2005

Analysis of Herding on the Internet - An Empirical Investigation of Online Software Download

Wenjing Duan
University of Texas at Austin, wenjing.duan@phd.mccombs.utexas.edu

Bin Gu
University of Texas at Austin, bin.gu@mccombs.utexas.edu

Andrews B. Whinston
University of Texas at Austin, abw@uts.utexas.edu

Analysis of Herding on the Internet – An Empirical Investigation of Online Software Download

Wenjing Duan
University of Texas at Austin
wenjing.duan@phd.mccombs.utexas.edu

Bin Gu
University of Texas at Austin
bin.gu@mccombs.utexas.edu

Andrew B. Whinston
University of Texas at Austin
abw@uts.cc.utexas.edu

ABSTRACT

Online shopping often requires consumers to choose among multiple products without detailed information about the quality. Herding is common in situations where consumers infer product quality from other consumers' choices and incorporate that information into their own decision-making process. The Internet affects the herding phenomenon in two ways. On the one hand, it provides more information about other consumers' choices, therefore making herding more feasible. On the other hand, it provides more details about product quality, thus making herding less desirable. This paper empirically examines those two effects in the context of online software downloading. We find significant herd behavior in our analysis, and, surprisingly, the provision of professional product reviews or user reviews does not significantly influence the herding phenomenon. This study contributes to the E-Commerce and the Internet marketing research by investigating online consumer behavior. This paper also contributes to the emerging literature of studying the impact of virtual communities.

Keywords

E-Commerce, herding, informational cascades, software download, virtual community, online user review.

INTRODUCTION

Herd behavior, i.e. everyone is doing what everyone else is doing, portrays many social and economic situations where individuals are influenced by the decisions of others. As documented in the *informational cascades* literature, when a sequence of individuals makes decisions with *incomplete* and *private* information, individuals may simply follow the choices of their predecessors even when their private information favors a different option. Herding has been observed in various social and economic contexts, such as in financial investment, technology adoption, firms' strategy decisions, political voting, and dining and fashion trends. While herding has often been observed in *local* environments (Bikhchandani et al. 1992), the pervasive use of the Internet and other information technologies may significantly change its influence. Online shopping often requires consumers to choose among multiple products without detailed information about the quality. Herding is common in such situations, as consumers infer product quality from other consumers' choices and incorporate that information into their own decision-making process. In spite of the apparent prevalence of herd behavior documented in non-digital environments, there are few studies addressing this phenomenon in digital contexts. In this study, we seek to fill this gap by analyzing herd behavior on the Internet.

The Internet affects herd behavior in two ways. On the one hand, the Internet and other digital channels provide much more information about other consumers' choices, therefore making herding more feasible. Many online shopping websites display products on the Web according to their popularity. The popularity information is an indicator of the preferences of earlier consumers, though the detailed sales information is unknown. On the other hand, more details about product quality can be obtained on the Internet, thus making herding less desirable. There are various forms of digital communities on the Web that enable consumers to exchange opinions and experiences regarding companies, products, and services. In addition, an increasing number of online sellers also provide spaces for customers' feedback on their products. In this paper, we empirically examine those two effects in the context of online software downloading. In particular, we analyzed how consumers' choices of software are affected by previous number of downloads and product reviews information. Important

insights can be drawn from this study to understand the online consumer behavior, evaluate the impact of virtual communities, and propose business strategies for firms.

LITERATURE REVIEW

Herd behavior has been well developed in the fields of economics, social, and financial markets (See Bikhchandani et al. 1996). Banerjee (1992) and Bikhchandani et al. (1992) independently introduced the *informational cascades* framework to interpret the herd behavior.[1] They showed that people place such significant weight on others' opinions that they may even ignore there own private information. An alternative justification for herd behavior was provided by Schafstein and Stein (1990). They attributed the herd behavior of managers' investment decisions to the essential agency problem between managers and stakeholders. Wermers (1999) theoretically and empirically identified herding in mutual fund markets and showed its impact on stock prices. Bikhchandani et al. (2001) provided a comprehensive review of the recent theoretical and empirical research on herd behavior in financial markets.

In addition to financial investment, empirical evidence of herding also has been documented in the realms of technology adoption and entertainment programming. Kennedy (2002) analyzed herd behavior in prime time television programming. Kennedy's study indicates that, even though on average imitative introduction underperforms differentiated introduction, major networks copy each other when introducing new programs. Simonsohn and Ariely (2004) tested their prediction of herd behavior in Ebay's online auction. They found that bidders often engage in non-rational herding by favoring auctions with more existing bids. Walden and Browne (2002) found that informational cascades play a significant role in firms' adoption of electronic commerce technologies. Kauffman and Li (2003) developed a framework to rationalize herd behavior in IT adoption. *Payoff externalities*, *asymmetric information*, and *managers' career concerns* were proposed to be the critical drivers in influencing managers' IT investment decisions.

Other than the studies on IT investment decisions, the research on herd behavior in other IT related fields is very limited. In this study, we are specifically interested in analyzing consumer herd behavior on the Internet, a topic which, to our knowledge, has not been explored.

HYPOTHESES

Our empirical study was conducted in the context of software downloading at CNET Download.com. We seek to gain some insights to our research question by identifying and analyzing the herd behavior of customers' software choices on the Internet. All software on this site can be downloaded without any charge, thus the price effect on consumer demand has been controlled by default. In addition, significant herd behavior has been documented in financial market and IT investment, which often involves extremely large amounts of capital. It is important to understand how influential the herding effect is in other environments.

Download.com updates download counts (*number of total downloads*) everyday for each software program and posts a most popular list every week. In addition, CNET also provides professional reviews for some of the software programs and solicits customer feedback as well. Download.com data have the unique advantage of providing the dynamics of customers' product choices, which is particularly important in capturing the herding process. Our daily panel data setting can control the intrinsic product quality, thus allowing the characterization of herding. Customers' choices were collectively measured by *daily download market share* in each individual *market* (i.e. in one software category). As predicted by prior informational cascades literature, people place significant weight on other people's choices especially when there is uncertainty of product quality. Customers may herd to a *hot* software program even when it has low ratings and bad reviews. The following hypotheses are developed to empirically test the herd behavior of customers' software choices.[2]

H1: There is a strong herding effect on software downloading, i.e. daily download market is positively affected by the previous number of downloads.

H2: Previous customers' ratings have no significant impact on daily download market share.

H3: CNET's ratings have no significant impact on daily download market share.

---

[1] The terms *informational cascades* and *herd* are used interchangeably in the literature. However, based on definition by Smith and Sørensen (2000), an *information cascade* occurs when individuals ignore their private information when making a decision, whereas *herd* takes place when all the individuals make an identical decision, not necessarily ignore their private information.

[2] The development and detailed explanation for each hypothesis will be documented in our complete paper.

H4: There is no significant difference in daily download market share between software programs with CNET ratings and those without CNET ratings.

H5: Daily download market share is negatively correlated with the number of days a software program has been posted.

H6: The longer a software program has been posted, the weaker the herding effect is, i.e. the interaction of time and previous number of downloads has a negative impact on daily download market share.

DATA AND VARIABLE MEASUREMENT

Our data were collected from CNET Download.com (CNETD: Http://www.download.com). CNETD is a library of over 30,000 free and free-to-try software programs for Windows, Macintosh, and Handheld devices. CNETD provides reviews and ratings for some of the software programs, with the emphasis on new and popular software. CNETD also offers a user feedback system for customers to share their experiences and opinions. A data acquisition program was developed to gather data from CNETD. We started collecting the data from November, 2004 on a daily basis.

CNETD provides a rank of the most popular titles in Windows each week, which includes the top 50 most-downloaded products for the past week. Based on this rank, we selected ten categories of software as our sample. On each day, for every software program listed in each category, we extracted the following information from CNETD: software name, description, date added, total download, last week download, CNET rating, number of user votes, thumbs-up (percentage of positive user feedback), and thumbs-down (percentage of negative user feedback).[3] We also collected general characteristics for each software program, including operating system requirement, file size, publisher, and license.

We constructed the measurement of daily market share for each individual software program, which reflects customers' choices in a particular category. Let $i = 1....I$ index the software in a specific category. We define $DAILYDOWNLOAD_{it}$ as the number of downloads of software i at day t. Hence, the daily download market share of software i at day t is

$$DAILYSHARE_{it} = \frac{DAILYDOWNLOAD_{it}}{\sum_{i=1}^{I} DAILYDOWNLOAD_{it}}$$

CNETD provides its own reviews and ratings for some of the software programs. They rate software programs on a scale of one through five, with one being the worst and five being the best. In addition to writing a detailed review, CNETD's user feedback system provides two options for user ratings: thumbs-up and thumbs-down. In order to make them comparable to CNET ratings, we constructed a new measurement of user ratings on a scale from one to five.

$$USERATING_{it} = \frac{THUMBUP_{it}}{THUMBUP_{it} + THUMBDOWN_{it}} \times 5$$

PRELIMINARY EMPIRICAL ANALYSIS

We analyzed our panel data for two categories of software, *Adware & Spyware Removal* and *Mp3 Search Tools*. Our sample includes daily data from January 4, 2005 to January 27, 2005. The following empirical model is estimated:

$$DAILYSHARE_{it} = \alpha_0 + \alpha_1 TOTALDOWNLOAD_{i,t-1} + \alpha_2 DAYS_{i,t-1} + \alpha_3 TOTAL\_DAYS + \alpha_4 USERRATING_{i,t-1} + \alpha_5 CNETRATING_{i,t-1} + \alpha_6 NUMSOFTWARE_{it} + \mu_i + \varepsilon_{it}$$

The fixed effect μ_i is incorporated to capture the idiosyncratic and time-constant unobserved characteristics associated with each software program. The intrinsic quality associated with software programs that may inherently affect their market share can be captured by the fixed effect. Fixed effect panel data estimation also allows the error term ε_{it} to be arbitrarily correlated with other exploratory variables, thus making the estimation results more robust. Since CNET ratings do not change for a

[3] Before January 28 2005, CNETD ask users to evaluate software only by choosing thumb up or thumb down. They redesigned the user rating system to a five-number scale since then. We used old system data for our analysis in this paper. The data collected from the new system will be analyzed in the full paper as a comparison. Results from the full data set analysis will be ready to present before May, 2005.

given software program, the effect of CNET ratings is captured by the fixed effect.[4] We use a dummy variable $CNETRATINGD_{i,t-1}$ to indicate if a CNET review has been added in the middle of the sampling period. The coefficient of this dummy variable tests H4. $DAYS_{i,t-1}$ controls the length of the time that a software has been added to CNETD. *TOTAL_DAYS* is the interaction term of $DAYS_{i,t-1}$ and $TOTALDOWNLOAD_{i,t-1}$. The coefficient of this interaction term tests H6. $NUMSOFTWARE_{it}$ denotes the total number of software programs posted in this category at day *t*, which captures the competition effect. Table 1 presents the fixed effect estimation results in the category of *Adware & Spyware Removal*.

Variable	Coef. (std. err.)	t	P > |t|
$TOTALDOWNLOAD_{i,t-1}$	4.54e-08 (3.75e-09)	12.1	.000***
$DAYS_{i,t-1}$	-.00004 (.00002)	-1.85	0.064*
TOTAL_DAYS	-6.96e-11 (6.72e-12)	-10.36	0.000***
$USERRATING_{i,t-1}$	.0002 (.0004)	0.49	0.621
$CNETRATINGD_{i,t-1}$	.001 (.003)	0.38	0.70
$NUMSOFTWARE_{it}$	-.00003 (.0001)	-0.24	0.81
$n = 1535, R^2 = 0.98$		*** p<.01 ** p<.05 *p<.10	

Table 1. Adware & Spyware Removal Fixed Effect Estimation

The panel data estimation results showed that daily download market share is significantly driven by previous number of downloads, suggesting a strong herding effect and lending support for H1. The negative coefficient of $DAYS_{i,t-1}$ shows marginal significance, rendering a weak support for H5. H6 is supported by observing the exceedingly significant negative effect of the interaction term *TOTAL_DAYS*. This finding implies that the herding effect dissipates over time. No significant effects are observed for other exploratory variables, which supports H2 and H4. We found qualitative equivalent results for the data in the category of *Mp3 Search Tools*, so the results will not be reported here.

CONCLUSION AND FUTURE RESEARCH

We find significant herd behavior in our data, and surprisingly, the provision of professional product reviews or user reviews does not have a significant influence on the herding phenomenon. Our preliminary results suggest that consumers choose to ignore their own information, but are in favor of information inferred from others' behavior. Our results are consistent with the prediction of the informational cascades literature. We will extend this analysis to further estimate whether the herding effect we have found is *rational* or *non-rational* in our future complete research. Current results also indicate that the vast amount of information provided on the internet may not have as big an impact on consumer decision-making as previously expected. Customers' significant disposition to herding enables online retailers to leverage their advantage of manipulating the products' popularity information. Customers may or may not be aware of such strategic manipulation, which will subsequently affect their choice behavior and product purchase decision. To identify and characterize the strategic interplay of the herding process on the Internet among different parties will also be a potentially interesting extension of this study. Our data collection process is still ongoing. We will next proceed to the full-scale data analysis, and more comprehensive empirical testing will follow. This study contributes to both practitioners and academic researchers in understanding online consumer behavior. This research also contributes to the investigation of the economic impact of virtual communities.

REFERENCES

1. Banerjee, A.V. (1992) A Simple Model of Herd Behavior, *Quarterly Journal of Economics*, 110, 3, 797-817.
2. Bikhchandani, S., Hirshleifer, D., and Welch, I. (1992) A theory of fads, fashion, custom, and cultural change as informational cascades, *Journal of Political Economy*, 100, 5, 992-1026.
3. Bikhchandani, S., Hirshleifer, D., and Welch, I. (1996) Informational Cascades and Rational Herding: An Annotated Bibliography, Mimeo, Yale University.
4. Bikhchandani, S., and Sharma, S. (2001) Herd Behavior in Financial Markets, *IMF Staff Papers*, 47, 3, 279-310.

[4] In order to test if higher CNET ratings lead to more downloads (H3), the estimated fixed effect coefficients can be regressed on CNET ratings and other non-time-varying software characteristics. This analysis will be conducted on the full data set and the results will be reported in our complete paper.

5. Kauffman, J.R., and Li, X. (2003) Payoff Externalities, Informational Cascades and Managerial Incentives: A Theoretical Framework for IT Adoption Herding, *Proceedings of the 2003 INFORMS Conference on IS and Technology, Atlanta, GA, October 2003.*
6. Kennedy, E.R. (2002) Strategic Fads and Competitive Convergence: An empirical Test for Herd Behavior in Prime-Time Television Programming, *The Journal of Industrial Economics*, L, 1, 57-84.
7. Scharfstein, D.S., and Stein, J.C. (1990) Herd Behavior and Investment, *American Economic Review*, 80, 3, 465-479.
8. Simonsohn, U., and Ariely, D. (2004) e-Bay's Happy Hour: Non-rational Herding in on-line Auctions, Working paper.
9. Smith, L., and Sørensen, P.N. (2000) Pathological Outcomes of Observational Learning, *Econometrica*, 68, 2, 371-398.
10. Walden, A.E., and Browne, J.G. (2002) Information Cascades in the Adoption of New Technology, *Proceedings of the 2002 Intl. Conf. on Information Systems, Barcelona, Spain, December 2002, 435-443.*
11. Wermers, R. (1999) Mutual Fund Herding and the Impact on Stock Prices, *Journal of Finance*, 54, 2, 581-622.