

# Word of Mouth and Marketing: Influencing and Learning from Consumer Conversations

By

Dina Mayzlin

B.S. Economics  
Massachusetts Institute of Technology, 1997

SUBMITTED TO THE DEPARTMENT OF MANAGEMENT IN PARTIAL FULFILLMENT  
OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY  
AT THE  
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

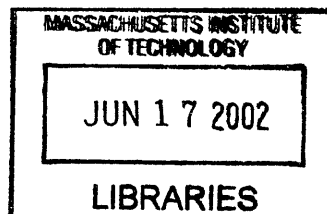
FEBRUARY 2002

© 2002 Massachusetts Institute of Technology. All rights reserved

Signature of Author: \_\_\_\_\_  
Department of Management  
January 9, 2002

Certified by: \_\_\_\_\_  
Birger Wernerfelt  
Professor of Management Science  
Thesis Supervisor

Accepted by: \_\_\_\_\_  
Birger Wernerfelt  
Professor of Management Science  
Chairman, Ph.D. Program



ARCHIVES

# Word of Mouth and Marketing: Influencing and Learning from Consumer Conversations

By

Dina Mayzlin

Submitted to the Department of Management on  
January 9, 2002 in partial fulfillment of the  
requirements for the degree of Doctor of Philosophy  
in Management

## Abstract

This thesis contains three separate essays that deal with word of mouth. In the first essay, "Promotional Chat on the Internet," we analyze the firms' incentives to anonymously supply positive reviews of products in chat rooms and other recommendation sites. This, in turn, lowers the credibility of word of mouth transmitted online. We develop a game theoretic model where an incumbent and an entrant that are differentiated in quality compete for the same online market segment. The consumers are uncertain about the entrant's quality, whereas the firms know the value of their products. The consumers hear messages online that make them aware of the existence of the entrant as well as help them decide which product is superior. We find a unique equilibrium where online word of mouth is informative despite the promotional chat activity by competing firms. In this equilibrium, we find that firms spend more resources chatting up inferior products. We also find that promotional chat may be actually more beneficial to consumers than a system with no promotional chat.

In the second essay, "Using Online Conversations to Measure Word of Mouth Communication," we test a long-held belief that word of mouth recommendations have a tremendous influence on the sales of new products. So far, there has been little empirical evidence to support this belief since, before the advent of the Internet, word of mouth recommendations were exchanged in private conversations that left no documentary evidence. The Internet provides a window into some of these private conversations and thus a means of measuring word of mouth activity. We pose the following pragmatic question: can we use these data to measure word-of-mouth and predict future product sales? We develop and test a model that predicts which metrics of online discussion activity should be correlated with long-run performance. Our empirical findings demonstrate that certain measures of online word-of-mouth are predictive of sales though their predictive power varies significantly over the show's lifetime.

Finally, in the third essay, "The Influence of Social Networks on the Effectiveness of Promotional Strategies," we examine the role of social network in word of mouth. The defining characteristic of "buzz" strategies is that the sellers approach the consumers directly, either in online chat rooms or in physical locations such as cafes or nightclubs. The goals of such strategies are twofold: to turn the approached consumer into a buyer and a missionary. The basic characteristic of buzz is that its diffusion is dependent on one's neighbors in the network in contrast to advertising, which allows the firm to communicate with consumers independently of their neighbors. We examine the network's moderating effect on the payoff from firm's investment to promote buzz. In addition, we compare the effectiveness of buzz promotion versus mass advertising as stand-alone marketing instruments and also analyze an advertising campaign that consists of both instruments.

Thesis Committee:

Birger Wernerfelt, Professor of Management Science (Chair)  
Duncan Simester, Associate Professor of Management Science  
Guido Kuersteiner, Assistant Professor of Economics

## Acknowledgments

First, I want to thank Aaron Cohen for being an excellent husband and a great friend. His love, support and companionship have turned my stay at MIT into a wonderful experience.

I also want to thank my parents for giving my brother and me the opportunity to live in this country. I would also like to thank them for investing in my education and supporting me in all my pursuits.

I benefited greatly from interaction with the entire faculty of the Marketing department while at MIT. I would like to thank Dan Ariely for his generosity in offering me financial (and moral) support for the past two years. Ely Dahan for useful job market advice and for the boat ride. John Hauser for great research advice and for keeping me informed on the latest pop music trends. Sandy Jap for great discussions on the nature of trust in organizations. John Little for advising me to pursue a PhD in marketing and for his constant encouragement while I was here. Drazen Prelec for encouraging my interests in behavioral research and for support throughout my time here. Duncan Simester for thoughtful and honest feedback on my research ideas and for giving me very helpful advice on the job market and beyond. Nader Tavassoli for my first marketing research seminar.

I would also like to thank Guido Kuersteiner for his help and advice on the second essay of this thesis. Sendhil Mullainathan for very helpful early advice on my empirical research. Florian Zettelmeyer for convincing me to apply to marketing PhD programs and for great job market advice.

I especially want to thank the chair of my committee, Birger Wernerfelt. From the very beginning of the program, he has been a true role model to me. His commitment and interest in research will always inspire me to seek interesting and important problems. There have been many instances when I came into his office feeling depressed about the progress of my research and left much happier and full of new ideas. I have always found Birger to be a dedicated advisor and a wonderful person.

My other great source of learning and motivation has been the other PhD students in the program. I would like to thank On Amir, Rebecca Hamilton, Hongmei Shang, Jiwoong Shin, and Robert Zeithammer for being great friends and colleagues. In particular, I would like to thank David Godes and Jose Silva for all the guidance and support that they have given since the very beginning of the program. Jose has always given me great research and dubious dietary advice. I could not list all the ways in which David has helped me. Most importantly, they always make me laugh.

## **Table of Contents**

<b>Abstract</b>	<b>2</b>
<b>Acknowledgments</b>	<b>3</b>
<b>Essay 1: Promotional Chat on the Internet</b>	
1. Introduction	5
2. Literature Review	8
3. Basic Model	11
a. Firms' Problem	14
b. Consumer' Problem	17
c. Bayesian Equilibrium	19
4. Consumer Welfare and Regulation	29
5. Extensions	31
a. Alternative Specifications	31
b. Price Signaling	36
c. Mixing Equilibrium	38
References	42
Appendix	44
<b>Essay 2: Using Online Conversation to Measure Word of Mouth Communication</b>	
1. Introduction	59
2. Related Literature	64
3. Theory	66
4. Model	69
5. Data	73
a. TV shows	73
b. Word-of-Mouth	75
6. Specification and Econometric Issues	78
a. Specification	78
b. Econometric Issues	79
7. Results	84
8. Conclusion	91
References	93
Appendix	96
<b>Essay 3: The Influence of Social Networks on the Effectiveness of Promotional Strategies</b>	
1. Introduction	105
2. Literature on Networks	107
3. A Model of Buzz	110
a. Network Structure	110
b. Transmission of Information in a Network: Buzz Promotion	112
c. The Firm's Investment into Buzz	116
4. A Model of Mass Advertising	123
a. The Choice Between Mass Ads and Buzz Promotion	124
b. Combining Mass Ads and Buzz Promotion	127
5. Future Research	131
6. Conclusion	133
References	134
Appendix	135

# Essay 1: Promotional Chat on the Internet

## 1. Introduction

In August of 1999, teenagers who frequented online bulletin boards of Britney Spears, a teen pop star, began to receive messages that recommended a new singer: Christina Aguilera. The authors of the messages frequently identified themselves by their first names only. Thus, Britney's fans had no means of distinguishing whether the messages they received came from other fans or from a marketing firm.

Some of the messages sent out did disclose that the authors were employees of Electric Artists, a promotional firm that specializes in online marketing. (The company's motto is "Shaping the Future of the Music on the Internet"). Electric Artists hired "posters" to surf various chat rooms and fan sites in order to generate online discussion and to provide information to potential fans. The campaign was ruled a success since Ms. Aguilera's album debuted at No. 1 on the charts and reached double platinum status. In early October of 1999, The Wall Street Journal devoted a front-page story titled "'Chatting' a Singer Up the Pop Charts" describing the various stages of the campaign.

A remarkable feature of this campaign was the means of communication used by the marketers: the Internet enabled the promoters to infiltrate and influence consumers' conversations. At first, this sounds like a very attractive strategy for marketers of many types of products. After all, in the past few years we have seen a proliferation of online communities: one community search engine (forumone.com) lists 310,000 web forums (which include "discussion forums, bulletin boards or message boards.") These forums

host conversations between consumers with diverse topics. One reason behind Electric Artists' success was the fact that consumers often offer unsolicited product recommendations online, which lent some credibility to chat about Christina. Following the success of Christina's promotion, Electric Artists has expanded its client list to include Tommy Hilfiger, the cellular provider Air Touch, YM, as well as Universal studios. Another Net promotion agency, M80 Interactive, employs similar marketing techniques. M80 was behind the marketing effort of the top-selling \*NSYNC album, "No Strings Attached." (Advertising Age; Chicago; May 1, 2000).

The marketers' ability to disguise their promotion as consumer recommendations is made possible by the anonymity enjoyed by participants of online communities. To quote Nirav Tolia, Epinions' co-founder, "The problem with most user-generated content on the Web is that there is no transparency, no context." (LA Times, 12/03/99, A-1, "Everyone Is a Critic in Cyberspace.") Some sites, such as Epinions, try to provide transparency by having users rate each other's reviews. However, a determined reviewer can enhance her ratings by having her friends contribute positive comments. There is even a term for such practice: "feedback abuse." (See "Building Stronger Brands Through Online Communities," Gil McWilliam) Ultimately, our identities as well as our incentives are obscure in the virtual world. Thus manufacturers can easily listen to the conversations that take place between what they think are consumers as well as actively participate in these discussions.

On the other hand, we might question the viability of such marketing efforts in the face of consumer skepticism. Consumers' awareness of the existence of such anonymous promotion (what we will call from now on "promotional chat") could cause them to discount online recommendations. Moreover, we would expect that the incumbent rival would engage in similar promotion to defend her market share. This paper poses the question whether promotional chat is a viable strategy in the long run.

More formally, this paper poses the following three research questions. First, we investigate conditions under which word of mouth online remains informative to the consumers in the presence of promotional chat by rival firms. Do we expect that anonymity, an aspect of the Internet that makes promotion so attractive, would be the undoing of promotional chat? Second, we ask whether promotional chat is most valuable for a firm whose product is more appealing than the competitor's product or a firm whose product is less appealing. Note that in previous advertising models firms spend more resources promoting their winners, which guarantees that advertising is a credible signal of quality. On the other hand, we also observe online recommendations of inferior products and questionable remedies: a post on [sci.med.prostate.cancer](http://sci.med.prostate.cancer) states, "New results from Shark Cartilage show how it has helped many people reduce and combat cancer." Third, we ask how consumer welfare is affected by anonymity and promotional chat compared to a context where promotional chat is not allowed.

We propose a game theoretic model where the incumbent and the entrant firms hold private information concerning the quality of their products. The firms send costly

recommendations to the consumer in order to influence his inference on the relative quality of the competing products. Consumers who have tried both products earlier also post online recommendations. Thus, online discussions are a mixture of unbiased recommendations as well as promotional activity by interested parties where the consumer is not able to tell apart the advertising from unbiased content. The consumer makes an inference on the quality of the new product based on the recommendation she receives. The consumer's inference will be affected by her knowledge that the firms engage in promotional chat.

We find that if the costs of engaging in promotional chat are sufficiently high, online chat remains informative. Thus, the firms' promotional activity does not turn chat rooms into noise: consumers are still more likely to hear the truth. Second, we find that promotional chat is more effective for products of low quality: firms lie. Note that the latter is the opposite of the signaling literature result where firms find it more profitable to promote their winners. The first and second results taken together are surprising: despite the firms' incentives to invest more into promoting the less appealing products, the consumers find chat informative. Third, we find that under certain conditions consumers may actually benefit from promotional chat due to the fact that it increases awareness of the new product.

## **2. Literature Review**

This paper relates to the existing literature on advertising and word of mouth communication. There is a rich literature in marketing on the sales response to advertising and the creation of an optimal advertising campaign. See, for example,

Vidale and Wolfe (1957), Little and Lodish (1969), Sasieni (1971), Little (1979), Simon (1982), Mahajan and Muller (1986), and Feinberg (1992). These are aggregate models in the sense that the consumer's decision is not modeled on an individual level, but the consumer's actions are summarized as a response function to sales. Thus, this literature is not concerned with questions of credibility of communication; questions that are central to this paper.

Another related stream of literature is new product diffusion through word of mouth and advertising. See, for example, Dodson and Muller (1978) and Mahajan, Muller and Kerin (1984). These papers model consumers as divided into segments where the consumers are either unaware of the product, aware of the product but had not yet purchased it or who have already purchased the product. Once again, these models do not address the questions of credibility. Instead, they take a somewhat mechanistic view of consumers' product choices: the flow between segments depends on the amount of advertising as well as word of mouth, which is in turn a function of relative sizes of the segments. Monahan (1984) addresses a similar problem using optimal control methodology.

In the economics literature, Ellison and Fudenberg (1995), Banerjee (1992) and Banerjee (1993) model the spread of information among consumers. McFadden and Train's (1996) model deals with consumers optimally choosing between trying a new product early on or waiting for others' reviews. Vettas (1997) examines the optimal timing of a monopolist's sales given word-of-mouth communication between consumers. Avery et.

al. (1999) deals with the organization of a market for product evaluations. Similarly to the approach taken in this paper, the papers above model consumers as Bayesian updaters (with the exception of Ellison and Fudenberg (1995) where more simple decision rules are used). In the models above, a consumer imperfectly learns from others' experience since there is heterogeneity in preferences or, in the case of Banerjee's model, uncertainty whether previous consumers acted on new information or "herded." However, none of the papers above model the firms' incentives to directly manipulate word of mouth.

Finally, there is a rich stream of literature that deals with the type of information that is conveyed by advertising. Dixit and Norman (1984) and Stegeman (1991) deal with the provision of hard information, namely, prices. Dixit et. al. (1984) finds that firms advertise excessively with respect to the social optimum, while Stegeman (1991) finds the opposite result : competitive firms usually underadvertise. Nelson (1974) and Kihlstrom and Riordan (1984) deal with the provision of soft information: the signaling value of advertising. Thus, in both of the models above advertising is a credible signal of quality. Advertising is a credible signal either due to the high quality firm's ability to recover the costs in repeat purchases or the high quality firm's lower costs of production. Note that in our model the costs of production do not correlate with quality and there is no possibility of a repeat purchase since we consider a one-period model. Unlike the models above, Horstmann and Moorthy (2000) show that advertising does not need to be monotone increasing in quality. They develop a model of service that includes a technological relationship between quality and capacity where high quality services cannot be provided by large capacity firms. Thus, in the presence of capacity constraints,

the high quality firms derive low value of advertising. This paper shows that even in the absence of capacity constraints, promotion need not be increasing in quality.

### **3. Basic Model**

The model in this section closely follows the example presented in the introduction. Here we present the set-up and the results of the basic model. In Section 3, we discuss an extension that deals with comparing consumer welfare across different game forms. In the last section, we discuss the results and present ideas for future research.

Let us first present a general overview of the model. There are two firms, an incumbent and an entrant, one risk-neutral uninformed consumer, and a segment of informed consumers. The firms offer substitute products of different quality. The firms observe which is the better product, but the uninformed consumer only observes the quality of the product offered by the incumbent firm. The segment of informed consumers have tried both products and thus also observe which is the superior product. Thus, we assume that the firms, through market research, perhaps, or experience, possess better knowledge of the industry than do the consumers who have not purchased the new product. (In the music industry, for example, audience tests can be conducted where a consumer is paid to rate snippets of various songs). More formally, we assume that the uninformed consumer expects a sure payoff of  $V^B$  from the incumbent's product (B) and an uncertain payoff of either  $\{V^C = V_H^C > V^B\}$ , which we call State 1, or  $\{V^C = 0\}$ , which we call State 2, from the entrant's product (C). Thus, in State 1, C is the more appealing product, whereas in State 2, C is the less appealing product.

Moreover, we assume that the uninformed consumer only becomes aware of the entrant through the promotional chat. That is, the consumer has a prior belief on the expected quality of an entrant, but cannot buy the entrant's product unless he hears a message mentioning the entrant's name. For instance, the consumer knows that 50% of new artists are appealing, but must learn from the chat the new artist's name before buying her CD. Thus, promotional chat serves two functions: awareness as well as recommendation on product choice.

Next, let us turn to the messaging that takes place online. There are three possible types of messages: messages that claim that the entrant's product is better than the incumbent's product, messages that claim that the incumbent's product is better than the entrant's product, and messages that pertain to the incumbent only. Alternatively, we can interpret these messages as positive word of mouth concerning the entrant, negative word of mouth concerning the entrant, or messages that do not mention the entrant. We model anonymity by assuming that consumers can't see the source of the message that they receive. This is a crucial assumption, and we later discuss how the presence of anonymity affects the results.

We also assume that the consumer observes only one message and makes an inference on the relative quality of the products based on that one message as well as his knowledge of the firms' actions in equilibrium. Thus, the consumer is unable to visit all the chat rooms, and observes a small subset of all the messages that are sent out. An alternative way to model this phenomenon would be to assume that the consumer receives a sample of total

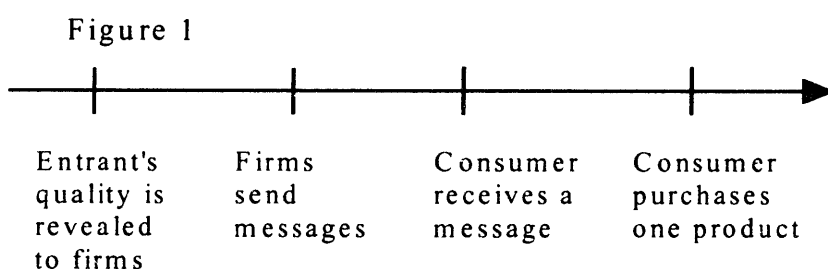
messages and updates his beliefs based on the ratio of positive messages received about the entrant. This alternative modeling technique would retain the flavor of the "one message" assumption: the more messages a firm sends, the more likely it is to convince the consumer of the high appeal of its product.

Let us next turn to the messages sent by the consumers. We assume that the informed segment of consumers sends  $N^U$  reviews that reveal the truth about the relative quality of the two products. We can think of this informed segment as consumers who have early knowledge on the quality of the entrant. For example, these are the teenagers who hear Christina's single and go online to talk about her. We also assume that the informed consumers send messages that contain information about B only. For example, these messages may discuss Britney's outfit at an awards ceremony. We assume that there are  $N^0$  of these irrelevant messages. Note that even though these messages are irrelevant from the perspective of product comparison, the participants may still enjoy sending as well as receiving these messages. We do not model the incentives of the informed unbiased consumers to post online. Instead, we implicitly assume that the posters are motivated by altruism, an assumption that is consistent with previous word of mouth literature. We do examine how the magnitude of these parameters affects our results.

On the firm's side, we assume that there are convex costs of messaging. This would model a situation where messages need to be personalized or it becomes more difficult to create an additional message since it has to visibly differ from the messages that come before it. We also discuss a variant of the model with linear costs. We focus on equilibria

with price pooling and show that, in fact, there does not exist a price signaling equilibrium.

The equilibrium concept used in the model is Bayesian Nash Equilibrium. Thus, the firms condition their actions on the State of the world (the quality of the entrant), and the consumer tries to infer the State of the world based on the message he receives. In addition, the consumer updates his prior taking into account the firms' strategies in both States of the world. Thus, both the consumer and the firms are fully strategic. We investigate the firms' actions, as well as the "informativeness" of the system to the consumer. Figure 1 presents the sequence of events of the game.



### 3a. Firms' Problem

To solve for the equilibrium, we need to specify the number of messages that the two firms will choose to send in the two states of the world as well as the inferences that the consumer makes. Let us first turn to the firms' problem. The strategy space available to firms is the number of messages that they send praising their product. As a reminder, we refer to incumbent as B(ritney) and to entrant as C(hristina). Also, let {State 1} be the state of the world where the entrant's product is superior, whereas {State 2} is the state of the world where the incumbent's product is superior. (Thus,  $N_1^B$  is the number of

messages sent by the incumbent in State 1;  $N_2^B$  is the number of messages sent by the incumbent in State 2;  $N_1^C$  and  $N_2^C$  are messages sent by the entrant;  $N^U$  is the number of unbiased messages praising the superior product; and  $N^0$  is the number of messages that are irrelevant to product comparison).

In the model, a consumer receives one message only. The three possible types of messages that can be received are:  $\{\mu^{B>C}, \mu^{C>B}, \mu^B\}$  where  $\mu^{B>C}$  stands for a message praising B over C,  $\mu^{C>B}$  stands for a message praising C over B, and  $\mu^B$  stands for a message that is irrelevant for product comparison. Note that there will be  $N^U + N_1^C$  messages praising C in State 1 since there will be  $N^U$  truthful unbiased messages and  $N_1^C$  messages sent by firm C. Similarly, there will be  $N^U + N_2^B$  messages praising B in State 2. We assume that the consumer picks a message at random from the existing pool of messages. Thus, the probability that a consumer observes  $\mu^{C>B}$  is the ratio of all messages praising C over the total number of messages. Hence, the probabilities that each type of message will be received by the consumer are summarized in Table 1 below.

**Table 1: Probability the consumer receives a particular message**

Type of Message	State 1	State 2
$\mu^{B>C}$	$\frac{N_1^B}{N_1^B + N_1^C + N^U + N^0}$	$\frac{N_2^B + N^U}{N_2^B + N_2^C + N^U + N^0}$
$\mu^{C>B}$	$\frac{N_1^C + N^U}{N_1^B + N_1^C + N^U + N^0}$	$\frac{N_2^C}{N_2^B + N_2^C + N^U + N^0}$
$\mu^B$	$\frac{N^0}{N_1^B + N_1^C + N^U + N^0}$	$\frac{N^0}{N_2^B + N_2^C + N^U + N^0}$

From the firm's perspective, each event above can result in two possible actions by the consumer: either the consumer buys or does not buy the firm's product (for a profit of either  $P$  or  $0$ ). Let  $\Pi_{[C>B]}^C$  stand for the profit that  $C$  derives following the consumer receiving a message  $\mu^{C>B}$ . (Likewise for all other messages as well as for Firm  $B$ ). Note that we assume that  $\Pi_{[B]}^B = P$  -- the consumer buys  $B$  if he does not become aware of the product's existence through the chat room. Thus, if he does not hear anything about the entrant (either negative or positive), he makes the default purchase: the incumbent's product. Also note that we assume that  $V^B > P$ , i.e. that the consumer derives positive value from the incumbent's product. We then can write out Christina's profit function in State 1 below:

$$\frac{N_I^C + N^U}{N_I^C + N_I^B + N^U + N^0} \Pi_{[C>B]}^C + \frac{N_I^B}{N_I^C + N_I^B + N^U + N^0} \Pi_{[B>C]}^C + \frac{N^0}{N_I^C + N_I^B + N^U + N^0} \Pi_{[B]}^C - a \frac{(N_I^C)^2}{2} \quad (1)$$

Consistent with the convexity assumption, we assume that the cost of sending messages is quadratic in the number of messages sent. We later discuss the results under the assumption of linear costs. We also include a parameter  $a$  on the cost function even though later on it will be more convenient to divide out the equation by  $a$ , and to talk about the profit/ cost ratio.

Since there are two firms and two states of the world, there are altogether four simultaneous maximization equations. We list the maximizations for State 1 below, while the whole system is listed in the Appendix.

1) For entrant ( $C$ ) in State 1:

$$\max_{N_1^C} \frac{N_1^C + N^U}{N_1^C + N_1^B + N^U + N^0} \Pi_{[C>B]}^C + \frac{N_1^B}{N_1^C + N_1^B + N^U + N^0} \Pi_{[B>C]}^C + \frac{N^0}{N_1^C + N_1^B + N^U + N^0} \Pi_{[B]}^C - a \frac{(N_1^C)^2}{2} \quad (2)$$

st.  $N_1^C \geq 0$

2) For incumbent (B) in State 1:

$$\max_{N_1^B} \frac{N_1^C + N^U}{N_1^C + N_1^B + N^U + N^0} \Pi_{[C>B]}^B + \frac{N_1^B}{N_1^C + N_1^B + N^U + N^0} \Pi_{[B>C]}^B + \frac{N^0}{N_1^C + N_1^B + N^U + N^0} \Pi_{[B]}^B - a \frac{(N_1^B)^2}{2} \quad (3)$$

st.  $N_1^B \geq 0$

For example, consider an equilibrium where the consumer buys C if he receives the message  $\mu^{C>B}$ , but buys B if he receives  $\mu^{B>C}$  or  $\mu^B$ . Since the price of the product is P, the firms' profit functions look like:

$$\text{Entrant (C) in State 1: } \frac{N_1^C + N^U}{N_1^C + N_1^B + N^U + N^0} \frac{P}{a} - \frac{(N_1^C)^2}{2}$$

$$\text{Incumbent (B) in State 1: } \frac{N_1^B + N^0}{N_1^C + N_1^B + N^U + N^0} \frac{P}{a} - \frac{(N_1^B)^2}{2}$$

### 3b. Consumer's Problem

Here the consumer is trying to infer which product will deliver higher value: B or C.

Thus, his strategy space is a decision on a purchase of one of the products based on the message received. Note that since in the model the two goods are substitutes, the consumer never chooses to buy both products.

If the consumer does not become aware of C, he buys B as default. Otherwise, the consumer updates his priors on the quality of the entrant, conditioning on the message received. His prior probability on the entrant delivering more value (State 1) is  $P(s1)$ . Let  $\theta(\{\text{message}\}) = P(\text{state1} | \{\text{message}\})$  -- consumer's posterior following a message. The

notation is the following:  $P(\mu^{C>B} | s1)$  is the likelihood that  $\mu^{C>B}$  message would be received in State 1. We next apply the Bayes' Rule to derive the updating of the consumer's beliefs:

$$\theta(\mu^{C>B}) = \frac{P(\mu^{C>B} | s1)P(s1)}{P(\mu^{C>B} | s1)P(s1) + P(\mu^{C>B} | s2)P(s2)} \quad (4)$$

$$\theta(\mu^{B>C}) = \frac{P(\mu^{B>C} | s1)P(s1)}{P(\mu^{B>C} | s1)P(s1) + P(\mu^{B>C} | s2)P(s2)} \quad (5)$$

where the probabilities of receiving each message in the two states of the world are summarized in Table 1 and presented once again below:

$$P(\mu^{C>B} | s1) = \frac{N_1^C + N^U}{N_1^B + N_1^C + N^U + N^0}; \quad P(\mu^{C>B} | s2) = \frac{N_2^C}{N_2^B + N_2^C + N^U + N^0}$$

$$P(\mu^{B>C} | s1) = \frac{N_1^B}{N_1^B + N_1^C + N^U + N^0}; \quad P(\mu^{B>C} | s2) = \frac{N_2^B + N^U}{N_2^B + N_2^C + N^U + N^0} \quad (6)$$

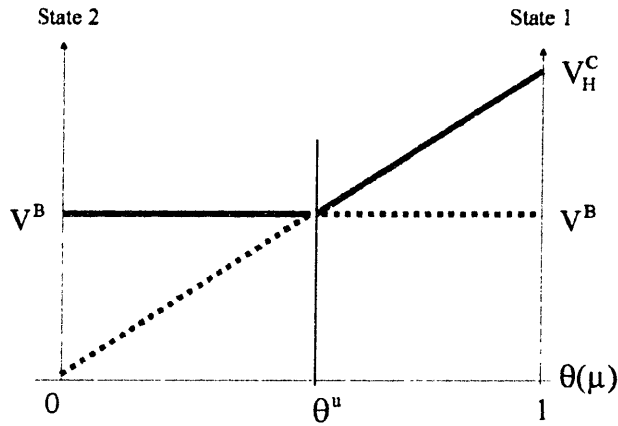
Thus, we can see that the consumer bases his decision taking into account the firms' optimal strategies in both states of the world. Note that he does not observe the firms' actions exactly, but instead calculates the equilibrium strategy.

Next we consider the relationship between the consumer's beliefs and his decision to purchase either C or B. Assuming risk-neutrality, Figure 2 below represents the decision that the consumer faces. The horizontal line represents the certain value from buying B. The positively sloped line represents the expected value of purchasing the entrant's product as a function of the posterior probability. The consumer maximizes his expected payoff by choosing the upper envelope of the two lines. If the consumer's posterior on State 1 is high enough ( $\theta(\mu) > \theta^U$ ), the consumer would rather go ahead and purchase C.

If, on the other hand, the posterior low enough ( $\theta(\mu) \leq \theta^u$ ), the consumer chooses to

purchase B. Note that  $\theta^u = \frac{V^B}{V_H^C}$ .

**Figure 2: Consumer's Optimal Value as a Function of the Posterior Belief**



For example, what are the conditions that are required to obtain an equilibrium where the users follow the recommendations that they receive online? (That is, when will the consumer buy B when he receives  $\mu^{B>C}$  and buy C when he receives  $\mu^{C>B}$ ?) This will

occur if  $\theta(\mu^{C>B}) \geq \frac{V^B}{V_H^C}$  and  $\theta(\mu^{B>C}) < \frac{V^B}{V_H^C}$ . After we make the appropriate substitutions,

we see that this is equivalent to the expressions below:

$$\frac{P(\mu^{C>B} | s1)}{P(\mu^{C>B} | s2)} \geq \frac{P(s2)V^B}{P(s1)(V_H^C - V^B)} \text{ and } \frac{P(\mu^{B>C} | s1)}{P(\mu^{B>C} | s2)} < \frac{P(s2)V^B}{P(s1)(V_H^C - V^B)} \quad (7)$$

The above conditions ensure that the signals received are informative: the consumer's beliefs change enough to change consumer purchase decision.

### 3c. Bayesian Equilibrium

Putting together the firms' and the consumer's problem together, we look for pure strategy equilibria where the consumer's beliefs and firms' actions are mutually consistent. Once again, the equilibrium consists of the firms' decisions on how many messages to send contingent on the state of the world as well as on the consumer's decision which product to buy contingent on the message he receives.

Let us first consider the consumer's problem. He has four possible strategies available to him. This is due to the fact that he may receive two possible messages and can choose whether to buy or not to buy the entrant following each message, where the decision not to buy the entrant's product is equivalent to a decision to buy the incumbent's product. By assumption, if the consumer receives an irrelevant message, his default purchase is the incumbent's product. Once the consumer's decision rule is fixed, we can derive the optimal messaging policies of the two firms. Finally, we have to check that the consumer's decision is optimal, given the firms' strategies. We find that a unique pure strategy equilibrium exists in a region where costs are above a certain cutoff. In this equilibrium, we find that the consumers will follow the online recommendations, but that the firms have an incentive to lie.

**Proposition 1:**

Let us define  $\rho = \frac{P}{a}$  (a ratio of profit to cost). For the set of parameters

$\{ P(s1), V_H^C, V^B, N^U, N^0 \}$  (where  $0 < P(s1) < 1$ ) there exists  $\hat{\rho}$  such that for all  $\rho \leq \hat{\rho}$

there exists a unique pure strategy Bayesian Equilibrium to the problem above. (No pure strategy equilibria exist in the region  $\rho > \hat{\rho}$ ).

- 1) In this equilibrium, the consumer buys C if he receives  $\mu^{C>B}$  and buys B if he receives  $\mu^{B>C}$ : recommendations online are informative.
- 2) The resulting promotional intensities are such that  $N_1^C < N_2^C, N_1^B > N_2^B$ : firms promote more heavily in the state of the world when their product is inferior.
- 3) The firm expects to make higher profits in the state of the world where its product is superior. See Appendix for the proof.

We next discuss and graphically illustrate the intuition behind the results. Let us fix the consumer beliefs to be such that he buys C if and only if he hears  $\mu^{C>B}$  and consider C's incentives to post messages. We also assume for now that the total number of messages of type  $\mu^{B>C}$  is fixed across the two states of the world:  $N_1^B = N_2^B + N^U \equiv N^{B1}$ , and we define an additional variable:  $T_i^C \equiv$  the total number of messages of type  $\mu^{C>B}$  in state  $i$  ( $T_1^C = N_1^C + N^U; T_2^C = N_2^C$ ).

Next, consider C's benefit and marginal benefit of messaging in State 1:

$$\text{Benefit} | s1 = P(\mu^{C>B} | s1)\rho = \frac{N_1^C + N^U}{N_1^B + N_1^C + N^U + N^0}\rho \equiv \frac{T_1^C}{N^B + T_1^C + N^0}\rho;$$

$$\frac{\partial \text{Benefit} | s1}{\partial N_1^C} = \frac{\partial \text{Benefit}}{\partial T_1^C} \frac{\partial T_1^C}{\partial N_1^C} = \frac{\partial \text{Benefit}}{\partial T_1^C}(1) = \frac{N^B + N^0}{(N^B + T_1^C + N^0)^2}\rho.$$

Similarly, in State 2,

---

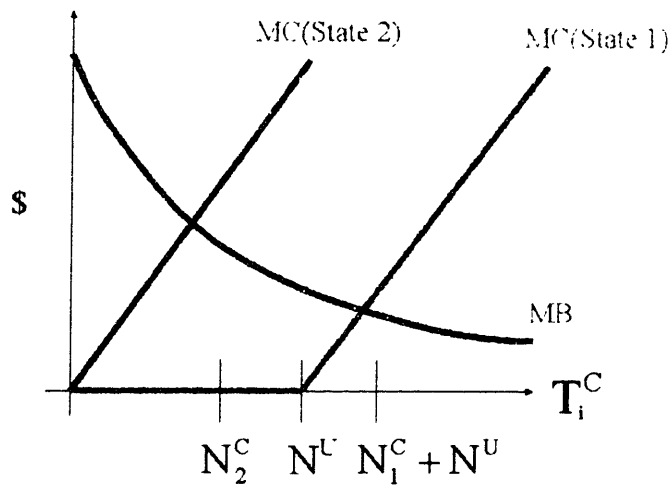
<sup>1</sup> Note that in equilibrium,  $N_1^B < N_2^B + N^U$ . In our illustration, we concentrate on the role of the cost structure. As will be shown below, this assumption allows us to fix the functional form of the marginal benefit across the two states of the world.

$$\text{Benefit} | s2 = P(\mu^{C>B} | s1)\rho \equiv \frac{T_2^C}{(N^B + T_2^C + N^0)} \rho; \frac{\partial \text{Benefit} | s2}{\partial N_2^C} = \frac{N^B + N^0}{(N^B + T_2^C + N^0)^2} \rho \quad (9)$$

We can see that the marginal benefit (change in probability) is decreasing in the total number of  $\mu^{C>B}$  messages sent. Note that in the example above the functional form of the marginal benefit as a function of  $T_i^C$  is the same for both states of the world.

Next, we turn to the marginal costs that C faces as a function of  $T_i^C$ . In State 1, C faces a marginal cost of 0 for  $N^U$  messages, and a linear marginal cost for each additional message. On the other hand, in State 2, C faces a linear marginal cost for all messages. The intersection of the marginal cost and marginal benefit curves determines the firm's strategy in both states of the world. Refer to the graph below for an illustration:

**Figure 3: MB/MC Trade-Off**

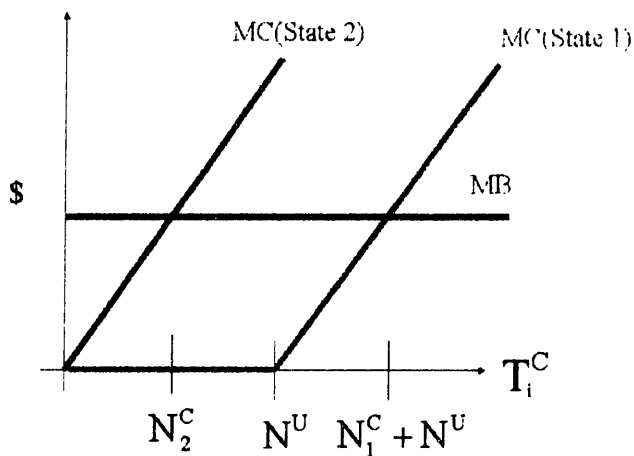


Thus, we can see that there will be more  $\mu^{C>B}$  messages in the state of the world where C is superior:  $N_1^C + N^U > N_2^C$ . This is the informativeness result. That is, consumers on average should believe the recommendations that they hear. On the other hand, we see that  $N_1^C < N_2^C$ , i.e. the firms spend more resources promoting their losers.

There are three aspects of the model that are driving the results: 1) the declining marginal benefit of messaging 2) the existence of unbiased reviews, and 3) the convexity of the costs. We consider these three aspects separately.

Let us first illustrate that the declining marginal benefit of messaging is crucial for the result that firms invest more heavily in their losers. Along with unbiased reviews, this ensures that the superior firm's marginal benefit of messaging is lower than the inferior firm's. Below, we re-draw Figure 3 with a constant marginal benefit curve. We can see that here  $N_1^C = N_2^C$  (with a linear marginal cost curve and a constant marginal benefit curve).

**Figure 4: Constant MB**



Of course, it is the micro model of consumer's drawing a message from a "bucket" of messages that results in the concavity of benefit. The assumption is meant to capture the reality that the more messages a firm sends, the more likely is a consumer to receive the promotional message.

Next, we turn to the assumption that there are unbiased truthful reviews out there. Note that this assumption is critical for obtaining the model's results. The difference between  $N_1^C + N^U$  and  $N_2^C$  decreases as  $N^U$  decreases. Thus, no informative equilibrium exists as  $N^U$  goes to zero. How reasonable is this assumption? One possible criticism is that here the incentives of the unbiased reviewers are not modeled. This is a possible extension of the model. Moreover, it is possible that the number of unbiased reviews differs across categories. High involvement categories (such as movies) are likely to have more reviewing, while low involvement categories (coffee) are likely to have less reviewing. We can make predictions on the level of promotional activity based on the involvement of the category. We argue that this is a reasonable assumption due to the fact that people differ in their patience level (discount factor). Thus, some consumers are

likely to quickly search out and try new products, while others are willing to wait and see how others react.

One of the attractive features of anonymous promotion on the Internet is the cost effectiveness of the campaign online compared to a similar campaign offline. Thus, the posters do not have to travel great distances to reach the consumers and can promote to an audience that is predisposed to buy a product in the category since the chat rooms are topic-based. Moreover, with the introduction of "chat bots" (programs that automatically generate chat) we might think that the costs of such a promotion are linear as opposed to convex. We thus consider how our results change if we assume that the costs are linear: each message costs a fixed amount,  $a$ . Once again, different values of  $a$  lead to different equilibria. In the Appendix, we formally define the different equilibria that are obtained in the various regions of the parameter space. In this section, we provide the intuition behind the results.

**Proposition 2:**

Consider the problem above with linear costs, where the cost per message is  $a$ . Let us

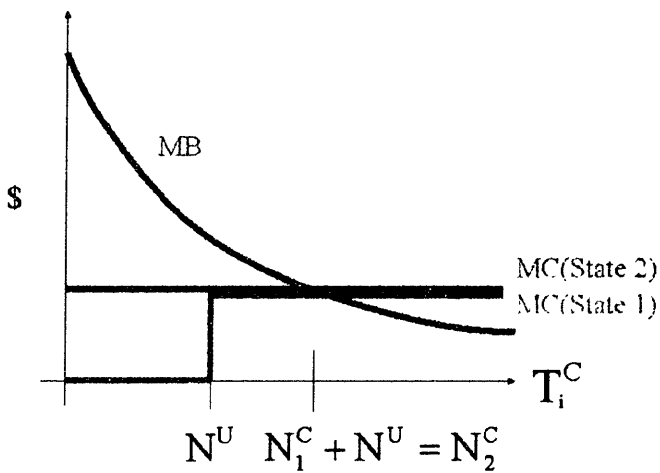
similarly define  $\rho = \frac{P}{a}$ . We find the following:

a) When  $N^U + N^O < \frac{\rho}{4}$  (relatively low costs), there is no informative pure strategy equilibrium.

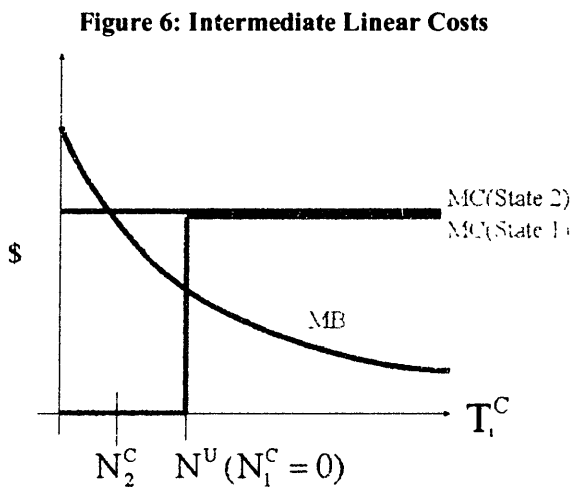
- b) When  $\frac{\rho}{4} < N^U + N^0 < \rho$  (intermediate cost range), there are potentially informative pure strategy equilibria in the sense that  $N_1^C + N^U > N_2^C$ ,  $N_2^B + N^U > N_1^B$  and where  $N_1^C < N_2^C$ ,  $N_2^B < N_1^B$  (firms lie).
- c) When  $N^U + N^0 \geq \rho$  (high cost range), the firms choose not to promote their products, and the only messaging is done by the unbiased source.

See Appendix for the formal proof. Instead, we turn to the intuition behind the results in this section. Essentially, there are three possible cost levels: high, intermediate, and low. If the costs are low, there is no informative equilibrium (see Figure 5 below). We can see that the total number of  $\mu^{C>B}$  messages is equal across the two states of the world. Thus, a message praising the product will not change the consumer's priors on the quality of the product, which renders the messages uninformative.

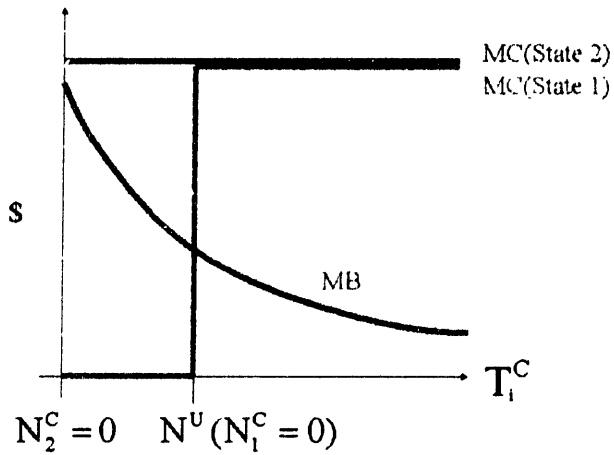
**Figure 5: Low Linear Costs**



However, if the costs are not too low, we run into corner solutions where either one or both of the firms choose not to expend any effort into promotion. Interestingly, these are informative equilibria. Figure 6 illustrates the situation where the costs are at an intermediate level. The superior firm chooses not to send any messages beyond  $N^U$ , whereas the inferior firm sends fewer than  $N^U$  messages. This is informative since a positive recommendation is an indicator that the product is of high quality. We can see once again that that firms lie (a firm spends more resources on its inferior products). Note that as we raise the costs even further (see Figure 7 below), we see that the firm chooses not to spend anything on promotion, regardless of the quality of their product. Here, no one promotes her product. All the recommendations a consumer sees online are sent by the unbiased source.

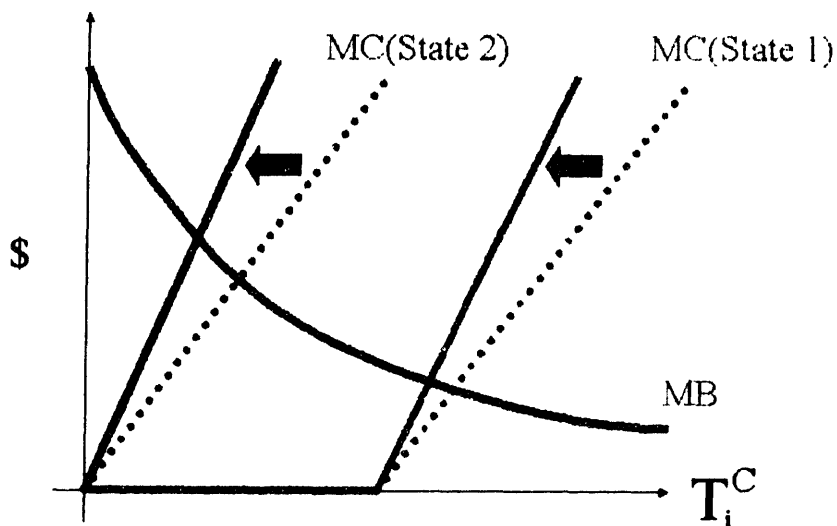


**Figure 7: High Linear Cost**



Note that with convex costs, we do not observe the discontinuities that we obtain with linear costs. Thus, firms never spend zero effort on promotion, no matter how high the cost parameter is. We again can demonstrate this graphically in Figure 8. Let us increase the  $a$  parameter. This rotates the marginal cost line upwards. This in turn decreases the number of messages sent by the firms, but will never decrease that number to 0 as long as  $a < \infty$ .

**Figure 8: Increase in Convex Costs**



Lastly, we contrast the finding of this paper with the findings of the advertising as signaling literature. As we mentioned, in that literature advertising provides soft information to the consumer. In that literature, the high quality firm can use advertising as a credible signal either due to differential costs or due to repeat purchase effects. That is, since the satisfied consumers will promote the high quality product to their friends, the firm with a better product benefits more from higher sales in the first period. Note that in our one-period, same cost model, we get the opposite result: the high quality firm views word of mouth as a substitute to advertising and advertises less than the low quality firm.

#### **4. Consumer Welfare & Regulation**

Next, consider to the issue of consumer welfare. In order to analyze how promotional chat and anonymity affect consumer welfare, we compare consumer welfare across two

different game forms. We compare consumer welfare in the basic model against a system where all advertising is banned.

**Proposition 3:**

We find that promotional chat may benefit the consumer compared to a system with no advertising in the following two scenarios:

- 1) If the prior,  $P(s1)$ , on the entrant being of high quality is high
- 2) If the system has a lot of irrelevant messages relative to unbiased consumers

Specifically, the consumer benefits from promotional chat if the following condition holds:

$$\frac{P(\mu^{C>B} | s1) - \frac{N^U}{N^U + N^0}}{P(\mu^{C>B} | s2)} \geq \frac{P(s2)V^B}{P(s1)(V_H^C - V^B)}. \text{ This simplifies to the following expression:}$$

$$\frac{N^0 - N^U}{N^0 + N^U} \geq \frac{P(s2)V^B}{P(s1)(V_H^C - V^B)}$$

See Appendix for proof.

Note that the result above is very intuitive. Thus, promotional chat is a good instrument for increasing awareness of the entrant. The consumer benefits from increased awareness only in the state of the world where the entrant is superior. (In the other state of the world the consumer follows the recommendation and is deceived). Hence, the prior on the entrant being superior is an important factor. On the other hand, the system of recommendations with no advertising sends out perfectly informative signals. However, the probability that such a signal is heard at all depends on the level of noise (number of irrelevant messages) relative to the level of unbiased messages. Thus, if noise is high

relative to the number of unbiased messages,  $\frac{N^0 - N^U}{N^0 + N^U}$  is high, the consumer is actually better off under a system where some advertising is allowed as opposed to a system with no advertising.

## 5. Extensions

In this section, we explore the robustness of the results in the main model by altering the specification of the model. In Section 5a, we examine several alternative specifications of the model. We consider a different process of message propagation that we call “seeding.” We also consider a specification where the consumer can receive multiple messages. In Section 5b, we turn to price signaling, and in Section 5c, we discuss a mixing equilibrium.

### 5a. Alternative Specifications

For the purposes of this section only, we consider symmetric specifications only. (The precise definition of a symmetric specification is provided below). That is, in our original specification, an irrelevant message favored the incumbent due to the awareness assumption. Here we assume that there are no irrelevant messages present ( $N^0 = 0$ ). We do this in order to explore how the asymmetry affects the results and also in order to simplify some of the later calculations. All other aspects of the model, including quadratic costs, remain the same.

To simplify the calculations, we define a few new variables. Following the notation above, we define  $T_i^C \equiv$  the total number of messages of type  $\mu^{C>B}$  in state  $i$ , and  $T_i^B \equiv$

the total number of messages of type  $\mu^{B>C}$  in state  $j$ . Below we summarize the probabilities that the consumer will receive either of the two messages in the two states of the world:

$$P(\mu^{C>B} | s1) \equiv f(T_1^C, T_1^B) = \frac{T_1^C}{T_1^C + T_1^B} ; P(\mu^{B>C} | s1) \equiv 1 - f(T_1^C, T_1^B) = \frac{T_1^B}{T_1^C + T_1^B} ;$$

$$P(\mu^{C>B} | s2) \equiv f(T_2^C, T_2^B) = \frac{T_2^C}{T_2^C + T_2^B} ; P(\mu^{B>C} | s2) \equiv 1 - f(T_2^C, T_2^B) = \frac{T_2^B}{T_2^C + T_2^B} ; \quad (10)$$

Note that the probability function is symmetric since  $f(T^C, T^B) = 1 - f(T^B, T^C)$ .

Next, let us consider the issue of the asymmetry in the messaging between a firm with a superior product and a firm with an inferior product. Specifically, let us investigate how a firm's effort relates to the total number of messages that end up circulating online. In the main model, we assume that  $T_1^C = N_1^C + N^U$ ,  $T_2^C = N_2^C$ ,  $T_1^B = N_1^B$ , and  $T_2^B = N_2^B + N^U$ . Another reasonable specification is to assume that in the state of the world where the product happens to be superior there is some "seeding" of information. The term "seeding" was suggested by Ken Krasner, the CEO of Electric Artists, in a private conversation. In this view of promotional chat, the information supplied by the firm is multiplied and propagated by the informed consumers. For example, this process can be achieved if the informed consumer forwards to the uninformed consumer the firm's messages praising the superior singer in addition to generating messages on his own.

Mathematically “seeding” yields a specification  $T_1^C = \kappa N_1^C + N^U$ ,  $T_2^C = N_2^C$ ,  $T_1^B = N_1^B$ , and  $T_2^B = \kappa N_2^B + N^U$  where  $\kappa > 1$ .<sup>2</sup>

With this specification once again we get results that are very similar to the main model. (See Appendix for a complete proof). That is, there exists  $\hat{\rho}$  such that for all  $\rho \leq \hat{\rho}$  we obtain a unique informative pure strategy equilibrium where  $T_1^C > T_2^C$  and  $T_1^B < T_2^B$ . In addition, the result that firms lie,  $N_1^C < N_2^C$  and  $N_1^B > N_2^B$ , holds as long as  $N^U > 0$ . (We do, however, find that as  $\kappa$  increases,  $N_2^C - N_1^C$  and  $N_1^B - N_2^B$  decrease). Similarly to the result obtained in the basic model, as  $N^U$  decreases,  $N_2^C - N_1^C$  and  $N_1^B - N_2^B$  decrease. Once again, the existence of independent word of mouth (unrelated to firm's actions) is essential for the result that firms lie.

Next, we explore a different specification for the probability function,  $f(T^C, T^B)$ .

Specifically, recall that in the main model we make an assumption that the uninformed consumer receives one message only. Let us relax that assumption, and instead, assume that the consumer can receive multiple messages praising B and C. Once again, for simplicity we assume that the function is symmetric. That is, there are no irrelevant messages.

---

<sup>2</sup> Note that when  $\kappa = 1$ , we have the basic model where  $N^U = 0$ .

In addition, let  $\tilde{X}(\mu^{C>B})$  be the number of messages praising C and let  $\tilde{X}(\mu^{B>C})$  be the number of messages praising B that the consumer receives. Since the consumer engages in sampling, the number of each particular message that he receives is stochastic. The firm controls the mean of the distribution of the number of messages received. Thus, assume that  $\tilde{X}(\mu^{C>B}) \sim \text{exponential with mean} = \bar{X}_i^C$

$\{ P([\tilde{X}(\mu^{C>B}) = x]) = \frac{1}{\bar{X}_i^C} \exp(-\frac{x}{\bar{X}_i^C}) \}$  where  $i$  is the State;  $\tilde{X}(\mu^{B>C}) \sim \text{exponential with mean} = \bar{X}_i^B$ . As before, the mean of the distribution of the number of messages received is a sum of the firm's effort  $(\lambda_i^C, \lambda_i^B)$  and the effort exerted by the unbiased recommenders  $(\lambda^U)$ :  $\{ \bar{X}_1^C = \lambda_1^C + \lambda^U, \bar{X}_2^C = \mu_2^C = \lambda_2^C, \bar{X}_1^B = \lambda_1^B, \bar{X}_2^B = \lambda_2^B + \lambda^U \}$ . Similarly to the previous specification, the costs that the firm bears is quadratic in its effort.

Upon receiving  $\{ \tilde{X}(\mu^{C>B}), \tilde{X}(\mu^{B>C}) \}$ , the consumer will buy C iff

$$\frac{P(\tilde{X}(\mu^{C>B}), \tilde{X}(\mu^{B>C}) | \text{state} = 1, \bar{X}_1^C, \bar{X}_1^B)}{P(\tilde{X}(\mu^{C>B}), \tilde{X}(\mu^{B>C}) | \text{state} = 2, \bar{X}_2^C, \bar{X}_2^B)} > \frac{P(s2)V^B}{P(s1)(V_{II}^C - V^B)}. \text{ To simplify the calculations}$$

below, assume that  $\frac{P(s2)V^B}{P(s1)(V_{II}^C - V^B)} = 1$ . Thus, the consumer will buy C iff

$$P(\tilde{X}(\mu^{C>B}), \tilde{X}(\mu^{B>C}) | \text{state} = 1, \bar{X}_1^C, \bar{X}_1^B) > P(\tilde{X}(\mu^{C>B}), \tilde{X}(\mu^{B>C}) | \text{state} = 2, \bar{X}_2^C, \bar{X}_2^B).$$

Substituting for the expression of the exponential p.d.f.,

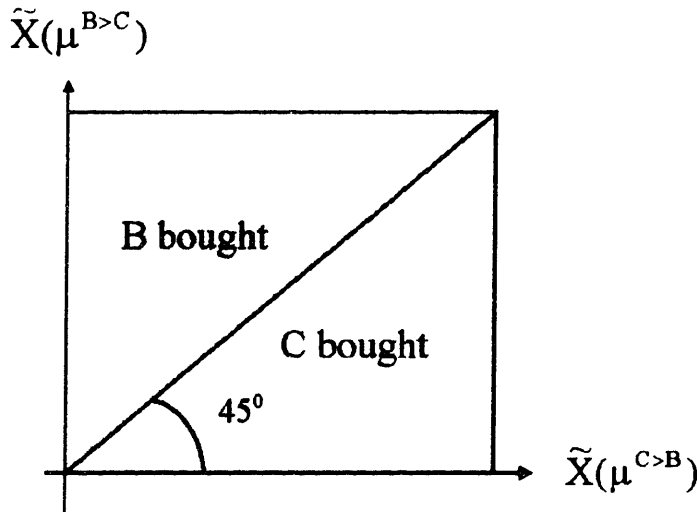
$$\frac{\bar{X}_2^B \bar{X}_2^C}{\bar{X}_1^C \bar{X}_1^B} \exp(-\frac{\tilde{X}(\mu^{C>B})}{\bar{X}_1^C} - \frac{\tilde{X}(\mu^{B>C})}{\bar{X}_1^B}) > \exp(-\frac{\tilde{X}(\mu^{C>B})}{\bar{X}_2^C} - \frac{\tilde{X}(\mu^{B>C})}{\bar{X}_2^B}). \text{ Let us focus on the}$$

symmetric equilibrium where  $\bar{X}_1^C = \bar{X}_2^B, \bar{X}_2^C = \bar{X}_1^B$ . This implies that the consumer

chooses C iff  $\tilde{X}(\mu^{C>B})(\frac{1}{\bar{X}_2^C} - \frac{1}{\bar{X}_1^C}) > \tilde{X}(\mu^{B>C})(\frac{1}{\bar{X}_1^B} - \frac{1}{\bar{X}_2^B})$ . It is easy to see that the only possible equilibrium is one where  $\{\bar{X}_1^C > \bar{X}_2^C, \bar{X}_1^B < \bar{X}_2^B\}$ .

To show that no other possible efforts can be an equilibrium, consider an equilibrium where  $\{\bar{X}_1^C < \bar{X}_2^C, \bar{X}_1^B < \bar{X}_2^B\}$ . This would imply that the consumer buys C whenever  $\tilde{X}(\mu^{C>B}) + \tilde{X}(\mu^{B>C}) < 0$ . This of course implies that the consumer always buys B. Thus, C finds it optimal not to expend any effort on promotion, which results in a contradiction:  $\bar{X}_1^C = \lambda^U > \bar{X}_2^C$ . The figure below demonstrates the region where the consumer chooses C:

**Figure 9: Consumer's Decision Based on Number of Messages Received**



If we integrate the region above, we see that  $\Pr(C \text{ is bought}) = \frac{\bar{X}_1^C}{\bar{X}_1^C + \bar{X}_1^B}$ . This is exactly equivalent to the symmetric specification we had discussed at the beginning of the section. Thus, we see that a different model that allows multiple messages to be

received can yield the same results as the simple symmetric model we had posed at the beginning.

### **5b. Price Signaling**

In the main model we assume that prices contain no information concerning the quality of the products (the state of the world). Thus, in our model the consumers can only infer quality from the chat and not from the prices. However, we have often seen in both the marketing as well as the economic literature that prices can serve as signals of quality. See, for example, Kalra, Rajiv, and Srinivasan (1998) and Anderson & Simester (2000). In our model, can prices possibly signal to consumers which products they should buy?

We next extend our main model to include possible price signaling, a separating equilibrium where firms post different prices depending on the state of the world. Thus, let us suppose that firms fix prices at the beginning of the game. A consumer may receive a message, as before, and draw an inference on the state of the world. With a positive probability, the consumer remains unaware of the entrant if he does not receive any messages. An aware consumer observes the incumbent's as well as the entrant's prices:  $[P_1^C, P_1^B]$  in State 1 and  $[P_2^C, P_2^B]$  in State 2 (the superscript refers to the player, and the subscript refers to the state of the world). The prices are informative only in case where  $[P_1^C, P_1^B] \neq [P_2^C, P_2^B]$ . Note that we do not require all elements of the vectors to differ. Thus, we are allowing the case where  $P_1^C \neq P_2^C$  but  $P_1^B = P_2^B$ . If we have separation in prices, the aware consumer becomes perfectly informed about the state of the world after viewing the prices. A consumer who is unaware of the entrant only

observes the price of the incumbent:  $P_1^B$  in State 1, and  $P_2^B$  in State 2. Note that if in equilibrium  $P_1^B \neq P_2^B$ , the unaware consumer can infer that an entrant exists and, moreover, becomes perfectly informed about the quality of the entrant.

We show that no price signaling equilibrium can exist in our model. The reason is that prices in our model are not credible signals of quality. Thus, in the price signaling literature, the high quality type is able to separate herself from the ghost low quality type by charging a higher price. Since the low quality type bears higher losses in increasing the price, high price can serve as a credible signal of quality. Our model, on the other hand, lacks any of the elements that can facilitate price signaling. Thus, in a one-period model with homogeneity of preferences and zero marginal cost, both types can only benefit from a hike in price. Thus, no separation exists. The only additional difficulty in our model as opposed to most of the is that signaling is usually done in the context of a monopolist firm. Here, however, we must consider possible price signaling in a duopoly.

#### **Lemma 1**

There does not exist a price signaling equilibrium in the model above. Thus, there does not exist a separating equilibrium where the consumer observes a pair of prices  $[P_1^C, P_1^B]$  in State 1 and a pair of prices  $[P_2^C, P_2^B]$  in State 2, where  $[P_1^C, P_1^B] \neq [P_2^C, P_2^B]$ .

See Appendix for details of the proof. The example below illustrates the idea behind the proof. Consider a price-separating equilibrium where the consumer sees prices

$[P_1^C, P_1^B] = [V_H^C - V^B, 0]$  in State 1, and  $[P_2^C, P_2^B] = [0, V^B]$  in State 2. Note that these are the duopoly prices that would prevail under perfect information. Thus, we see that

the aware consumer buys C in State 1, and B in State 2. The unaware consumer only observes B's prices, but can nonetheless infer the state of the world and makes the same purchase decisions as the aware consumer since  $P_1^B \neq P_2^B$ .

Let us examine off-path beliefs that can support the equilibrium above. Note that to keep C from deviating in State 2, the consumer must infer that the state of the world is 2 after observing prices  $[V_H^C - V^B, V^B]$ . However, to keep B from deviating in State 1, the consumer must infer that the state of the world is 1 following observing prices  $[V_H^C - V^B, V^B]$ . Since it is impossible to maintain beliefs that satisfy both of these conditions, we arrive at a contradiction. Thus, a separating equilibrium in prices does not exist in our model.

### 5c. Mixing Equilibrium

Let us focus more intently on the result that the firms lie in equilibrium or spend more resources promoting an inferior product. Let us suppose that we are in a situation where the ex-ante probability that the entrant is superior is low,  $P(s1)$  is low, as is  $N^U$ . Thus, there are few unbiased reviews and the entrant is most certainly of inferior quality. Would that imply that in expectation C would invest lots of resources into disseminating false information?

This does not turn out to be the case. Note that when  $P(s1)$  is very low, it will be "hard"

to convince a consumer to buy C. More formally,  $\frac{P(\mu^{C>B} | \text{state} = 1)}{P(\mu^{C>B} | \text{state} = 2)} < \frac{P(s2)V^B}{P(s1)(V_H^C - V^B)}$ .

Since no pure strategy equilibrium exists, we can consider a mixing equilibrium.

Specifically, let us consider an equilibrium where the consumer mixes between buying C and B upon receiving  $\mu^{C>B}$ , and buys B otherwise. Thus, the consumer buys C with probability  $\delta$  upon receiving  $\mu^{C>B}$ . In our model, this is equivalent to multiplying the profit parameter by the mixing equilibrium  $= \rho\delta$ . In this equilibrium,

$$\frac{P(\mu^{C>B} | \text{state} = 1)}{P(\mu^{C>B} | \text{state} = 2)} = \frac{P(s2)V^B}{P(s1)(V_H^C - V^B)}. \text{ We can show that the expression}$$

$$\frac{P(\mu^{C>B} | \text{state} = 1)}{P(\mu^{C>B} | \text{state} = 2)}$$

is decreasing in the profit parameter. As we decrease  $P(s1)$ , the

expression on the right increases. This implies that the profit parameter must decrease, or  $\delta$  must decrease. This in turn implies a dampening of promotional activity by both B and C. This agrees with our intuition that we should not get inundated with false messages by the firm.

There are a number of extensions that can be pursued in the future:

- 1) We can further think about endogenizing some other parameters. One natural extension would be to endogenize the prior,  $P(s1)$ , parameter by postulating that there is a cost of entry. If we consider an extension where there are many types of quality entrants, we will get the result that the very low quality types will not enter since they will make very little profits.
- 2) Another natural candidate to be treated as an endogenous variable is the number of unbiased reviews. Thus, we might think that in a multi-period model firms can

influence the size of  $N^U$  by free trial. (Consider a recent news item where a movie studio refused to screen a movie for critics ostensibly not to reveal important plot points.)

- 3) In our model there is homogeneity of preferences. We can show that preference heterogeneity does not change the results as long as the niche segment is not too large.
- 4) We can look at another game form: a game form with advertising but no anonymity. Note that this is the setting that is equivalent to offline advertising. This is a signaling model where consumers interpret advertising as a signal of quality. Note that in our set-up, a firm has more incentive to advertise its inferior product. Thus,  $\mu^{C>B}$  message is actually an indicator of poor quality of C. However, if  $P(s1)$  is high enough, and as long as the signal is sufficiently imprecise, the consumers may still choose to buy C following  $\mu^{C>B}$ . Thus, this format would motivate C to advertise simply to increase product awareness in the context where consumers are positively pre-disposed to entrants. Note that the logic here is different from a classic signaling model since here advertising is an imprecise negative signal, but the firm still chooses to advertise to increase awareness.

Let us conclude with the following thought. What is the fundamental question that this paper addresses? We explore a new advertising context: a setting where advertising and word of mouth become perfect substitutes since to the consumer they appear indistinguishable. We find that in this context consumers still benefit from chat, despite

the fact that firms choose to lie. Moreover, the consumers may benefit from such a system compared to a regulated system.

## References

- 1) Anderson, E.; Simester, D. (2000) "Price Discrimination as an Adverse Signal: Why an Offer to Spread Payments May Hurt Demand," MIT working paper.
- 2) Avery, C.; Resnick, P.; Zeckhauser, R. (1999) "The Market for Evaluations," *American Economic Review*. 89 (June), 564-84.
- 3) Banerjee, A. (1992) "A Simple Model of Herd Behavior," *The Quarterly Journal of Economics*. 107 (3), 797-817.
- 4) Banerjee, A. (1993) "The Economics of Rumours," *The Review of Economic Studies*. 60 (2), 309-27.
- 5) Dixit, A.; Norman, V. (1978) "Advertising and Welfare," *The Bell Journal of Economics*. 9 (1), 1-17.
- 6) Dodson, J.; Muller, E. (1978) "Models of New Product Diffusion Through Advertising and Word-of-Mouth," *Management Science*. 24 (15), 1568-1578.
- 7) Ellison, G.; Fudenberg, D. (1995) "Word-of-Mouth Communication and Social Learning," *The Quarterly Journal of Economics*. 110 (1), 93-125.
- 8) Feinberg, F. (1992) "Pulsing Policies for Aggregate Advertising Models," *Marketing Science*. 11 (3), 221-234.
- 9) Horstmann, I.; Moorthy, S. (2000) "Advertising and Quality for Services," *Working paper*, University of Toronto.
- 10) Kihlstrom, R.; Riordan, M. (1984) "Advertising as a Signal," *Journal of Political Economy*. 92 (3), 427-450.
- 11) Kalra, A.; Rajiv, S., Srinivasan, K. (1998) "Response to Competitive Entry: A Rationale for Delayed Defensive Reaction," *Marketing Science*. 17 (4), 380-405.
- 12) Little, J. (1979) "Aggregate Advertising Models: The State of the Art," *Operations Research*. 27 (4), 629-667.
- 13) Little, J.; Lodish, L. (1969) "A Media Planning Calculus," *Operations Research*. 17 (1), 1-35.
- 14) Mahajan, V.; Muller, E., Kerin, R. (1984) "Introduction Strategy for New Products With Positive and Negative Word-of-Mouth," *Management Science*. 30 (12), 1389-1404.

- 15) Mahajan, V.; Muller, E. (1986) "Advertising Pulsing Policies For Generating Awareness for New Products," *Marketing Science*. 5 (2), 89-106.
- 16) McFadden, D.; Train, K. (1996) "Consumers' Evaluation of New Products: Learning from Self and Others," *Journal of Political Economy*. 104 (4), 683-703.
- 17) Monahan, G. (1984) "A Pure Birth Model of Optimal Advertising With Word-of-Mouth," *Marketing Science*. 1 (2), 169-178.
- 18) Nelson, P. (1974) "Advertising as Information," *Journal of Political Economy*. 81 (July/August), 729-754.
- 19) Simon, H. (1982) "ADPULS: An Advertising Model with Wearout and Pulsation," *Journal of Marketing Research*. 19 (August), 352-63.
- 20) Sasieni, M. (1971) "Optimal Advertising Expenditure," *Management Science*. 18 (4), 64-72.
- 21) Stegeman, M. (1991) "Advertising in Competitive Markets," *The Rand Journal of Economics*. 81 (1), 210-223.
- 22) Vettas, N. (1997) "On the Informational Role of Quantities: Durable Goods and Consumers' Word-of-Mouth Communication," *International Economic Review*. 38 (4), 915-944.
- 23) Vidale, M.; Wolfe, H. (1957) "An Operations-Research Study of Sales Response To Advertising," *Operations Research*. 5, 370-381.

## Appendix

### Proposition 1

Since the consumer's decision is binary (whether or not to buy the entrant based on a message), let us start with the four possible equilibrium strategies by the consumer. Note that the consumer buys B following  $\mu^B$  by assumption.

- 1) Suppose that the consumer buys the entrant's product following both  $\mu^{B>C}$  and  $\mu^{C>B}$ .  
This implies that the incumbent only makes profit when  $\mu^B$  message is drawn. But this would imply that the incumbent would not have any incentives to message, which in turn implies that  $\mu^{B>C}$  message is perfectly informative. (This is due to the fact that  $\mu^{B>C}$  message has to come from the unbiased source). Since  $\mu^{B>C}$  message is perfectly informative, the consumer should buy B following  $\mu^{B>C}$  which contradicts our initial assumption. This set of beliefs cannot be consistent.
- 2) Let us next suppose that the consumer buys the incumbent's product following both  $\mu^{B>C}$  and  $\mu^{C>B}$ . Similarly, this cannot be in equilibrium since the entrant would not message, and  $\mu^{C>B}$  becomes perfectly informative.
- 3) Consider an equilibrium where the consumer buys C following  $\mu^{B>C}$  and buys B following  $\mu^{C>B}$ . Here we get a corner solution since both the B & C would not want to message. (Note that here the assumption that firm B can only send  $\mu^{B>C}$  message & firm C can only send  $\mu^{C>B}$  message is important). However, since neither biased sources are messaging, the signals becomes perfectly informative which would imply that the consumer should buy B following  $\mu^{B>C}$  and should buy C following  $\mu^{C>B}$ . This is a contradiction of our earlier assumption.
- 4) Next, consider an equilibrium where the consumer buys B following  $\mu^{B>C}$  and buys C following  $\mu^{C>B}$ . This results in the following four maximization problems for the firms (where  $G_i^C$  is C's net profit function in State i and  $G_i^B$  is B's net profit function in State i):

$$\begin{aligned} \text{Entrant (C) in State 1: } \max_{N_1^C} G_1^C(N_1^C, N_1^B) &\equiv \max_{N_1^C} \frac{N_1^C + N^U}{N_1^C + N^U + N_1^B + N^0} \rho - \frac{(N_1^C)^2}{2} \\ \text{s.t. } N_1^C &\geq 0 \end{aligned} \quad (1)$$

$$\begin{aligned} \text{Incumbent (B) in State 1: } \max_{N_1^B} G_1^B(N_1^C, N_1^B) &\equiv \max_{N_1^B} \frac{N_1^B + N^0}{N_1^C + N^U + N_1^B + N^0} \rho - \frac{(N_1^B)^2}{2} \\ \text{s.t. } N_1^B &\geq 0 \end{aligned} \quad (2)$$

$$\begin{aligned} \text{Entrant (C) in State 2: } \max_{N_2^C} G_2^C(N_2^C, N_2^B) &\equiv \max_{N_2^C} \frac{N_2^C}{N_2^C + N^U + N_2^B + N^0} \rho - \frac{(N_2^C)^2}{2} \\ \text{s.t. } N_2^C &\geq 0 \end{aligned} \quad (3)$$

$$\begin{aligned} \text{Incumbent (B) in state 2: } \max_{N_2^B} G_2^B(N_2^C, N_2^B) &\equiv \max_{N_2^B} \frac{N^U + N_2^B + N^0}{N_2^C + N^U + N_2^B + N^0} \rho - \frac{(N_2^B)^2}{2} \\ \text{s.t. } N_2^B &\geq 0 \end{aligned} \quad (4)$$

Note that we constrain the firms' actions to be positive. This could introduce the complications of corner solutions. However, it turns out that we can show that if  $N_1^B \geq 0$ , C would choose  $N_1^C \geq 0$  and vice versa. (We can show this by looking at the firms' reaction functions). Thus, we do not need to worry about corner solutions in the relevant region. In addition, we can show that

$$\begin{aligned} \frac{\partial^2 G_1^C}{\partial^2 N_1^C} &= -\frac{2N_1^B}{(N_1^C + N_1^B + N^U + N^0)^3} \rho - 1 < 0; \quad \frac{\partial^2 G_1^B}{\partial^2 N_1^B} = -\frac{2(N_1^C + N^U)}{(N_1^C + N_1^B + N^U + N^0)^3} \rho - 1 < 0 \\ \frac{\partial^2 G_2^C}{\partial^2 N_2^C} &= -\frac{2(N_2^B + N^U + N^0)}{(N_2^C + N_2^B + N^U + N^0)^3} \rho - 1 < 0; \quad \frac{\partial^2 G_2^B}{\partial^2 N_2^B} = -\frac{2N_2^C}{(N_2^C + N_2^B + N^U + N^0)^4} \rho - 1 < 0 \end{aligned}$$

This insures a unique maximum in the relevant region. Hence, our strategy for obtaining the solution is to solve using FOCs and ignore the constraint. We later discard any negative solutions as infeasible.

The four resulting FOCs are listed below:

$$\text{Entrant (C) in State 1: } (N_1^B + N^0)\rho = N_1^C(N_1^B + N_1^C + N^U + N^0)^2 \quad (5)$$

$$\text{Incumbent (B) in State 1: } (N_1^C + N^U)\rho = N_1^B(N_1^B + N_1^C + N^U + N^0)^2 \quad (6)$$

$$\text{Entrant (C) in State 2: } (N_2^B + N^U + N^0)\rho = N_2^C(N_2^B + N_2^C + N^U + N^0)^2 \quad (7)$$

$$\text{Incumbent (B) in State 2: } N_2^C\rho = N_2^B(N_2^B + N_2^C + N^U + N^0)^2 \quad (8)$$

We next show that there exists a unique positive solution to the equations (5) & (6). If we add (5) & (6) and simplify, we get the expression  $\rho = (N_1^B + N_1^C)(N_1^B + N_1^C + N^U + N^0)$  (9)

$$\text{From this we can solve to get, } N_1^B + N_1^C = \frac{-(N^U + N^0) + \sqrt{(N^U + N^0)^2 + 4\rho}}{2} = w > 0 \quad (10)$$

This is the only positive solution. On the other hand, if we divide (5) by (6) (we can do this since we can see that  $N_1^B = N_1^C = 0$  is not a solution to the equation), we get

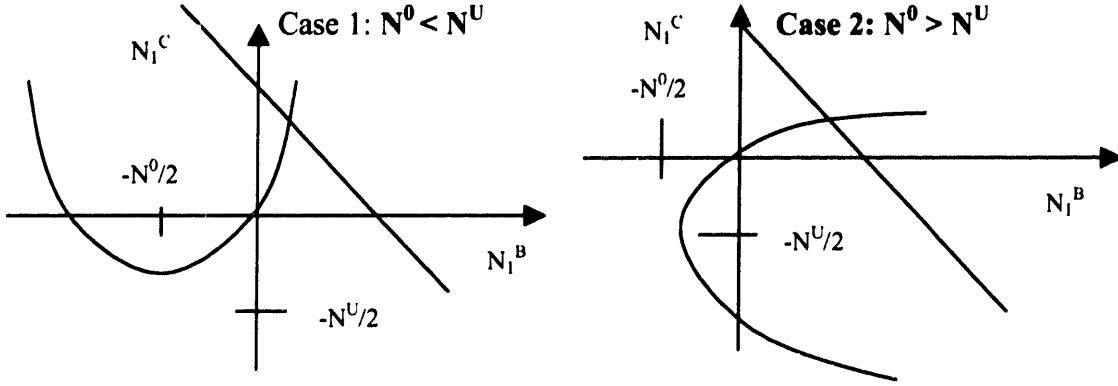
$$N_1^B(N_1^B + N^0) = N_1^C(N_1^C + N^U) \quad (11)$$

Note that any solution that satisfies (5) & (6) must also satisfy (10) & (11). Thus, we are not losing any solutions. Graphically, we see that (10) describes a line on the  $(N_1^B, N_1^C)$  plane, whereas (11) defines a hyperbola. (To see this more clearly, we can re-write (11)

$$\text{as } \frac{(N_1^B - (-N^0/2))^2}{(N^0)^2 - (N^U)^2} - \frac{(N_1^C - (-N^U/2))^2}{(N^0)^2 - (N^U)^2} = 1) \text{ The intersection of (10) \& (11) give us the}$$

equilibrium values. The hyperbola can have two orientations depending on the relative size of  $N^U$  and  $N^0$  parameters. From the graphs below, we see that there exists a unique positive solution:

**Figure 1A: Existence and uniqueness of the solution**



This is guaranteed since the hyperbola goes through the origin & the line

$N_1^B + N_1^C = w > 0$  has a positive intercept term. Similarly, we can show uniqueness and existence for  $(N_2^B, N_2^C)$ .

The explicit solutions are listed below.

$$N_1^B = \frac{1}{2} \left[ \frac{2\rho + (N^0)^2 + N^U N^0 - N^0 \sqrt{(N^U + N^0)^2 + 4\rho}}{\sqrt{(N^U + N^0)^2 + 4\rho}} \right]; \quad (12)$$

$$N_1^C = \frac{1}{2} \left[ \frac{2\rho + (N^U)^2 + N^U N^0 - N^U \sqrt{(N^U + N^0)^2 + 4\rho}}{\sqrt{(N^U + N^0)^2 + 4\rho}} \right]; \quad (13)$$

$$N_2^B = \frac{1}{2} \left[ \frac{2\rho + (N^U + N^0)^2 - (N^0 + N^U) \sqrt{(N^U + N^0)^2 + 4\rho}}{\sqrt{(N^U + N^0)^2 + 4\rho}} \right]; \quad (14)$$

$$N_2^C = \left[ \frac{\rho}{\sqrt{(N^U + N^0)^2 + 4\rho}} \right] \quad (15)$$

All the expressions above are greater than 0. Note the following two results:

$$1) \quad N_2^C - N_1^C = \frac{1}{2} N^U \left[ 1 - \frac{N^U + N^0}{\sqrt{(N^U + N^0)^2 + 4\rho}} \right] = \Delta \quad \text{Note that } 0 < \Delta < \frac{N^U}{2}. \quad \text{Thus, the}$$

entrant (C) spends more on promotion when his product is inferior, but

$N_2^C < N_1^C + N^U$ . That is, the consumer is more likely to hear  $\mu^{C>B}$  when the entrant is superior.

$$2) \quad N_1^B - N_2^B = \frac{1}{2} N^U \left[ 1 - \frac{N^U + N^0}{\sqrt{(N^U + N^0)^2 + 4\rho}} \right] = \Delta. \text{ Similarly, the incumbent (B) spends}$$

more on promotion when his product is inferior, but  $N_1^B < N_2^B + N^U$ . That is, the consumer is more likely to hear  $\mu^{B>C}$  when the incumbent is superior. From this we clearly see that the firm makes more profit when its product is superior since then it spends less money on promotion and is more likely to sell.

Next, let us turn back to consumer's problem and check that his decision rule is optimal, given the firms' actions. As we previously discussed, the consumer will optimally follow the recommendation system iff

$$\frac{P(\mu^{C>B} | s1)}{P(\mu^{C>B} | s2)} \geq \frac{P(s2)V^B}{P(s1)(V_H^C - V^B)} \quad \& \quad \frac{P(\mu^{B>C} | s1)}{P(\mu^{B>C} | s2)} < \frac{P(s2)V^B}{P(s1)(V_H^C - V^B)} \quad (7)$$

(Note that if the inequalities in (7) do not hold, the firm's action will not affect the consumer's decisions. This contradicts our initial assumption and is not an equilibrium).

Let us substitute the conditional probabilities that we obtain above. Here we see that

$$\frac{P(\mu^{C>B} | s1)}{P(\mu^{C>B} | s2)} = \frac{N_1^C + N^U}{N_2^C} \quad (16)$$

$$\frac{P(\mu^{B>C} | s1)}{P(\mu^{B>C} | s2)} = \frac{N_1^B}{N_2^B + N^U} \quad (17)$$

(We use the result that  $N_1^B + N_1^C = N_2^B + N_2^C$  to simplify the fraction above). Note that since we showed that  $N_1^B < N_2^B + N^U$  &  $N_2^C < N_1^C + N^U \rightarrow (16) > (17)$ . We can also show that

$$\frac{\partial \left[ \frac{N_1^C + N^U}{N_2^C} \right]}{\partial \rho} < 0, \text{ while } \frac{\partial \left[ \frac{N_1^B}{(N_2^B + N^U)} \right]}{\partial \rho} > 0. \text{ Thus, as long as } 0 < P(S1) < 1 \text{ \& } V_H^C > V^B,$$

we can find  $\hat{\rho}$  s.t.  $\forall \rho \leq \hat{\rho}$ , (7) is guaranteed to hold. Thus, we showed the existence of a unique equilibrium. QED.

## Proposition 2: Linear Costs

Steps 1-3 of the proof of the proposition do not depend on the convexity of costs and can be repeated here. Let us here explore the remaining possible equilibrium where the consumer buys C following  $[C > B]$  and buys B following  $[B > C]$ . The firms' maximizations are:

$$\text{Entrant (C) in state 1: } \max_{N_1^C} \frac{N_1^C + N^U}{N_1^C + N^U + N_1^B + N^0} \rho - N_1^C \quad (18)$$

$$\text{Incumbent (B) in state 1: } \max_{N_1^B} \frac{N_1^B + N^0}{N_1^C + N^U + N_1^B + N^0} \rho - N_1^B \quad (19)$$

$$\text{Entrant (C) in state 2: } \max_{N_2^C} \frac{N_2^C}{N_2^C + N^U + N_2^B + N^0} \rho - N_2^C \quad (20)$$

$$\text{Incumbent (B) in state 2: } \max_{N_2^B} \frac{N^U + N_2^B + N^0}{N_2^C + N^U + N_2^B + N^0} \rho - N_2^B \quad (21)$$

Of course, once again we impose the constraint that none of the firms' actions are negative. If we ignore the constraint, then we can obtain the firms' reaction functions from the maximizations above. The reaction functions are the following:

$$N_1^C = \sqrt{(N_1^B + N^0)\rho} - N_1^B - N^U - N^0 \quad (22)$$

$$N_1^B = \sqrt{(N_1^C + N^0)\rho} - N_1^C - N^U - N^0 \quad (23)$$

$$N_2^C = \sqrt{(N_2^B + N^U + N^0)\rho} - N_2^B - N^U - N^0 \quad (24)$$

$$N_2^B = \sqrt{N_2^C \rho} - N_2^C - N^U - N^0 \quad (25)$$

The above equations have one (possibly) positive solution:

$$\{ N_1^C = \frac{\rho}{4} - N^U; N_1^B = \frac{\rho}{4} - N^0; N_2^C = \frac{\rho}{4}; N_2^B = \frac{\rho}{4} - N^0 - N^U \}. \text{ We can see that this is a}$$

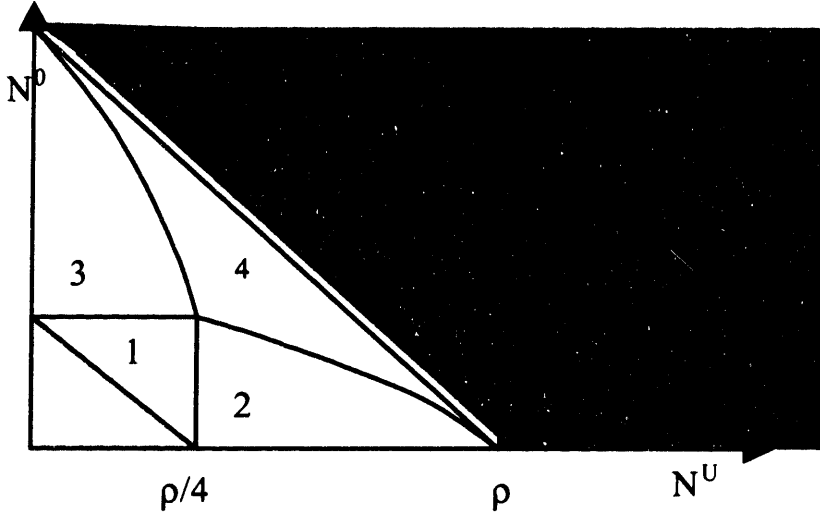
positive solution when  $N_2^B = \frac{\rho}{4} - N^0 - N^U > 0$ . Note that this is the case when  $\rho$  is

relatively large. This is analogous to stating that the costs are relatively low. We can see that in this region there is no informative equilibrium since

$N_1^C + N^U = N_2^C; N_1^B = N_2^B + N^U$ . Thus, the content of the message has no informational value since the message is equally likely to have arrived from either of the two states of

the world. In the language used earlier, we consider this a low cost region. In the Figure below, this region is the triangle in the lower left corner of the graph.

**Figure 2A: Linear Costs**



Next, let us consider the high cost region. In the region where  $N^U + N^0 > \rho$ , all firms choose to send 0 messages in both states of the world. (This is the region on the right side of the Figure).

In the intermediate cost region,  $\frac{\rho}{4} < N^U + N^0 < \rho$ , we have the following 4 regions:

Region 1:  $\frac{\rho}{4} > N^U; \frac{\rho}{4} > N^0$

$$N_2^C = \sqrt{(N^U + N^0)\rho} - (N^U + N^0) < \frac{\rho}{4}; N_2^B = 0; N_1^C = \frac{\rho}{4} - N^U; N_1^B = \frac{\rho}{4} - N^0$$

We can show that  $N_2^C > N_1^C$  ( $\frac{\rho}{4} < N^U + N^0$  and  $\frac{\rho}{4} > N^0 \rightarrow N_2^C > \frac{\rho}{4} - N^U = N_1^C$ ) Thus, we can see that here the incumbent only spends on promotion when her product is of inferior quality, as does the entrant. Moreover, we can see that  $N_2^C < N_1^C + N^U$  and

$N_1^B < N_1^B + N^U$  (since  $\frac{\rho}{4} < N^U + N^0$ ) Thus, the messages are potentially informative.

Region 2:  $\frac{\rho}{4} < N^U; \frac{\rho}{4} > N^0; \sqrt{N^U\rho} - (N^U + N^0) > 0$

$$N_2^C = \sqrt{(N^U + N^0)\rho} - (N^U + N^0) < \frac{\rho}{4}; N_2^B = 0; N_1^C = 0; N_1^B = \sqrt{N^U\rho} - (N^U + N^0)$$

Here again we see that the firms lie. We can also see that there is potential for

informativeness since  $N_2^C < \frac{\rho}{4} < N^U = N_1^C + N^U$ . We also see that

$$N_1^B < N_2^C < N^U = N_2^B + N^U$$

**Region 3:**  $\frac{\rho}{4} > N^U; \frac{\rho}{4} < N^0; \sqrt{N^0\rho} - (N^U + N^0) > 0$

$$N_2^C = \sqrt{(N^U + N^0)\rho} - (N^U + N^0) < \frac{\rho}{4}; N_2^B = 0; N_1^C = \sqrt{N^0\rho} - (N^U + N^0); N_1^B = 0$$

Here again we see that the entrant lies. We can easily see that here  $N_1^B < N_2^B + N^U$ . We

can also show that here  $N_2^C < N_1^C + N^U$ . We need to show that

$$\sqrt{(N^U + N^0)\rho} - \sqrt{N^0\rho} < N^U \rightarrow f = \sqrt{(N^U + N^0)\rho} - \sqrt{N^0\rho} - N^U < 0. \text{ Note that}$$

$$\frac{\partial f}{\partial N^U} = \frac{\sqrt{\rho}}{2\sqrt{(N^U + N^0)}} - 1 < 0 \text{ since } N^U + N^0 > \frac{\rho}{4}. \text{ Also note that}$$

$$\frac{\partial f}{\partial N^0} = \frac{\sqrt{\rho}}{2\sqrt{(N^U + N^0)}} - \frac{\sqrt{\rho}}{2\sqrt{N^0}} < 0. \text{ Thus, } f \text{ decreases as } N^0 \text{ and } N^U \text{ increase. Let's pick}$$

the smallest  $N^U$  and  $N^0$  in the region:  $N^0 = \frac{\rho}{4}, N^U = 0$ . At this point,  $f=0$ . This is the

maximum of  $f$ . Thus, we can see that in the relevant region,  $f < 0$ . This demonstrates informativeness.

**Region 4:**  $\sqrt{N^0\rho} - (N^U + N^0) < 0; \sqrt{N^U\rho} - (N^U + N^0) < 0$

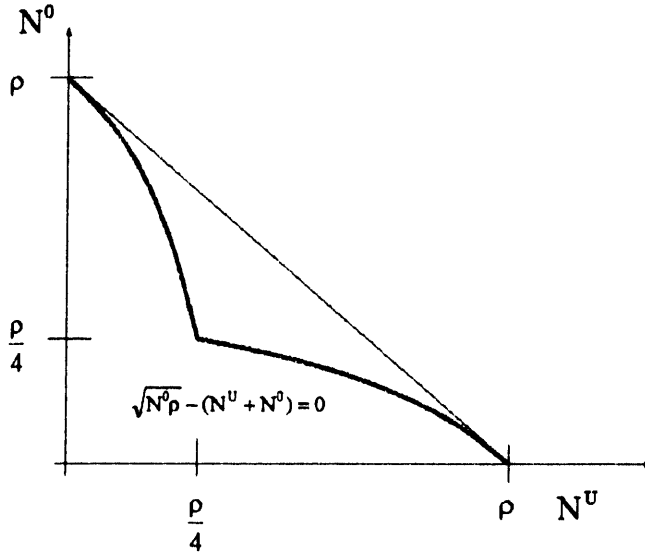
$$N_2^C = \sqrt{(N^U + N^0)\rho} - (N^U + N^0); N_2^B = N_1^B = N_1^C = 0$$

Here again we see that the entrant lies. We can easily see that here  $N_1^B < N_2^B + N^U$ . We

can also show that  $N_2^C < N_1^C + N^U = N^U$ . Similarly to the proof above, we need to show

that  $g = \sqrt{(N^U + N^0)\rho} - 2N^U - N^0 < 0$ . The only difficulty here lies in the irregularity of region 4. It is graphed below in more detail:

**Figure 3A: Region 4**



Our strategy is to find the maximum value of  $g$  at the region above and show that it is 0.

We can show that  $\frac{\partial g}{\partial N^0} = \frac{\sqrt{\rho}}{2\sqrt{(N^U + N^0)}} - 1 < 0$  since  $N^U + N^0 > \frac{\rho}{4}$ . Thus, for any point

in the region 4, there exists a point on the border between A and C or A and B s.t. the point on the border results in a higher value of  $g$ . We thus can look for the maximum point on that border.

Let us next find the maximum value of  $g$  on that border curve. Let us first focus on the border between A and C. The equation for that curve is  $\sqrt{N^0 \rho} - (N^U + N^0) = 0$ . Let us solve for  $N^U$  in terms of  $N^0$ :  $N^U = \sqrt{N^0 \rho} - N^0$  and plug this into  $g$  and take a derivative w.r.t.  $N^0$ :

$$\frac{\partial g / \text{border}}{\partial N^0} = \frac{3}{4} \sqrt{\frac{\rho}{N^0}} - 2 \sqrt{\frac{\rho}{N^0}} + 4$$

Since  $\frac{\rho}{4} < N^0 < \rho$ , we see that the expression above  $> 0$ . Thus, the  $g$  function is

increasing in  $N^0$  on the border. The maximum point is, therefore at  $N^0 = \rho, N^U = 0$ . Of course, at this point  $g = 0$ .

Next, let us consider the curve between A and B. Once again, the equation for that curve is  $\sqrt{N^U \rho} - (N^U + N^0) = 0$ . We can solve for N in terms of  $N^U$ :  $N^0 = \sqrt{N^U \rho} - N^U$ .

$N^U = \sqrt{N^0 \rho} - N^0$ . Let us plug this into g and take a derivative w.r.t.  $N^U$ :

$$\frac{\partial g / \text{border}}{\partial N^U} = \frac{3}{4} \sqrt{\frac{\rho}{N^U}} - 2 \sqrt{\frac{\rho}{N^U}} - 4 < 0 \text{ since } \frac{\rho}{4} < N^U < \rho. \text{ Thus, } g \text{ is decreasing in } N^U,$$

and the maximum of g is at  $N^0 = \frac{\rho}{4}, N^U = \frac{\rho}{4}$ ; g at that point is  $< 0$ . QED.

### Proposition 3

Let us compare the customer's expected value between the two game forms: the basic model (which we denote by *BM*) versus a game form where no advertising by either firm is allowed (which we denote by *NA*).

$$\text{Value}^{BM} = P(s1)(P^{BM}(\mu^{C>B} | s1)V_H^C + (1 - P^{BM}(\mu^{C>B} | s1))V^B) + P(s2)(1 - P^{BM}(\mu^{C>B} | s1))V^B - P \quad (26)$$

$$\text{Value}^{NA} = P(s1)(P^{NA}(\mu^{C>B} / s1)V_H^C + (1 - P^{NA}(\mu^{C>B} / s1))V^B) + P(s2)V^B - P \quad (27)$$

Note that in the basic model, in equilibrium the consumer can hear signal  $\mu^{C>B}$  in either state of the world. Of course, hearing this signal in state 2 will cause the consumer to buy C and derive no value from the purchase. On the other hand, in the system with no advertising, the consumer will never hear  $\mu^{C>B}$  in state 2. On the other hand, the consumer may also be less likely to hear signal  $\mu^{C>B}$  in state 1 in the system with no advertising. (Since C will not be promoting its product in state 1). Thus, it is the tradeoff between these two forces that determines which system is better.

$$\frac{P^{BM}(\mu^{C>B} / s1) - P^{NA}(\mu^{C>B} / s1)}{P^{BM}(\mu^{C>B} / s2)} \geq \frac{P(s2)V^B}{P(s1)(V_H^C - V^B)} \quad (28)$$

In order to simplify the expression above, we need to substitute for the expressions below:

$$P^{BM}(\mu^{C>B} | s1) = \frac{N_1^C + N^U}{N_1^C + N_1^B + N^U + N^0} \quad (29)$$

$$P^{BM}(\mu^{C>B} | s2) = \frac{N_2^C}{N_2^C + N_2^B + N^U + N^0} \quad (30)$$

$$P^{NA}(\mu^{C>B} | s2) = \frac{N^U}{N^U + N^0} \quad (31)$$

After simplification, we obtain the following expression for the left-hand side of (28):

$$\frac{P^{BM}(\mu^{C>B} | s1) - P^{NA}(\mu^{C>B} | s1)}{P^{BM}(\mu^{C>B} | s2)} = \frac{N^0 N_1^C - N^U N_1^B}{(N^U + N^0) N_2^C} = \frac{N^0 - N^U}{N^U + N^0}$$

Thus, as long as  $\frac{N^0 - N^U}{N^U + N^0} > \frac{P(s2)V^B}{P(s1)(V_H^C - V^B)}$ , the consumer expects to be better off in

the game where advertising is allowed. Note that  $\frac{N^0 - N^U}{N^U + N^0}$  measures the signal-to-noise ratio in the model with no advertising.

### Alternative Specifications

*Symmetric Specification: "Seeding"*

Consider the "seeding" specification where  $T_1^C = \kappa N_1^C + N^U$ ,  $T_2^C = N_2^C$ ,  $T_1^B = N_1^B$ , and  $T_2^B = \kappa N_2^B + N^U$ . Parts 1-3 of proof of Proposition 1 still hold. The only candidate for equilibrium is one where the consumer buys C following  $\mu^{C>B}$  and B following  $\mu^{B>C}$ .

Thus, we can derive the probabilities of getting either of the messages:

$$\begin{aligned} P(\mu^{C>B} | s1) &\equiv f(T_1^C, T_1^B) = \frac{T_1^C}{T_1^C + T_1^B}; \quad P(\mu^{C>B} | s2) \equiv f(T_2^C, T_2^B) = \frac{T_2^C}{T_2^C + T_2^B} \\ P(\mu^{B>C} | s1) &\equiv f(T_1^C, T_1^B) = \frac{T_1^B}{T_1^C + T_1^B}; \quad P(\mu^{B>C} | s2) \equiv f(T_2^C, T_2^B) = \frac{T_2^B}{T_2^C + T_2^B} \end{aligned} \quad (32)$$

The maximization problems for the two players in state 1 (after a change of variables) are:

$$\max_{T_1^C} G_1(T_1^C; T_1^B) = \max_{T_1^C} \frac{T_1^C}{T_1^C + T_1^B} \rho - \frac{(T_1^C - N^U)^2}{2k}$$

$$\text{s.t. } T_1^C \geq N^U$$

$$\max_{T_1^B} F_1(T_1^B; T_1^C) = \max_{T_1^B} \frac{T_1^B}{T_1^C + T_1^B} \rho - \frac{(T_1^B)^2}{2}$$

$$\text{s.t. } T_1^B \geq 0$$

Note that  $\frac{\partial^2 G_1}{\partial^2 T_1^C} = -2T_1^B / (T_1^C + T_1^B)^3 - \frac{1}{\rho k^2} < 0$  as long as  $T_1^B \geq 0$ . (The function is

strictly concave in the relevant region). Similarly,  $\frac{\partial^2 F_1}{\partial^2 T_1^B} < 0$  as long as  $T_1^C \geq 0$ .

In order to demonstrate the existence and uniqueness of the solution, we need to graph the best response functions of both players. Due to the concavity result, we can

$$\frac{(T_1^C - T_1^B) / (T_1^C + T_1^B)^4}{2(T_1^C + T_1^B)T_1^B / (T_1^C + T_1^B)^4 + \frac{1}{\rho k^2}} < 0$$

Thus, once again we turn our attention to the FOCs. (We later show that there are no corner solutions and that  $\frac{\partial^2 G_1^C}{\partial^2 N_1^C} < 0$  in the relevant range).

Due to the symmetry of the specification, we can see that  $N_1^C = N_2^B$ ,  $N_1^B = N_2^C$ . Thus, we can consider the FOCs for State 1 only. We express the results in terms of  $T_1^C$  and  $T_1^B$ :

$$\text{From C's maximization: } k\rho \frac{T_1^B}{(T_1^C + T_1^B)^2} = \left( \frac{T_1^C - N^U}{k} \right) \quad (33)$$

$$\text{From B's maximization: } \rho \frac{T_1^C}{(T_1^C + T_1^B)^2} = T_1^B \quad (34)$$

If we assume that  $N^U > 0$ , this implies that  $T_1^C > 0$ . From (34) we can see that this would imply that  $T_1^B > 0$ . Since  $T_1^B > 0$ , according to (33),  $N_1^C = \left( \frac{T_1^C - N^U}{k} \right) > 0$ .

If we divide (33) by (34) (since  $T_1^B > 0$ , we are dividing by a non-zero expression), we obtain the expression

$$k^2(T_1^B)^2 = T_1^C(T_1^C - N^U) \quad (35)$$

We add  $\frac{(N^U)^2}{2}$  to both sides of (35) to complete the square. After we substitute the expressions for  $T_1^B$  and  $T_1^C$ , we get

$$(N_2^C)^2 = (N_1^C)^2 + \left(\frac{N^U}{k}\right)N_1^C \quad (36)$$

From this, we can see that  $N_2^C > N_1^C$  as long as  $N^U > 0$ . Also,  $N_2^C - N_1^C$  decreases as  $k$  increases, and  $N_2^C - N_1^C$  increases as  $N^U$  increases. Note also that  $\lim_{k \rightarrow \infty} N_2^C - N_1^C = 0$ .

(Same results hold for B by symmetry).

We can similarly demonstrate uniqueness of the solution.

### **Lemma 1**

The following is a proof by contradiction. Thus, suppose that there exists a separating equilibrium in prices: aware consumer observes  $[P_1^C, P_1^B]$  in state 1 and  $[P_2^C, P_2^B]$  in state 2, where  $[P_1^C, P_1^B] \neq [P_2^C, P_2^B]$ . The unaware consumer observes  $[P_1^B]$  in state 1 and  $[P_2^B]$  in state 2. Let us assume that the unaware consumer can make an inference on the state of the world iff  $P_2^B \neq P_1^B$ .

Let us next consider the profits that the two firms make in the two states of the world:  $\{\Pi_1^C, \Pi_1^B\}$  and  $\{\Pi_2^C, \Pi_2^B\}$ . Since C can deliver no value to the consumer in state 2,  $\Pi_2^C = 0$ . Let us apply the Pareto optimality criterion and only consider equilibria where B is bought in state 2. (Thus, we are ruling out equilibria where B charges such a high price in state 2 that the consumer buys neither B nor C. Such equilibria are Pareto dominated by an equilibrium where  $P_2^B \leq V^B$  since both the consumer and B are better

off, while C is indifferent). Thus, we have  $\{\Pi_2^C, \Pi_2^B\} = \{0, P_2^B\}$ . Similarly, let us rule out equilibria where neither good is bought in state 1 since these equilibria are Pareto dominated by those where at least one good is bought in state 1.

1) Let us suppose that an aware consumer buys C in state 1. This implies that

$V^B - P_1^B < V_H^C - P_1^C$  since buying B always remains a choice for the consumer. Let's call the fraction of unaware consumers in state 1,  $\alpha$ . Note that the consumer only buys B if  $P_1^B = P_2^B = P^B$ . Otherwise, he becomes aware of the entrant, learns the state of the world and chooses to buy C. (Note that under no beliefs would B be better off by charging  $P_1^B > V^B$  in state 1 since even under most favorable beliefs this results in zero profit for B). Thus, if B pools on prices and charges  $P_1^B = P_2^B$ , the payoffs are  $\{\Pi_1^C, \Pi_1^B\} = \{(1 - \alpha)P_1^C, \alpha P^B\}$ ,  $\{\Pi_2^C, \Pi_2^B\} = \{0, P^B\}$ . If B separates on price and charges  $P_1^B \neq P_2^B$ , the payoffs are  $\{\Pi_1^C, \Pi_1^B\} = \{P_1^C, 0\}$ ,  $\{\Pi_2^C, \Pi_2^B\} = \{0, P_2^B\}$ . We can see that in both cases, B is weakly better off in state 2, while C is weakly better off in state 1.

Suppose that B pools on prices:  $P_1^B = P_2^B = P^B$ . Note that there are no beliefs that support separation since C prefers to charge  $P_1^C$  in both states of the world and lead the consumer to infer that the true state of the world is state 1. Thus, we arrive at a contradiction.

Suppose that B does not pool on price:  $P_1^B \neq P_2^B$ . Note that since C weakly makes more profits in state 1, in order to support the equilibrium the consumer must believe that the **state of the world is 2** when encountering the off-path pair of prices  $[P_1^C, P_2^B]$ . Otherwise, C would charge  $P_1^C$  in state 2. To keep B from deviating and charging  $P_2^B$  in state 1, since B weakly makes more profit in state 2, the consumer must believe that **state 1** is in effect upon seeing  $[P_1^C, P_2^B]$ . Thus, we arrive at a contradiction.

- 2) Let us suppose that an aware consumer buys B in state 1. Thus C is never bought. This implies that B sets  $P_1^B = P_2^B = V^B$ . (Note that under no beliefs is B better off by charging  $P_2^B > V^B$  in state 2 since even under most favorable beliefs this results in zero profit for B. Similarly, B can always improve his payoff by raising his price). The only way to obtain a separation here is by having C signal the state of the world. Since C makes 0 in both states of the world, let us assume that it prefers to pool on price. Thus, we see that no separating equilibrium can exist.

## Essay 2: Using Online Conversations to Measure Word of Mouth Communication

### 1. Introduction

There is a long-held belief in marketing that word of mouth recommendations have tremendous influence on the sales of new products. Kotler (2000) emphasizes the importance of the personal communication channel in building a business: in a study of 7,000 consumers in seven European countries, 60 percent said they were influenced to buy a new brand by family and friends. Bass (1969) also points out the importance of word of mouth as a driver of new product diffusion. This belief is supported by findings from survey data, where consumers routinely cite word of mouth recommendations as an important influence on product choice.

While the belief that word-of-mouth is “important” is nearly universal, it has nonetheless been difficult to quantify its impact precisely since the information is exchanged in private conversations. As a result, marketers have either relied on consumer recall or have inferred the process of information exchange from aggregate diffusion data. One fascinating and important implication of the rise of online communities is that this development makes feasible the observation by marketers of consumer-to-consumer conversations. In this paper, we investigate the potential that these conversations present as an opportunity to measure word of mouth. Specifically, our task here is two-fold. We ask what measures of word-of-mouth communication are predictive of future product sales and apply these measures to the question of how the impact of word-of-mouth changes over time.

The first question is of both theoretical and managerial interest. At its core, a “conversation” is comprised of a set of statements between people. Our challenge is to somehow transform these

statements into a shape that the firm can use. Which of these transformations are meaningful is a theoretical question; a question that has not to our knowledge been addressed in the marketing literature. From a managerial perspective, firms are clearly interested in on-line conversations due to their potential to offer a cost effective window into people's conversations. *Which* measures offer the clearest window, however, is not known. Currently, the most-common approach is to use simple counts. This is akin to the typical news clipping services in which firms monitor how many times their products are "mentioned." For example, Yahoo! Buzz Index keeps track of how many times users query a particular topic on the Yahoo! Search engine. We will compare this naïve measure to other measures of word of mouth that are suggested by the extant literature and test the ability of each to predict future product sales. We are particularly interested in the predictive power of measures of the *dispersion of conversations across different communities*. Based on the sociological research on weak vs. strong ties (Granovetter (1973)), we hypothesize that conversations that are focused within a more narrow audience are likely to have less of an impact than those that occur across a wider spectrum of communities.

The second question, the change in the impact of word-of-mouth communication over time, is also of interest to both researchers and practitioners. From a theoretical perspective, we view this inquiry as a first step toward understanding the motivation behind word-of-mouth communications. We hypothesize that there are predominantly two motivations to having a conversation: (i) to inform the other person about something or to reduce their uncertainty, or (ii) to derive consumption value from the conversation itself. The two motivations differ in several important ways. First, since there is little awareness of the product's existence or quality

early in its life, we'd expect that the amount of information-motivated conversations would reach a peak early on and decline over time. On the other hand, there is no reason to believe that the pleasure associated with discussing a product would similarly decrease over time. In fact, since the number of previous or current users of the product is non-decreasing over time, we might even expect that the number of consumption value-motivated conversations to increase. Either way, these two combined suggest that the proportion of information-motivated discussions, as compared with the consumption value-motivated discussion, should decrease over time. A second important difference is that information-motivated discussions (e.g. "Did you hear about a cool new show?") are more likely to involve non-customers while consumption-value-motivated discussions ("Do you think that Max is an alien?") are more likely to involve current customers. Thus, we'd expect that the former has a bigger impact on future sales. Combining these two ideas, our theory would predict that the *impact of today's word-of-mouth on tomorrow's sales should decrease over time*. We will test this theory using the measures developed above.

The practice of "buzz marketing" is not significantly different from traditional PR in which the firm attempts to borrow the credibility of a communications medium. In the traditional case, the medium is a newspaper, for example, that might publish a story culled from a company press release. In buzz marketing, the medium is a consumer passing on information she may have heard either from another consumer or from a person hired by the company. Marketers have specifically targeted online communities since they act as influential sources of information<sup>1</sup> while providing anonymity to their participants. As shown by Mayzlin (2001), this process of

---

<sup>1</sup> In a 1999 Forrester survey of online communities' users, more than half of respondents to a 1999 survey said that they found online comments in communities to be at least "Important" in their online buying decision.

“seeding” word-of-mouth communications by posing as a customer can be beneficial to some firms even though consumers know that they are potentially being duped by a company shill. This study would inform managers’ ability to formulate optimal marketing strategies of this type including the timing and focus of the campaign.

The product category we have selected for our inquiry is new TV shows during the 1999/2000 seasons. Word of mouth appears to be especially important for entertainment goods: a recent Forrester report concludes that approximately 50% of young Net surfers rely on word of mouth recommendations to purchase CDs, movies, Videos/DVDs and games.<sup>2</sup> Thus, television is a natural candidate to use as a laboratory for testing the dynamic nature of word-of-mouth. In addition, the “purchase” of a TV show is a repeat purchase. This is interesting in this context because one’s consumption experience in period  $t$  will affect not only her decision to talk about it but also her consumption experience in period  $t+1$ . As a result, it is likely that people will talk about TV shows they’re watching for both informational *and* consumption value reasons. On the other hand, it seems less likely that the pure consumption value motivation for word-of-mouth is a significant factor in durable goods category (movies, for example). Our source of word-of-mouth information is drawn from Usenet: a collection of thousands of newsgroups with very diverse topics. (For more information on the history and the structure of Usenet, see Appendix). We do not argue that Usenet on its own is a very influential source of opinions. Instead, we claim that this particular collection of newsgroups presents us with a convenient measure of over-all word-of-mouth activity.

---

<sup>2</sup> Source: (Forrester Report (2000)).

1

It is important to note that we confine ourselves in this study to very simple measures of word of mouth activity. Specifically, we avoid here any content analysis of the posts. We are very comfortable limiting ourselves in this way for several reasons. First, it is much less costly to obtain count measures. Second, it is our belief that this approach introduces noise into our measures which would bias the effects of interest towards insignificance. Third, the simple measures proposed by us are easily replicable in other settings.

Among our main findings is the somewhat surprising insight that the simple “counts” that are so prevalent in practice are not necessarily informative. While many firms both on- and off-line pay a great deal of attention to the “number of mentions” that they or their products have garnered in the press, their attention may be better spent on other measures. One measure that we find to be particularly powerful is the *entropy* of posts across newsgroups. This is simply a measure of the dispersion of conversations across communities. Specifically, we find that product sales tend to be *increasing* in entropy. This is consistent with one of the fundamentals of the theory of weak ties: that information moves quickly within communities but slowly across. Finally, we find that these insights do not necessarily hold throughout a product’s life. In the case of these new television shows, we find that after the first 7 or 8 episodes, our measures of word-of-mouth add nothing to the firm’s ability to predict sales. We interpret this as meaning that word-of-mouth is effective at increasing awareness of a new product and at decreasing uncertainty about its quality. After a certain time, awareness is sufficiently high and quality known with sufficient certainty that word-of-mouth is no longer valuable in this respect.

## 2. Related Literature

Marketers have long recognized the importance of somehow incorporating word-of-mouth into their normative and descriptive models. The three typical methods of measuring WOM have been (i) estimation, (ii) survey, and (iii) social network analysis. The Bass (1969) model is the best-known of the models that estimates, in some regard, the impact of WOM. He specifies the following model of the adoption of a new product:

$$n(t) = p \cdot [m - N(t)] + \frac{q}{m} \cdot N(t) \cdot [m - N(t)] \quad (2.1)$$

where  $n(t)$  is the number of new adopters in time  $t$ ,  $N(t)$  is the cumulative adoption by time  $t$ , and  $m$  is the market potential. The parameters that are estimated are typically  $m$ ,  $p$  and  $q$ . The latter is called by Bass the “coefficient of imitation” and later called the “coefficient of internal influence,” (Mahajan, Muller and Srivastava (1990)). This is meant to capture the geometric growth of the user base for a new product due to interpersonal communication and observation. Note that the measure proposed in the Bass model, the coefficient  $p$ , could actually be a function of many different factors other than WOM, including pure imitation, network externalities and risk reduction. In fact, it is likely that  $p$  is a function of all of these factors to a different extent, depending on the particular industry. In contrast, we propose more “fine-tuned” measures of WOM.

There have been many surveys of WOM communication in the marketing literature. Richins (1983), for example, investigates the antecedents and consequents of *negative* word-of-mouth via a survey. However, as argued by Granovetter (1973), Frenzen and Nakamoto (1993), Reingen and Kernan (1986) and Reingen, et al. (1984), to understand WOM, one must truly understand the social structure in which the information transmission is taking place. At the very least, one

should attempt to study the phenomenon in a context in which *group membership* can be either ascertained or inferred. In the Theory Section of this paper, we discuss in more detail the application of this research to our work.

Finally, Reingen and Kernan (1986) propose a social-network approach for measuring interpersonal communications. They apply the extant research from sociology to specify what they call “referral networks” in which a firm might track the flow of information about its products/services among its existing and potential competitors. In addition, they allow for exogenous information sources, such as other firms, to affect the system. They provide an illustration of their methodology in which they analyze the referral network for a piano tuner. Their method calls for the estimation of two networks, the referral network and the social network, in order to understand their relationship. This was done via phone surveys. Their major insight is that stronger ties are more likely to be activated as a referral source.

We can think of the social-networks literature as an enrichment of the Bass model. In principle, survey methods could provide us with a very rich picture of the impact of word of mouth on product diffusion. The limitation of this methodology as applied to the questions posed here is that it is extremely difficult to capture the data necessary to estimate models of this type. For example, it is necessary to be able to track all of the relationships and conversations that each actor had within and between the groups in which he has membership. The advantage of the methods proposed in this paper is the low cost of obtaining the information on social networks and the elimination of reliance on recall.

### 3. Theory

As described in the first two sections, word-of-mouth activity has an important influence on the sales of a new product. We propose three hypotheses that deal with the impact of word-of-mouth activity on product diffusion.

Based on the previous research, as described in the section above, we expect WOM measures to have explanatory power in the estimation of future sales:

#### Hypothesis 1

WOM measures explain a statistically significant portion of variance in future sales beyond the information contained in the specific product effects and current sales. In particular, we expect the volume of conversations to be one such measure of WOM.

The seminal work of Granovetter (1973) provides rich insight into the structural underpinnings of word-of-mouth communication. He characterizes the ties between people according to their “strength.” He assumes that strength is a combination of, and is increasing in, “the amount of time, the emotional intensity, the intimacy, and the reciprocal services” that characterize the relationship. Examples of strong ties might be spouses, co-workers in the same department or members of the same high school “clique.” It is weak ties, however, that Granovetter sees as the key to the diffusion of word-of-mouth. His reasoning is that if information is available to one member of a community, then it will quickly spread via strong ties to the other members of that community. However, for it to spread to members of other communities, there must exist weak ties between members of these communities. Without them, information stays trapped within a community. Granovetter (1970) provides some empirical support for this phenomenon. In a

vivid example, he looks at people who had recently changed jobs and asks them about the source of the “lead” for their new position. Only 17% reported that the person who told them about the job was someone whom they saw “often” while 55% were seen “occasionally” and 28% were seen “rarely.”<sup>3</sup> That is, weak ties were the source of the new information more often than strong ties. This reinforces the idea that those with whom we have strong ties are more likely to have the same information that we have and are therefore less able to provide us with anything new.

Within the context of our study, we are interested in a specific implication of this work: *information travels quickly within communities but slowly across*. This follows directly from the definition of tie strength as (partially) determined by the time that the parties spend together. Put simply, even though weak ties are activated less often, they are the source of inter-community diffusion. This would imply that information that is scattered across communities is more likely to diffuse broadly than would information that is concentrated within a small subset of communities. This is because this “scattering” will have taken care of the slower weak-tie-based inter-community diffusion while the concentrated information would have to traverse these arcs. Thus, it would seem that the probability of the future widespread diffusion of information would be decreasing in the relative concentration of that information at a given point in time. We will exploit this idea by measuring this concentration and showing that, in fact, it may be negatively associated with future product success.

Thus, it seems natural that the future popularity of a show might depend on whether conversations about the show are spread out across different communities or confined to a few

---

<sup>3</sup> He used the following coding scheme: “often” = at least twice a week, “occasionally” = less than twice a week but more than once a year, and “rarely” = once a year or less.

communities. If the concentration is higher, the information would be more likely to remain within the community. We'll call the degree to which the set of conversations about a product is dispersed across communities the "breadth" of the conversations.

### Hypothesis 2

The impact of word of mouth on future sales is increasing in the breadth of the conversations.

Finally, we expect that as time goes by, the uncertainty concerning the appeal of the show will decrease with time. Thus, we expect for WOM recommendations to have diminished impact with time. Thus, the proportion of conversations that are recommendations decreases with time, while the volume of conversations among fans may actually increase.

### Hypothesis 3

The impact of word of mouth on sales is decreasing with time.

We can motivate the hypotheses above using a simple example. Suppose that a guest at a party with 10 other people spends the whole time talking about a movie, "Panic," that he watched last night. All 10 of the guests at that particular party will become aware of that movie. However, from the perspective of maximizing awareness, the film producers would prefer that the movie be discussed for half an hour at two different parties with 10 people each. Moreover, if the same 10 people got together again in a month's time, his discussions on the topic will have less impact on the other guests' likelihood of seeing the movie.

#### 4. Model

The model we present is based on a simple theory of product diffusion. We posit that the sales of product  $i$  at time  $t$  are given by the following stochastic process:

$$\text{Sales}_{i,t} = \eta_i + \beta_t \text{Sales}_{i,t-1} + \epsilon_{i,t} \quad (1)$$

$$\epsilon_{i,t} = \omega_{i,t} + \xi_{i,t} \quad (2)$$

Thus, the current sales of a product are a sum of a product-specific constant, a linear (possibly time-variant) function of last period's sales and a noise variable. Further, the stochastic aspect of sales,  $\epsilon_{i,t}$ , can be separated into a measurable part,  $\omega_{i,t}$ , and pure noise,  $\xi_{i,t}$ . We argue below that measures of word-of-mouth activity are correlated with  $\omega_{i,t}$ .

Note that the model above can be thought of as reduced form. For illustration purposes, we present a process that motivated our specification in (1)-(2). Suppose that under complete information, a product would enjoy sales of  $\text{Sales}_{i,\infty}$ . Before this equilibrium share is reached, the show may be “over-sold” ( $\text{Sales}_{i,t} > \text{Sales}_{i,\infty}$  (current product's share is higher than the share in the limit) or “under-sold” ( $\text{Sales}_{i,t} < \text{Sales}_{i,\infty}$ ). Due to the information revealed by trial, as well as word of mouth, the next period's sales adjust, eventually reaching the stable share.<sup>4</sup> In addition, there is an effect of buzz that is decreasing in the cumulative awareness of the show.

Combining these two effects, we obtain:

$$\text{Sales}_{i,t} = \text{Sales}_{i,\infty} + \beta(\text{Sales}_{i,\infty} - \text{Sales}_{i,t-1}) + \gamma(1 - \text{awareness}_{i,t-1})\text{WOM}_{i,t-1} + \epsilon_{i,t} \quad (3)$$

From the equation above, we can see the following:

- 1) The cumulative awareness of the show increases with time. In turn, this would imply a *decrease* of the impact of WOM with time.
- 2) With time, the sales of a show tend to the equilibrium sales level,  $Sales_{i,\infty}$ . In order to illustrate this point, consider a simpler equation (the result holds for the original equation as well):

$$Sales_{i,t} = Sales_{i,\infty} + \beta(Sales_{i,\infty} - Sales_{i,t-1}) \quad |\beta| < 1 \quad (4)$$

Suppose that  $Sales_{i,1} = Sales_{i,\infty} + \epsilon$  (the case of “over-sell” relative to the equilibrium share). We can see that the following sequence would follow:

$$Sales_{i,2} = Sales_{i,\infty} - \beta\epsilon,$$

$$Sales_{i,3} = Sales_{i,\infty} + \beta^2\epsilon,$$

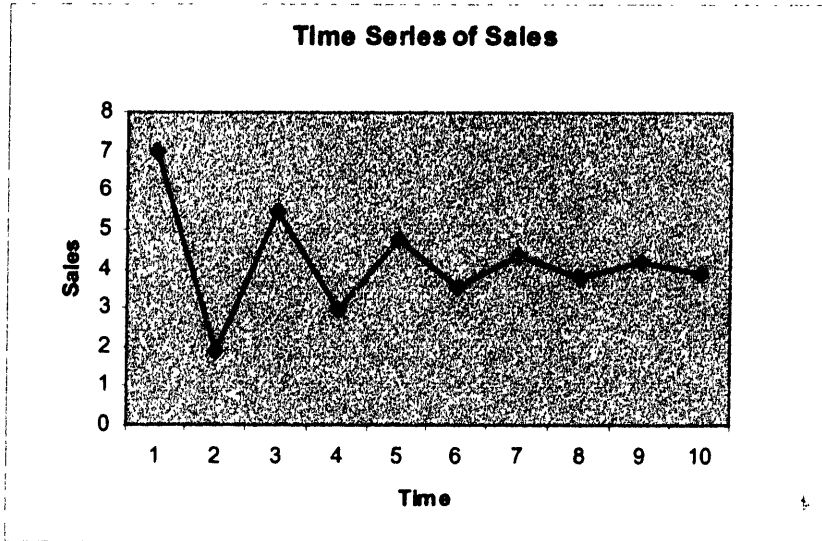
$$Sales_{i,4} = Sales_{i,\infty} - \beta^3\epsilon$$

We graph this time series for a numerical example in a graph below:

---

<sup>4</sup> Note that this is similar to the partial adjustment model described in Green (p. 798) In that model,

$$Sales_{i,t} = (1 - \lambda)Sales_{i,\infty} + \lambda Sales_{i,t-1} + \epsilon_{i,t}$$



Let us once again turn to equations (1)-(2), the reduced form of the model. The product-specific constant  $\eta_i$  describes differences in the inherent appeal of show  $i$  which we assume to be constant over time. For example, some shows are targeted at a broader audience segment and claim a higher proportion of the TV-watching audience. The lagged sales component captures the serial correlation in sales.

Note that we would expect the specific effect and the lagged sales variable to explain a lot of the variance in the current sales of a product. However, we are primarily interested in whether measures of word of mouth add information *beyond* the information contained in lagged sales and the specific effect. We argue that measures of word of mouth allow us to obtain an estimate of  $\omega_{i,t}$ , the measurable component of the noise term. We propose two WOM measures that we expect to be associated with positive shocks in ratings. The first is a simple summary measure of over-all chat activity (we can think of it as a measure of over-all “buzz.”) The second measure is a measure of the spread of this activity across different communities: the breadth of the conversations.

Note that there are two ways in which word of mouth can affect future sales. First, an increase in word of mouth activity raises awareness of the product. Second, these conversations may contain information on the suitability of the product to the potential consumers. Thus, theoretically, an increase in word of mouth activity may not necessarily result in higher sales if most of the recommendations are actually negative in content. We make no distinction between positive and negative chat. This is generally in keeping with the approach taken by most of the commercial news-clipping services, for example, which solely count mentions. Moreover, it reflects our goal of developing a useful tool for managers to implement easily and affordably. This approach necessarily biases our analysis against finding anything meaningful in the data.

As we discussed in the Theory Section, we expect an increase in the total amount of buzz to result in an increase in awareness of the new product. The second measure of interest, the “breadth” of conversations, measures the dispersion of conversations across communities. Moreover, we hypothesize a positive correlation between  $\omega_{i,t}$  and the two proposed measures of WOM and, therefore, between sales and the two measures of WOM. We would expect this correlation between  $\omega_{i,t}$  and our WOM measures to decrease over time. That is, as more people learn about which shows they like, we would expect the fixed effect and last period's ratings would provide sufficient information to predict next period's performance.

To summarize, we obtain the following specification:

$$\text{Sales}_{i,t} = \eta_i + \beta_t \text{Sales}_{i,t-1} + \varphi_t \text{TotalChat}_{i,t-1} + \alpha_t \text{Breadth}_{i,t-1} + \xi_{i,t} \quad (5)$$

We restate the hypotheses of the Theory Section in terms of the specification above.

### Hypothesis 1

WOM measures explain some of the variance in future sales beyond the information contained in the specific product effects and current sales: *we can reject that  $\varphi_i = 0$  and  $\alpha_i = 0$ .*

### Hypothesis 2

The impact of word of mouth on future sales is increasing in the breadth of the conversations:  $\alpha_i > 0$ .

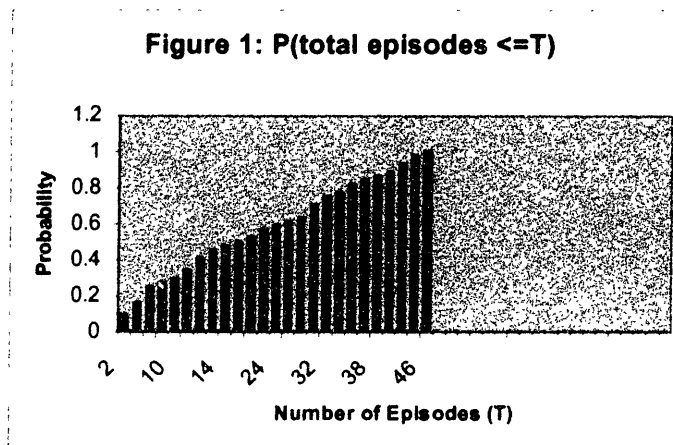
### Hypothesis 3

The impact of word of mouth on sales is decreasing with time:  $|\varphi_{\text{Late}}| < |\varphi_{\text{Early}}|$  and  $|\alpha_{\text{Late}}| < |\alpha_{\text{Early}}|$ .

## 5. Data

### 5.1 TV shows

Our data set consists of 44 TV shows that premiered in the 1999/2000 seasons. Our sample includes only the shows of the major networks: ABC, CBS, NBC, FOX, UPN, and WB. Among the 44 new shows, only 14 shows survived into the 2000/2001 seasons. A few of the shows were



cancelled very quickly. For example, 4 shows were cancelled after only 2 episodes, and half of all shows were shown fewer than 17 times. In Figure 1, we graph the distribution of total

episodes of a new show. The performance measure that we use is the weekly Nielsen ratings as reported in *Broadcasting & Cable* magazine, which reflect the percentage of households who watched the show that week<sup>5</sup>. In Table 1A in the Appendix, we include the break-out of new shows in 1999-2000 seasons by network, as well as the average duration of a show by network. For a detailed listing of the titles of the new shows, see Table 2A in the Appendix. Tables 1 and 2 presents descriptive data on show premieres. The variance in the size of the audience in our sample is very high. For example, 13.2 million households watched the premiere of CBS's "Judging Amy," while 1.6 million households watched the premiere of WB's "DC." Also note that while most of the shows premiered in late September or early October 2000 following the Sydney Summer Olympics, some shows were mid-season replacements, among them the very successful "Malcolm in the Middle" on Fox.

**Table 1: Top 5 first episode ratings**

Show	Network	Day of Week	Date	Rating (% of TV Universe)	TV Homes (millions)
Judging Amy	CBS	Sun	9/19/99	13.5	13.2
Stark Raving Mad	NBC	Thur	9/23/99	12.3	12.1
Once and Again	ABC	Tue	9/21/99	12.3	12.1
Malcolm in the Middle	FOX	Sun	1/9/00	12.1	11.9
West Wing	NBC	Wed	9/22/99	12.1	11.9

**Table 2: Bottom 5 first episode ratings**

Show	Network	Day of Week	Date	Rating (% of TV Universe)	TV Homes (millions)
DC	WB	Sun	4/2/00	1.6	1.6
Mission Hill	WB	Tue	9/21/99	1.8	1.8
The Beat	UPN	Tue	3/21/00	2.2	2.2
The Strip	UPN	Tue	10/12/99	2.3	2.3
Popular	WB	Thur	9/30/99	2.5	2.5

---

<sup>5</sup> We also used the show's "Share," which is the proportion of households that watched the show conditional on having watched t.v. The results are qualitatively equivalent so we report only the results using Rating.

## *5.2 Word-of-Mouth*

The Internet provides a unique opportunity to "listen in" on consumer conversations. We argue here that online conversations are an imperfect but meaningful measure of all word of mouth recommendations. However, even though we limit ourselves to online conversations only, we are still confronted with an infinite set of possible measures. We have chosen to use Usenet newsgroups to measure word of mouth activity for several reasons. First, an historical archive of Usenet newsgroups is readily and easily available to the public on Google.com (previously [deja.com](http://deja.com)). Moreover, there is a wide breadth of topics covered on Usenet, from [rec.autos.sport.nascar](#) to [alt.fan.noam-chomsky](#). Thus, it seems that this is a fertile area for both managers and academics to conduct research on on-line word-of-mouth.

A Usenet posting contains the moniker of the author, a subject line, the name of the newsgroup to which the post was sent, the date of the post, as well as the text of the message. The archive is searchable by subject, author, group, etc. Posts are organized into "threads" which contain posts on roughly the same topic. We might think of a thread as the on-line analog of a "conversation." Very often, all posts in a thread contain the same subject line. For an example of a complete thread, see the Appendix.

We restricted our analysis to newsgroups with names beginning with either "alt.tv" or "rec.arts.tv." It is our belief that searching the entire Usenet database for posts of interest is neither useful nor efficient. To identify a post as being "about" a show, we looked for the name of the show in the subject header. This is a conservative approach as there are likely to be a fair number of posts about shows which do not include the show's name on the subject line. We found 169 different groups that contained messages about the shows in our sample. The groups'

focus varies from television in general (for example, "rec.arts.tv") to specific shows (for example, "alt.tv.x-files" is attended by fans of Fox's TV show "The X-Files."). Of course, the subscribers of "x-files" often chat about other shows that they find interesting. Thus, in the Appendix, we present a thread that deals with WB's "Roswell" that takes place on the "x-files" chat room. This is not particularly surprising since both are science fiction shows. Note that it takes time for fans to assemble a newsgroup devoted to a new show such as "Roswell." In the initial period following the debut of a show, the chat is dispersed among groups that are devoted to other shows. Below, we present the top 20 groups that had the most postings dealing with shows in our sample. It is interesting to note that none of these groups bear the names of the shows in our sample.

**Table 3: Number of posts in our sample per chat room**

<b>Group</b>	<b>Number of posts</b>
rec.arts.tv	9649
alt.tv.game-shows	2892
alt.tv.law-and-order	1621
alt.tv.party-of-five	1013
alt.tv.homicide	932
alt.tv.buffy-v-slayer	764
rec.arts.tv.mst3k.mis	578
alt.tv.simpsons	533
alt.tv.star-trek.voya	527
alt.tv.dawsons-creek	498
alt.tv.x-files	440
alt.tv.er	391
alt.tv.emergency	326
alt.tv.millennium	311
alt.tv.newsradio	258
alt.tv.real-world	236
alt.tv.highlander	176
alt.tv.3rd-rock	162
alt.tv.twin-peaks	153

We then constructed measures along the lines of the theoretical constructs discussed in the Model Section of the paper. Below, we describe our method of constructing the three key measures presented in the paper:

$TotPosts_{i,t}$ : Number of posts that (a) contain the title of the show in the subject line and (b) were posted between episode  $t$  and episode  $t-1$ . This is the on-line analog of the naïve “count” measure.

$Entropy_{i,t}$ : A measure of how diffuse the chat is across different groups. We expect that a broader pattern results in more rapid spread of information, resulting in higher awareness and viewership of the show. Specifically,  $entropy_{i,t} = - \sum_{rooms} \frac{posts_{i,t}(room_i)}{TotPosts_{i,t}} \log(\frac{posts_{i,t}(room_i)}{TotPosts_{i,t}})$ .

We choose this measure over, say, variance because it is independent of the total volume of posts. Like variance, however, if the posts are all concentrated in one newsgroup, entropy is 0. Entropy is maximized when posts are evenly distributed across different groups. Note that we assign a value of 0 to entropy in the case when there are no posts dealing with the show in that period. We have done the estimation with other measures of breadth and have obtained similar results.

There are three shows that we excluded from the sample: "Angel," "Harsh Realm," and "Grownups." We exclude "Angel" because we found too many posts that contained the word "angel" in the subject line. Most of the posts are unrelated to the show. On the other hand, there are no posts that contained the words "grownups" and "harsh realm" in the subject line. This demonstrates that our technique of extracting posts is imperfect. This is especially the case when shows' names contain common words such as "angel" or involve shows that generated very little buzz. Finally, while most shows run on the same time every week, this is not always the case. Some shows have episodes separated by more than a week. This might be due to special programming such as the World Series or a Presidential debate. Moreover, some shows run more than once a week, particularly early in their life. First, the results we present below do not control for these factors. We have estimated alternative specifications which consider posts per

intervening days rather than total posts between episodes. The results are qualitatively equivalent so we do not present them here. However, we do control for the fact that sometimes two episodes of the same show are run on the same day. In this case, we use the ratings from the *first* episode run that day and exclude the second.

Tables 4 and 5 summarize the variable definitions and provide the means of the variables used in the study. Note that, since Rating is a percentage, we use a logarithmic transformation.

**Table 4: Variable Definitions**

Variable	Description
$R$	$\ln(\text{ratings}_t)$
$TotPosts$	Total number of posts w/ subject header containing the show title
$Entropy$	Variability of posts across different groups
$Episode$	Episode Number

**Table 5: Summary information**

Variable	Mean	Std. Dev.	Min	Max
R	1.53	0.65	-0.36	2.65
TotPosts	27.8	41.20	0	261
Entropy	0.49	0.66	0	3.00

## 6. Specification and Econometric Issues

### 6.1 Specification

As discussed above, we estimate the following equation:

$$\text{Sales}_{i,t} = \eta_i + \beta_t \text{Sales}_{i,t-1} + \varphi_t \text{TotalChat}_{i,t-1} + \alpha_t \text{Breadth}_{i,t-1} + \xi_{i,t} \quad (6)$$

We are also interested in analyzing the differential effect of word-of-mouth communications across time. Below, we list the different specifications that will allow us to measure these effects:

$$R_{i,t} = \eta_i + \beta R_{i,t-1} + \varphi \text{TotPost}_{i,t-1} + \alpha \text{Entropy}_{i,t-1} + \xi_{i,t}, \quad t \in [2, \bar{T}] \quad (7)$$

Here we estimate the equation using only the data in the initial period following the debut of the show. This captures the impact of word-of-mouth early in the life of the show. Since we are not certain about the appropriate length of this period – what “early” might mean -- we estimate the above equation for a range of introductory periods:  $T=\{4,5,6\}$ . Note that this specification allows us to test Hypotheses 1–2: the explanatory power of WOM measures and, specifically, the impact of the entropy measure.

In an alternative specification, we estimate the equation using the data available for the entire season. We allow the effects of chat measures as well as lagged ratings to vary over time:

$$R_{i,t} = \sum_{j \in \{\text{Early}, \text{Late}\}} \eta_{ij} + \beta_j R_{i,t-1} + \phi_j \text{TotPost}_{i,t-1} + \alpha_j \text{Entropy}_{i,t-1} + \xi_{i,t} \quad (8)$$

In this specification, we allow for our independent variables to have a different effect in the “early” period and the “late” period. Again, not knowing exactly where the dividing line ought to be, we estimate the model for a range of values. This specification allows us to test Hypothesis 3: the change in impact of the WOM measures over time.

It is useful to consider these two specifications from the network executive’s perspective. The first specification, (7), is an estimation that can be done with very little data. The executive can use it to also estimate the fixed effect of the show as well as measure the impact of word-of-mouth in order to make a better decision on a scheduling policy. The second specification, (8), on the other hand, implies that the executive has information about the show’s quality beyond what could be gleaned from the first few shows.

## 6.2 Econometric Issues

While (7) and (8) are very straightforward models, they nonetheless present us with several econometric issues. Most of the difficulty is created by the presence of the lagged endogenous variable  $R_{i,t-1}$ . In addition, our panel is unbalanced since some shows "live" longer than others. This raises the possibility of selection bias. We will discuss the potential endogeneity problem in the second part of this Section.

First, we re-write (7) as:

$$R_{i,t} = \eta_i + \beta R_{i,t-1} + \gamma \vec{\text{Chat}}_{i,t-1} + \varepsilon_{i,t} \quad (9)$$

Thus, last period's ratings,  $R_{i,t-1}$ , is the lagged endogenous variable, and we combine the exogenous measures of chat into a vector of variables we call  $\vec{\text{Chat}}_{i,t-1}$ . This equation belongs to a family of models called *dynamic panel data models*. In the absence of serial correlation in the error term, one can easily show that a fixed effects estimator has a bias for any finite sample of data but is consistent as  $T \rightarrow \infty$ . Note that a fixed effects estimator on the original data is equivalent to an OLS estimator on a data from which individual means have been subtracted. Following the transformation, we can re-write (7) as

$$R_{i,t} - \bar{R}_i = \beta(R_{i,t-1} - \bar{R}_i) + \gamma(\vec{\text{Chat}}_{i,t-1} - \vec{\bar{\text{Chat}}}_i) + (\varepsilon_{i,t} - \frac{\varepsilon_{i,1} + \dots + \varepsilon_{i,T}}{T}). \quad \text{Now, we can}$$

clearly see that  $R_{i,t-1}$  is not orthogonal to  $(\varepsilon_{i,t} - \frac{\varepsilon_{i,1} + \dots + \varepsilon_{i,T}}{T})$ . At the same time, the bias goes to zero as  $T \rightarrow \infty$ . Several studies, including Kiviet (1995), estimate the magnitude and direction of the bias.

Another strategy to estimate the equation above is to first-difference the data<sup>6</sup>:

$$\begin{aligned} R_{i,t-1} - R_{i,t-2} &= \beta(R_{i,t-2} - R_{i,t-3}) + \rho(\vec{Chat}_{i,t-2} - \vec{Chat}_{i,t-3}) + \varepsilon_{i,t-1} - \varepsilon_{i,t-2} \\ R_{i,t} - R_{i,t-1} &= \beta(R_{i,t-1} - R_{i,t-2}) + \rho(\vec{Chat}_{i,t-1} - \vec{Chat}_{i,t-2}) + \varepsilon_{i,t} - \varepsilon_{i,t-1} \end{aligned} \quad (10)$$

From (8), we can see that even under all standard assumptions, such as no serial correlation and strictly exogenous variables, we need an instrument for the  $R_{i,t-1} - R_{i,t-2}$  term since it is not orthogonal to  $\varepsilon_{i,t} - \varepsilon_{i,t-1}$ . We can, however, use lagged *levels* of the endogenous variable to instrument for the differenced endogenous variable. That is, we can use  $\{R_{i,t-2}, R_{i,t-3}, R_{i,t-4}, R_{i,t-5}\}$ , etc. as instruments for  $R_{i,t-1} - R_{i,t-2}$ . This is a strategy employed by Arellano and Bond (1991), as well as by a number of other earlier papers, including Anderson and Hsiao (1982). Note that this estimator is consistent only as long as there is no serial correlation in the error term,  $\varepsilon_{i,t}$ . If the errors are serially correlated,  $R_{i,t-2}$  is not orthogonal to  $\varepsilon_{i,t} - \varepsilon_{i,t-1}$ . Arellano and Bond propose tests of the no serial correlation hypothesis, which we discuss in more detail in Section 6. In addition, the Arellano and Bond estimator can be used to estimate models with unbalanced panel, a result that we will make use of. We will present estimates of our models using a standard fixed effects estimator and use the Arellano and Bond estimator to test the robustness of our results.<sup>7</sup>

---

<sup>6</sup> Another type of transformation is called "orthogonal deviations." (See Arellano and Bover (1995)). In this transformation, we express each observation as the deviation from the average of future observations for that individual. Each deviation is then weighted to standardize the variance:

$x_{i,t}^* = (x_{i,t} - \frac{x_{i,t+1} + \dots + x_{i,T}}{T-t}) (\frac{T-t}{T-t+1})^{1/2}$  for  $t=1, \dots, T-1$ . We present the results of both types of transformations.

<sup>7</sup> In certain models, one can make the additional assumption that the instruments are uncorrelated with the fixed effects. In this case, the additional information in levels improves the efficiency of the estimator. For example, in such a model one could use  $R_{i,t-1} - R_{i,t-2}$  as an instrument for  $R_{i,t-1}$ . This is an approach taken by Arellano and Bover (1995). Another innovation in their paper is an introduction of a different transformation: an orthogonal deviations transformation. Both Arellano and Bond and Arellano and Bover are generalized method of moments

Another issue with which we must contend is that of show cancellations. Note that above we have modeled the problem without taking into account the network executive's decision-making. Let us next consider the effect of an additional actor, the manager, on our estimation. We argue that the network executives' decision to cancel a show is primarily motivated by the show's past ratings. For example, a New York Times article describes the story behind the cancellation of *Wonderland*, one of the shows in our sample, "The series, which was heavily promoted and earned strongly favorable reviews, was abruptly canceled after its second episode on April 6 because of a precipitous decline in ratings."<sup>8</sup> This quote suggests that the change in ratings has a strong effect on an executive's decision rule.

We do not want to claim that the past ratings are the sole input into an executive's decision. For example, networks may be more inclined to keep shows by certain powerful producers, "But the size of an audience watching a series is only one factor in the decision making. Television executives say, for example, that NBC may have a hard time canceling the potentially vulnerable series 'Third Watch' ... because its creator, John Wells, has an entrenched relationship with the network ..."<sup>9</sup> However, there is little doubt that low ratings make a new show more vulnerable to cancellations, and highly rated new shows are never cancelled. When an executive weighs the decision to cancel a show, he must consider the stream of future ratings. Thus, the manager must consider the discounted net present value of future ratings. This would imply that the manager

---

(GMM) estimators that use the orthogonality conditions as linear moment restrictions. We are reluctant to make the extra assumption used in Arellano & Bover since the past growth in ratings may very well be correlated with the size of the fixed effect since both may be related to the quality of the show.

<sup>8</sup> Source: (The New York Times, April 27, 2000, "Network Offers No Asylum; Acclaim Couldn't Assure a Home for Dark 'Wonderland'")

<sup>9</sup> Source: (The New York Times, May 1, 2001, "For TV's Vulnerable Shows, It's Fear-and-Lobbying Season.")

faces an optimal stopping rule problem, where the state variable is the vector of past ratings, and there is an opportunity cost of airing a poor performer. At time  $t-1$ , the executive cancels the show if

$$E_{t-1}[R_{i,t} | R_{i,t-1}, R_{i,t-2}, \dots] + \delta \max_{\text{keep, cancel}} V_t < \Pi^A \quad (11)$$

Here the manager cancels the show if the expected ratings next period plus the discounted next period's value is less than the alternative profit from airing the best alternative show.

We claim that under reasonable assumptions on the manager's selection rule there is no selection bias in our estimation. In order to demonstrate this, we consider a simpler selection rule, but our claims are valid in the context of a more general rule as well. Suppose that the manager is myopic in that he only considers the following selection rule: he cancels the show if  $E_{t-1}[R_{i,t} | R_{i,t-1}, R_{i,t-2}, \dots] < \Pi^A$ . If we substitute for the expression of next period's ratings, this would imply that the show is cancelled if  $E_{t-1}[\eta_{i,t}] + \beta R_{i,t-1} + \gamma E_{t-1}[\vec{\text{Chat}}_{i,t-1}] + E_{t-1}[\epsilon_{i,t}] < \Pi^A$ . We apply the expectation operator to the fixed effect and the chat since the manager may not perfectly observe or measure these effects.

Next, let us consider next period's expected error term:  $E_{t-1}[\epsilon_{i,t}] \equiv \mu_{i,t-1}$ . This turns out to be the crucial term that determines whether we have a selection bias or not, and it depends on the manager's ability to predict future shocks to the ratings. We will make the assumption that it is very difficult to predict the future shocks: that  $\text{COV}(\mu_{i,t-1}, \epsilon_{i,t}) = 0$ . In this case, our main equation can be estimated with no bias since the manager selects on the right-hand-side variable, last period's ratings. (Recall that the main equation of interest

is  $R_{i,t} = \eta_i + \beta R_{i,t-1} + \rho \vec{\gamma} \text{Chat}_{i,t-1} + \varepsilon_{i,t}$ . A show with high  $R_{i,t-1}$  is more likely to remain in the sample. If there is no serial correlation in the error term, this is selection on an independent variable and introduces no bias into our estimates of the equation). However, in the presence of serial correlation, selection on past ratings also implies selection on future shocks. This would, of course, imply that we would have selection bias. This again demonstrates the importance of testing for the presence of serial correlation.

## 7. Results

We first present our estimation of (7) using a fixed effects estimator. In this specification, we only use data from the first few episodes. This also coincides with the intuition that for forecasting purposes it is reasonable to use the initial sample only. We then test the robustness of the results, and correct for a potential bias, by re-estimating the same specification using the Arellano and Bond GMM estimator. The Arellano and Bond procedure also allows us to test for the presence of serial correlation in the error term. We then turn to the estimation of  $S_2$ .

### 7.1 Fixed Effects Estimation of (7)

Table 6 presents the results of the fixed effects estimation on a sub-sample of the data that only includes a few initial episodes. As mentioned above, we are uncertain about the length of the "initial" period. Thus, we repeat the estimation for slightly different samples.

**Table 6: Fixed Effects Estimator, Dependent variable =  $R_{i,t}$ , no repeat**

	$t \in [2,6]$ n=168	$t \in [2,5]$ n=138	$t \in [2,4]$ n=109
$R_{t-1}$	<b>-0.249</b> <b>(-3.25)</b>	<b>-0.286</b> <b>(-3.28)</b>	<b>-0.375</b> <b>(-3.72)</b>
TotPosts	0.000301 (1.59)	0.000320 (0.57)	0.000178 (0.28)
Entropy	0.040 (1.62)	<b>0.0503</b> <b>(1.72)</b>	<b>0.085</b> <b>(2.13)</b>
Episode	<b>-0.030</b> <b>(-3.32)</b>	<b>-0.037</b> <b>(-2.79)</b>	<b>-0.057</b> <b>(-2.81)</b>
R-squared	0.75	0.80	0.78
F-test $Entropy=0, TotPosts = 0$ (p-level)	<b>F(2,123)=3.18</b> <b>(0.05)</b>	F(2,93)=2.20 (0.12)	<b>F(2,64)=2.67</b> <b>(0.08)</b>
F-test $Entropy = 0$ (p-level)	F(1,123)=2.63 (0.11)	<b>F(1,93)=2.95</b> <b>(0.09)</b>	<b>F(1,64)= 4.53</b> <b>(0.04)</b>

**Note:**

- 1) **Bold = significant at 10% confidence level**
- 2) **Numbers in parentheses are t-stats**

In order to test Hypothesis 1, for each sub-sample we perform two F-tests. In the first test, we test the unrestricted model with WOM measures against a model where we restrict the coefficients on the WOM measures to be zero. In the second test, we restrict only the coefficient on the entropy measure to be zero. The tests seem to provide support for Hypothesis 1. Out of the 6 different tests, in 4 tests we can reject the linear restrictions at 10% significance. In two of the remaining tests, the F-statistic is significant at a p-level slightly above 10%. Thus, WOM measures provide have explanatory power in the specification.

Second, we see that the only significant variable among the chat measures is the *entropy* of the chat. The insignificance on the *TotPosts* measure is somewhat surprising, while the positive and significant coefficient on *entropy* presents support for Hypothesis 2: the breadth of the

conversations has an impact on future sales. Also, the magnitude of the coefficient declines as we expand the sample to include later episodes. This provides at least partial support for Hypothesis 3: we expect the impact of chat variables to decline with time. On average, an increase in entropy by 25% results in 2% increase in future ratings. Since each percentage point in Nielsen ratings represents almost a million households, this is a sizable effect. The results confirm our theory that the breadth of conversations is an early indicator of the success of a show. Moreover, we show that the effect is very robust to different specifications.

Note also that the coefficient on the lagged ratings variable is negative. Negative serial correlation in the ratings implies that initially there is a lot of up and down movement in the ratings after we control for other effects, which is consistent with the motivation presented in the Model Section.

We discussed in Section 6 the possible finite sample bias of the fixed effects estimator. In addition, in the presence of serial correlation in the errors, the fixed effects estimator is inconsistent even for large samples. Thus, we re-estimate the equation using the Arellano and Bond estimator and present the results of various specification tests.

### *7.2 Arellano and Bond Estimation using initial data only*

In the Appendix, we present a brief summary of the Arellano and Bond estimation, as well as an overview of the specification tests presented in this paper. In Table 7, we present the results of the estimation. As explained in the appendix, Arellano and Bond estimation requires at least 3 periods of data. Thus, the 4 shows that were canceled after 2 episodes are dropped from the sample, which leaves us with 37 shows. For each sub-sample of the data, we present two

different specifications: one in which we assume that all the chat variables are exogenous, in the sense that the chat measures are orthogonal to all future and present shocks in ratings (See the Appendix for more detail on strict exogenous vs predeterminedness). In an alternative specification, we model *total posts* as predetermined since it would be reasonable to assume that the total volume of chat that follows a show is a function of the ratings of the show. The estimates are done using all available lags of endogenous and pre-determined variables. In addition, the estimates are one-step and are robust to heteroskedasticity.

We first note that the results are qualitatively similar to the fixed effects results. The coefficient on lagged ratings is initially negative and becomes smaller in magnitude as we consider longer samples of data. As with fixed effects, the coefficient on the *total posts* variable is not significant. The entropy coefficient is positive and significant (for all specifications as long as  $t \leq 6$ , with the exception of  $S_2$ ). We believe that these results support the robustness of the finding that simply counting word-of-mouth activity is not necessarily optimal and that measuring the dispersion of the activity may be more powerful. The only notable difference is in the sign of the time trend. Whereas in the fixed effects estimation the time trend was negative, here it is positive.

In all cases, the Sargan test (see Appendix) *does not* reject the null of valid specification.<sup>10</sup> (In all cases, the p-value is greater than 0.1) Thus, we can't reject the validity of a model in which *total posts* does not depend on last period's sales. There are two possible explanations for this result. First, it might very well be the case that our measures of total posts are too noisy. This is likely

---

<sup>10</sup> Note that the Sargan statistic is obtained from a two-step homoskedastic Arellano and Bond estimator. See Appendix for more detail.

to be a bigger problem in the Arellano and Bond estimation than with standard Fixed Effects because the former eats up more degrees of freedom in performing the relevant tests. Second, it may in fact be true that the counting measure, as we have constructed it, is independent of previous ratings. This would be very surprising. Further studies into the motivation behind word of mouth recommendations are needed to differentiate between these two possible explanations.

Finally, it is important to note that none of the specifications reject the null of no serial correlation in the error term. The results of this test are given in the second to last row of the table. The procedure is described more thoroughly in the Appendix. This is an important result since the consistency of both our fixed effects and Arellano and Bond estimates depend on this assumption. The Table also presents a Wald test ( $df = 5$ ) where the null is that all the coefficients except the constant are zero. The null is rejected under all specifications. We also perform a Wald test to test the linear restriction that the coefficient on the entropy variable should be set to zero. We reject this restriction in all cases where the coefficient on entropy is significant.

**Table 7\*: ARELLANO & BOND Estimator, Dependent variable =  $R_{it}$ , No repeat, Robust Errors**

	$t \in [3,6], n=122$		$t \in [3,5], n=93$		$t \in [3,4], n=65$	
	S6 $pre(Post_{-1})$	S5 All exogenous	S4 $pre(Post_{-1})$	S3 All exogenous	S2 $pre(Post_{-1})$	S1 All exogenous
$R_{t-1}$	<b>-.476</b> (-2.14)	<b>-.442</b> (-1.87)	<b>-.482</b> (-1.97)	<b>-.477</b> (-1.99)	<b>-.507</b> (-2.14)	<b>-.497</b> (-2.16)
TotPosts	-.0012 (-1.36)	.000446 (1.07)	-.00150 (-0.85)	.00014 (0.35)	-.00211 (-0.97)	-.000146 (0.33)
Entropy	<b>.0716</b> (2.43)	<b>.0370</b> (2.09)	.0662 (1.55)	<b>.0346</b> (1.70)	.120 (1.52)	.065 (1.50)
Episode	<b>.0290</b> (1.73)	.0202 (1.38)	<b>.0493</b> (1.94)	<b>.0453</b> (1.97)	.0607 (1.11)	.053 (1.40)
Cons	<b>-.185</b> (-1.73)	<b>-.137</b> (-1.76)	<b>-.266</b> (-2.17)	<b>-.238</b> (-2.21)	-.305 (-1.46)	-.26 (-1.72)
Sargan (p-value)	$\chi^2(18) =$ 17.06 (0.52)	$\chi^2(9) =$ 7.58 (0.58)	$\chi^2(10) =$ 3.14 (0.98)	$\chi^2(5) =$ 2.96 (0.71)	$\chi^2(4) =$ 1.91 (0.75)	$\chi^2(2) =$ 2.53 (0.282)
AB test z (p-value)	0.69 (0.49)	0.88 (0.38)	1.05 (0.29)	0.71 (0.48)	NA	NA
Wald Test $\chi^2(4)$	8.74	6.56	5.45	6.02	6.43	7.14
Wald Test $\chi^2(1)$ (Entropy=0)	<b>5.89 (0.02)</b>	<b>4.36 (0.04)</b>	2.40 (0.12)	<b>2.87 (0.09)</b>	2.31 (0.13)	2.25 (0.13)

Note: the numbers in parentheses are z-statistics

In Table 8, we present the results of an estimation performed with an orthogonal deviations transformation of the data. The reason we employ this transformation is due to a finding in Alavarez and Arellano (1998) that the orthogonal transformation may be less biased than the first difference transformation, “for fixed T the IV estimators in orthogonal-deviations and first-differences are both consistent, whereas as T increases the former remains consistent but the latter is inconsistent.” (p. 29) Note that in all other respects the estimation procedure is the same as the one employed by the Arellano and Bond estimator. We see that the change in transformation has little effect on the magnitude of the coefficient on Entropy. The TotPosts coefficient similarly remains insignificant, but there is a slight decrease in the magnitude of the

coefficient on the lagged ratings. Thus, we can see that the results are robust under several reasonable econometric specifications.

**Table 8\*: ARELLANO & BOND Estimator with orthogonal deviation transformation,**  
All independent variables are exogenous  
Dependent variable =  $R_t$ , No repeat,  
Robust Errors

	$t \in [3,6]$	$t \in [3,5]$	$t \in [3,4]$
$R_{t-1}$	<b>-0.396</b> (-1.72)	<b>-0.391</b> (-1.73)	<b>-0.387</b> (-1.98)
TotPosts	0.000718 (1.39)	0.000294 (0.55)	0.000078 (0.15)
Entropy	<b>.0344</b> (1.76)	<b>.0470</b> (1.65)	0.0861 (1.61)
Episode	<b>-0.0271</b> (-1.75)	<b>-0.0414</b> (-1.82)	-0.0381 (1.42)
Cons	<b>.178</b> (2.06)	<b>0.218</b> (2.00)	0.19 (1.70)
Sargan (p-value)	$\chi^2(10) =$ 10.55 (0.31)	$\chi^2(5) =$ 4.41 (0.492)	$\chi^2(2) =$ 2.70 (0.259)
AB test z (p-value)	0.852 (0.39)	1.062 (0.29)	NA
Wald Test $\chi^2(4)$	10.4	9.82	9.38
Wald Test $\chi^2(1)$ (Entropy=0)	3.11 (0.078)	2.73 (0.099)	4.58 (0.03)

**Note:** the numbers in parentheses are z-statistics

### 7.3 Fixed Effects Estimation of (8)

In the previous Section, we show that our Fixed Effects estimation is fairly robust since we obtain similar results with a consistent estimator developed by Arellano and Bond. Also, we found no evidence of serial correlation in the error term. In this section, we present the results of the fixed effects estimation on the entire data set. In this version of  $S_2$ , we group episodes 2 to 5

into the *early* period and episode 6 and up into the *late* period. Once again, we want to see if we observe differences in the early versus late effects in order to test Hypothesis 3. Again, we consider this to be a test of our theory within the context of a more fully-informed network executive.

The results are shown in Table 9. As we observed earlier, the coefficient on TotPosts is insignificant. Once again, the coefficient on early entropy is positive and significant. On the other hand, the impact of entropy is not significant in the later periods. This is entirely consistent with Hypothesis 3.

**Table 9: Full Sample, Fixed Effects, Dependent variable =  $R_{it}$ , n=688, repeat=0, different show dummies for late and early periods, robust**

	<i>early</i> = $t \in [2,5]$
$R_{it-1} \times \text{early}$	-.186 (-1.57)
$R_{it-1} \times \text{late}$	<b>.361</b> <b>(6.53)</b>
TotPosts x early	.000562 (1.03)
TotPostsx late	-.000402 (-1.35)
Entropy x early	<b>.0605</b> <b>(2.28)</b>
Entropy x late	-.0182 (-1.08)
Episode	<b>-.00509</b> <b>(-5.30)</b>
R-squared	0.95

## 8. Conclusion

In this paper, we demonstrate the importance of information on the dispersion of word of mouth across different communities for predicting product sales. We also find that the total volume of conversations, a measure that is used by many businesses such as Yahoo and Lycos, may not be a very meaningful measure. We show that the measures of chat are only important in the initial period following the debut of a show in a period of uncertainty.

This paper raises many issues that would benefit from further work. We see a lot of benefit in behavioral research that would explore what motivates people to post recommendations. This would contribute to forming more meaningful measures of word of mouth activity. Empirically, it would be very interesting to estimate the models presented above on products from other categories. We are especially interested in exploring whether these effects would survive in categories where expertise is involved, such as cars or stereos. We are also interested in the interaction between online and off-line word of mouth. Thus, survey data that would address this interaction would be very interesting.

The Internet has provided us with an unprecedented opportunity to listen and learn from consumer conversations. This paper is a first step in the direction of exploring this opportunity.

## References

1. Ahn, Seung and Schmidt, Peter (1995), "Efficient Estimation of Models for Dynamic Panel Data," *Journal of Econometrics*, 68, 5-27.
2. Alvarez, Javier and Arellano, Manuel (1998), "The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators," CEMFI Working Paper No. 9808.
3. Anderson, T. and Hsiao, C. (1982), "Formulation and Estimation of Dynamic Models Using Panel Data," *Journal of Econometrics*, 18, 47-82.
4. Arellano, Manuel and Bond, Stephen (1991), "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, 58, 277-297.
5. Arellano, Manuel and Bond, Stephen (1998), "Dynamic Panel Data Estimation Using DPD98 for Gauss: A Guide for Users."
6. Arellano, Manuel and Bover, Olympia (1995), "Another Look at the Instrumental Variable Estimation of Error-Components Models," *Journal of Econometrics*, 68, 29-51.
7. Baltagi, Badi and Li, Qi (1995), "Testing AR(1) Against MA(1) Disturbances In An Error Component Model," *Journal of Econometrics*, 68, 133-151.
8. Banerjee, Abhijit (1992), "A Simple Model of Herd Behavior," *Quarterly Journal of Economics*, 110, 797-817
9. \_\_\_\_\_ (1993), "The Economics of Rumours," *Review of Economic Studies*, 60, 309-327.
10. Bass, F. (1969), "A New Product Growth Model for Consumer Durables," *Management Science*, 15 (January), 215-27.

11. Bhargava, A and Franzini, L (1982), "Serial Correlation and the Fixed Effects Model," *Review of Economic Studies*, XLIX, 533-549.
12. Bikhchandani, Sushil, D. Hirshleifer and I. Welch (1991) , "A Theory of Fads, Fashions, Custom and Cultural Change as Information Cascades," *Journal of Political Economy*, 100, 992-1026.
13. Durbin, J. (1970), "Testing For Serial Correlation In Least-Squares Regression When Some of the Regressors Are Lagged Dependent Variables," *Econometrica*, 38, 410-421.
14. The Forrester Report: Entertaining Young Net Surfers, March 2000.
15. Franzen, J. and K. Nakamoto (1993), "Structure, Cooperation, and the Flow of Market Information," *Journal of Consumer Research*, 20, December 1993, 360-375.
16. Granovetter, M (1973), "The Strength of Weak Ties," *American Journal of Sociology*, 78 (6), 1360-1380.
17. \_\_\_\_\_ (1970) "Changing Jobs: Channels of Mobility Information in a Suburban Community," Doctoral Dissertation, Harvard University.
18. Holtz-Eakin, Douglas, Newey, Whitney, and Rosen, Harvey (1988), "Estimating Vector Autoregressions with Panel Data," *Econometrica*, 56, 1371-1395.
19. Keane, Michael and Runkle, David (1992), "On The Estimation of Panel-Data Models With Serial Correlation When Instruments Are Not Strictly Exogenous," *American Statistical Association*, 10, 1-29.
20. Kiviet, Jan (1995), "On Bias, Inconsistency, and Efficiency of Various Estimators in Dynamic Panel Data Models," *Journal of Econometrics*, 68, 53-78.
21. Kotler, Philip (2000), "Marketing Management," Prentice Hall, p. 560.

22. Li, Q. and Hsiao, C. (1998), "Testing Serial Correlation In Semiparametric Panel Data Models," *Journal of Econometrics*, 87, 207-237.
23. Mahajan, V., E.Muller and R. Srivastava (1990), "Determination of Adopter Categories by Using Innovation Diffusion Models," *Journal of Marketing Research*, 27 (37-50).Stata User's Manual, 291-308.
24. Reingen, P. and J. Kernan (1986) , "Analysis of Referral Networks in Marketing: Methods and Illustration," *Journal of Marketing Research*, 23, 370-8.
25. \_\_\_\_\_, B. Foster, J. Johnson Brown and S. Seidman (1984), "Brand Congruence in Interpersonal Relations: A Social Network Analysis," *Journal of Consumer Research*, 11 (December 1984), 771-83.
26. Resnick, P., Zeckhauser, R. (2001), "Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System," Version for Review by NBER workshop participants, [http: www.si.umich.edu ~presnick papers ebayNBER index.html](http://www.si.umich.edu/~presnick/papers/ebayNBER/index.html).
27. Richins, M. (1983), "Negative Word-of-Mouth by Dissatisfied Customers: A Case Study," *Journal of Marketing*, 47, 68-78.
28. Urban, Glen and Hauser, John (1993), "Design and Marketing of New Products," Second Edition, Prentice Hall.

## Appendix

(The following text is taken from [HTTP://GROUPS.GOOGLE.COM/GOOGLEGROUPS/BASICS.HTML](http://groups.google.com/googlegroups/basics.html).)

### Where did Usenet come from?

Before the Web and web browsers, and before email became ubiquitous, online communication meant posting text messages on electronic bulletin boards where others could read and reply to them. Usenet began as a collection of these bulletin boards (now called discussion forums or newsgroups) started in 1979 by Steve Bellovin, Jim Ellis, Tom Truscott, and Steve Daniel at Duke University. Over the years, the number of such newsgroups has grown to the thousands, hosted all over the world and covering every conceivable topic about which humans converse.

While there was incredibly valuable information available in the discussions taking place on newsgroups, finding that information could be an exercise in frustration and futility. Someone would start a topic with a posting on a bulletin board. Someone else would reply. This initial post and response now constituted a "thread" on the topic. The thread might grow to include dozens or even hundreds of individuals responding to the first post or any that came after it. They might start threads of their own as offshoots of the original discussion. Think of that initial post as a single cell dividing again and again, mutating and expanding geometrically with no predefined direction. The result is likely something you'd find in a low budget horror movie. Finding a specific bit of information in Usenet was an equally horrific task. To make it more complex, almost all newsservers expire messages after a few days or, at most, a few weeks. Expired messages are deleted from the live discussion forums and aren't viewable or searchable by users.

In 1995, Deja News was created to provide a user-friendly interface to Usenet. Deja began archiving and indexing messages so they could be searched and sorted, turning an ephemeral and unmanageable resource into a reference tool that was fairly easy to use.

In February 2001, Google Inc., a company dedicated to providing access to all information online, acquired the Usenet discussion service from Deja.com, including its entire Usenet archive of more than 500 million messages -- over a terabyte of human conversation. Google has expanded accessibility to the Usenet database through deployment of improved search and browsing tools and integration of the full archive with more recent postings. The combined database of more than 650 million messages already constitutes the largest collection of Usenet data on the Web and is growing at a rapid pace.

### Navigating Usenet

Usenet is like a river with thousands of tributaries. The main forks in the river lead to the top-level discussion categories (such as "alt"). Follow one of the river's forks and you'll come to smaller branches (such as alt.animals), which lead to tributaries containing messages divided into even more specific topics (such as alt.animals.dogs). Ultimately, your journey will take you to the smallest part of the data stream: the part containing messages from people who are interested in one particular topic (such as alt.animals.dogs.beagles).

The different parts of a newsgroup's name are always separated by a period, a traditional categorization symbol in the computer world. Each newsgroup contains threads made up of messages (also referred to as 'articles' or 'postings') that look like e-mail between one user and another, but can be read by anyone accessing that particular newsgroup.

**Table 1A: New TV shows in the 1999/2000 seasons**

**\*Note:** the classification of the shows is taken from an online database: [imdb.com](http://imdb.com)

Showname	Type	Network	Runtime	Number of
----------	------	---------	---------	-----------

			(mins)	times aired
Battery Park	Comedy/ Crime	NBC	30	4
Action	Comedy	FOX	30	12
Love & Money	Comedy	CBS	30	3
Get Real	Comedy/ Drama	FOX	60	21
Greed	Game Show	FOX	60	46
Stark Raving Mad	Comedy	NBC	30	21
Once and Again	Drama	ABC	60	26
Work with me	Comedy	CBS	30	4
Caulfield	Drama	FOX	60	2
Movie Stars	Comedy	WB	30	27
Mission Hill	Comedy/ Animation	WB	30	2
Malcom in the Middle	Comedy	FOX	30	27
Ladies Man	Comedy	CBS	30	27
City of Angels	Drama	CBS	60	12
Cold Feet	Drama	NBC	60	4
DC	Drama	WB	60	4
Family Law	Drama	CBS	60	24
Freaks and Geeks	Drama/ Comedy	NBC	60	12
God, Devil & Bob	Comedy/ Animation	NBC	30	3
WWF Smackdown	Action/ Adventure	UPN	120	44
Wonderland	Drama	ABC	60	2
West Wing	Drama	NBC	60	32
Judging Amy	Drama	CBS	60	35
Now and Again	Action/ SciFi	CBS	60	25
Odd Man Out	Comedy	ABC	30	13
Oh Grow Up	Comedy	ABC	30	11
The Mike O'Malley Show	Comedy	NBC	30	2
The Parkers	Comedy	UPN	30	43
Popular	Comedy/ Drama	WB	60	44
Roswell	Drama/ SciFi	WB	60	35
Safe Harbor	Drama	WB	60	17
Shasta McNasty	Comedy	UPN	30	26
Snoops	Drama/ Crime	ABC	60	10
Law and Order: Special Victims Unit	Drama/ Crime	NBC	60	34
The Beat	Drama	UPN	60	6

Talk to me	Comedy	ABC	30	3
Then came you	Comedy	ABC	30	6
The others	SciFi	NBC	60	14
The Strip	Drama	UPN	60	16
Third Watch	Drama	NBC	60	32
Time of your life	Drama	FOX	60	13
Angel	Action/ Drama	WB	60	41
Harsh Realm	Drama/ SciFi	FOX	60	3
Grown-Ups	Comedy	UPN	30	43

**Table 2A: Duration by network**

Network	Number of new shows	Min episodes	Max episodes	Average episodes (per show)
ABC	7	2	26	10.1
CBS	7	3	35	18.6
NBC	10	2	34	15.8
FOX	7	2	46	17.7
UPN	6	6	44	29.7
WB	7	2	44	24.3
Total	44	2	46	18.9

**A thread on Usenet dealing with a WB show "Roswell"**

(Note: we have deleted the signatures to shorten the posts, but all else, including the grammar, is unaltered)

From: Spooky Alex (mflulder@mindspring.com)

Subject: OT: Roswell on the WB

Newsgroups: alt.tv.x-files

Date: 1999/10/06

did anyone see this show? it was like a cross between 'dawsons creek' and '3rd rock from the sun'. so what do you guys think of it?

From: Steven Weller (az941@lafn.org)

Subject: Re: OT: Roswell on the WB

Newsgroups: alt.tv.x-files

Date: 1999/10/07

In another thread, I dubbed it Dawson's Crash, so I think we probably agree on it.

From: Jeff Burden (jeff922@aol.com.x-files)  
Subject: Re: OT: Roswell on the WB  
Newsgroups: alt.tv.x-files  
Date: 1999/10/07

I watching Dawsons Creek (yes I like, so :P) and saw the previews for it, it looked good, but I was to geeked to watch "The West Wing" which was very good tonight.

From: Phil R. (TwoCentsWorth\_@webtv.net)  
Subject: Re: OT: Roswell on the WB  
Newsgroups: alt.tv.x-files  
Date: 1999/10/07

What a great little show. I was so impressed. The acting, the settings and the characters were all great. And the music was fitting as well. I loved the song 'Crash' by The Dave Mathews Band at the end of the show. I will definitely be catching this one weekly. I like the blonde alien <raising eyebrows up and down>.

Phil

From: C. Morgan (mako1@herald.infi.net)  
Subject: Re: OT: Roswell on the WB  
Newsgroups: alt.tv.x-files  
Date: 1999/10/08

I forgot to mention the music! I liked the use of Crash and the Garbage song, but especially Sarah McLachlan's "Fear." That was extremely effective.

Connie

### **Arellano and Bond Estimator**

First, let us assume that there are no exogenous variables. The basic equation to be estimated becomes:

$$R_{i,t} = \eta_i + \beta R_{i,t-1} + \varepsilon_{i,t}, \quad t \in [2, \bar{T}] \quad (A1)$$

Assume:

$$1) |\alpha| < 1$$

$$2) E[\epsilon_{i,t}] = E[\epsilon_{i,t}\epsilon_{i,s}] = 0 \quad t \neq s$$

Taking first difference of A1):  $\Delta R_{i,t} = \beta \Delta R_{i,t-1} + \Delta \epsilon_{i,t}$  (A2) Note once again that we cannot assume that  $\Delta R_{i,t-1} \perp \Delta \epsilon_{i,t}$ . However, due to assumption 2, we can see that  $R$  lagged two periods or more can be used as instruments for (A2). The linear moment restrictions implied by the model are:

$$E[(\Delta R_{i,t} - \beta \Delta R_{i,t-1}) R_{i,t-j}] = E[\Delta \epsilon_{i,t} R_{i,t-j}] = 0 \quad (j = 2, \dots, (t-1), t = 3, \dots, T) \quad (A3)$$

Note that this implies that we can only use those shows with at least 3 episodes in order to estimate the equation above. Also note that there are more instruments available for later periods (See DPD98 User's Guide):

Equations	Instruments Available
$\Delta R_{i,3} = \beta \Delta R_{i,2} + \Delta \epsilon_{i,3}$	$R_{i,1}$
$\Delta R_{i,4} = \beta \Delta R_{i,3} + \Delta \epsilon_{i,4}$	$R_{i,1}, R_{i,2}$
$\Delta R_{i,5} = \beta \Delta R_{i,4} + \Delta \epsilon_{i,5}$	$R_{i,1}, R_{i,2}, R_{i,3}$

An optimal GMM estimator based on sample moments is available and is derived in Arellano and Bond (p. 279) (where  $Z$  is the instrument matrix):

$$\hat{\beta} = \arg \min_{\beta} ((\Delta \bar{\epsilon})' Z) A_N (Z' (\Delta \bar{\epsilon})) \quad (A4)$$

The optimal choice of  $A_N$  is the inverse of the sample variance-covariance matrix. Thus, there are two possible estimators:

- A) One-step estimator,  $\hat{\beta}_1$ , where  $A_N = (N^{-1} \sum_i Z_i' H Z_i)^{-1}$ . Here  $H$  is a  $(T-2)$  square matrix which has twos in the main diagonal, minus ones in the first subdiagonal and zeroes elsewhere.
- B) Two-step estimator,  $\hat{\beta}_2$ , where  $A_N^{-1} = N^{-1} \sum_i Z_i' (\Delta \bar{\epsilon}_i)(\Delta \bar{\epsilon}_i)' Z_i$ . Here the first-step residuals are estimated using  $\hat{\beta}_1$ .

Under the assumption of homoskedasticity and independence of the error terms,  $\hat{\beta}_1$  and  $\hat{\beta}_2$  are asymptotically equivalent. However, Arellano and Bond recommend using the one-step results for inference since two-step standard errors have a downward bias, which we proceed to do in this paper. We can also calculate standard errors that are robust to heteroskedasticity, which we also do in this paper.

Next, consider the addition of exogenous variables into our main equation:

$$R_{i,t} = \eta_i + \beta R_{i,t-1} + \phi \text{TotPost}_{i,t-1} + \alpha \text{Entropy}_{i,t-1} + \epsilon_{i,t} \quad (A5)$$

There are two possible cases that we need to consider. First, suppose that the chat measures are *strictly exogenous*. That is,

$$E[\text{TotPost}_{i,s}\epsilon_{i,t}] = 0, E[\text{Entropy}_{i,s}\epsilon_{i,t}] = 0, E[\text{Depth}_{i,s}\epsilon_{i,t}] = 0, \forall s, t.$$

In this case, the exogenous variables can be used as their own instruments. (The differenced chat measures are entered into the instrument matrix,  $Z$ ). This is one of the specifications that we present in the paper.

Next, we need to consider the case where some exogenous variables are predetermined. That is, the variables are orthogonal to the future shocks, but not the past or present shocks. According to a reasonable specification, this is the case for the *total posts* variable:

$$E[\text{TotPost}_{i,t-1}\epsilon_{i,s}] = 0 \text{ if } s \geq t, \text{ but } E[\text{TotPost}_{i,t-1}\epsilon_{i,s}] \neq 0 \text{ if } s < t.$$

This is due to the fact that the total chat is a function of the ratings of the episode that precedes it. But *entropy* is strictly exogenous since both are independent of the volume of posts according to our model.

We need to instrument for the predetermined exogenous variables. (Since we cannot assume that  $(\text{TotPost}_{i,t-1} - \text{TotPost}_{i,t-2}) \perp (\epsilon_{i,t} - \epsilon_{i,t-1})$ ). In this case, we can use  $\text{TotPost}_{i,t-2}$  or earlier lags as instruments, since  $\text{TotPost}_{i,t-2} \perp (\epsilon_{i,t} - \epsilon_{i,t-1})$ .

Arellano and Bond results also extend to analysis of unbalanced panel.

## Arellano and Bond: Specification Tests

### 1) Arellano Bond Test for Serial Correlation

In order to simplify the notation, suppose that we want to estimate

$$\Delta y = \Delta x\delta + \Delta \varepsilon \quad (A6)$$

where all the  $x$  are potentially correlated with the fixed effect. Given a consistent estimator under the null of no serial correlation, we have an expression for estimated residuals and for

lagged residuals. Thus,  $\Delta \hat{\varepsilon} = \Delta y - \Delta x\hat{\delta} = \begin{bmatrix} \hat{\varepsilon}_{i,3} - \hat{\varepsilon}_{i,2} \\ \dots \\ \hat{\varepsilon}_{i,t-1} - \hat{\varepsilon}_{i,t-2} \\ \hat{\varepsilon}_{i,t} - \hat{\varepsilon}_{i,t-1} \\ \hat{\varepsilon}_{i,t+1} - \hat{\varepsilon}_{i,t} \\ \dots \end{bmatrix}$ . Also, define  $\Delta \hat{\varepsilon}_{-1}$  as estimated

residuals lagged once and  $\Delta \hat{\varepsilon}_{-2}$  as estimated residuals lagged twice. Under the null hypothesis of no first order correlation in the error term, we expect that  $E[\Delta \hat{\varepsilon} \Delta \hat{\varepsilon}_{-2}] = 0$  since  $(\hat{\varepsilon}_{i,t+1} - \hat{\varepsilon}_{i,t}) \perp (\hat{\varepsilon}_{i,t-1} - \hat{\varepsilon}_{i,t-2})$ . Arellano and Bond (p.282) propose a test statistic based on this assumption. The statistic's asymptotic distribution is standard normal. Note that this test is only define for  $T \geq 5$ .

## 2) Sargan Test of Over-Identifying Restrictions

When the number of instruments is greater than the number of regressors, we can perform the Sargan test. The test is based on the assumption that the instruments are orthogonal to the error vector:  $E[Z' \Delta \epsilon] = 0$ . We calculate the sample errors using a two-step estimates. The test has a limiting chi-square distribution. The degrees of freedom is the number of columns of  $Z$  minus the number of regressors. The null of valid specification cannot be rejected if the statistic is small enough.

A Sargan test based on one-step estimates that is also robust to heteroskedasticity is not available. The test is only valid when errors are i.i.d. over time. In other circumstances, the serial correlation test must be used.

Note that we can potentially use the Sargan test to compare the validity across different specifications. For example, if we are not sure if an exogenous variable is pre-determined or strictly exogenous, we could use the Sargan to test between the two specifications.

## Essay 3: The Influence of Social Networks on The Effectiveness of Promotional Strategies

### 1. Introduction

In the past two years, marketers have been turning to unorthodox communication strategies, such as “buzz marketing” in order to promote their products. According to Renie Dye of McKinsey, lots of brands are engaging in the practice, “In the past couple of years, it has gone from being a fringe marketing strategy favored by small shops to a mainstay of the Fortune 500. Virtually every major U.S. brand -- from the staid (Ford, General Electric, Volvo) to the hip (Nike, Tommy Hilfiger, Palm) -- is dabbling in some form of the practice.”<sup>1</sup>

The defining characteristic of these strategies is that the sellers approach the consumers directly, either in online chat rooms or in physical locations such as cafes or nightclubs. For example, in the summer of 2001, Vespa had hired good-looking models to showcase their scooters in LA cafes. These hired actors recommended boutiques where Vespas were available. Similarly, Hebrew National hired “mom squads” consisting of Jewish mothers to grill hot dogs in neighborhood barbecues.<sup>2</sup> The goals of such strategies are twofold: to turn the approached consumer into a buyer and a missionary. That is, it is hoped that the consumer will spread the word further. In this paper we ask when such buzz marketing is more efficient than traditional advertising in which buyers are approached en masse.

---

<sup>1</sup> “The Cool Kids are Doing It. Should You?” (John Gaffney), *Business 2.0*, October 2001.

The unique characteristic of buzz is that its diffusion is dependent on one's neighbors in the network. Advertising, on the other hand, allows the firm to communicate with consumers independently of their neighbors. Thus, we assume that a "buzz" message diffuses from one neighbor to another, while an ad may reach an individual independently of the neighbor. At the same time, however, the consumer is less likely to pass on an ad message to his neighbor.

We analyze the differential effectiveness of buzz marketing and traditional advertising in the context of a network. We examine how several parameters affect the differential effectiveness of the two promotional tools; variables such as: the structure of the network, the size of the community and the target segment, the salience of the message, and the ability to target traditional advertising.

In the basic model, we compare diffusion of a message across three simple types of social networks (a uni-directional ray, a bi-directional ray, and a circle) that vary in the degree of connectedness. We then examine how the firm's investment influences the expected diffusion of information. We find that there are two types of investments that have qualitatively different effects on the diffusion of buzz. The first type of investment – one that involves hiring multiple marketers to spread the message – exhibits decreasing marginal returns. On the other hand, an investment that involves changing the consumer's interaction pattern may exhibit increasing marginal returns.

---

<sup>2</sup> "Buzz Marketing," (Gerry Khermouch and Jeff Green), Business Week, July 30, 2001.

Finally, we make two types of comparisons between a buzz campaign and mass advertising. First, we compare the effectiveness of buzz and mass ads as stand-alone instruments. Second, we examine a model where the firm can choose to invest in both a buzz campaign and mass advertising. We examine how the share of the firm's budget varies with the parameters of interest.

## **2. Literature on Networks**

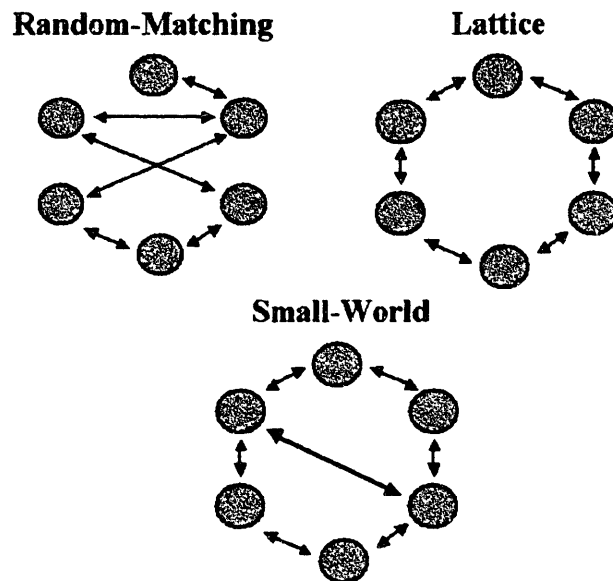
This paper relates to the existing literature on social networks and epidemiology. The properties of networks are of interest to very diverse fields: the papers discussed here span the fields of sociology, economics, and physics. The examination of the effects of social networks in economics is a very active field, as well as is the study of small-world networks.

There are three broad classes of networks that have been considered. (See Figure 1 for some examples<sup>3</sup>). In the first class, any pair of customers may communicate with each other based on random facts. Such networks are often studied in epidemiological literature, as well as in economics literature as we discuss later. These networks have the property of short average path lengths between any two individuals. In the second class, much more structure is imposed: consumers communicate with their immediate neighbors only. Such networks are usually described in terms of multi-dimensional lattices and may more accurately mimic the geographic layout of a community. The small-world networks (Watts and Strogatz (1998)) are a hybrid of these two classes and exhibit characteristics of both: short average path length and some structure.

---

<sup>3</sup> Figure 1 in Watts and Strogatz (1998)

**Figure 1: Three Classes of Networks**



We examine how the relevant literature fits into the three classes above. Random matching models are often studied in epidemiology. A more recent literature that deals with the AIDS epidemic has extended the random matching model to include a more realistic structure where there are different patterns of infection within and across groups of people with different sexual practices (see, for example, Kaplan et al (1989)).

In the economics literature, the random matching model is often used to describe the inter-agent interactions within a community. As an example, consider Kandori (1992). This paper shows how repeated games self-enforcing contracts can be maintained even if the agent can only observe own experience. The nature of the interaction involves random pairings of individuals who proceed to play a game with each other.

Lattice structures are also often employed in the economic literature. See, for example, Mobius (2001). Interestingly, a paper by Bala and Goyal (2000) derive the lattice structure as an equilibrium of social network formation. In their model, the benefit of each link is a function of the well-connectedness of one's neighbors, but the cost of maintaining a link is incurred by the initiator of the link only. Thus, the authors can employ the Nash equilibrium concept of non-cooperative game theory. The authors find that the limiting networks have simple architectures, such as a wheel or a star. The networks considered in this paper are closely related to the wheel network.

Next, we turn to small-world networks. These networks were first pioneered in Watts and Strogatz (1998). In that study, the authors show that the neural network of a worm, the power grid of western US, and collaboration network of movie actors are small-world networks. More recent studies have explored the analytic properties of these networks. For example, see Newman and Watts (1999).

Finally, we briefly describe the findings of a recent paper by Pastor-Satorras and Vespignani (2001). The authors analyze networks where the probability that a node has  $k$  connections follows a scale-free distribution:  $P(k) \sim k^{-\gamma}$  (this essentially implies that some nodes have a very large number of connections). The authors argue that such a model applies to the Internet. The study concludes that such networks are prone to spreading and persistence of infections (as is apparently the case with computer viruses). The results are very intriguing, even though it is not clear whether this network model may apply to social networks.

### **3. A Model of Buzz**

We model the main differences between traditional advertising and buzz promotion in the following way. The basic characteristic of buzz is that its diffusion is dependent on one's neighbors in the network. On the other hand, advertising allows the firm to communicate with consumers independently of their neighbors. Thus, we assume that a buzz message diffuses from one neighbor to another, while an ad may reach an individual independently of the neighbor. At the same time, however, the consumer will not pass on an ad. (Similarly, we could assume that a buzz message is passed on with a higher probability than an ad).

In a buzz campaign, the diffusion of information is determined by the size and structure of the social network and the underlying salience and interest in the product. It is important to note that we assume that the underlying network structure is outside of the marketer's control. In an extension, we also allow the marketer's investment to influence how widely the consumer passes on the message.

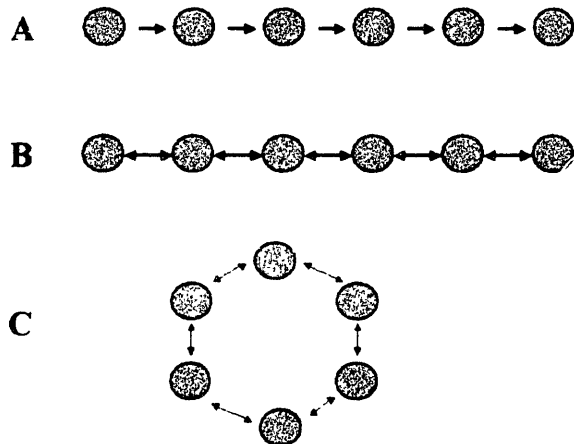
#### **3a. Network Structure**

The goal of the paper is to contrast buzz and traditional advertising as communication mechanisms. Since the structure of the network plays a crucial part in this comparison, we begin by exploring the effect of various network structures on the effectiveness of buzz. In this section, we define the network structures we consider in this paper. In the next section, we compare the diffusion of a message across the networks.

In this paper, we deal with the second class of networks, (as defined earlier): an ordered lattice. This structure allows us to model the basic trade-off between ads and buzz. We choose this class over the more random networks since a lattice may more reasonably describe interaction in a community. Note that even in online communities there may still be some geographic co-location since people often exhibit a lot of inertia by going to the same set of communities. In that sense, visitors to the same site may be considered neighbors. We do not consider small-world networks since they are much less analytically tractable.

Within the class of lattice networks, we define three distinct structures (see Figure 2). A *uni-directional ray* structure consists of nodes positioned on a line, where each node may pass information on to his right-hand neighbor. A *bi-directional ray* consists of nodes in a line, where each node may pass information on to nodes that are immediately adjacent to him. A *circle* consists of nodes in a circle, where each node that may pass information on to nodes that are immediately adjacent to him. Note that the circle and the bi-directional ray are equivalent in the limit, as  $N$  approaches infinity.

**Figure 2: Lattice Structures: (A) Uni-directional Ray, (B) Bi-directional Ray, (C) Circle**



Note that the above structures allow us to see how changes in connectedness influence the diffusion of buzz messages. The basic structure can be further modified to make the network more connected by allowing each consumer to be connected to non-adjacent neighbors.<sup>4</sup> This modification will be discussed in a later section. In the following section, we compare the diffusion of a message cross the three network structures.

### 3b. Transmission of Information in a Network: Buzz Promotion

While there has been a large literature in economics that deals with networks (both strategic and non-strategic), there has been little work done on the “infiltration” of a network by firms. The only paper, to our knowledge, is “Death through Success: The Rise and Fall of Local Service Competition at the Turn of the Century” (Markus Mobius, January 15, 2001). This paper deals with network dynamics and demonstrates how minority networks can thrive in a low state of development.

Here we assume that a firm infiltrates the network at an opening (such as a nightclub) and passes the message to the adjacent nodes/customers. In this simple model, we allow only one such opening to exist. One can consider adding more openings to the network, an extension that we consider in a later section.

We assume that there are  $N$  people in the network and two segments in the market. Under full information, consumers who belong to segment A would choose to purchase the product, while consumers who belong to segment B would not choose to purchase the product. We assume that there are and that each customer belongs to segment A with probability  $\theta$ . Assume that the customers (but not the firm) know their neighbors’

---

<sup>4</sup> A similar extension to a simple circle lattice is discussed in Watts, Strogatz (1998).

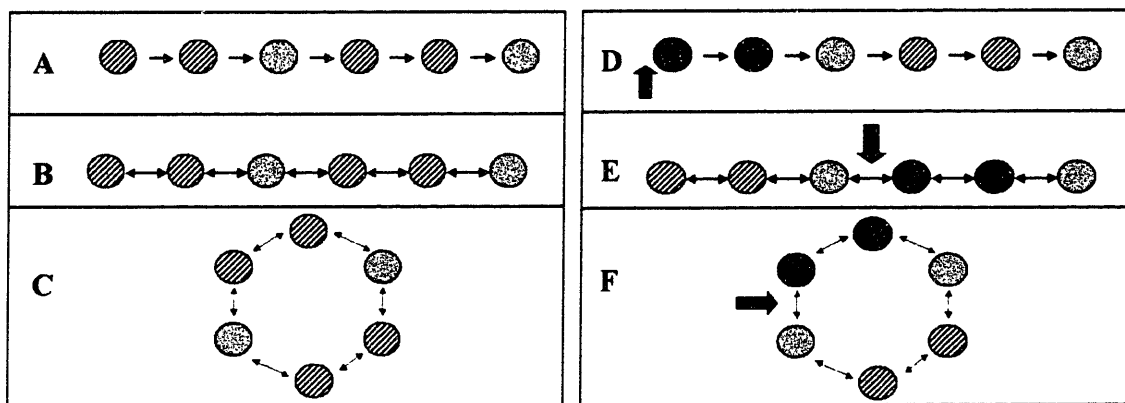
preferences. That is, the firm does not observe the preferences of each node, but it does observe the *structure* of the network (e.g.: it observes whether the network is a ray and the location of the endpoints of the ray). Here we assume that a customer will only pass on a buzz message to her neighbor if the neighbor would be interested in the message.

We also assume that people are randomly arrayed in the lattice with respect to their interest in the product. Thus, each consumer belongs to segment A with probability  $\theta$ , independent of the neighbors' preferences. Note that if the network formation were related to the consumers' interest in the product, this assumption would not hold. We argue that while consumers do have some discretion in forming social networks, some of the network formation is geography or blood-based. Thus, there is a limit on the dependencies in neighbors' preferences.

In addition to not passing on information that may not interest a neighbor, a consumer may not pass on a message due to other factors. The probability that the message will be passed may depend on the salience of the product, which we model by stating that a consumer passes on the message with probability  $s$ . This probability may also be a function of the strength of the bond in the network.<sup>5</sup> That is, a stronger bond would imply that the consumers have more exposure to their neighbors, implying a higher likelihood of information being passed on. We model the propensity to pass on information with parameter  $0 < s < 1$ . In future research, we plan on differentiating between the two factors described above.

Below in Figure 3, we illustrate the outcome of the diffusion. In Figure 3, A-C, the solidly shaded gray nodes represent consumers who are not interested in the product. The hatched nodes are the potential buyers of the product. In Figure 3, D-F, we show the diffusion of the message across the three different networks. Here, the solidly shaded black nodes have been exposed to the buzz message (by assumption, these nodes are also interested in the product). As illustrated below, the presence of a single non-interested node (gray node) stops the flow of the buzz message, as can a failure to pass on the message due to forgetting.

**Figure 3: Diffusion of a message across A) Uni-directional ray B) Bi-directional ray C) Circle**



Note that the placement of “buzzers” we assume (and as portrayed in the Figure above) is in fact the optimal placement. Thus, the firm optimally places a “buzzer” at the left endpoint of a uni-directional ray, and a “buzzer” in the middle of a bi-directional ray. The firm is indifferent with respect to the placement in a circle. Note that if the firm were not able to observe the structure of the network and thus performed a random entry, the effectiveness of the buzz campaign in the two rays would decrease relative to the effectiveness in the circle network.

<sup>5</sup> See Granovetter (1973).

**Proposition 1:** Let  $\rho = s\theta$ . The expected number of people exposed to the buzz message above is:

A) For a *uni-directional ray* network,  $V_R = \frac{\rho}{1-\rho}(1-\rho^N)$  (1)

B) For a *bi-directional ray* network,  $V_S = \frac{2\rho}{1-\rho}(1-\rho^{N/2})$  (2)

C) For a *circle* network,  $V_C = \frac{2\rho}{1-\rho}(1-\rho^N) - N\rho^N$ . (3)

These functions are concave and increasing in  $N$  for all  $0 < \rho < 1, 0 < s < 1, N > 2$ . They are increasing in  $\rho$  for all  $0 < \rho < 1, 0 < s < 1, N > 1$ . Assuming a fixed one-time cost (wage of hiring a chatter, for example), a zero marginal cost and a price of  $\pi$ , the profit from a buzz strategy is  $\pi V_i - w$  where  $V_i \in \{V_R, V_S, V_C\}$ .

See the Appendix for the proof of the proposition above. Our intuition would dictate that the expected diffusion would be highest in the circle since it is the most connected of the three networks. Similarly, we would expect the diffusion to be least in a uni-directional ray since it is the least connected. The next Lemma bears out this intuition.

**Lemma 1:**

$$V_R < V_S < V_C < 2V_R \text{ for } 2 < N < \infty. \quad (4)$$

Note that  $\lim_{N \rightarrow \infty} V_S = \lim_{N \rightarrow \infty} V_C = \frac{2\rho}{1-\rho}$  and  $\lim_{N \rightarrow \infty} V_R = \frac{\rho}{1-\rho}$ . Thus, the distinction between a

circle and a bi-directional ray vanishes, while the expected informed nodes in a ray are

half that of a circle in the limit. This agrees with our intuition since in the circle and in the bi-directional ray each customer talks on average to twice as many people as customers in a uni-directional ray, and with an infinite many nodes the structure of the circle is equivalent to the structure of the bi-directional ray.

### 3c. The Firm's Investment into Buzz

In the previous section, we allow only one entry into the social network. Here, we allow the firm to invest into making multiple entries into the network. Thus, consider the example of Hebrew National hiring “mom squads” to promote hot dogs to their neighbors. Presumably, Hebrew National can set the intensity of promotion by deciding how many moms they hire to target a particular neighborhood. We also consider a very different investment, where we allow the firm to change the consumer's interaction pattern. Thus, we allow the firm to influence the number of neighbors with whom a consumer may share the buzz message.

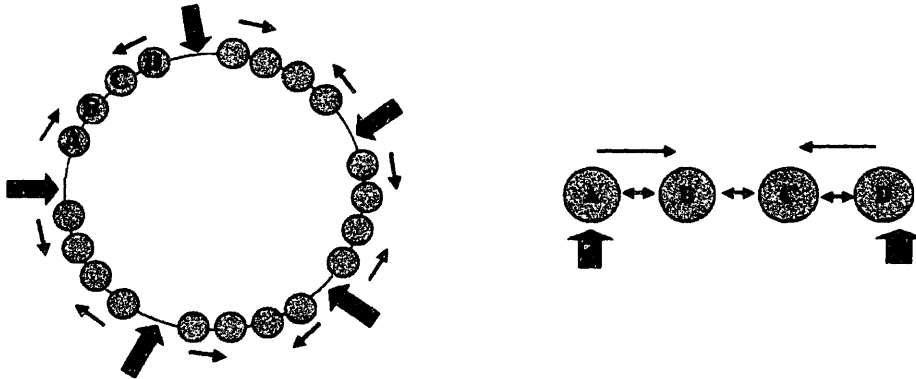
One helpful way in which we can differentiate between the two types of investments is to think about the former as an investment into the intensity of the promotion, while the latter is an investment into the message itself. Thus, to once again use the Hebrew National example, a catchy campaign such as “mom squads” may result in more “buzz per consumer” versus a less successful campaign where hot dogs are given out at a mall.

First, we address the investment into multiple openings (the intensity of the push). Just as we had done for the single entry model, we now derive the expected number of affected individuals for each of the three networks:

### **Circle**

A circle lattice structure with  $k$  evenly spaced entries consists of a set of  $k$  sections of size  $\frac{N}{k}$ . The expected number of aware individuals in each section is equivalent to the expected number of aware nodes in a circle of size  $\frac{N}{k}$  with one entry (see Appendix for illustration). In Figure 4, on the left we illustrate a circle with 20 nodes and 5 entries. A section of the circle between the two entries is pictured on the right. It has the same expected number of informed nodes as a circle with 4 nodes and one entry.

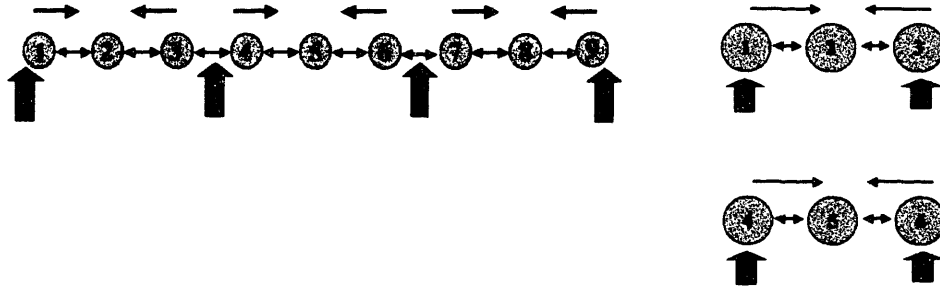
**Figure 4: Multiple entries in a circle network**



### **Bi-directional ray**

For simplification purposes, we assume that the firm chooses to place marketers at the endpoints of the bi-directional ray as long as  $k > 1$ .<sup>6</sup> A bi-directional ray lattice structure with  $k$  evenly spaced entries consists of a set of  $k-1$  sections of size  $\frac{N}{k-1}$ . The expected number of influenced nodes in each section is equivalent to the expected number of aware nodes in a circle with one entry of size  $\frac{N}{k-1}$ . For an illustration of a bi-directional ray with 9 nodes and 4 entries, see Figure 5. Once again, two sections of size 3 are pictured on the right.

**Figure 5: Multiple entries in a bi-directional ray**



### **Uni-directional ray**

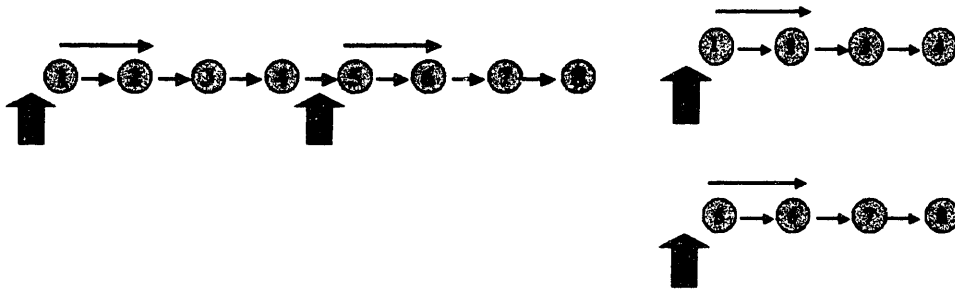
A uni-directional ray lattice structure with  $k$  evenly spaced entries consists of a set of  $k$  sections of size  $\frac{N}{k}$ . The expected number of aware individuals in each section is

---

<sup>6</sup> We have also analyzed the scenario where the entries are evenly spaced, but no entries are made at endpoints. This complicates the proofs but does not qualitatively change the results in Proposition 2.

equivalent to the expected number of aware nodes in a uni-directional ray of size  $\frac{N}{k}$  with one entry. In Figure 6, on the left we show a uni-directional ray of size 8 with two entries. On the right we show the two sections of size 4.

**Figure 6: Multiple entries in a uni-directional ray**



Since we are interested in the return to investment in buzz, we need to find the expected number of informed nodes as a function of  $k$  and  $N$ . By summing the expected number of informed nodes in each section, we arrive at the following expressions:

**Proposition 2:** The expected number of people exposed to the buzz message with  $k$  entries is:

A) For a *uni-directional ray* network,  $V_R^k = k \frac{\rho}{1-\rho} (1 - \rho^{\frac{N}{k}})$  (5)

B) For a *bi-directional ray* network,

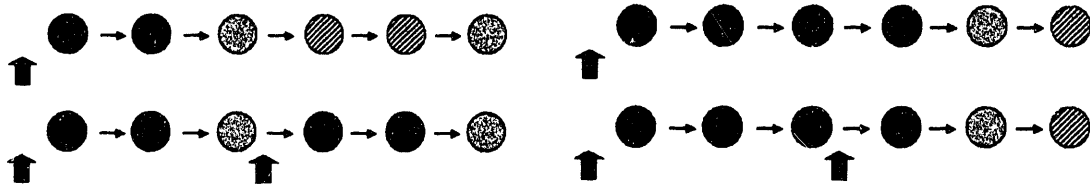
$$V_s^k = \begin{cases} \frac{2p}{1-p} (1 - \rho^{\frac{N}{2}}) & \text{if } k = 1 \\ (k-1) \left[ \frac{2p}{1-p} (1 - \rho^{\frac{N}{k-1}}) - \frac{N}{k-1} \rho^{\frac{N}{k-1}} \right] & \text{if } k > 1 \end{cases} \quad (6)$$

$$\text{C) For a circle network, } V_c^k = k \left[ \frac{2p}{1-p} (1 - \rho^{\frac{N}{k}}) - \frac{N}{k} \rho^{\frac{N}{k}} \right]. \quad (7)$$

The above functions are increasing and concave in  $k$  as long as  $1 \leq k < \frac{N}{2}$ .

We provide an example that illustrates the decreasing marginal returns result. Let's consider a uni-directional ray with  $N$  nodes. We want to show that increasing number of entries from one to two less than doubles the expected number of informed nodes. Let's set  $s = 1$  and consider an unbroken chain of interested nodes, starting from the left. If the length of the chain is less than  $\frac{N}{2}$ , then adding another entry in the middle can potentially double the number of informed nodes (see left side of Figure 7). On the other hand, if the length of the chain is  $\geq \frac{N}{2}$  (see right side of Figure 7), adding entry has no effect on the number of informed nodes. Thus, adding another marketer less than doubles the expected number of informed nodes.

**Figure 7: Uni-directional ray -- 1 v. 2 entries**



Next, let us consider whether the company's investment in promotional buzz always exhibits decreasing marginal returns. After all, most of the pop business literature seems to imply that there are non-linearities involved in this type of promotion – a clever enough buzz promotion may suddenly turn into an avalanche of publicity.<sup>7</sup> This seems to imply that an investment above a certain threshold produces a much larger payoff compared to an investment below the threshold. In order to address this question, we consider an investment that changes the fundamentals of the diffusion process. Once again, this can be thought of as an investment into the effectiveness of the message (a more memorable campaign, etc).

We allow the firm to make a different type of investment. Let us consider a generalized uni-directional ray network, where consumers can talk to  $m$  of their adjacent neighbors (see Figure 8). Let us also, for analytical tractability, assume that  $N$  is infinite.

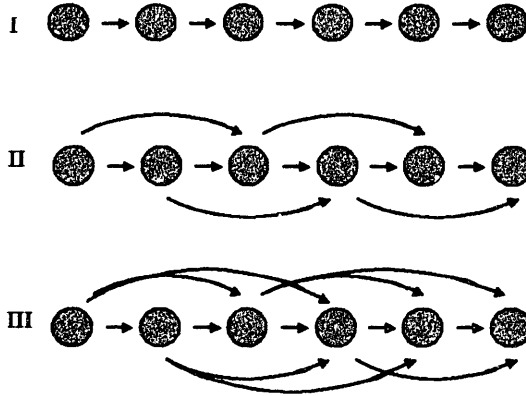
<sup>7</sup> See, for example, "The Tipping Point," Gladwell, Malcolm. Little, Brown and Company, 2000.

**Figure 8: A uni-directional ray structure where each node may pass info on to  $m$  adjacent right-hand-side neighbors**

**I: each node talks to 1 neighbor ( $m = 1$ )**

**II:  $m = 2$**

**III:  $m = 3$**



Suppose that each person is in principle connected to up to  $M$  of his neighbors. We can imagine that the further connections are more distant. After hearing the message, a consumer can decide how many of these neighbors he should contact with this information, perhaps based on some function of the information (such as entertainment value, etc). Here we assume that he would contact the closer neighbors first. This allows us to analyze the change in the expected number of nodes as a function of  $m$ .

**Proposition 3:**

In a uni-directional ray network,  $N = \infty$ , where each consumer contacts  $k$  neighbors, the

expected number of exposed consumers is:  $\frac{\rho}{(1-\rho)^m}$ . (8)

This function is convex in  $m$ .

Here we see that changing  $m$  would result in an exponential growth. Of course, changing the diffusion process may in fact prove to be a very difficult task, whereas hiring more

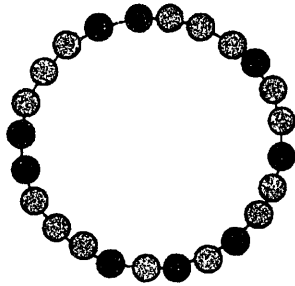
“buzzers” is relatively simple. This simple analysis would allow us to calculate the trade-offs between these two types of investments.

#### **4. A Model of Mass Advertising**

We proceed to contrast buzz promotion with mass advertising. Examples of mass advertising include TV and radio ads, magazine ads, and highway billboards. We allow the firm to optimally set the level of exposure to advertising. Thus, the firm buys an optimal number of minutes of TV or radio ads.

Note that even though in both cases the firm is ignorant of the consumers’ preferences, a buzz campaign is more targeted than an ad campaign since consumers pass on a buzz message only to other interested consumers. An ad, however, may be ignored or may reach an uninterested consumer. We vary the degree of targeting that a firm is able to do in an ad campaign with the parameter  $0 < q < 1$  – the probability that an interested individual is influenced by the ad. The diffusion pattern for a mass advertising campaign is shown in Figure 9. From the illustration, we can see that in our model the probability that an ad influences a node is independent of the node’s neighbors.

**Figure 9: The Diffusion of Advertising Within a Community**



We choose a simple model of advertising that allows us to model the trade-off between ads and buzz. We allow the firm to choose the level of exposure ( $\mu$ ) of each individual to the ad. The benefit to the firm of such exposure is simply  $\pi N \theta \mu q$  -- price  $\times$  expected number of interested individuals  $\times$  level of exposure  $\times$  ability to target an ad.

The cost of the ad campaign is linear in the number of individuals exposed. This is definitely the case in TV ads: sponsors pay based on viewer points – ratings of the show.

We also assume that the costs are convex in the level of exposure. Thus, it becomes increasingly more difficult to set the level of exposure to be very high. Thus, the cost

function of the ad placement is  $= N \frac{\mu^2}{2}$ .

#### 4a. The Choice Between Mass Ads and Buzz Promotion

First, we assume that the firm must choose *between* buzz and a mass ad campaign. This is not a very realistic assumption, but it allows us to analyze the relative strengths of either approach. In the next section, we analyze a more realistic problem where the firm can allocate portions of its budget between ads and buzz.

**Proposition 4**

In the mass ad medium only, the firm faces the following maximization:

$$\max_{\mu} \pi N \theta \mu q - \frac{\mu^2}{2} N \quad (9)$$

The optimal advertising level is  $\mu^* = \min[1, \pi \theta q]$ . In order to ensure an interior solution, let us assume that  $\pi \theta q < 1$ . The optimal profit from an ad-only campaign is

$$\frac{\pi^2 \theta^2 q^2 N}{2} = \Pi_{\text{Ads}}^* \quad (10)$$

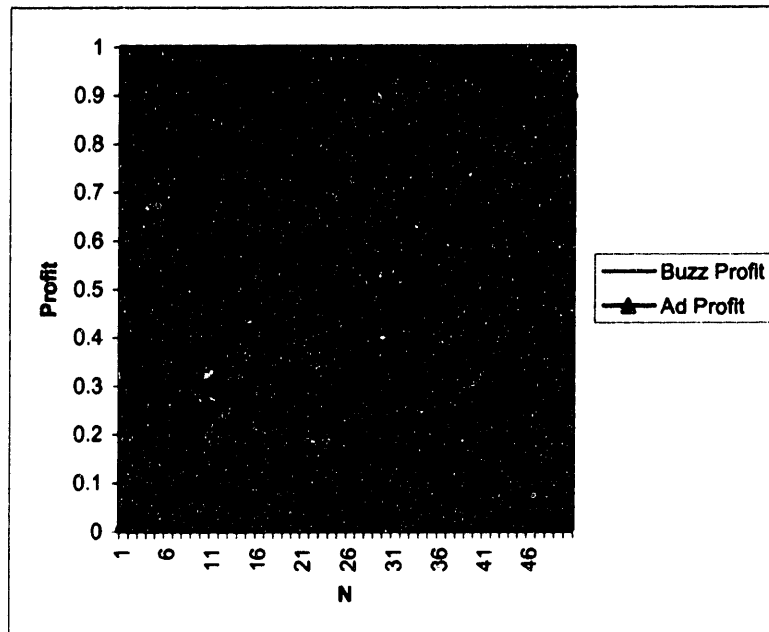
Next, let us compare the value of advertising as a tool with the value of buzz, where for the purposes of this section, we look at networks with one entry only. As we showed earlier, the profit from buzz is  $\max[\pi V_i - w, 0] = \Pi_{\text{Buzz}}^*$ .

**Lemma 2:** If the firm had a choice between employing a buzz strategy or an advertising strategy, depending on  $N$ , there are two possible outcomes:

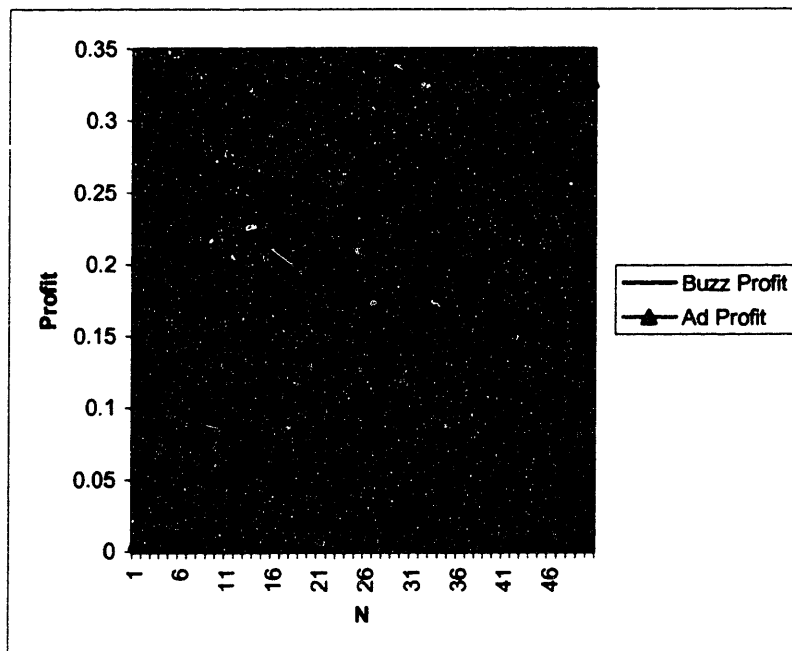
- 1) If  $\max_N \Pi_{\text{Buzz}}^* - \Pi_{\text{Ads}}^* < 0$ : the firm always chooses ads over buzz.
- 2) If  $\max_N \Pi_{\text{Buzz}}^* - \Pi_{\text{Ads}}^* > 0$ : the firm chooses to “buzz” in an intermediate region of  $N$ ,  $[N_L, N_U]$ , where  $0 < N_L < N < N_U$  and  $N_L, N_U$  satisfy  $\Pi_{\text{Buzz}}^* - \Pi_{\text{Ads}}^* = 0$ . The size of the region where buzz is chosen is decreasing with  $q$ .

Below, we graph the two cases:

**Figure 9: Case 1 – Ads Dominate Buzz As Stand-Alone Strategies**



**Figure 10: Case 2 – Buzz Dominates Ads in Intermediate Region**



Note that here the results are driven by the concavity of the buzz function in  $N$ . Thus, if the market size is very small, the fixed cost of buzz makes it unattractive option, whereas

ads are relatively cheap. As the market size becomes very large, ads are more effective since they provide a coverage that cannot be reached with buzz alone. Thus it is in the middle region where buzz is more effective.

#### 4b. Combining Mass Ads and Buzz Promotion

Next, we look at combining buzz and advertising. Once again, we allow the firm to set  $v$  (we use a symbol other than  $\mu$  to differentiate this from the previous section)-- be the probability that an individual sees an ad. Unlike the previous section, we now have to take into account consumers who are exposed to buzz and to ads. Here we assume that conditionally on being exposed to buzz, seeing an ad does not change behavior. In Future Research Section, we discuss an extension that deals in more depth with the interaction between buzz and ads.

In order to calculate the expected number of affected nodes, we need to account for all nodes that are exposed either to buzz **or** to ads. That is, we need to calculate the union of the set of nodes that are exposed to ads and buzz. In the table below, we calculate such a union:

**Table 1: Exposure to promotional message through ads and buzz**

Promotion type	Exposed to message	Interested in product
Ad	$Nqv$	$Nqv\theta$
Buzz	$V_i$	$V_i$
Buzz & Ad	$V_i qv$	$V_i qv$
Total	$V_i + Nqv - V_i qv =$ $V_i(1 - qv) + Nqv$	$V_i + Nqv\theta - V_i qv =$ $V_i(1 - qv) + Nqv\theta$

The sales of the product are thus  $V_i(1 - qv) + Nqv\theta$ : the weighted average of expected numbers of nodes influenced by ads and nodes influenced by buzz. As  $qv$  increases, a bigger share of consumers will be influenced through ads. Also note that the firm essentially pays a fixed cost only in order to implement a buzz campaign. (This cost is  $w$ , the cost of hiring a promoter). Thus, if  $w$  is too high, the firm may choose not to employ the buzz strategy at all. We take this discontinuity into account in Proposition 4 below:

**Proposition 4:** When faced with a choice of mass ad medium and buzz advertising, the firm faces the following maximization:

$$\max \left[ \max_v \pi[(1 - qv)V_i + Nqv\theta] - \frac{v^2}{2} N - w, \Pi_{\text{Ads}}^* \right]$$

That is, the firm decides between having an ad-only campaign, which yields a profit of  $\Pi_{\text{Ads}}^*$ , or a campaign that optimally combines ads and buzz. The size of  $w$ , the fixed cost of buzz, determines whether or not the firm will engage in a buzz campaign. The optimal advertising level, given that both instruments are used, is

$$v^* = \frac{\pi q[N\theta - V_1]}{N}, \text{ where } 0 < v^* < 1. \quad (11)$$

Finally, we examine the optimal allocation of budget between ads and a buzz campaign.

Suppose that both instruments are used. Let us define  $S_B = \frac{w}{w + \frac{\pi^2 q^2 [N\theta - V_1]^2}{2N}}$  -- the

share of the firm's advertising budget that is devoted to buzz. Thus, since the investment into chat is fixed,  $S_B$  summarizes the optimal investment into advertising, taking into account the interaction between ads and buzz. We can analyze how the  $S_B$  varies with the parameters in the model. See Lemma 3 for the results.

**Lemma 3**

$$1) \frac{\partial S_B}{\partial N} < 0 \quad (12)$$

$$2) \frac{\partial S_B}{\partial s} > 0 \quad (13)$$

$$3) \frac{\partial S_B}{\partial q} < 0 \quad (14)$$

$$4) \frac{\partial S_B}{\partial w} > 0 \quad (15)$$

Let us discuss the four results in turn. The share of budget in buzz decreases as the size of the community grows. Thus, as market size grows, the firm invests into buzz *and* increases the investment into advertising. This result essentially hinges on the concavity of  $V_1$  in  $N$ , as can be seen in the Appendix.

Thus, we would expect products that are targeted at bigger communities to be more heavily mass advertised. This seems to be consistent with what we observe in practice. For example, movies that may appeal to bigger communities (action movies) are more heavily advertised than movies that appeal to smaller communities (foreign movies).

Second we see that as the salience of the product or the strength of the bonds increase, the share of budget in buzz increases. This is due to the fact that an increase in salience or bond strength makes buzz more attractive, which in turn reduces the need for the firm to advertise. Hence, we would expect for a firm with a salient product to rely more heavily on buzz as compared to ads. Similarly, we would expect that buzz would be more effective in tightly-knit communities. This again is consistent with what we observe in practice. That is, buzz has been more heavily used in entertainment product categories such as music and film, and strongly bonded communities, such as science fiction fans, are often targeted for buzz.<sup>8</sup>

Third, a decrease in the ability to target an ad campaign increases the share in buzz. It is interesting to note that this result echoes some of the comments justifying the firms' disenchantment with TV ads. The complaint deals with the marketers' inability to target the teen audience with TV ads: "Expensive, conventional approaches such as network-TV ads seem less and less able to move the needle. That's particularly true when it comes to younger consumers, who because they are still forming their brand preferences,

---

<sup>8</sup> A lot of the promotion for the movie "Lord of the Rings" was targeted at online science fiction fans. ("**Rings' Has Two Targets** Plan Is to Grab the Geeks So the Masses Will Follow," (Frank Ahrens), Washington Post, December 19, 2001.

are coveted by marketers. They spend less time planted in front of the tube and are more skeptical of the messages they receive there.”<sup>9</sup>

Finally, the fourth result is pretty straightforward since an increase in a fixed cost of buzz necessarily implies an increase in share of budget devoted to buzz. Of course, as  $w$  increases, at some point the firm would choose not to invest in buzz at all. We take into account this discontinuity by allowing the firm to invest in advertising only.

## **5. Future Research**

The work presented in this paper is an initial exploration of the effect of social network on the effectiveness of promotional strategies. Here we briefly outline a few extensions that we intend to explore in the near future.

First, as we noted earlier, in this paper we assume that there is no interaction between exposure to both an ad and a buzz promotion. However, being exposed to an ad may change the probability that a consumer would pass on a buzz message, and, similarly, being exposed to a buzz message might change the probability that one pays attention to an ad. This of course would require us to pay attention to the order in which the messages are received. Furthermore, it is not clear whether ads would enhance or interfere with a buzz campaign. On one hand, after seeing an ad, the consumer may be more likely to pay attention to a buzz message. On the other hand, a consumer may not pass on a buzz message if he thinks that his friend is already aware of the product through ads. This would imply that an ad campaign would decrease the effectiveness of a

---

<sup>9</sup> “Buzz Marketing,” (Gerry Khermouch and Jeff Green), Business Week, July 30, 2001.

simultaneous buzz campaign, and would imply that the firm would prefer to temporally separate a buzz and an ad campaign.

Second, an interesting problem that is not addressed in this study is the strategic interaction between competing firms. Thus, suppose that Warner Brothers and DreamWorks plan to release two competing action movies in early December: “The Lethal Bunny” (LB) and “The Tragic Robot” (TR). Given that mass advertising as well as buzz promotion is available to both companies, what do we expect to be the equilibrium of the interaction? The answer would depend on the consumer’s reaction to hearing messages from different sources. For example, one can imagine that there are constraints on how much information can be conveyed in a conversation. This would imply that buzz surrounding LB would displace some of the buzz surrounding TR. Thus, perhaps TR would prefer to engage in mass advertising, given that LB chooses to engage in buzz promotion. However, perhaps TR would benefit from the buzz surrounding LB since more consumers would engage in movie-related conversations. This of course would have different implications for the choice of promotional strategies.

Another interesting aspect of the model that is left unexplored in this study is the issue of consumer welfare. Note that a buzz promotion is more targeted: most of the people who are exposed to the buzz are in fact interested in the product. This is not the case in mass advertising: interested as well as non-interested individuals are exposed to the message. Thus, to calculate consumer welfare, we would have to assign an “annoyance” cost to receiving messages concerning unwanted products.

Finally, in the current study we do not draw distinctions between the different consumers. However, there exists a lot of evidence to suggest that certain consumers exert more influence than others. This may be especially true in certain product categories such as fashion. Thus, marketers would be especially interested in targeting these consumers. This of course would introduce heterogeneity into the network structure.

## **6. Conclusion**

The problem of optimal communication is a very general issue that applies in many contexts. For example, consider a new CEO of an organization who intends to implement a cultural change. In deciding the best way to communicate this change, she faces exactly the problem that is described above. She has to decide whether she should influence a few managers and hope that they, in turn, will influence their subordinates. On the other hand, she could also try to communicate more directly with the rank-and-file. Similarly, a political candidate must optimally allocate his time and budget between mass ads and more personal encounters that he hopes would spur good buzz.

The contribution of this paper is in defining the problem and introducing the tools of networking into the problem. We show that the word of mouth strategies are fundamentally different from the conventional advertising methods. We also show that the specifics of the network and product characteristics must be taken into account in order to optimize the communication strategy.

## References

- 1) Bala, Vankatesh; Goyal, Sanjeev (2000) "A Noncooperative Model of Network Formation" *Econometrica*. 68, 5, 1181-1229.
- 2) Gallager, Robert (1996), *Discrete Stochastic Processes*, Kluwer Academic Publishers, 119-120.
- 3) Gladwell, Malcolm (2000), *The Tipping Point: How Little Things Can Make a Big Difference*, Little, Brown & Co.
- 4) Granovetter, M (1973), "The Strength of Weak Ties," *American Journal of Sociology*, 78 (6), 1360-1380.
- 5) Kandori, Michihiro. (1992) "Social Norms and Community Enforcement," *The Quarterly Journal of Economics*. 59, 63-80.
- 6) Kaplan, Edward; Cramton, Peter C.; Paltiel, A. David (1989) "Nonrandom Mixing Models of HIV Transmission," *Lecture Notes in Biomathematics*, 83:218-241.
- 7) Mobius, Markus (2001) "Death through Success: The Rise and Fall of Local Service Competition at the Turn of the Century", Harvard working paper.
- 8) Newman, M. E. J.; Watts, Duncan J. (1999) "Scaling and Percolation in the Small-World Network Model," *Physics Review E*. 60, 7332-7342.
- 9) Pastor-Satorras, Romualdo; Vespignani, Alessandro. (2001) "Epidemic Spreading in Scale-Free Networks," *Physical Review Letters* 86, 14, 3200-3203.
- 10) Watts, Duncan and Strogatz, Steven (1998) "Collective Dynamics of Small-World Networks" *Nature*, 393 (June 4).

## Appendix

Throughout the Appendix, I use the fact that for  $0 < \rho \leq 1$ ,  $-\ln(\rho) > 1 - \rho$ . Here we show a

quick proof of this statement. Let  $w(\rho) = -\ln(\rho) - (1 - \rho)$ .  $\frac{\partial w(\rho)}{\partial \rho} = 1 - \frac{1}{\rho} \leq 0$ ;

$\frac{\partial^2 w(\rho)}{\partial^2 \rho} = \frac{1}{\rho^2} > 0$ . From this we can see that the unique minimum is of  $w$  is at  $\frac{\partial w(\rho)}{\partial \rho} = 0$

at  $\rho = 1 \rightarrow w(1) = 0$ . This proves that  $w(\rho) > 0$  for  $0 < \rho \leq 1$ .

## Proposition 1

### Circle Network

In the circle configuration, the message is passed to the adjacent neighbor with probability  $\rho$ . If the circle had an infinite number of nodes, we could think of the total number of affected nodes as the sum of two geometric variables, which is a negative binomial variable. Since the number of nodes is finite, we have to figure out an expression that is closely related to the expectation of a truncated negative binomial variable.

More formally, if we let  $X_1$  and  $X_2$  be the number of nodes to the left and to the right of the injection of buzz that are exposed to the message. Let  $Z = X_1 + X_2$ . Then

$$P(Z = m) = \begin{cases} (m+1)\rho^m(1-\rho)^2 & \text{if } m \leq N-2 \\ (m+1)\rho^m(1-\rho) & \text{if } m = N-1 \\ \rho^m & \text{if } m = N \end{cases} \quad (\text{A1})$$

$$E(X_1 + X_2) = \sum_{m=0}^{N-2} m(m+1)\rho^m(1-\rho)^2 + (N-1)N\rho^{N-1}(1-\rho) + N\rho^N \quad (\text{A2})$$

In order to find a closed-form solution for (A2), we need to compute the first term in the expression. Below, we outline the series of steps that allow us to compute this term.

First, we need to calculate the following expression that is related to the expectation of a negative binomial variable:

$$\begin{aligned} \sum_{y=b}^{\infty} yP(Y = y) &= \sum_{y=b}^{\infty} (y+1)y\rho^y(1-\rho)^2 = \sum_{y=b}^{\infty} (n+b+1)(n+b)\rho^{n+b}(1-\rho)^2 = \\ &\rho^b \left[ \sum_{n=0}^{\infty} (n+1)n\rho^n(1-\rho)^2 + b \sum_{n=0}^{\infty} (n+1)\rho^n(1-\rho)^2 + \rho b \sum_{n=0}^{\infty} n\rho^{n-1}(1-\rho)^2 + b^2 \sum_{n=0}^{\infty} \rho^n(1-\rho)^2 \right] \end{aligned}$$

The expression above can be simplified. Note that  $\sum_{n=0}^{\infty} (n+1)n\rho^n(1-\rho)^2$  is the

expectation of a negative binomial random variable. Thus,  $\sum_{n=0}^{\infty} (n+1)n\rho^n(1-\rho)^2 = \frac{2\rho}{1-\rho}$ .

Also,  $\sum_{n=0}^{\infty} (n+1)\rho^n(1-\rho)^2 = \sum_{n=0}^{\infty} n\rho^{n-1}(1-\rho)^2 = 1$ . (Since we are summing up a pdf, and

the first term is zero).

$$\text{Thus, } \sum_{y=b}^{\infty} yP(Y = y) = \rho^b \left[ \frac{2\rho}{1-\rho} + b(1+\rho) + b^2(1-\rho) \right] \rightarrow$$

$$\sum_{y=0}^{b-1} yP(Y=y) = \frac{2\rho}{1-\rho} - \rho^b \left[ \frac{2\rho}{1-\rho} + b(1+\rho) + b^2(1-\rho) \right] \quad (\text{A3})$$

Next, we can substitute  $N=b+1$  in (A3) to obtain

$$\sum_{y=0}^{N-2} yP(Y=y) = \frac{2\rho}{1-\rho} - \rho^{N-1} \left[ \frac{(N-1)(N-2)\rho^2 - 2N(N-2)\rho + N(N-1)}{1-\rho} \right] \quad (\text{A4})$$

Combining (A2) and (A4), and after some simple algebra, we can now obtain a closed-form expression for the expectation:

$$V_C = E(X_1 + X_2) = \frac{2\rho}{1-\rho} (1 - \rho^N) - N\rho^N \quad (\text{A5})$$

Next, we show that  $V_C$  is increasing and concave in  $N$ .

$$\frac{\partial V_C}{\partial N} = \rho^N \left( -\frac{2\rho}{1-\rho} \ln(\rho) - 1 - N \ln(\rho) \right) \quad (\text{A6})$$

Since  $-\ln(\rho) > 1 - \rho$ , we see that

$$\frac{\partial V_C}{\partial N} = \rho^N \left( -\frac{2\rho}{1-\rho} \ln(\rho) - 1 - N \ln(\rho) \right) > \rho^N (2\rho + N(1-\rho) - 1) > 0 \text{ for } N > 1.$$

Similarly, we can show concavity:

$$\frac{\partial^2 V_C}{\partial^2 N} = \rho^N \ln(\rho) \left( -\frac{2\rho}{1-\rho} \ln(\rho) - 2 - N \ln(\rho) \right) \quad (\text{A7})$$

Once again, we use the fact that  $-\ln(\rho) > 1 - \rho$ . Thus,

$$-\frac{2\rho}{1-\rho} \ln(\rho) - 2 - N \ln(\rho) > \rho^N (2\rho + N(1-\rho) - 2) > 0 \text{ for } N > 2. \text{ This, in turn, implies that}$$

$$\frac{\partial^2 V_C}{\partial^2 N} < 0 \text{ for } N > 2.$$

Last, let us show that  $V_c$  is increasing in  $\rho$ .

$$\frac{\partial V_c}{\partial \rho} = \frac{1}{\rho(1-\rho)^2} [2\rho - N^2\rho^N + 2N^2\rho^{N+1} - 2N\rho^{N+1} - 2\rho^{N+1} - N^2\rho^{N+2} + 2N\rho^{N+2}] \quad (A8)$$

We need to show that

$$g = 2\rho - N^2\rho^N + 2N^2\rho^{N+1} - 2N\rho^{N+1} - 2\rho^{N+1} - N^2\rho^{N+2} + 2N\rho^{N+2} = \\ 2\rho - \rho^N [N^2(1-\rho)^2 + 2N\rho(1-\rho) + 2\rho] > 0$$

Since  $\rho < 1$ , we know that

$$N^2(1-\rho)^2 + 2N\rho(1-\rho) + 2\rho < N^2(1-\rho)^2 + 2N(1-\rho) + 2 = [N(1-\rho) + 1]^2 + 1 \quad (A9)$$

In turn, this allows us to put a lower bound on  $g$ :

$$g > b_L(N) = 2\rho - \rho^N ((N(1-\rho) + 1)^2 + 1) \quad (A10)$$

Note that at  $N=1$ ,  $b_L(1) = 0$ . Also,

$$\frac{\partial b_L}{\partial N} = \rho^N (-\ln(\rho) [N(1-\rho) + 1]^2 + 1) - 2(1-\rho)(N(1-\rho) + 1) \quad (A11)$$

$$\text{Since } -\ln(\rho) > 1 - \rho, \text{ we know that } \frac{\partial b_L}{\partial N} = \rho^N (N(1-\rho))^2 > 0. \quad (A12)$$

Thus,  $g > b_L > 0 \rightarrow \frac{\partial V_c}{\partial \rho} > 0$ . QED.

### **Uni-directional ray Network**

Here, we have to find the expectation of a truncated geometric variable.

$$P(X = m) = \begin{cases} \rho^m(1-\rho) & \text{if } m \leq N-1 \\ \rho^m & \text{if } m = N \end{cases} \quad (A13)$$

$$E(X) = \sum_{m=0}^{N-1} m\rho^m(1-\rho) + N\rho^N \quad (A14)$$

Using the technique from Lemma 1, we do the following to find a closed-form for the expression above:

$$\sum_{y=b}^{\infty} y\rho^y(1-\rho) = \sum_{n=0}^{\infty} (n+b)\rho^{n+b}(1-\rho) = \rho^b \sum_{n=0}^{\infty} n\rho^n(1-\rho) + b\rho^b(1-\rho) \sum_{n=0}^{\infty} \rho^n = \rho^b \left( \frac{\rho}{1-\rho} + b \right)$$

→

$$\sum_{y=0}^{b-1} y\rho^y(1-\rho) = \frac{\rho}{1-\rho}(1-\rho^b) - b\rho^b \quad (\text{A15})$$

If we substitute  $N=b$ , we can obtain the expression

$$\sum_{y=0}^{N-1} y\rho^y(1-\rho) = \frac{\rho}{1-\rho}(1-\rho^N) - N\rho^N. \quad (\text{A16})$$

Combining this with (A14), we get that

$$V_R = E(X) = \frac{\rho}{1-\rho}(1-\rho^N) \quad (\text{A17})$$

It is easy to show that (A17) is increasing and concave in  $N$  since  $\frac{\partial V_R}{\partial N} = -\frac{\rho^{N+1}}{1-\rho} \ln(\rho) > 0$

$$\text{and } \frac{\partial^2 V_R}{\partial^2 N} = -\frac{\rho^{N+1}}{1-\rho} [\ln(\rho)]^2 < 0$$

We next show that  $\frac{\partial V_R}{\partial \rho} > 0$ . First, note that

$$V_R = E(X) = \frac{\rho}{1-\rho}(1-\rho^N) = \frac{\rho(1-\rho)(1+\rho+\rho^2+\dots+\rho^{N-1})}{(1-\rho)} = \rho + \rho^2 + \dots + \rho^N. \text{ This}$$

function is clearly increasing and convex in  $\rho$ .

### **Bi-directional ray Network**

Once a firm infiltrates a bi-directional ray, the information will spread in both directions. (See (B) in Figure 4) It is as if we now have 2 uni-directional rays of possibly unequal size. Since we showed the expected number of exposed nodes in a uni-directional ray is a concave function of  $N$ , to maximize the number of exposed nodes, the marketer will infiltrate the bi-directional ray network in the middle in order to maximize the expected number of exposed consumers. Thus, the total number of exposed nodes in a bi-

$$\text{directional ray network is } V_S = \frac{\rho}{1-\rho}(1-\rho^{\frac{N}{2}}) + \frac{\rho}{1-\rho}(1-\rho^{\frac{N}{2}}) = \frac{2\rho}{1-\rho}(1-\rho^{\frac{N}{2}}) \quad (\text{A18})$$

We know from the proof above that  $V_R$  is increasing and concave in  $N$ . Since a bi-directional ray consists of two uni-directional rays, we see that  $V_R$  is similarly increasing and concave in  $N$  and increasing in  $\rho$ .

#### **Lemma 1**

Let us show that  $V_R < V_S < V_C < 2V_R$  if  $N > 2 \rightarrow$

$$\frac{\rho}{1-\rho}(1-\rho^N) < \frac{2\rho}{1-\rho}(1-\rho^{\frac{N}{2}}) < \frac{2\rho}{1-\rho}(1-\rho^N) - N\rho^N < \frac{2\rho}{1-\rho}(1-\rho^N) \quad (\text{A19})$$

$$1) \text{ We can easily see that } \frac{2\rho}{1-\rho}(1-\rho^N) - N\rho^N < \frac{2\rho}{1-\rho}(1-\rho^N) \rightarrow V_C < 2V_R. \quad (\text{A20})$$

$$2) \text{ By concavity of } V_R \text{ in } N, \text{ we can show that } \frac{\rho}{1-\rho}(1-\rho^N) < \frac{2\rho}{1-\rho}(1-\rho^{\frac{N}{2}})$$

$$\rightarrow V_R < V_S. \quad (\text{A21})$$

$$3) \text{ Next, we need to show that } V_S < V_C \rightarrow \frac{2\rho}{1-\rho}(1-\rho^{\frac{N}{2}}) < \frac{2\rho}{1-\rho}(1-\rho^N) - N\rho^N$$

$$\rightarrow \frac{2\rho}{1-\rho}(1-\rho^N) - N\rho^N - \frac{2\rho}{1-\rho}(1-\rho^{\frac{N}{2}}) > 0 \rightarrow \rho^{\frac{N}{2}} \left( \frac{2}{1-\rho}(1-\rho^{\frac{N}{2}}) - N\rho^{\frac{N}{2}-1} \right) > 0.$$

$$\text{Let } Q(N) = \frac{2}{1-\rho}(1-\rho^{\frac{N}{2}}) - N\rho^{\frac{N}{2}-1}. \quad (\text{A22})$$

We need to show that  $Q(N) > 0$ . Note that at  $N=2$ ,  $Q(2) = \frac{2}{1-\rho}(1-\rho) - 2\rho^0 = 0$ .

$$\text{Further, } \frac{\partial Q(N)}{\partial N} = -\rho^{\frac{N}{2}-1} \left( \frac{\rho \ln(\rho)}{1-\rho} + 1 + \frac{N \ln(\rho)}{2} \right). \quad (\text{A23})$$

If we can show that  $\frac{\rho \ln(\rho)}{1-\rho} + 1 + \frac{N \ln(\rho)}{2} < 0$  for  $N > 2$ , then  $\rightarrow \frac{\partial Q(N)}{\partial N} > 0$ ,

$$\rightarrow Q(N) > 0 \rightarrow V_s < V_c \text{ for } N > 2. \quad (\text{A24})$$

By re-arranging some terms we see that we need to show that  $\frac{\rho \ln(\rho)}{1-\rho} + 1 + \frac{N \ln(\rho)}{2} < 0$  or,

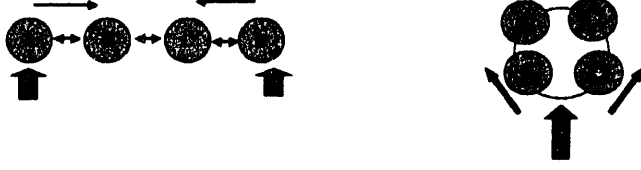
after further re-arrangement, we need to show that  $-\ln(\rho) > \frac{2}{2\rho + N(1-\rho)}(1-\rho)$ . Note

that this is the case since  $-\ln(\rho) > (1-\rho)$  and  $2\rho + N(1-\rho) > 2$  for  $N > 2$ . QED.

## Proposition 2

We first need to show that expected number of aware individuals in a bi-directional ray of size  $n$ , with entries at each endpoint is equivalent to the expected number of aware nodes in a circle of size  $n$  with one entry. (See Figure 1A) The answer can be shown if we compute the expectation, but is also clear from the picture. We call this Result 1A.

**Figure 3A: A bi-directional ray with entries at each endpoint = a circle with one entry of same size**



### Circle Network

Using Result 1A, we can show that  $V_c^k = k \left[ \frac{2\rho}{1-\rho} (1 - \rho^{\frac{N}{k}}) - \frac{N}{k} \rho^{\frac{N}{k}} \right]$  (A25)

(Each section has expected value of  $\frac{2\rho}{1-\rho} (1 - \rho^{\frac{N}{k}}) - \frac{N}{k} \rho^{\frac{N}{k}}$ , and we sum over the  $k$  sections to obtain (A25)).

a) We first show that  $V_c^k$  is increasing in  $k$ .

$$\frac{\partial V_c^k}{\partial k} = \frac{2k^2 \rho [1 - \rho^{\frac{N}{k}}] + N \rho^{\frac{N}{k}} \ln(\rho) [2k\rho + N(1-\rho)]}{1-\rho} \quad (A26)$$

Also, note that  $\frac{\partial V_c^k}{\partial k} \Big|_{(N=2k)} = \frac{2\rho[1-\rho^2 + 2\rho \ln(\rho)]}{1-\rho}$  (A27)

Let us define  $h(\rho) = 1 - \rho^2 + 2\rho \ln(\rho)$ . We can show that  $h(\rho) > 0$  since  $h(1) = 0$  and

$h'(\rho) = 1 - \rho + \ln(\rho) < 0$ . This implies that  $\frac{\partial V_c^k}{\partial k} \Big|_{(N=2k)} > 0$  (A28)

Next, we show that

$$\frac{\partial^2 V_c^k}{\partial k \partial N} = \frac{N \rho^{\frac{N}{k}} \ln(\rho) [2k(1-\rho) + \ln(\rho) [2k\rho + N(1-\rho)]]}{k^3(1-\rho)} > 0 \quad (A29)$$

if  $N > 2k$ .

In order to prove (A29), it is sufficient to show that

$$2k(1 - \rho) + \ln(\rho)[2k\rho + N(1 - \rho)] < 0 \quad (\text{A30})$$

We can show that this holds as long as  $N \geq 2k$ .  $\rightarrow \frac{\partial^2 V_c^k}{\partial k \partial N} > 0$  if  $N \geq 2k$ . Combining this

with (A28), we can show that  $\frac{\partial V_c^k}{\partial k} > 0$  as long as  $N \geq 2k$ . QED.

b) Next, we show that  $V_c^k$  is concave.

$$\text{First, note that } \frac{\partial^2 V_c^k}{\partial k^2} = \frac{N^2 \rho^{\frac{N}{k}} \ln(\rho) \left[ -\ln(\rho) \frac{2\rho}{1-\rho} - 2 - \frac{N \ln(\rho)}{k} \right]}{k^3} \quad (\text{A31})$$

$$\text{In order to show concavity, it is sufficient to show that } -\ln(\rho) \left( \frac{2\rho}{1-\rho} + \frac{N}{k} \right) - 2 > 0 \quad (\text{A32})$$

Using the fact that  $-\ln(\rho) > 1 - \rho$ , we see that

$$-\ln(\rho) \left( \frac{2\rho}{1-\rho} + \frac{N}{k} \right) - 2 > 2\rho + \frac{N}{k}(1 - \rho) - 2 > 0 \text{ as long as } N \geq 2k. \text{ This, in turn, implies}$$

concavity. QED.

### **Bi-directional ray Network**

Using Result A1, we can show that

$$V_s^k = \begin{cases} \frac{2\rho}{1-\rho} (1 - \rho^{\frac{N}{2}}) & \text{if } k = 1 \\ (k-1) \left[ \frac{2\rho}{1-\rho} (1 - \rho^{\frac{N}{k-1}}) - \frac{N}{k-1} \rho^{\frac{N}{k-1}} \right] & \text{if } k > 1 \end{cases} \quad (\text{A33})$$

We show that  $V_s^k$  is increasing and concave in  $k$ .

a) Show for  $k = 1$ :

$$V_s^2 = V_c > V_s^1 = V_s \text{ by Lemma 1} \rightarrow \text{increasing at } k, \text{ when } k = 1.$$

We can also use Lemma 1 to show that  $V_s^2 = V_c < 2V_s^1 = 2V_s \rightarrow \text{concave at } k = 1.$

b) Show for  $k > 1$ :

We perform a change of variables:  $k' = k - 1$ . We see that

$$V_s^k = k' \left[ \frac{2\rho}{1-\rho} (1 - \rho^{\frac{N}{k'}}) - \frac{N}{k'} \rho^{\frac{N}{k'}} \right]. \quad (\text{A34})$$

We can now apply the results of the previous section to show that the function is

increasing and concave in  $k'$  as long as  $N > 2k' \rightarrow N > 2(k-1)$ .

### **Uni-directional ray Network**

Summing up the expected values of the  $k$  sections, we obtain  $V_R^k = k \frac{\rho}{1-\rho} (1 - \rho^{\frac{N}{k}})$  (A35)

We first show that  $V_R^k$  is increasing in  $k$ .

$$\frac{\partial V_R^k}{\partial k} = \frac{\rho [1 - \rho^{\frac{N}{k}} + \frac{N}{k} \ln(\rho) \rho^{\frac{N}{k}}]}{1-\rho} \quad (\text{A36})$$

$$\frac{\partial^2 V_R^k}{\partial k \partial N} = \frac{\rho N \ln(\rho)^2 \rho^{\frac{N}{k}}}{(1-\rho)k^2} > 0. \quad (A37)$$

$$\frac{\partial V_R^k}{\partial k} \Big|_{(N=0)} = 0. \quad (A38)$$

Combining (A36) and (A37), we see that  $\frac{\partial V_R^k}{\partial k} > 0$ .

$$V_R^k \text{ is concave since } \frac{\partial^2 V_R^k}{\partial k^2} = -\frac{\rho N^2 \ln(\rho)^2 \rho^{\frac{N}{k}}}{(1-\rho)k^3} < 0. \text{ QED.} \quad (A39)$$

### Proposition 3

In order to prove the proposition, we need to employ some results from the finite state Markov chains literature. The basic idea is the following. We follow the flow of information. We want to calculate how many interested nodes will be exposed to the information before the flow stops. The flow is interrupted when there are at least  $m$  uninterested nodes in a row. We define a finite state Markov chain, assuming that the first customer is interested. (To correct for this assumption, we have to multiply the resulting expectation by  $\rho$ ). The states are:

$S_1$ : last node is interested in the product (potential buyer)

$S_2$ : last node is uninterested in the product, but previous node is interested

$S_3$ : last 2 nodes are uninterested in the product, but previous node is interested

...

$S_{m+1}$ : last  $m$  nodes are uninterested in the product

The transitional probabilities (which comprise the probability transition matrix  $\mathbf{P}$ ) are:

$$P_{i,i} = \rho$$

$$P_{j,j-1} = \rho \quad 1 < j < m+1$$

$$P_{j,j+1} = 1 - \rho \quad 1 \leq j < m$$

$$P_{m+1,m+1} = 1$$

(Everywhere else the probabilities are zero).

For an illustration, see Figure 2A. Note that  $S_{m+1}$  is an absorbing state. The expected number of interested aware nodes is equivalent to the expected number of visits to  $S_1$  before the chain enters the trapping state. In other words, let us define a reward of 1 every time the chain enters  $S_1$ , and a reward of 0 in every other state (reward vector

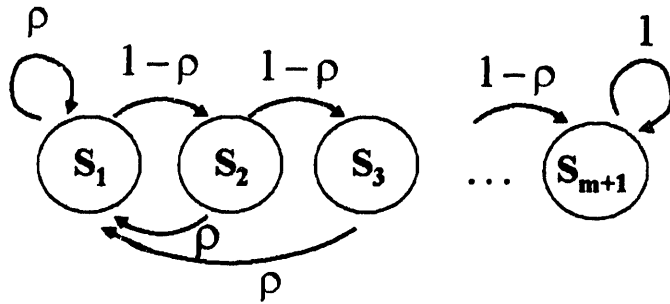
$$r = \begin{bmatrix} 1 \\ 0 \\ \dots \\ 0 \end{bmatrix}).$$

We can then use expected first passage time to calculate the number of visits

to  $S_1$  before the flow of information is interrupted.

$$\begin{bmatrix} v_1 \\ v_2 \\ \dots \\ v_m \\ v_{m+1} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ \dots \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} \rho & 1-\rho & 0 & \dots & 0 \\ \rho & 0 & 1-\rho & & 0 \\ \dots & & & & \\ \rho & 0 & 0 & 0 & 1-\rho \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \dots \\ v_m \\ v_{m+1} \end{bmatrix}$$

**Figure 4A: The Markov Chain**



We can solve for  $v_i$  – expected number of steps to reach  $S_{m+1}$  if start at  $S_i$ , where  $i < m+1$ .

We are primarily interested in  $v_1$ . In order to solve for  $v$ , we have to solve the following linear system:

$$v = r + [P]v ; v_{m+1} = 0. \quad ^{10} \quad (A40)$$

Or, after some re-arrangement, (can also eliminate last row and column in  $P$  since  $v_{m+1} = 0$ ):

$$\begin{bmatrix} 1 \\ 0 \\ \dots \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1-\rho & \rho-1 & 0 & \dots & 0 \\ -\rho & 1 & \rho-1 & & 0 \\ \dots & & & & \\ -\rho & 0 & 0 & 1 & \rho-1 \\ -\rho & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \dots \\ v_{m-1} \\ v_m \end{bmatrix} \quad (A41)$$

We can solve the system of equations above to obtain:

$$v_1 = \frac{1}{(1-\rho)^m}. \quad (A42)$$

<sup>10</sup> Gallager, Robert. "Discrete Stochastic Processes," Page 120.

The expected number of interested and aware nodes =  $\rho v_i = \frac{\rho}{(1-\rho)^m}$ . (A43)

$$\frac{\partial^2 \frac{\rho}{(1-\rho)^m}}{\partial^2 m} = \frac{\rho \ln(1-\rho)^2}{(1-\rho)^m} > 0 \quad (\text{A44})$$

QED.

#### Proposition 4 & Lemma 2

The firm chooses the level of advertising:

$$\max_{\mu} \pi N \theta \mu q - \frac{\mu^2}{2} N \quad (\text{A45})$$

Using FOC, we have  $\mu^* = \min[1, \pi \theta q]$ . We are guaranteed an internal solution since we assume that  $\pi \theta q < 1$ . Thus,

$$\frac{\pi^2 \theta^2 q^2 N}{2} = \Pi_{\text{Ads}}^* \quad (\text{A46})$$

$$\max[\pi V_i - w, 0] = \Pi_{\text{Buzz}}^* \quad (\text{A47})$$

By definition,  $\Pi_{\text{Ads}}^*$  is linear in  $N$ . We showed earlier that  $V_i$  is concave in  $N$ .

Therefore,  $\Pi_{\text{Buzz}}^*$  has is zero for low  $N$ 's and concave for higher  $N$ s. When comparing between the profits from the two strategies, there are 2 possible cases, as described in Lemma 2 and pictured in Figure 8 and Figure 9.

#### Proposition 4 & Lemma 2

Here the firm makes two choices:

- 1) Whether to employ chat at all
- 2) The optimal level of advertising

Thus, the optimization problem is given below:

$$\max_v \left[ \max_v \pi[(1-qv)V_R + Nqv\theta] - \frac{v^2}{2} N - w, \Pi_{Ads}^* \right] \quad (A48)$$

Once again, the optimal advertising level in the case when buzz & ads are used,

$$v^* = \frac{\pi q[N\theta - V_i]}{N}, \text{ where } 0 < v^* < 1. \quad (A49)$$

Note that  $N\theta > V_i$  since the expected number of nodes affected by buzz ( $V_i$ ) is necessarily less than the expected number of interested nodes ( $N\theta$ ). Since  $q < 1$ , we are guaranteed to have  $0 < v^* < 1$ .

Finally, we demonstrate the comparative statics in Lemma 2:

$$S_B = \frac{w}{w + \frac{\pi^2 q^2 [N\theta - V_i]^2}{2N}} \quad (A50)$$

a) We can show that

$$\frac{\partial S_B}{\partial N} = -a \left[ N \left[ \theta - 2 \frac{\partial V_i}{\partial N} \right] + V_i \right] \text{ where } a > 0. \quad (A51)$$

In order to show that  $\frac{\partial S_B}{\partial N} < 0$ , it is sufficient to show that

$$N \left[ \theta - 2 \frac{\partial V_i}{\partial N} \right] + V_i > 0. \quad (A52)$$

$$N[\theta - 2\frac{\partial V_i}{\partial N}] + V_i = N\theta + V_i - 2N\frac{\partial V_i}{\partial N} > 2V_i - 2N\frac{\partial V_i}{\partial N} = 2\left(V_i - N\frac{\partial V_i}{\partial N}\right) \quad (A53)$$

Thus, to demonstrate (A52), it is sufficient to show that

$$d(N) = V_i - N\frac{\partial V_i}{\partial N} > 0. \quad (A54)$$

Note that this is the case since  $d(0) = 0$  and

$$d'(N) = \frac{\partial V_i}{\partial N} - N\frac{\partial^2 V_i}{\partial N^2} - \frac{\partial V_i}{\partial N} = -N\frac{\partial^2 V_i}{\partial N^2} > 0 \text{ due to concavity of } V_i. \quad (A55)$$

$$\rightarrow \frac{\partial S_B}{\partial N} < 0. \text{ QED.}$$

$$\text{b) We can show that } \frac{\partial S_B}{\partial s} < 0 \text{ since } \frac{\partial S_B}{\partial s} = a \left( \frac{\pi^2 q^2 [N\theta - V_i] \frac{\partial V_i}{\partial \rho} \frac{\partial \rho}{\partial s}}{N} \right) > 0 \quad (A56)$$

This is since  $\frac{\partial V_i}{\partial \rho} > 0$  (as we showed earlier) and  $\frac{\partial \rho}{\partial s} = \rho > 0$ .

$$\text{c) We can show that } \frac{\partial S_B}{\partial w} > 0 \text{ since } \frac{\partial S_B}{\partial w} = b \left( \frac{\pi^2 q^2 [N\theta - V_i]^2}{2N} \right) > 0 (b > 0). \quad (A57)$$

$$\text{d) We can show that } \frac{\partial S_B}{\partial q} < 0 \text{ since } \frac{\partial S_B}{\partial w} = -c \left( \frac{q \pi^2 [N\theta - V_i]^2}{2N} \right) < 0 (c > 0). \quad (A58)$$