HERDING BEHAVIOR AS A NETWORK EXTERNALITY

Completed Research Paper

Gordon Burtch
Fox School of Business
Temple University
1810 N 13th Street
Philadelphia, PA, 19122
gburtch@temple.edu

Abstract

Crowd-funding markets have recently emerged as a new avenue for entrepreneurs to raise funds. In these markets, any individual can pitch ideas and interested others can then invest in them. These markets provide investors with rich information on the investment decisions of prior others, thus they are rife with the potential for social influence. With this in mind, I examine a crowd-funded market for fashion, empirically evaluating some implications of Banerjee’s “simple model of herd behavior.” Specifically, I examine the implication that an increase in the number of observable decision makers within a marketplace should drive an increase in herding, because a greater number of inexperienced deciders will lower the average level of private knowledge in the market. I find evidence that supports this notion. Further, I identify a negative association between herding and the optimality of investor decision-making, supporting the characterization of herding as a negative network externality.

Keywords: Crowd-funding; herding behavior; network effects; electronic commerce
Introduction

With the emergence of online crowd-funding markets, entrepreneurs have recently gained access to a new pool of potential investors (Kappel 2008; Schwienbacher and Larralde 2010). These markets operate via a peer-to-peer lending model in which any individual can pitch a project idea to the crowd (the marketplace), and interested others can then invest their funds in the proposed idea to help see it to fruition (and perhaps capitalize on its success). Contributors in these marketplaces often directly benefit from investing. For example, proposers typically incentivize contribution with offers of sales discounts on initial units of production. In some of these markets, investors can even earn dividends on each unit sold once the project enters production, creating the potential for a long-term stream of returns if investors can identify a high-performing project in the investment stage.

A key novelty of crowd-funded markets is the presence of rich, publicly observable information on prior others’ investment decisions. This information is recorded and published for consideration by later deciders. While a number of recent studies, such as that by Duan et al. (2009), have considered electronic marketplaces that provide users with popularity indicators (e.g., software download rankings), the nature of crowd-funded markets is such that the popularity indicators are deeper in this setting, being generally more granular. For example, consider the crowd-funding context from which data was gathered for the present study. This marketplace, which deals with crowd-funded fashion, provides potential investors not only with an indication of whether others invested in a design, but also the timing of that investment.

Because of such rich information signals, crowd-funding markets are rife with the potential for social influence, making them ideal for the study of phenomena rooted in such mechanisms (e.g., herding behavior). Leveraging published information on individuals’ investment decisions, I seek to evaluate some previously untested implications of Banerjee’s (1992) seminal work on herd behavior in this ideal setting. These implications, which are discussed further in later sections, suggest that a growth in the number of observable deciders will be associated with an increase in the prevalence of herding behavior, because first-time deciders will tend to have a lower level of private information. Notably, a related implication here is that a greater number of observable adopters can also increase the chances that one's own private information will be overwhelmed.

The identification of a positive association between network growth and the prevalence of herding behavior would effectively constitute a novel form of negative network externality, given that herding behavior has been found to have negative consequences for consumer decision making. For example, Dholakia and Soltysinski (2001) have shown that bidders in online auctions often herd to their detriment, as they overlook more attractive auctions merely because alternatives have received greater numbers of bids. Similarly, Hanson and Putler (1996), discussing the self-reinforcing nature of download-based popularity measures for software products online, conclude that such signals are inefficient as they are easily manipulated; that is, a spurt of artificial demand can drive herding effects in the market. As a significant portion of investors’ utility in crowd-funded markets is tied to their ability to identify high-performing projects in the investment stage, this is an important question, not least because crowd-funded markets have recently boomed and expanded into many other industry sectors. Where once these markets facilitated relatively small transactions, they are now dealing with much larger sums of money. By way of example, Kickstarter.com raised more than $53 million between April, 2009 and March, 2011 (Strickler 2011), while ProFounder.com has reported average individual investments on the order of

---

1 Importantly, this is the case for the marketplace from which data was gathered for the present study's empirical analysis.

2 As detailed in later sections, the entrepreneurs in this market have joined the website in order to pitch t-shirt designs to raise the funds necessary to enter them into production. The contributors then benefit from shirts being produced, as subsequent sales generate dividends for them. To attract funds, the t-shirt designers create a t-shirt image, along with an optional explanation of the inspiration for the idea and its meaning. The designers and potential investors can then communicate with one another by posting upon the design’s comment thread. The crowd-funding platform earns all of its money off t-shirt sales, which it takes a percentage of the revenue from, thus the purveyor of the website also has an interest in optimizing investors decision outcomes.
$1,300, with entrepreneurs raising upwards of $30,000 from the broader community toward an average project.

With the above in mind, in this work I seek to extend the nascent body of literature on crowd-funding (e.g., Agarwal et al. 2010), as well as the literature on network externalities. I do this by demonstrating, in a crowd-funding context, that herding behavior is in fact a previously unconsidered form of negative network externality. Stated simply, this study addresses the following research question: How does network scale influence the prevalence of herding behavior in an electronic marketplace?

A secondary question I also consider here is the quantification of the expected negative association between herding behavior and investor decision optimality. Addressing this secondary question is useful because it helps buttress the claim that herding is inherently likely to have a negative effect on investors in crowd-funding contexts (i.e., the ‘negative’ portion of negative network externality).

In the following sections, I begin with a review of the relevant literature pertaining to crowd-funding markets, herding behavior and negative network effects. I then discuss Banerjee’s (1992) simple model of herd behavior and demonstrate how trivial extensions lead to the implication that a growth in the number of deciders will be associated with an increased prevalence of herding. I then detail my methodology and dataset before presenting my results, both in regard to the main research question as well as my secondary analysis of the influence of herding on the optimality of investment decision-making. Finally, I conclude the paper by discussing the implications of my findings for both practitioners and scholars dealing with crowd-funding markets, noting some limitations of the study, and proposing some potential avenues for future research.

**Literature Review**

**Crowd-funding**

There is a small, growing stream of literature pertaining to the concept of crowd-funded markets. Crowd-funding is defined as “the financing of a project or a venture by a group of individuals instead of professional parties” (Schwienbacher and Larralde 2010). Schwienbacher and Larralde, drawing on an earlier review of the concept by Belleflamme (2010), note that numerous online crowd-funding platforms have emerged in recent years. Well-known examples of crowd-funding platforms include: Sellaband.com (Agarwal et al. 2010), a crowd-funded platform for musicians in which they can raise money to produce musical albums; Spot.us, a crowd-funded market for online journalism (Aitamurto 2011); and the peer-to-peer lending site Prosper.com (Lin et al. 2009).

Much of the prior literature in this space has sought to explore the motivations and mechanisms underlying the behavior of participants in these markets. For example, drawing on data from Prosper.com, Lin et al. (2009) seek to identify novel forms of information that individuals consider when making lending decisions. These authors conclude that the likelihood of credit being issued is greater when the borrower exhibits greater social capital (e.g., a larger social network), as lenders appear to take this as a sign of credibility or trustworthiness. Aitamurto (2011) explores the motivations of contributors in the Spot.us marketplace via a qualitative analysis, finding that these funders perceive that they are contributing to a greater public good. Lastly, Agarwal et al. (2010) draw on data from the Sellaband marketplace to re-evaluate the “flat world hypothesis” – whether an online investment context eliminates any role of physical distance between the investor and musician.

While the literature on the topic of crowd-funding is still growing, there is a wide body of related work, such as the stream of research examining the effects of popularity indicators on later arrivers’ adoption decisions. Tucker and Zhang (2011), examining how online consumers’ wedding vendor adoption decisions are influenced by the provision of prior adoption statistics, find that these statistics cause the formation of a ‘steep tail’, attracting new outsiders to what was previously a niche product choice. Duan et al. (2009), considering users’ downloads of free software from download.com, observe severe changes in software download rates when published product rankings shift, suggesting that consumers take prior downloads as a signal of product approval and quality. Thus, published indicators of prior users’ preferences have been found to have a major influence on subsequent deciders.
**Herding Behavior**

It is often logical to mimic the behavior of others. For example, by listening to, reading, or watching the same things, actors can pursue common interests, around which they can interact, fostering a sense of belonging or community (Adler 2006; Salganik et al. 2006). Because of these sorts of benefits, popular products might be expected to become more popular, leading to ‘cumulative advantage’ or a ‘winner-take-all’ scenario. Thus, social influence can have a very large effect on consumer decisions and product success. Supporting this line of reasoning, Salganik et al. (2006) and Hanson and Putler (1996) employ Internet- and lab-based experiments to show that individuals are more than willing to follow others in their decision making behavior online, reacting to published information about others’ adoption decisions.

Mimicking others’ decisions may also make sense in the presence of sufficient uncertainty. Uncertainty is something that individuals must often contend with, given that we are inherently limited in our ability to process information (March and Simon 1958). To address this limitation, actors often rely on proxy indicators of the privately held information of others. Key examples of such indicators are others’ decisions to purchase or adopt a product. Reliance upon such indicators in the face of uncertainty is the traditionally accepted source of herding behavior and informational cascades (Banerjee 1992; Bikhchandani and Hirshleifer 1992; Bikhchandani et al. 1998; Kübler and Weizsacker 2004), and the mechanism I focus on in the present work.

Banerjee’s simple model of herding behavior (1992) describes a scenario in which a group of actors make the same decision (selecting from between two restaurants), in sequence. Each decision maker is capable of observing the choices of prior decision-makers, however, they are not directly aware of prior deciders’ private information. That is, they may only infer the nature of such information based on the decisions that others have taken. Banerjee analytically demonstrates that in many scenarios the information provided about the decisions of prior actors in the sequence will supersede any private information held by the focal decider, particularly when a sufficient number of prior deciders have made the same selection. In turn, this will be true of all following deciders, thus private information becomes meaningless after a point, as all actors begin to herd toward a common decision.

While Banerjee’s model is simplistic, it does explain behavior that has been found to manifest in a variety of decision-making contexts. For example, Borenstein and Netz (1999) determine that increased competition amongst airlines produces a degree of complexity and uncertainty that makes strategizing difficult, which results in herding in the scheduling of flight departure times. Similarly, Kennedy (2002) finds that television broadcasters will tend to imitate each other in the introduction of new television programming. As noted in the introduction, Duan et al. (2009) demonstrate that herding manifests with respect to online software downloads. These authors identify significant jumps and drops in software download rates as the publicly observable software download rankings shift, referring to these jumps and drops as informational cascades. Lastly, Dholakia and Soltysinski (2001) find that herding behavior is exhibited in online auctions, as bidders will tend to gravitate toward “popular” auctions which have received a large volume of bids.

Two of these studies are particularly relevant to the present work. Specifically, the work by Kennedy (2002) and by Dholakia and Soltysinski (2001) is directly related to the present study because these authors identify that herding leads to detrimental outcomes in their respective contexts. In Kennedy’s (2002) case, herding amongst television broadcasters results in sub-optimal profits. Similarly, in Dholakia and Soltysinski’s (2001) case, they note that herding causes bidders to overlook more attractive auctions. Based on these prior studies, and the proposed idea that there will be a positive association between herding and network scale, I am, in short, describing a novel form of negative network externality. With this in mind, in the following section I discuss the literature on negative network externalities.

**Negative Network Externalities**

Katz and Shapiro (1985), in explaining network externalities, state the following: “There are many products for which the utility that a user derives from consumption of the good increases with the number
of other agents consuming the good.” Further, “the utility that a given user derives from a good depends upon the number of other users who are in the same network.” Katz and Shapiro clearly characterize these effects in relation to product adoption and usage, however, Shapiro and Varian (1998) provide a similar definition that is more general, pertaining simply network membership (i.e., a definition that includes electronic markets). These authors state that all networks have a common, fundamental characteristic: “the value of connecting to a network depends on the number of other people already connected to it.”

The nature and influence of positive network externalities have been widely considered in the literature. Kauffman et al. (2000) identify the influence of network externalities on network adoption in the context of electronic banking, empirically demonstrating their positive influence on adoption rates. Liu et al. (2011) examine pseudo-standardization of products through conversion technology in IT markets that are subject to network effects. The authors find that a threshold of network effects from product usage will always exist, below which the providing firm will agree to offer conversion technologies for free, in order to facilitate adoption. Gallaugher and Wang (2002) empirically examine a variety of factors in two-sided software markets, notably identifying an association between the price premium that providers are able to charge and the size of their market share, thereby providing evidence of network externalities in such markets.

However, network externalities are not uniformly positive, as a number of researchers have considered a variety of negative network externalities. For example, Hellofs and Jacobson (1999) examine the association between customers’ perceptions of product quality and product market share, finding that quality can actually decrease with growing market share if part of the utility customers derive from the product stems from exclusivity. Liebowitz and Margolis (1996) discuss the relationship between positive network externalities and the emergence of monopolies. One implicit suggestion of this discussion is that another source of diminishing returns to network scale lies in the hosting firm gaining attention from antitrust regulators. However, probably the most commonly identified form of negative network externality in the literature is network congestion. Examples of this include Rysman (2004), who discusses congestion in the Yellow Pages, noting that advertisers will suffer from a greater “congestion effect” in larger public directories, as their advertisements are less likely to be seen. Similarly, Riggins et al. (1994) discuss the negative network externalities experienced by suppliers joining inter-organizational systems hosted by large buyers, as they find it increasingly difficult to differentiate themselves from competition.

In the information systems literature, network congestion has seen significant consideration, particularly in regard to peer-to-peer (P2P) file-sharing networks. Exploring this concept, Asvanund et al. (2004) empirically examine the effects of network externalities in P2P network growth, determining that the optimal size of such networks is bounded due to diminishing returns to scale. As the network grows, the authors find that network congestion increases, offsetting any positive benefits from a greater volume or diversity of shared content. Johar et al. (2011) examine the same concept, looking at contribution to P2P file sharing networks in the presence of congestion, exploring the use of different congestion measures.

These last two pieces of work are particularly relevant to the present paper, as they explicitly consider the actions of users within the network following their decision to join, considering that such actions have the potential to impose negative network externalities. However, while the literature has acknowledged that the actions (or lack thereof) of a given user can have indirect consequences on the utility of other users in the network (e.g., free-riders add little value and their decision to join can lead to congestion), it has not considered the influence of increasing numbers of observable others in a decision-making context. As social influence is a clear possibility in most online markets, an increase in the number of observable others may negatively influence the behaviors of a focal decider in the network by driving them to herd.

In the following section, I clarify exactly how growth in the number of observable others can translate into herding behavior. I do this by referring to Banerjee’s (1992) simple model of herd behavior, applying trivial extensions that account for growth in the number of decision makers over time.
Adjusting the Model

Banerjee’s (1992) simple model of herding behavior suggests that in a given network of sequential decision makers, there is some non-zero probability that herding behavior will manifest\(^3\). Thus, to begin my empirical analysis, I will first attempt to identify the presence of herding behavior.

The second focus of my empirical analysis is based upon an extension to Banerjee’s (1992) analytical model. His model assigns actors two probabilistic values, \(\pi\) and \(\beta\). Here, \(\pi\) represents the probability of a given actor having a signal, and \(\beta\) represents the probability of that signal being accurate. The suggestion that all actors in a network have uniform probabilities of having a signal (i.e., private information of which decision is optimal) is obviously somewhat unrealistic. As such, in this work, the model will be modified, such that, when an actor initially joins the network, his or her \(\alpha\) is initially assigned a value of 0. This value is then allowed to increase with each new period, toward its limit, Banerjee’s original conceptualization of \(\alpha\). Assuming that the next decision-maker will be drawn from the remaining sample of undecided nodes with uniform probability, then, as additional nodes enter the network, the likelihood that the next decider will have a signal will decrease to some value lying between 0 and \(\alpha\). Banerjee notes that the original model bears the implication that the probability of no one in the market making the correct decision is expressed by 
\[
1 - \alpha(1 - \beta) - (1 - \alpha)(1 - \beta)
\]

a function that is decreasing in both \(\alpha\) and \(\beta\). Thus, as alpha falls, this probability will rise. As the above-proposed modification to the model implies a decrease in \(\alpha\) as nodes are added to the network, then the probability of the marketplace herding around the wrong decision will rise, at least for a time. The logical implication of this is that the probability of any herding taking place, whether comprised of the entire network or not, will increase with network expansion.

With the above in mind, I will next attempt to identify an interaction between network size and the prevalence of herding behavior. It is important to point out here that this analytical implication has already found some traction in related literature dealing with the voting process in political elections. A key finding of Ali and Kartik (2006) is that the probability of herding taking place around a particular candidate increase to 1.0 as the size of the electorate grows. Finally, in addition to the above implication, it is important to determine whether herding has a negative effect on investor decision-making (else this is not truly a negative network effect). As such, I will then test whether herding behavior leads to investment in poorer performing projects.

In the following section, an overview is provided of the study context and methods used to empirically validate these hypotheses. I begin by articulating the workflow of the investment and sales process in this particular crowd-funded marketplace, and I then describe the dataset used in operationalizing the relevant variables.

Empirical Evaluation

Study Context

The market I consider is a now defunct crowd-funding platform (Kappel 2008; Schwienbacher and Larralde 2010) that was based in Chicago, Illinois. The site enabled artists to propose t-shirt designs and to raise the capital necessary to enter those designs into production and sale. The site functioned as follows: individuals would first register on the site, joining the community. Members would then establish a profile page, where they could provide details about themselves, along with a photograph and a link to a personal web page. Any member of the community could then choose to create and post a t-shirt design. Along with the design graphics, participants were also allowed to provide a description of the design’s inspiration and explanation. The web site’s management team would then evaluate the content, after which other members of the community could choose to invest in the design by purchasing ‘shares’ in it, up to a specified budgetary requirement threshold. Once the threshold was reached (i.e., once the required budget was raised), the design was then entered into production. At this point, investors would

\(^3\) For the sake of brevity, it is left to the reader to reference this work for the complete details of the model.
receive a discount on their purchase of the t-shirt and, for every unit sold, the investors and the designer would receive monetary dividends.

In this business model, the funding of designs provides direct utility to investors as they obtain dividends on sales, as well as a discount on their first purchase of the shirt. As such, investors might feasibly wish to rely on others’ investment decisions as signals of their private information about a design or designer’s quality (e.g., perhaps certain investors have prior knowledge of a designer’s ability to garner sales or perhaps certain designs have characteristics that are attractive to many individuals, which some investors can identify and others cannot), and then herd around those decisions.

However, before one can conclude that the alleviation of uncertainty is a reasonable cause of herding in the present setting, alternative explanations must be ruled out, such as whether the investment action is driven by a positive network effect (Shapiro and Varian 1998) or if sanctions upon deviants exist (Bikhchandani et al. 1998). With respect to network effects, our scenario is made easier by the fact that the behavior in question (investing in a particular design) does not appear to be subject to them. That is, adoption or investment by one individual does not feasibly enhance the relative value of investing for others. With respect to the second driver of herding (sanctions upon deviants), there exist no punitive responses to withholding investment. Based on this, it seems logical to conclude that herding in this crowd-funding marketplace, if identified, is driven by the mechanisms outlined in Banerjee’s simple model.

**Model Formulations**

I will begin the model formulation by noting the dependent variables that are of interest in the main and secondary analyses: the number of investors that arrive for a particular design, on a particular day, and the number of t-shirts sold, for a particular design, on a particular day. The latter of these two is intended to capture project performance. The remainder of the section is devoted to detailing the independent variables that will influence these outcomes.

In considering recent literature that has examined herding behavior in online contexts, it is readily apparent that signals emitted by contributors in our setting are relatively more complex. This is because the vast majority of empirical studies that have examined herding have looked at adoption decisions, where information on those decisions is presented to observers without any indication of the timing of those decisions (e.g., Duan et al. 2009; Tucker and Zhang 2011). In our case, however, an observer is quite likely to consider the temporal aspects of the contribution decision, given that this information is easily accessible. Upon viewing a design, a user is provided with a clear indication of how long that design has been posted for consideration. Further, the observer is also presented with a summary of the number of investors date. Based on these two values (the posting date and the number of investors), a potential contributor is capable of forming an impression of how quickly investors are arriving and how many are involved.

With this in mind, I define a composite measure of herding that is appropriate to this context, capturing the observable temporal and aggregate information. I define investment frequency as the number of investors arriving, per period, over the duration of a design’s funding in the marketplace. This measure incorporates both the total number of investors and the duration of funding, simultaneously. Thus, investment frequency reflects the rate at which funds are being raised by the design, as of a given point in time, and how many investors were involved in that fundraising. Based on the above, a greater frequency of investment may signal the quality or value of the design to other investors. If investors in this market herd, observing higher frequencies of prior investment should result in an increased volume of additional investment, at a particular point in time. To clarify the definition of this variable, a project that has been posted for 3 days and has received funding from 9 investors, would an investment frequency of $3^4$. In addition to frequency, a second key predictor here is network scale. The operationalization of network scale is relatively straightforward, as this value is simply the number of users in the marketplace, as of a given point in time.

---

4 It should be noted that the focus in this analysis is on the number of investors, as opposed to investment dollars, because of the nature of the available dataset.
Behavior similar to that observed by Duan et al. (2009), with respect to informational cascades in software downloads, is expected. That is, Duan et al.’s observation that changes in download statistics signal product value to other potential software downloaders on download.com is expected to apply here as well. So, a greater frequency of prior investor arrival should signal other potential investors to the quality of the investment alternative, resulting in an increased likelihood of subsequent investment taking place. Further, a greater frequency of investment over the course of a project’s investment phase should also reflect the occurrence of herding, due to the reinforcing effect of herding. That is, when herding takes place, investment should be completed more quickly. Thus, if herding is detrimental to decision making, a greater frequency during the investment stage should be associated with lower t-shirt sales.

Based on the above, the aforementioned expected relationships can be situated in the present context as follows. I expect that: i) investment frequency will have a positive effect on the number of subsequent investors, ii) the number of users in the marketplace will positively moderate the effect of investment frequency on the number of subsequent investors, and iii) investment frequency will have a negative effect on sales.

In addition to these key relationships, I expect a number of other factors to influence the number of investors for a given t-shirt design, on a particular day. First, I will include the lag of the number of investors as a predictor, as this is expected to capture unobservable word of mouth effects (Duan et al. 2009). While the focus of this study is on the influence of preference indicators and the predictive power of prior investor behavior, I also consider some important project characteristics associated with investment behavior in my analysis. This is done in order to control for their effects, thereby resulting in a more comprehensive model. In particular, I include the project’s budget outstanding, the number of comments that have been posted to a particular design by community members, as well as the total number of other designs available for investment in the marketplace, as of a given day. The impact of the project’s outstanding budget on the arrival of additional investment is difficult to predict. A greater volume of funds required may result in a desire to assist the designer. Alternatively, this may be perceived as somewhat of an insurmountable goal, and thus designs with larger outstanding amounts may have a harder time attracting funding. It is important to note here that I reviewed a random sample of 30 comments from this marketplace and found that the vast majority of them were positive in sentiment (i.e., few community members post negative comments indicating the poor quality of a design; rather, they simply refrain from commenting at all). Based on this, I focus on the volume of comments, rather than content. Lastly, the total number of other designs available for investment should have a negative impact on subsequent investment, as the disposable income available to the community may be spread more thinly over the available investment opportunities.

In order to ensure the model is identified, it is also important to account for exogenous sources of influence on investor behavior (i.e., correlated demand shocks, such as marketing efforts on behalf of the platform owner or individual t-shirt designers). For example, it is possible that some designers are well known and draw a greater following by advertising their new designs on personal blogs. Similarly, the platform owner may instigate occasional e-mail marketing efforts that are not observable in the data. I address these potential confounds by considering the number of web page referrals to the crowd-funding marketplace from external sources, such as search engines, Facebook links, e-mail links and the like, at the time of a given observation. Such referrals are in contrast to those page visits that stem from individuals manually accessing the marketplace by directly typing the URL into their web browser. A greater number of external referral visits would reflect greater societal popularity of the project content posted on the marketplace website.

This referral variable is again used in the secondary analysis considering the association between herding and t-shirt sales. The referral variable serves the same purpose here as it does in the main investment analysis. In addition, I incorporate the lag of sales to capture word-of-mouth effects. I do not incorporate the other variables noted above in this analysis, as they do not have direct relevance. This is because the sale of t-shirts is not limited to members of the community; rather, anyone and everyone can purchase units of these t-shirts via the marketplace “storefront,” which is a separate component of the market.

---

5 This is based on the notion that herding will lead to more frequent investment (due to the winner-take-all effect).
maintained in parallel to the crowd-funding platform. So, for example, the number of designs undergoing funding (other_designs), the number of users in the marketplace, and the frequency of investment have no bearing in this second regression. All of the variables identified in this section are summarized in Table 1, below, along with their definitions.

<table>
<thead>
<tr>
<th>Table 1. Variable Definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. investors</td>
</tr>
<tr>
<td>2. budget_remaining</td>
</tr>
<tr>
<td>3. frequency</td>
</tr>
<tr>
<td>4. comments</td>
</tr>
<tr>
<td>5. other_designs</td>
</tr>
<tr>
<td>6. referrals</td>
</tr>
<tr>
<td>7. users</td>
</tr>
<tr>
<td>8. sales</td>
</tr>
</tbody>
</table>

**Analytical Technique**

For a given design, \( i \), on a given day, \( t \), the final, complete model evaluating the antecedents of herding (the direct effect of frequency and the moderating effect of user volumes) is captured by equation (1). Here, the final two terms represent a vector of design-level fixed effects, addressing unobservable heterogeneity between designs, and an error term, respectively. As the investors variable constitutes count data (the number of investors on a given day), my evaluation of this model is conducted via time series regression employing the negative binomial estimator with a fixed effects dispersion parameter (Hausman et al. 1984). Subsequent robustness checks are performed using the time series Poisson estimator, which produces qualitatively similar results in terms of signs, significance and economic magnitude of effects.

\[
\text{investors}_i = \beta_1 \log(\text{investors}_{(t-1)}) + \beta_2 \log(\text{budget\_remaining}_i) + \beta_3 \log(\text{frequency}_i) + \\
\beta_4 \log(\text{comments}_i) + \beta_5 \log(\text{other\_designs}_i) + \beta_6 \log(\text{referrals}_i) + \\
\beta_7 \log(\text{users}_i) + \beta_8 \log(\text{sales}_i * \text{users}_i) + \beta_9 \log(\text{cumulative\_investors}_i * \text{users}_i) + \phi_i + \epsilon_i \tag{1}
\]

The secondary analysis of the association between herding and sales is less complex. This model is presented below in equation (2). Here, I again employ the negative binomial estimator, given that sales is a count measure of the number of purchases for a particular design on a particular day.

\[
\text{sales}_i = \beta_1 \log(\text{sales}_{(t-1)}) + \beta_2 \log(\text{referrals}_i) + \phi_i + \epsilon_i \tag{2}
\]

---

6 It is important to note that the original range of frequency values lies roughly between 0 and 0.30. As such, given that all other variables are, roughly speaking, positive integer values, I have multiplied all frequency values by 100 to maintain some degree of consistency in variable ranges. These multiplied values are what is presented in the descriptive statistics that follow in the Dataset section.

7 I do not incorporate designer fixed effects into my model, as the vast majority of designers in my sample have designed only a single t-shirt. As such, designer fixed effects are highly collinear with design fixed effects and lead to qualitatively similar results.
With the exception of the dependent variables (investors and sales), all non-categorical variables in my analyses will be log transformed, thereby allowing me to identify percentage changes in effect. This was deemed appropriate primarily because the number of investors that arrive to different projects on different days varies quite widely, as does the popularity of different designs captured by website referrals. Thus, understanding contribution effects in percentage terms is significantly more useful (Keene 1995). The dependent variable is not transformed because prior work has noted that doing so impedes the usefulness of count estimators (O’Hara and Kotze 2010).

Given the presence of a lagged dependent variable, which is potentially endogenous, an additional robustness check is performed in the first model, using the Arellano-Bond/Blundell-Bover System GMM estimator (Arellano and Bover 1995; Blundell and Bond 1998). This estimator allows one to instrument for a lagged dependent variable using available higher order lags, while simultaneously addressing fixed effects through the use of a difference equation. Admittedly, higher order lagged variables might not be ideal instruments since it is possible to have common demand shocks correlated over time, in which case lagged variables would be correlated with the current period demand shock. However, common demand shocks correlated over time are conceptually similar to trends. Hence, a suitable control for correlated demand shocks or trends can alleviate this problem in the GMM estimation (Archak et al. 2011). The web site referrals control variable thus alleviates any potential concerns in this regard. In addition, it should also be noted that, in order to alleviate any concerns about a high instrument count in the SGMM estimation (Roodman 2009), I used a collapsed instrument set comprised only of low order lags. This approach was deemed appropriate as it has been shown to produce more reliable results in scenarios where the instrument count is on the higher side (Mehrhoff 2009).

**Dataset**

The data used in this study was collected from multiple sources, though the bulk of it is proprietary in nature. The data is predominantly comprised of web traffic statistics, access to which was provided to the author by one of the company owners. This data was retrieved programmatically from the company’s Google Analytics account. The author developed a small software application in Java that integrated with the Google Data Export API to retrieve time series data on web traffic, by URL, by day, for a specified period of time. Investor arrival is therefore operationalized by the number of unique visitors that accessed the “successful investment” URL for a particular t-shirt design, which investors were only directed to upon completing the investment process for a particular t-shirt design. Similarly, sales are operationalized by the number of unique visitors to the purchase checkout URL associated with a particular t-shirt design. Access to the company’s Google Analytics account was also used to retrieve time series data about website referral rates (i.e., visitor source URLs), which, to reiterate, reflects the exposure of marketplace content on the broader Internet. That is, it stands to reason that the more links that exist on the Internet, on Facebook, Search Engines or in personal e-mails, the greater the marketing effort associated with marketplace content, and other similar, correlated demand shocks. This web traffic data was also supplemented by additional proprietary data. The author was supplied with programmatic access to a web server database containing details on t-shirt design information (e.g., budget threshold) and posting dates. Further, the same database included information on community comments and forum posts, as well as the number of users within the marketplace, along with their individual profile information. The descriptive statistics of all variables are presented in Table 2.

The data spans a 22-month period, beginning in June of 2008 and ending in 2010. Of the 319 t-shirt designs in the sample, 29 completed funding within the period of observation. Of those designs, the average duration of funding was 78.93 days, though there is considerable variation in this regard, with the fastest design having completed funding in 1 day, and the longest taking a many as 301 days. 4,428 unique investments events took place over the period of observation.
Results

The results of the negative binomial regression are presented below in Table 3. As per Cameron and Trivedi (2007), coefficients associated with log transformed independent variables in count regressions are interpretable as elasticities. The first key point is that each regression is highly significant, over all (based on the Wald statistic).

Looking at the model in column 1, it can be seen that there is clear support for the presence of herding in this marketplace, as the coefficient on frequency is highly significant, positive and quite strong. This provides evidence that Banerjee’s model holds practical significance in the context of this crowd-funded marketplace. In particular, the estimates suggest that an increase of 1% in the frequency of investor arrival will be associated with a 0.32% increase in the number of investors in the subsequent period. Further, the lag of investor arrival is also significant and positive, indicating that a 1% increase in the number of prior period investors is associated with a 1.18% increase in the number of investors in the subsequent period. Lastly, and surprisingly, there is a negative effect from comments. There are a number of possible explanations for this finding. For example, this relationship may be reflective rather than causal, in that it is possible that individuals make comments only after they have invested in a design (e.g., “I just invested in your design because I really liked it”).

Moving on to evaluating the second relationship, the interaction effect between the number of users in the marketplace and the prior frequency of investor arrival, there is evidence in support of this. It can be seen in column 4 that the coefficient associated with the frequency*users term has a significant, positive effect (i.e., the number of users amplifies the effect of herding signals on subsequent investor arrival). This effect is over and above the base effect of additional users simply being present in the marketplace. In short, this finding suggests that as the number of users in the marketplace grows, the tendency to rely on prior investment frequency as a signal, and thus the tendency for investors to herd, does increase.

With respect to the various control variables in the model, we can see a number of other interesting results. First, the outstanding budget for a design has a consistently positive effect. This suggests that investment in a particular design tends to taper off as a design nears its funding threshold. It is possible that this reflects waning interest in a design if it is posted for too long.

---

It is important to note here that statistical significance at the p < 0.001 is not particularly surprising given the size of the sample. However, what is important is that the identified effects are also economically significant as well.

---

8 It is important to note here that statistical significance at the p < 0.001 is not particularly surprising given the size of the sample. However, what is important is that the identified effects are also economically significant as well.
Next, the number of other designs present in the marketplace has a consistently negative effect on investment in a focal design. This is fairly intuitive, as it suggests that the probability of a design receiving investment falls as potential investors are presented with a greater number of investment options. Lastly, the referrals variable, which captures the absolute number of incoming website referrals from elsewhere on the Internet (e.g., Facebook, e-mails, search engines or blogs), has a positive effect on investor arrival. This is important, as the referrals variable is key to the identification strategy of this study. This variable allows me to control for exogenous correlated demand shocks over time, due, for example, to the website purveyor’s marketing efforts or societal interest in the subject of crowd-funding. These exogenous factors need to be controlled for as they present potential confounds to the relationship understudy (i.e., herding as a network externality). The positive effect of this variable is again intuitive, as it suggests that greater marketing effort on the part of the website purveyor has a positive impact on website traffic and, as a result, investment.

All of the identified effects are highly consistent across the hierarchical regressions, thus the estimates appear stable and reliable. In order to ensure robustness, however, I re-estimated the full model, taking a number of different approaches. In particular, the model was re-estimated as a time series Poisson regression, and again using the Arellano-Bover/Blundell-Bond SGMM estimator. As noted previously, the latter estimator addresses design-level fixed effects by using a difference equation, while also addressing the potential endogeneity of the lagged dependent variable by instrumenting for it using higher order lags. The coefficient estimates from both regressions were equivalent to those reported above. In the SGMM estimation, I employed a collapsed instrument set comprised of the 2nd through 4th order lags. The results of this regression were consistent with those reported above. Further, the test for 2nd order autocorrelation, AR(2), was insignificant, validating the usage of 2nd order+ lags as instruments. Further, the Hansen tests of over-identification and the difference in Hansen tests were also insignificant in this estimation.

I next undertake the secondary analysis, in an attempt to quantify the negative effect of herding on investor decision-making. To reiterate, such an analysis is useful as it helps to buttress the argument that the emergence of herding behavior as a network externality is indeed an example of a negative network effect. The results of this secondary analysis are presented below in Table 4. The dependent variable in Table 3. Results of Negative Binomial Regression (Investors)

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>investors(t-1)</td>
<td>1.18*** (0.04)</td>
<td>0.93*** (0.04)</td>
<td>-0.90*** (0.04)</td>
<td>0.87*** (0.04)</td>
</tr>
<tr>
<td>frequency</td>
<td>0.32*** (0.04)</td>
<td>0.94*** (0.05)</td>
<td>0.85*** (0.05)</td>
<td>0.46** (0.11)</td>
</tr>
<tr>
<td>comments</td>
<td>-0.94*** (0.05)</td>
<td>-0.34*** (0.06)</td>
<td>-0.31*** (0.06)</td>
<td>-0.21*** (0.06)</td>
</tr>
<tr>
<td>budget_outstanding</td>
<td>-</td>
<td>0.13*** (0.01)</td>
<td>0.11*** (0.01)</td>
<td>0.09*** (0.01)</td>
</tr>
<tr>
<td>other_designs</td>
<td>-</td>
<td>-1.37*** (0.05)</td>
<td>-4.51*** (0.30)</td>
<td>-4.52*** (0.30)</td>
</tr>
<tr>
<td>referrals</td>
<td>-</td>
<td>0.76*** (0.03)</td>
<td>0.90*** (0.03)</td>
<td>0.91*** (0.03)</td>
</tr>
<tr>
<td>users</td>
<td>-</td>
<td>--</td>
<td>3.42*** (0.33)</td>
<td>3.29*** (0.03)</td>
</tr>
<tr>
<td>frequency*users</td>
<td>-</td>
<td>--</td>
<td>--</td>
<td>0.12*** (0.03)</td>
</tr>
<tr>
<td>Observations</td>
<td>100,525</td>
<td>100,525</td>
<td>100,525</td>
<td>100,525</td>
</tr>
<tr>
<td>Wald Chi²</td>
<td>1,671.51 (3)</td>
<td>2,697.40 (6)</td>
<td>2,933.31 (7)</td>
<td>2,976.23 (8)</td>
</tr>
<tr>
<td>-2LL</td>
<td>-12,081.644</td>
<td>-11,324.96</td>
<td>-11,266.07</td>
<td>-11,257.29</td>
</tr>
</tbody>
</table>

Notes:

*** p < 0.001; ** p < 0.01; * p < 0.05; + p < 0.10.
Bootstrap standard errors in brackets for coefficients (50 iterations)
Degrees of freedom in brackets for test statistics.
this regression, which is again uses the fixed effects negative binomial estimator, is t-shirt sales. The measure of herding that I use is the final investment frequency associated with a design as of the completion of the funding process (notably, this variable is a static value, rather than dynamic). Logically, those designs that are subject to a greater level of herding in the investment stage will be funded more quickly. As such, a higher final frequency value should indicate a greater degree of herding, all things being equal.

In this model, I also incorporate two other basic factors: the number of website referrals, again controlling for exogenous correlated demand shocks (or demand trends), as discussed earlier, and the lag of sales, controlling for word of mouth. These variables are operationalized in the same manner as described earlier. Once more, all independent variables in this analysis are log transformed.

First of all, the overall model is clearly significant, based on the Wald statistic. Further, however, I find a significant negative effect on sales from the final frequency value. A 1% increase in the frequency of investor arrival during the funding stage is associated with a 0.06% decrease in sales following the design’s entering into production. Though the sample size in this analysis is comparatively small (being associated with only 42 of the 319 designs from the original analysis), this analysis provides some credibility to the suggestion that the identified network externality is indeed negative in nature. In short, those investors that herd around prior, frequent investment, end up committing to t-shirt designs that perform less well once they go on sale. As such, these investors earn fewer dividends than might otherwise be the case. Beyond this key finding, I also find a significant positive effect from prior period sales, as well as referrals, both of which can be interpreted in a manner similar to the first analysis effort reported in Table 3. That is, these factors can be interpreted as capturing word of mouth effects, and the effects of exogenous drivers of demand, such as marketing efforts.

### Discussion & Implications

This paper solidifies the important influence of herding signals in the process of consumer decision-making and, further, demonstrates that this behavior is more likely to manifest in larger marketplaces. Intuitively, herding behavior might be expected to manifest more strongly simply because an individual is provided with a much stronger signal of the crowd’s approval. Thus, if an individual observes 300 others making a particular decision, this may have a much stronger effect on the focal decider than observing only 10. The existence of this effect has interesting implications for sellers and consumers in electronic markets, and crowd-funded markets in particular. One primary concern is whether the outcomes of this information processing are beneficial. Thus, a key question I have attempted to address is whether herding around frequent investment results in optimal project selection. I have presented a secondary finding which seems to indicate that herding behavior in this context is indeed associated with poorer decision-making by investors, however, future work could likely improve on this analysis by drawing on a larger dataset, and exploring this relationship in the context of other product types.
When one considers the growing prevalence of open innovation mechanisms (where crowd-funding platforms might be argued to be an example of such) and the use of related platforms like prediction markets, these findings have direct implications for their viability and optimality. If these markets incorporate preference indicators, or information on prior others’ decision making, the findings of the present study imply that purveyors of such platforms and marketplaces need to take the size of their participant network into consideration. It appears likely that an optimal size exists at which the market should be bounded. This is conceptually similar to the discussion presented by Asvanund et al. (2004) in their discussion of the optimal size of P2P file sharing networks, given the countervailing influence between positive network effects (in terms of volume and diversity of content) and network congestion.

It should also be pointed out that, from a project proposer’s standpoint, herding in investment might actually be viewed as a good thing, as it improves the speed and likelihood of a project coming to fruition. With this in mind, a project proposer might be best served by encouraging numerous participants (e.g., friends and family) to join the marketplace and demonstrate their support, to improve their chances of instigating a herding effect. Further, those wishing to encourage herding might seek to subsidize or match others’ investment to reach a point of critical mass at which herding or information cascades takes over.

There are also some potential implications from the identification of a significant negative association between herding in investment and project performance. In particular, the presence of such a significant effect is unlikely to be causal in nature. Rather, this relationship is better characterized as predictive. That is, this relationship implies that these indicators of investment behavior could potentially be relied upon in a predictive capacity, to anticipate project performance in advance. There is a relatively new stream of work that explores these types of relationships, attempting to forecast such things as unemployment (Choi and Varian 2009) and real estate value (Wu and Brynjolfsson 2009) based on the observed behaviors of individuals (in this case, based on Google Search trends). The line of work I am referring to would therefore fit well with that stream of literature. Beyond simply exploring the nature of the predictive capability of these indicators, such work might also experimentally explore how to fine-tune or improve upon these signals, both from a prediction standpoint, as well as from the standpoint of streamlining the investment process and improving investor decision-making.

Limitations & Future Work

The exploration, design and implementation of marketplace mechanisms to control for the manifestation and influences of consumers tendency to herd would appear to be a prudent next step, following this study. While this work has identified one influential factor (i.e., network size), other contingencies likely exist. However, this study is also subject to a number of limitations that should be noted. First, generalizability of these findings remains an issue, as the user community is largely concentrated in the continental United States, and is rather young, with a mean age of roughly 23 years. Further, herding effects will quite likely vary with product type. One must consider such things as the ease with which products can be evaluated, both before and after consumption (i.e., search, experience and credence goods) (Nelson 1970), as this will likely have an influence on investors willingness to take others’ behaviors as a relevant signal of quality. Further, one must consider whether the product in question is a form of public good, as contribution toward such products would be more likely to exhibit substitutive rather than reinforcing investment behavior (i.e., anti-herding might be observed). A more robust analysis of the association between herding and investment performance is also likely warranted, perhaps drawing on a data set from a more popular crowd-funded market dealing with greater transaction volumes. If the significance and consistency of a negative association can be identified again, this would, as noted above, have further implications from a managerial standpoint, as these indicators could be used in a predictive capacity to anticipate project performance and to adapt accordingly in other markets and contexts.

Conclusion

By leveraging a novel dataset comprised of proprietary web traffic statistics and firm data, this work presents evidence supporting the hypothesis that herding behavior does manifest in crowd-funding contexts, and, further, that the tendency to herd is moderated by the number of users in the marketplace. That is, as more users join the market, the prevalence of herding increases. Further, this work also presents some preliminary evidence that such herding is also associated with poorer investment
decisions, as a negative association is identified between the frequency of investment at the completion of funding (a proxy for herding), and subsequent product sales once the project enters production. These findings have direct managerial implications, both for entrepreneurs seeking capital and for operators of crowd-funding platforms. With this in mind, I have suggested a number of aspects that these stakeholders can consider going forward, in addition to highlighting some avenues for future research that are likely to prove fruitful. The nascent body of literature on crowd-funding is likely to benefit from the further exploration of these ideas.

References


