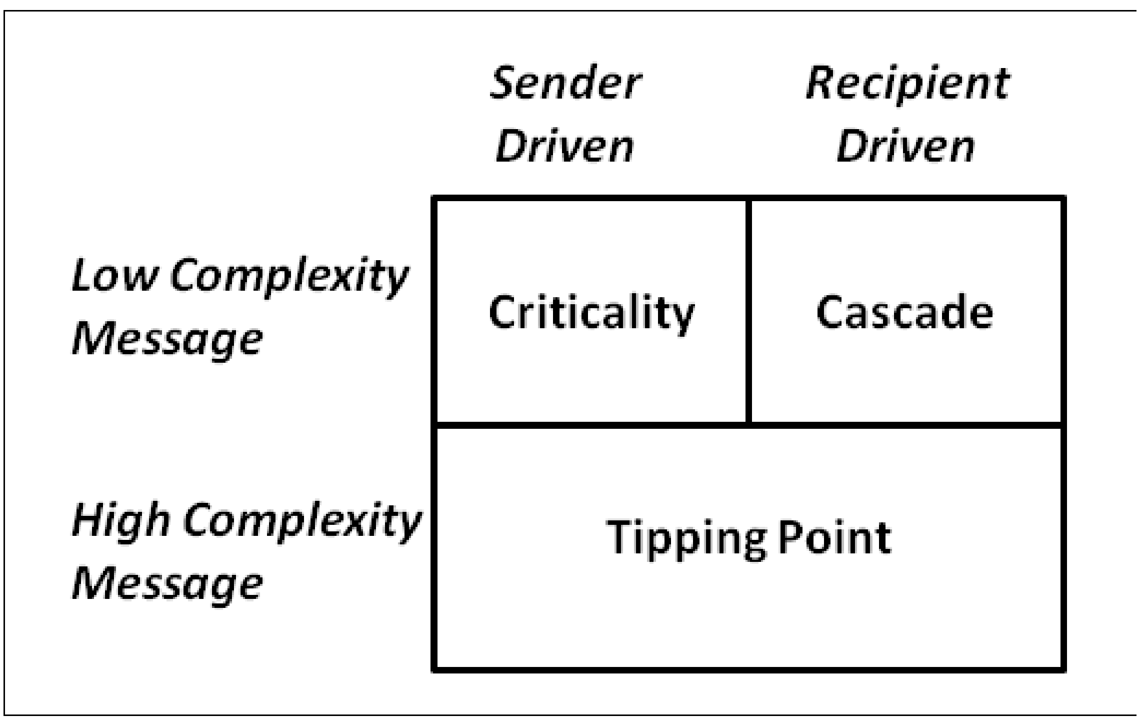


Figure 11: Mapping model domains to who drives informing and message complexity

of the models varies across two dimensions: 1) whether it is driven by client-senders (individuals wishing to spread the information within the client community) or client-receivers (individuals who want to acquire the information within the client community) and, 2) the level of complexity of the information being conveyed. The most appropriate domains are summarized in Figure 11.

There are many areas for future research that are suggested by the current paper. One such area involves the imposition of formal patterns of communications on top of the basic model. Informing, for the most part, is likely to take place in contexts where the random patterns of informing assumed in the models presented here are not terribly good representations of reality. Extraordinary progress is currently being made in the development of mathematically based models (e.g., Barabasi, 2002; Watts, 2003) that help us better understand how particular patterns of communication evolve. They have, however, such a large number of parameters and assumptions that they are not yet particularly useful in predicting what structure will evolve in a particular real world setting.

On the other hand, in real world settings we often have existing structures to work with. Organizations have hierarchies and social networks leading to patterns of communications that are far from random, and sometimes even documented (at least in general terms). Even within informal informing settings, such as the classroom, groups tend to form that can have a major impact on the learning processes of their members. The problem with incorporating these patterns into a general discussion, such as the one conducted in the present paper, is that there are so many possible patterns it is hard to know what meaningful generalizations can be made about their impact. That does not make them unimportant. Rather, it suggests that network connection dynamics will need to be superimposed on top of the basic models of client-to-client informing on a case by case basis. How to accomplish such merging of the networking and client-to-client informing models is an important topic that deserves to be investigated further.

How technology impacts client-to-client informing has been given little treatment in the present paper, with the minor exception of proposing a criticality model particularly suited for understanding email blizzards. For the informing sciences, with its strong focus on the role of technology in the informing process, this presents a great opportunity. Questions such as "Have electronic communications changed the nature of word of mouth in client-to-client informing?" or "How does media richness of available channels in diffusion processes?" are well worth further investigation.

A particularly interesting area for future research involves the intersection of technology and network models. In this area, a substantial amount of data has already been gathered on the structure of internet communities—it is central to the validation of both the small world (Watts, 2003) and scale free (Barabasi, 2002) network models. What appears to be less explored is the qualitative nature of changes to the diffusion of research that has been enabled by the availability of global connectivity—presuming that such changes have actually occurred. Longitudinal studies tracking the flow of specific ideas pre- and post-web would provide a useful empirical complement to the mathematical formulations that dominate the rapidly growing field of network theory.

To restate the central point of this paper: client-to-client informing processes are very important in many informing systems. The study of these processes—in the form of research into the diffusion of innovations—has been one of the most successful research themes in the social sciences. Unlike many areas, academic researchers have played an important role in developing practical findings; these findings are frequently well-supported, have useful implications, and, most surprisingly, yield insights not immediately obvious to those involved in practice. Unfortunately, client-to-client informing processes are also an area where academic researchers often fail to apply what they know. In the activities that we engage in within our own disciplinary informing system, we are very much aware of their power. Thus, we value communications with our colleagues through papers, conferences, and informal discussions in the hallway. When it comes to external clients—such as practitioners and researchers in other disciplines—however, we are inclined to assume that our findings should somehow magically diffuse into these communities based upon the short discussion of "relevance to practice" that we dutifully tack on to the end of our papers. What this paper has argued is that tossing a paper over the giant walls that separate us from practice and from other disciplines is unlikely to have any impact—unless, of course, the content of that paper is a simple, sticky, and juicy rumor. If we are interested in complex informing processes, we must learn to better understand and manage the client-to-client informing processes that are absolutely essential to diffusion. This is, however, a research area that is well-suited to the transdisciplinary character of the informing sciences, our familiarity with technology, and our profound interest in the process of informing.

References

Anderson, L. R. (2001). Payoff effects in information cascade experiments. *Economic Inquiry, 39*(4), 609-615.

Anderson, L. R., & Holt, C. A. (1997). Information cascades in the laboratory. *American Economic Review, 87*(5), 609-615.

Arthur, W. B. (1988). Self-reinforcing mechanisms in economics. In P. W. Anderson, K. J. Arrow, & D. Pines, *The economy as an evolving complex system*. Reading, MA: Addison-Wesley.

Arya, A., Glover, J., & Mittendorf, B. (2006). Hierarchical reporting, aggregation, and information cascades. *Managerial and Decision Economics, 27,* 355–362.

Bak, P. (1996). *How nature works: The science of self-organized criticality*. New York, NY: Springer-Verlag.

Banerjee, A. V. (1992). A simple model of herd behavior. *Quarterly Journal of Economics, 107*(3), 797-817.

Barabasi, A. L. (2002). *Linked*. New York, NY: Plume.

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

Bowden, M., & McDonald, S. (2008). The impact of interaction and social learning on aggregate expectations. *Computational Economics, 31*, 289–306.

Daley, D. J., & Kendall, D. G. (1965). Stochastic rumours. *Journal of the Institute of Mathematics and its Applications, 1*, 42-55.

De Vany, A., & Lee, C. (2001). Quality signals in information cascades and the dynamics of the distribution of motion picture box office revenues. *Journal of Economic Dynamics & Control, 25*, 593-614.

Dietz, K. (1967). Epidemics and rumours: A survey. *Journal of the Royal Statistical Society, Series A (General), 130*(4), 505-528.

DOE. (1993). *DOE fundamentals handbook: Nuclear physics and reactor theory* (Vol. 2 Of 2). U.S. Department of Energy, Publication DOE-HDBK-1019/2-93.

Dunstan, R. (1982). The rumour process. *Journal of Applied Probability, 19*(4), 759-766.

Gill, T. G. (2008). The single client resonance model: Beyond rigor and relevance. *Informing Science: the International Journal of an Emerging Transdiscipline, 11,* 281-310. Retrieved from http://inform.nu/Articles/Vol11/ISJv11p281-310Gill222.pdf

Gill, T. G., & Bhatacherjee, A. (2007). The informing sciences at a crossroads: The role of the client. *Informing Science: the International Journal of an Emerging Transdiscipline, 10*, 17-39. Retrieved from http://inform.nu/Articles/Vol10/ISJv10p017-039Gill317.pdf

Gladwell, M. (2000). *The tipping point*. New York: Back Bay Books.

Goeree, J. K., Palfrey, T. R., Rogers, B. W., & McKelvey, R. D. (2007). Self-correcting information cascades. *Review of Economic Studies, 74*, 733–762.

Granovetter, M. S. (1978). Threshold models of collective behavior. *American Journal of Sociology, 83*, 1420-1443.

Grebe, T., Schmid, J., & Stiehler, A. (2008). Do individuals recognize cascade behavior of others? – An experimental study. *Journal of Economic Psychology, 29*, 197–209

Heath, C., & Heath, D. (2007). *Made to stick.* New York, NY: Random House.

Kubler, D., & Weizsacker, G. (2003). Information cascades in the labor market. *Journal of Economics, 80*(3), 211–229.

Kubler, D., & Weizsacker, G. (2004). Limited depth of reasoning and failure of cascade formation in the laboratory. *Review of Economic Studies, 71*(247), 425-441.

Lefevre, C., & Picard, P. (1994). Distribution of the final extent of a rumour process. *Journal of Applied Probability, 31*, 244-249.

Mahajan, V., Muller, E., & Bass, F. M. (1991). New product diffusion models in marketing: A review and directions for research. In N. Nakicenovic & A. Grubler (Eds), *Diffusion of technologies and social behavior* (pp. 125-177). New York, NY: Springer-Verlag.

Maki, D. P., & Thompson, M. (1973). *Mathematical models and applications*. Englewood Cliffs, NJ: Prentice-Hall.

Nekoveea, M., Morenob, Y., Bianconic, G., & Marsilic, M. (2007). Theory of rumour spreading in complex social networks. *Physica A, 374*, 457–470.

Pittel, B. (1990). On a Daley-Kendall model of random rumours. *Journal of Applied Probability, 27*, 14-27.

Porter, M. E. (1980). *Competitive strategy*. New York, NY: The Free Press.

Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). New York: Free Press.

Shiller, R. J. (1995). Conversation, information, and herd behavior. *American Economic Review, 85*(2), 181-185.

Watts, D. J. (2003). *Six degrees: The science of a connected age*. New York, NY: Norton.

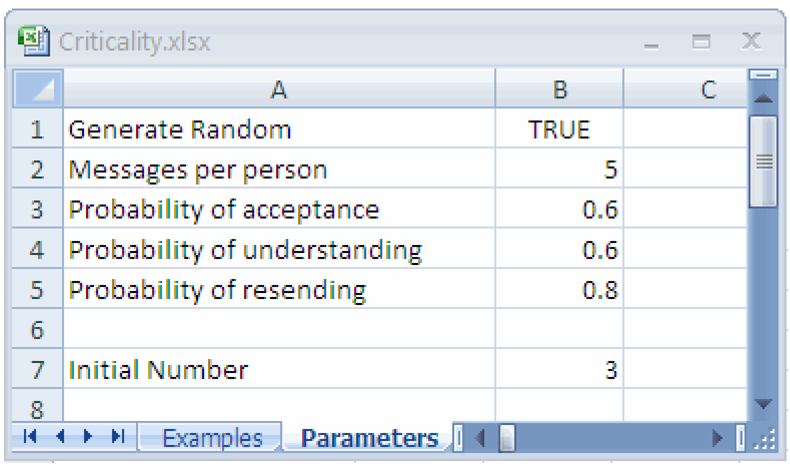
Appendix A:
Construction of the Spreadsheet Criticality Model

The purpose of these appendices is to provide the interested reader with the details of the models discussed in the paper. In addition, specific formulas are presented so as to make it possible for the *unusually* interested reader to construct their own personal versions of the simulations. To serve both these purposes, an attempt has been made to make the models as simple as possible.

The criticality model is the simplest of the three models. It is based upon the one shot informing model, which assumes that after client is informed the message is retransmitted, on average, a specified number of times in the following period. (This differs from the economic rumor models, which assume the client keeps pairing up with another client every subsequent period after being informed; a process that stops only when a client who is already aware of the rumor is encountered).

Parameters

There are two types of parameters to the model, those which are model specific and those which are common to all three simulation models (which will only be described here). The full set of criticality parameters is presented in Exhibit 1.



	A	B	C
1	Generate Random	TRUE	
2	Messages per person	5	
3	Probability of acceptance	0.6	
4	Probability of understanding	0.6	
5	Probability of resending	0.8	
6			
7	Initial Number	3	
8			

Exhibit 1: Criticality Model Parameters

The two common parameters are *Generate Random* (TRUE/FALSE) and *Initial Number*. The Generate_Random parameter allows us to use a consistent set of random numbers in testing different parameter values or to generate a new set for each run. It is implemented with three separate worksheets, all of which have the same number of rows and columns populated:

1. **RandomFormulas**: This worksheet consists of an array of cells each containing =RAND()
2. **RandomNumbers**: This worksheet consists of an array of randomly generated values. (Normally, it is created by doing a copy values from the original RandomFormulas worksheet).
3. **RandomValues**: This worksheet contains the following formula in each cell (with the cell address—A1—varying appropriately):

=IF(Generate_Random=TRUE,RandomFormulas!A1,RandomNumbers!A1)

The Initial_Number parameter is used to seed the process (the same role played by a neutron gun in supplying the initial neutrons to a reactor). It allows values (0=not informed, 1=informed) to be probabilistically determined. Specifically, the first period of the results spreadsheet will have the following formula in each cell:

```
=IF(RandomValues!A1<Initial_Number/150,1,0)
```

The 150 represents the total number of clients in the simulation and, since **RandomValues!A1** is uniformly distributed between 0 and 1, this gives us a probability distribution that will—on average—produce Initial_Number values of 1 for the 150 cells. Naturally, there will be some binomially distributed variation, which is why a value such as 3 is generally used, increasing the likelihood of at least some values in the initial period.

The specific parameters of the model are discussed in the body of the paper. The input values correspond to these, except that *Probability of Acceptance* is $(1 - P_{IGNORED})$.

Model Construction

The computations in the criticality model involve two worksheets. The **State** worksheet holds values signifying if a client is informed (1) or not informed (2) for each period. The **NewValues** identifies value changes from the previous period a formula. For example, the formula in cell **NewValues!B1** would be:

```
=IF(AND(State!A1=0,State!B1=1),1,0)
```

All new values for each period are summed in row 151 of the **NewValues** worksheet.

The **State** worksheet takes the messages sent from all the **NewValues** in the previous periods and tests if they inform a particular client. This is done probabilistically with a formula (shown for cell **State!B1**):

```
=IF(A1=1,1,IF(POWER((1-Probability_of_acceptance*Probability_of_understanding),
            NewValues!A$151*Messages_per_person*
                  Probability_of_resending/150) >
            RandomValues!B1,0,1))
```

The arguments of the POWER() function are described in the body of the paper. It takes the probability that a given message will not inform the client and raises it to the power of the average number of messages that particular client can expect to receive. That average number is, in turn, computed by taking the number of new values from the previous period (found in **NewValues** row 151, as previously noted) and multiplying it by messages per person and the likelihood of resending. This determines the likelihood that the cell will remain uninformed, which is compared with a random number that determines whether the cell is 0 (stays uniformed) or a 1 (newly informed).

The IF() function surrounding the POWER() function ensures that cells already turned on remain on.

50 periods are computed in the original spreadsheet, although it turns out that we virtually never need more than 20.

Appendix B: Construction of the Spreadsheet Information Cascade Model

The information cascade model turns out to be the most complex of the three simulations owing to the need to choose between options. It also includes a number of tuning parameters whose relationship to real world values is somewhat indeterminate.

Parameters

The parameter worksheet is presented in Exhibit 2. These are discussed in the body of the text.

Cascades.xlsx

	A	B
1	Optimal Occupancy Percent:	50%
2	Percent Deviation Cost:	0
3	Private Observation Error:	0%
4	Taste Difference Factor:	2
5	Behavior Constant	0.01
6	Observation Weight:	0.6
7	Threshold Factor:	1
8	Urgency Factor:	0.1
9	Allow Switch:	FALSE
10	Initial State Count	5
11		
12	Use Formulas	TRUE

Parameters | Exam

Exhibit 2: Cascade Parameter Worksheet

The *Use Formulas* and *Initial State Count* parameters correspond to the general parameters mentioned in Appendix A.

Model Construction

Much of the complexity of the cascade model stems from the fact that two decision options are possible, not including the third (undecided) option. Thus, our 0 and 1 states now become 0 (undecided), 1 (option 1) and 2 (option 2).

There are two sources of randomness in the simulation. The first is randomness related to the individual client and that does not change from period to period. This includes two items:

1. **RandomThreshold**-*Column A*: A *threshold value* that determines how large the difference in preference between option 1 and option 2 needs to be before a decision is made.
2. **RandomThreshold**-*Column B*: A *taste preference* for option 1 or 2, determined by subtracting 0.5 from a random number, with negative results favoring option 1, positive results favoring option 2.

The remaining random numbers, included in **RandomObs**, provide period-to-period random changes in the perceived attractiveness of the options, from the individual's perspective. Both sources of randomness have underlying number and formula worksheets that are controlled by the same *Use Formulas* parameter, precisely as described in the parameter section of the criticality model (Appendix A).

The **Behavior** worksheet controls the simulation. It is organized as follows:

- *Row 1*: Contains the period number. This is used for the purpose of reducing the threshold where urgency is set.
- *Rows 2-151*: Contains the values for the states of each individual.
- *Row 153*: Contains the current number of option 1 selections. This is computed using the formula: =COUNTIF(A$2:A$151,=1)
- *Row 154*: Contains the current number of option 2 selections. This is computed using the formula: =COUNTIF(A$2:A$151,=2)
- *Row 156*: Contains the adjusted value for option 1, based upon the *Optimal Occupancy Cost* and *Percent Deviation Cost* parameters. The precise formula is:

```
100-Percent_Deviation_Cost*ABS(A153/150-Optimal_Occupancy_Percent)
```

- *Row 157*: Contains the adjusted value for option 2, based upon the *Optimal Occupancy Cost* and *Percent Deviation Cost* parameters. The precise formula is:

```
100-Percent_Deviation_Cost*ABS(A154/150-Optimal_Occupancy_Percent)
```

The last two rows allow for externalities based upon occupancy to be factored into the computation.

The state rows (2-151) are generated as follows. The first column uses the *Initial State Count* parameter to set initial values probabilistically, similar to how it was done for the criticality model (Appendix A). It also uses the individual's taste value (**RandomThreshold**-Column B) to determine if the value is 1 or 2. The formula used is the following:

```
=IF(Initial_State_Count/150<RandomObs!A1,0,IF(RandomThreshold!B1>0.5,2,1))
```

The remaining columns use a complex formula that is constructed as follows (for cell B2):

```
=IF(
        OR(Allow_Switch=TRUE,Behavior!A2=0),
                IF(
                        Observation_Weight*
                                (A$156-A$157+(0.5-RandomThreshold!$B1)*
                                Taste_Difference_Factor +
                                (RandomObs!B1-0.5)*Private_Observation_Error)+
                        (1-Observation_Weight)*
                                Behavior_Constant*(Behavior!A$153-
                                        Behavior!A$154)>
                                MAX(Threshold_Factor*RandomThreshold!$A1-
                                        Urgency_Factor*Behavior!B$1,0),
                        1,
                        IF(
                                Observation_Weight*
                                        (A$157-A$156-(0.5-RandomThreshold!$B1)*
                                        Taste_Difference_Factor-
                                        (RandomObs!B1-
                                        0.5)*Private_Observation_Error)+
                                (1-Observation_Weight)*
                                        Behavior_Constant*(Behavior!A$154-
```

```
                    Behavior!A$153)>
          MAX(Threshold_Factor*RandomThreshold!$A1-
               Urgency_Factor*Behavior!B$1,0),
          2,
          Behavior!A2)
     ),
  Behavior!A2)
```

The outer IF simply determines if a zero state in the cell to the left (previous period) is present or if switching is allowed. If neither of these is true, then the value of the cell to the left is used.

The two inner IF() constructs are mirror images of each other: the first testing if the threshold for option1 has been reached, the second testing if the threshold for option 2 has been reached. The option scoring is in two parts. The *Observation Weight* multiplies the components of the privately observed value of the option. These consist of:

- Occupancy-induced fitness differences (the difference between rows 156 and 157 for the previous period)
- Taste-based differences (using column B of the **RandomThreshold** worksheet) times a *Taste Difference Factor* can be used to weight it equivalently to the other differences
- Observations errors for the period, pulled from the **RandomObs** worksheet, weighted by the *Private Observation Error* parameter.

The (1 – *Observation Weight*) factor takes the Behavior Constant (another parameter available to make values equivalent) and multiplies it times the difference between observed adoption counts (from rows 153 and 154).

The total of the *Observation Weight* and (1 – *Observation Weight*) terms is then compared with the *Threshold Factor* times the individual's threshold value (from column A of the **Random-Threshold** worksheet) from which the *Urgency Factor* times the period number (in row 1) is subtracted, thereby reducing the threshold value as time goes on. The MAX() function surrounding the term is used to prevent negative thresholds—which could happen if the Urgency Factor is large enough—that could cause option 1 to be selected even if it was less attractive than option 2. The nested IF() functions are set up so that if neither threshold is reached, the previous column value carries forward.

Appendix C: Construction of the Spreadsheet Tipping Point Model

The tipping point model is relatively straightforward to construct because so many of the parameters are well defined by the analysis in the book (i.e., Gladwell, 2000). Because the model suggests the need for inter- and well as intra-community informing, 10 regions of 150 clients each are assumed in the model.

Parameters

The parameter values for the model are presented in Exhibit 3. They are further discussed in the body of the paper.

TippingPointFinal.xlsx

	A	B
1	Generate Random:	TRUE
2	Maven Percent:	0.05
3	Persuader Percent:	0.05
4	Connector Percent:	0.05
5		
6	Maven Factor:	5
7	Persuader Factor:	5
8	Connector Factor:	5
9		
10	Threshold Adjust	1.4

RandomFormulas

Exhibit 3: Tipping Point Parameters

For the tipping point simulation, rather than providing a single parameter for initial informing (e.g., such as *Initial Number* in Exhibit 1), target informing levels are provided by region and additional signals can be sent in during any period—although testing was limited to the initial period. This required a parameter screen devoted entirely to setting informing levels, shown as Exhibit 4. In this illustration, each of the 10 regions is seeded with 3 informing signals in the first period.

TippingPointFinal.xlsx

	A	B	C	D	E	F	G	H	I
1	3	0	0	0	0	0	0	0	0
2	3	0	0	0	0	0	0	0	0
3	3	0	0	0	0	0	0	0	0
4	3	0	0	0	0	0	0	0	0
5	3	0	0	0	0	0	0	0	0
6	3	0	0	0	0	0	0	0	0
7	3	0	0	0	0	0	0	0	0
8	3	0	0	0	0	0	0	0	0
9	3	0	0	0	0	0	0	0	0
10	3	0	0	0	0	0	0	0	0

Parameters | FinalTallys | Seeds | Randomizer

Exhibit 4: Tipping Point Seed Parameters

The seeds are accommodated by the model by including by including a seeding formula during each period. For the first period, that term is:

=IF(Seeds!A$1/150>SeedRand!A1,1,0)

Where column E of the **SeedRand** table is specifically set up for seeding values in each period. The **Seeds!A$1** row changes every 150 rows. In other words, rows 1-150 use A$1, rows 151-300 use A$2, rows 301 to 450 use A$3, etc.

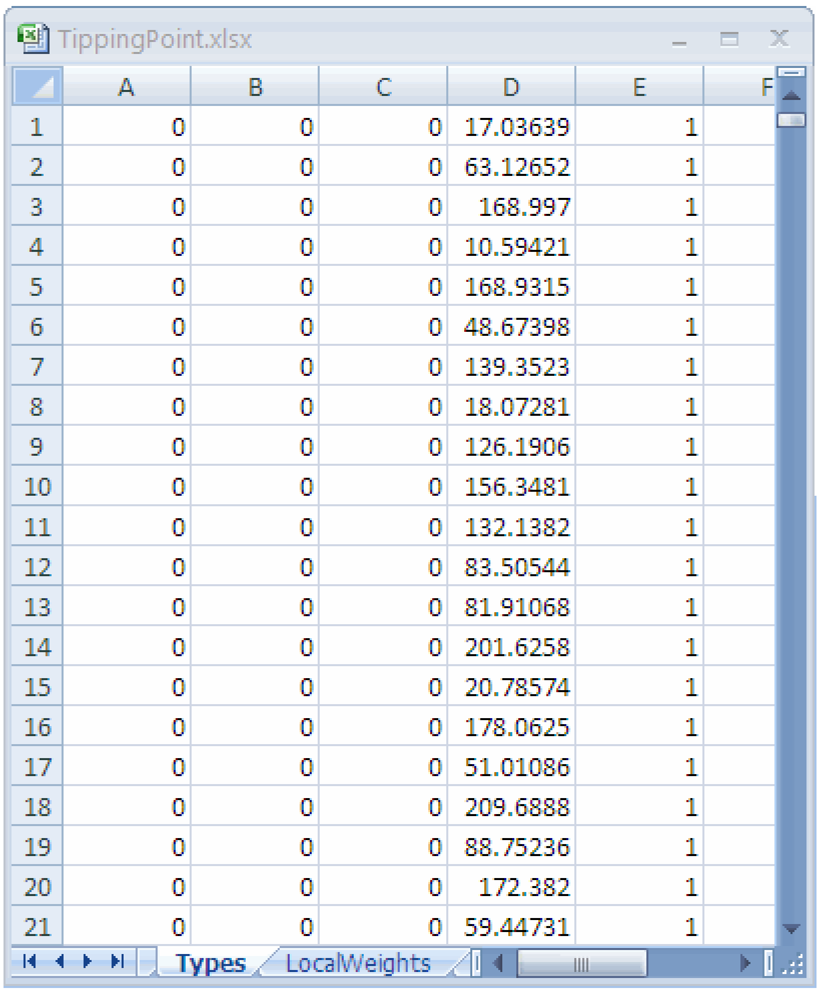
For subsequent periods, the seeding term is included as part of the overall state expression.

Model Construction

The model construction involves two separate processes, building the individual characteristics for each client and computing informing states. The individual characteristics, listed by column in the **Types** worksheet, consist of:

A. *Maven* (0 or 1)
B. *Persuader* (0 or 1)
C. *Connector* (0 or 1)
D. *Adjusted Threshold* (a random threshold value * the Threshold Factor * 150, identifying the number of individuals in a region who must adopt before the client adopts)
E. *Individual weight* (adjusted upwards if an individual is a persuader).

A section of the Types worksheet is presented in Exhibit 5.



	A	B	C	D	E	F
1	0	0	0	17.03639	1	
2	0	0	0	63.12652	1	
3	0	0	0	168.997	1	
4	0	0	0	10.59421	1	
5	0	0	0	168.9315	1	
6	0	0	0	48.67398	1	
7	0	0	0	139.3523	1	
8	0	0	0	18.07281	1	
9	0	0	0	126.1906	1	
10	0	0	0	156.3481	1	
11	0	0	0	132.1382	1	
12	0	0	0	83.50544	1	
13	0	0	0	81.91068	1	
14	0	0	0	201.6258	1	
15	0	0	0	20.78574	1	
16	0	0	0	178.0625	1	
17	0	0	0	51.01086	1	
18	0	0	0	209.6888	1	
19	0	0	0	88.75236	1	
20	0	0	0	172.382	1	
21	0	0	0	59.44731	1	

Exhibit 5: Section of Types Worksheet

The first 4 columns each map to a corresponding column of the **Randomizer** worksheet, which holds random number by individual client. The formulas for the top row, which should be relatively self-explanatory, are listed by column as follows:

A. =IF(Randomizer!A1<Maven_Percent,1,0)
B. =IF(Randomizer!B1<Persuader_Percent,1,0)
C. =IF(Randomizer!C1<Connector_Percent,1,0)

D. =Threshhold_Adjust*IF(A1=1,150*Randomizer!E1/Maven_Factor,150*Randomizer!E1)
E. =IF(B1=1,Persuader_Factor,1)

Three worksheets interact to compute state values. The **State** worksheet keeps track of what clients are informed (1) and not informed (0). The **LocalWeights** spreadsheet takes the state worksheet and adjusts by the weight from column E in the **Types** worksheet (which is 1 unless the individual is a persuader, in which case it is whatever the persuader factor is). An example of the formula used in **LocalWeights!A1** is:

=State!A1*Types!$E1

The **GlobalWeights** worksheet is used to distribute the impact of connectors across the regions. If the client is a connector, it takes the computed local weight and probabilistically adds the local weight (based upon whether or not the random number used for the task meets the desired threshold). If the connector weight is 10 or greater, it will always add the connector. The formula for cell **GlobalWeights!A1** is:

=LocalWeights!A1*IF(Types!$C1=0,0,IF(10*Randomizer!$F1<Connector_Factor,1,0))

What this formula does, in effect, is to keep track of all globally visible clients. What would, perhaps have been a better implementation would have been to probabilistically determine what connectors are visible in each region. That would, however, have required a very large grid (1500 by 10) and would not have necessarily added much value to the simulation.

Once local and global values have been computed for one period, the state values, in the **State** worksheet, can be determined for the next period. The formula used for this (for cell **State!B1**) is:

```
=IF(
        OR(
                A1>0,
                Seeds!B$1/150>SeedRand!B1,
                SUM(LocalWeights!A$1:A$150)+SUM(GlobalWeights!A$1:A$1500)-
                        SUM(GlobalWeights!A$1:A$150)>Types!$D1
        ),
        1,
        0
)
```

Within the OR() function, the first test checks for the previous cell and the next checks to see if a seed is present with a probabilistic test to see if it should be used (the A$1 varies by region, with rows 1-150 using A$1, rows 151-300 using A$2, rows 301-450 using A$3, etc.). The third test is a threshold test. It takes the sum of the local and global weights, then subtracts out the portion of the global weights that have already been counted in the local region (again, the rows summed change every 150 rows). It then compares the total count with the threshold value established in the **Types** worksheet. If any of these tests succeed, the value of 1 is set for state, 0 otherwise.

Biography

Grandon Gill is an Associate Professor in the Information Systems and Decision Sciences department at the University of South Florida. He holds a doctorate in Management Information Systems from Harvard Business School, where he also received his M.B.A. His principal research areas are the impacts of complexity on decision-making and IS education, and he has published many articles describing how technologies and innovative pedagogies can be combined to increase the effectiveness of teaching across a broad range of IS topics. Currently, he is an Editor of the *Journal of IT Education* and an Associate Editor for the *Decision Science Journal of Innovative Education.*