DOES THE OPINION OF THE CROWD PREDICT COMMERCIAL SUCCESS? EVIDENCE FROM THREADLESS

Anirban Mukherjee
Assistant Professor of Marketing
Lee Kong Chian School of Business, Singapore Management University
50 Stamford Rd. #05 - 0 1, Singapore,178899
+65-68281932
anirbanm@smu.edu.sg

Ping Xiao[1]
Senior Lecturer in Marketing
UTS Business School, University of Technology Sydney
14-28 Ultimo Road, Ultimo, NSW 2007
+61-02-95147563
ping.xiao@uts.edu.au

Li Wang
Assistant Professor of Marketing
Shanghai University of Finance and Economics
No. 777 National Road, Yangpu, Shanghai, China, 200433
+86-21-65906973
wang.li@mail.shufe.edu.cn

Noshir Contractor
Jane S. & William J. White Professor of Behavioral Sciences
McCormick School of Engineering & Applied Science, the School of Communication and the
Kellogg School of Management at Northwestern University
2145 Sheridan Road, Tech D241, Evanston IL, USA, 60208-3119
+1-847-491-3669
nosh@northwestern.edu

[1] The first two authors are listed in alphabetical order and contributed equally to the study. This material is based upon work supported by the National Science Foundation under grant no. IIS-1514427. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect those of the National Science Foundation.

Does the Opinion of the Crowd Predict Commercial Success? Evidence from Threadless

Abstract

Crowdsourcing new products involves an open call for creative ideas. To select among submissions, crowdsourcing portals ask the community (the "crowd") to voice its opinion. Does the voice of the crowd predict the commercial success of a new product? This is an open question because over a half a century of research in consumer behavior is inconclusive on how peoples' expressed attitudes predict their behavior. We study this question on a pioneering crowdsourcing portal, Threadless.com. We collect and examine a large-scale dataset tracking about 150,000 designs from 45,000 designers that received almost 150 million votes from 600,000 users between 2004 and 2010. We find that the counts of positive and neutral votes are consistent predictors of sales. However, the count of negative votes is an inconsistent predictor of sales – *receiving more negative* votes is associated with *higher* sales from the users who cast the votes, but *lower* sales from the users who did not cast the votes. These findings are consistent with users strategically voting down their best competitors to improve their odds of being selected.

Keywords: crowdsourcing, crowd-voting, new product development, creative designs.

A crucial challenge in marketing is evaluating the commercial potential of an innovative new product design prior to its commercialization (Åstebro and Michela 2005; Manceau et al. 2014; Toubia and Netzer 2016). This challenge is amplified in crowdsourced new product development, because membership in crowdsourcing communities is unrestricted and the crowdsourcing portal has an open call for submissions (Ogawa and Piller 2006). Therefore, it is common for the portal to receive many submissions of varying quality (Poetz and Schreier 2012; Bayus 2013). Furthermore, crowdsourced submissions are often very creative and not constrained by traditional product and category norms (Kornish and Ulrich 2014). Taken together, this makes it challenging to understand and foresee the commercial potential of submitted product ideas.

To select among submissions, crowdsourcing portals ask the community to vote on submissions (henceforth referred to as crowd-voting). For example, 99designs is a crowdsourcing portal that allows brands to hold graphic design contests. To help decide on submissions, 99designs allows the brand (the contest creator) to hold a poll on the submitted designs.

However, despite its prevalence, the efficacy of crowd-voting is uncertain. This is because consumer behavior research is inconclusive on the extent to which individuals' attitudes are consistent with behavior. This phenomenon is termed "attitude-behavior [in]consistency" (Wicker 1969; c.f. Sutton 1998 for a review) and is documented in a diverse set of purchase contexts including contractual services (Bolton 1998; Wirtz et al. 2014), grocery (Chandon et al. 2005), apparel (Seiders et al. 2005), and automobiles (Mittal and Kamakura 2001). In particular, in traditional pre-launch product-idea screening, some evidence suggests that individuals: (a) provide inaccurate predictions about their future behavior, and (b) exhibit behavior that does not accord with their stated preferences (Seiders et al. 2005). However, other evidence suggests that if individuals' attitudes are elicited at the appropriate level of specificity, they do correspond closely with behavior (Ajzen and Fishbein 1977; Fazio 1990).

In addition, crowdsourcing participants may be dishonest in voting. Selection by a crowdsourcing portal is a prize to the submitting user; crowdsourcing involves a contest, such that "individuals expend irretrievable resource[s] to win [a] valuable prize" (Chowdhury and Gurtler, 2015). Users may potentially engage in sabotage, "a deliberate and costly act of damaging a rival's likelihood of winning the contest" (Chowdhury and Gurtler, 2015), and strategically vote down their best competitors (Luca and Zervas 2016). Thus, the votes of the crowd reflect a distorted voice of the crowd.

We study crowd-voting on an iconic crowdsourcing portal, Threadless.com (henceforth Threadless). In 2001, a t-shirt design contest on Dreamless.org, a forum for graphic/web designers and programmers, inspired two design enthusiasts—Jake Nickell and Jacob DeHart—to start a biweekly t-shirt design contest, with winning designs offered for sale to the public. The contest was a runaway success and led to the creation of Threadless, a dedicated crowdsourcing portal for t-shirt design. Inspired by their success, Howe coined the term "crowdsourcing" in a *Wired* magazine article to describe the business model that Nickell and DeHart had accidentally stumbled upon (Howe 2006).

Our data tracks all designs and votes submitted on Threadless between January 2004 and July 2010: over 150,000 designs submitted by more than 45,000 designers that received almost 150 million votes from over 600,000 users. Importantly, the data includes the revenues of designs selected for retail by Threadless. Primarily due to a paucity of data, only a handful of studies have examined the success of selected or implemented crowdsourced ideas. Amongst these studies, our data is unique in allowing us to examine the relationship of crowd-voting ("what the crowd says") with sales ("what the crowd purchases").

A summary of our empirical findings is as follows. More positive votes for a design consistently predict higher sales of a t-shirt based on the design. More neutral votes for design consistently predict lower sales of a t-shirt based on the design. More negative votes predict sales differently across different users and time frames. Specifically, *more negative* votes predict *higher* sales from individuals who voted on the design, but *lower* sales from individuals who did not vote on the design. In addition, while positive and neutral votes maintain their relationship with sales in the long-run, negative votes do not predict long-run sales.

In the remainder of this article, we proceed as follows. First, we discuss our conceptual framework. Second, we discuss our institutional context, data, and variable operationalization. Third, we describe our empirical model. Fourth, we describe our results. The final section concludes with implications of our findings and directions for future research.

CONCEPTUAL DEVELOPMENT

Crowdsourcing is the sourcing of organizational functions (for e.g., new product design) from the "crowd": a large, undefined community of a firm's consumers, partners, and collaborators (Howe 2006; p. 226, Bayus 2013). Crowdsourcing, similar to crowdfunding and crowdlending, is a means of soliciting resources from the crowd. However, while crowdfunding and crowdlending pertain primarily to obtaining financial resources (by disbursing rewards, equity, and debt), crowdsourcing pertains to sourcing creative resources (for e.g., new product ideas) and tasks (for e.g., evaluating new product ideas) from the crowd.

The literature on crowdsourcing is significantly smaller than the literature on crowdfunding (for a recent review on crowdfunding, see Kuppuswamy and Bayus, 2015) and crowdlending (Butler et al. 2016). Early papers on crowdsourcing provide a theoretical overview of the economics of crowdsourcing and discuss several of its prominent exemplars, including Threadless (Ogawa and Piller 2006; Poetz and Schreier 2012). The subsequent empirical work on crowdsourcing can be classified into four broad streams (Table 1 is a comprehensive list of all empirical studies in crowdsourcing published in the top 50 business journals[2]).

--- INSERT TABLE 1 HERE ---

The first stream considers participation and success in the crowdsourcing of new ideas (for e.g., Bayus 2013; Huang et al. 2014; Bockstedt et al. 2015; Hutter et al. 2015; Piezunka and Dahlander 2015; Bauer et al. 2016; Schemmann et al. 2016). The second stream focuses on the role of social and network structures in idea generation (for e.g., Chua et al. 2015; Stephen et al. 2016). The third stream investigates the quality of crowdsourced ideas (for e.g., Kornish and Ulrich 2014; Liu et al. 2014; Dissanayake et al. 2015; Blohm et al. 2016; Jame et al. 2016).

While the first three streams study the crowdsourcing of ideas, the fourth stream studies design crowdsourcing: the crowdsourcing of functional product designs (Nishikawa et al. 2017; Allen et al. 2017). For example, Nishikawa et al. (2017) find that communicating that a product is based on a crowdsourced design at point of purchase increases sales as consumers perceive the product as better able to meet their needs, while Allen et al. (2017) examine the antecedents of design crowdsourcing and the sales performance of products based on crowdsourced designs.

Our paper contributes to this stream of the literature by examining the efficacy of crowd-voting, i.e., crowdsourcing the task of evaluating (crowdsourced) designs. The challenge of

[2] We use the list of top 50 business journals developed by the Financial Times, and conduct a keyword search for "crowdsourcing" and "co-creation" in each journal.

selecting among many submissions is unique to crowdsourcing amongst crowd activities. This is because crowdsourcing involves a single portal or a single focal firm sourcing a business process from the crowd. Thus, the onus is on the crowdsourcing portal or focal firm to select among crowdsourced submissions. In contrast, in crowdfunding and crowdlending, the crowd decides whether to fund an idea and a loan respectively and the portal does not select among submissions. Crowd-voting is thus unique to crowdsourcing.

Theory suggests that crowd-voting may not be effective. A vote expresses the attitude of an individual towards a submission. For example, in our context, users vote on the extent to which they like (or dislike) a design. The commercial potential of a submission, in contrast, depends on the behaviour of the user. For example, in our context, the commercial potential of a design depends on how many users purchase t-shirts of that design if it is selected and commercialized (the next section provides further institutional details). Therefore, the relationship between the crowd's votes and the commercial potential of a submission depends upon the underlying relationship between attitude and behaviour. However, the literature is generally inconclusive on how attitudes relate to behaviour. For example, some research suggests that individual's attitudes reflect an anticipation of the social norm (Cialdini et al. 1991). Specifically, individuals' votes, may in part reflect how they anticipate the crowd to vote. This may lead to a "bandwagon effect", where individuals engage in group-think when rating products (Lee et al. 2015). Particularly, the bandwagon effect is likely to be consequential for designs that challenge the status quo, as is common on Threadless where designers often propose avant-garde designs (Ho et al. 2017).

In addition, despite the importance and prevalence of crowd-voting in crowdsourcing, relatively little is empirically known about it. Extant empirical papers focus primarily on crowdsourcing applications where crowd-voting plays no role or a modest role. For example, some studies examine applications where users attempt to solve a well-specified problem or complete a well-specified task (e.g., as in Topcoder, Innocentive, and Zooppa, c.f. Lakhani et al. 2010; Boudreau and Lakhani 2011; Zhang et al. 2016). In these applications, experts usually decide the winning solution (Chen et al. 2016). Other studies focus on applications where firm(s) solicit suggestions to improve their long-standing business processes (e.g., Bayus 2013 and Huang et al. 2014, examine IdeaStorm.com, a website to crowdsource ideas for Dell's products and services; c.f. Haas et al. 2015). In these applications, the person or firm posting the problem/task or sourcing suggestions for its business processes is relatively well positioned to judge the submitted ideas.

An important exception is a working paper by Chen et al. (2016). The authors compare the selection of submissions by experts and by peers on Zooppa.com, a crowdsourcing platform that helps corporate clients acquire user-generated advertisements. They focus on how four factors – selector expertise, herding, incomplete evaluation, and social favoritism – shape expert and crowd opinion. Our paper differs from theirs in considering how the crowd both selects and buys designs. Therefore the main theoretical consideration in our paper is the extent of attitude and behavior [in]consistency. In addition, we examine a context where the voting mechanism is designed to minimize interference from the distortive influences that are the subject of Chen et al.'s study.

Finally, despite surface similarities, crowd-voting is distinct from consumer reviews. Crowd-voting relates to an idea for a product or service that does not exist. In contrast, a consumer's rating reflects his/her experience with a product or service (Yadav and Pavlou 2014). In addition, crowd-voting is conducted in a specific voting period, using a specific template. In contrast, consumer reviews may be posted by a consumer any time, in several different forms (including numerical ratings), and on several different media. Finally, voting in crowd-voting is directed at the crowdsourcing firm or portal to help it select among submissions and typically not shared publicly. In contrast, consumer reviews are typically directed at other consumers and posted in open online public forums (Floyd et al. 2014).

To analyze the crowd's votes, we turn to the literature on psychometric scales (Bearden et al. 1997). On Threadless, attitudes are expressed as a vote on a discrete (0 to 5) scale. In studies involving (discrete) attitude scales, participants often demonstrate various responses biases (c.f. Weijters et al. 2010). For example, they may center their responses at the midpoint of the scale (directional bias) or favor extremal over midpoint values (extremal bias). Therefore, it is preferable to interpret responses to discrete scales as ordinal rather than interval data (Martilla and Carvey 1975; Jamieson 2004). However, interval data allows parametric testing, which is statistically more efficient than the non-parametric testing required by ordinal data. Due to efficiency concerns, empirical studies involving similar scales, therefore, often adopt an interval data interpretation and a parametric modeling paradigm (c.f. p. 225 Aaker et al. 2013; You et al. 2015).

In our application, efficiency is not a concern as we observe about 150 million votes cast on about 150,000 designs. In addition, users may display sabotage behavior. This would increase the number of negative (extremal) votes and may taint the parametric analysis. Indeed, we observe more negative than midpoint votes in our data, as described in the next section. Thus, both theory and empirical observations suggest that we should adopt an ordinal data interpretation. Furthermore, this reasoning is in line with the recommendations of Babić Rosario et al. (p. 301, 2016) who conduct a meta-analysis of the sales implications of consumer feedback on similar discrete scales. Drawing on analyses of 92 studies spanning over a decade, Babić Rosario et al. suggest the use of "composite valence-volume", which is the count (volume) of ratings of different valence (see p. 301). Following their recommendations, we, therefore, construct three variables: positive, neutral, and negative votes, which are the counts (volume) of positive extremal, midpoint, and negative extremal votes submitted for each design respectively.

Positive, neutral, and negative votes

More positive votes are likely to predict higher sales. This is because the marketing literature suggests that higher evaluations relate to higher consumer evaluation and satisfaction, which, *inter alia*, predicts higher brand loyalty, purchase incidence, and purchase quantity (for e.g., Kumar et al. 2007; Lovett et al. 2013; Baker et al. 2016).

Prior theory and empirical evidence are mixed on how neutral votes relate to sales. Forman et al. (2008) argue that neutral votes represent an equivocal opinion rather than an unequivocal opinion (unlike positive and negative votes), and are therefore intrinsically less informative of sales. However, Clemons et al. (2015) argue that, in highly differentiated, mature product categories, people purchase only what they love rather than that to which they are indifferent. Therefore, more neutral votes are likely to predict lower sales.

Expectations are also mixed on how negative votes relate to sales. Some studies have found that negative votes correspond to lower sales (Sun 2012). This finding is consistent with negative votes corresponding to lower consumer evaluations and satisfaction, and therefore to lower purchase incidence and purchase quantity (Richins 1983; Anderson et al. 2004; Lamberton and Stephen 2016). However, other studies find the opposite empirical relationship (for e.g. Doh and Hwang, 2009). Importantly, in a crowdsourcing contest, voting may also be strategic (Aral 2014; Luca and Zervas 2016). In particular, users vote on their competitor's designs. To increase the chances of their own designs being chosen by Threadless, users may strategically vote down their best competitors. Importantly, as users are most likely to sabotage designs that they find the most attractive, negative votes may act as a signal of higher rather than lower evaluations.

Predicting the behavior of those who did not vote

On Threadless, there are three groups of potential customers for commercialized designs: individuals who are members of the community and voted for the design, individuals who are members of the community but did not vote for the design, and individuals who are not registered users on Threadless. Of these, votes are likely to be most informative of the purchase behaviour of the individuals who cast them because (a) they directly articulate the attitudes of these voters and (b) in casting their vote, users morally commit themselves to actions consistent with their vote (Ogawa and Piller 2006).

It is unclear if the opinion of users who voted on a design, will systematically predict the behaviour of those who did *not* vote on the design. Users self-select into both being a member of the community and voting on a design. The preferences of users who voted on a design may, therefore, be different from users who did not vote on the design but are members of the community (e.g., Li and Hitt 2008). Moreover, purchasing behaviour of users who voted on a design may be different from individuals who are not members of the community and did not vote on the design (You et al. 2015).

In addition, the relationship of positive, neutral, and negative votes to sales, may vary differently across users. For example, on one hand, the relationship of positive and negative votes with purchase may be amplified for users who voted, relative to those who did not vote on the design, as these consumers may feel obliged to behave in a manner consistent with their voting (Ogawa and Piller 2006). On the other hand, consumers who were equivocal of their preferences, and cast a neutral vote, may not feel a similar moral obligation.

INSTITUTIONAL SETTING, DATA AND VARIABLE OPERATIONALIZATION

We study Threadless[3], a pioneering crowdsourcing portal. Threadless sources new t-shirt designs from the crowd. Specifically, Threadless has an open call for new t-shirt designs. In response to this call, registered users (registration is free and open to the public) submit new designs. A submission is the digital image and a title of a new design. Excluding the title, the image is stand-alone: There is no text description or explanation of a submitted design. The submission process is trivially simple as it only involves uploading the digital image of, and providing a title for, a submission.

Submitted designs, subject to basic moderation to ensure they adhere to basic community norms, are put up for voting when they are submitted (Figure 1 provides a timeline for the design). A design is voted on for a period of seven days from the time it is put up for voting. The start of the voting period is determined by the submission date. Therefore, only designs that are submitted relatively close to one another, overlap in voting.

--- INSERT FIGURE 1 HERE ---

To ensure designs receive a fair vote, Threadless shows designs that are open for voting in random order to users. This ensures that all designs get a fair chance of being voted on, regardless of what votes were cast prior. Furthermore, users are unaware of the distribution of previously cast votes, when voting. This mitigates herding.

[3] The information in this section describes Threadless as at the end of our data period (February 2011). Since February 2011, Threadless has made some minor changes (for e.g., it now uses a numerical scale from 1 to 5 rather than 0 to 5). Most crucial aspects of Threadless' business model, however, remain as described in this section.

Each registered user (excluding the user who submitted the design) may vote once on a submitted design. Users vote on a numerical scale of 0 to 5, with 0 corresponding to “I don’t like this design” and 5 corresponding to “I love this design.” Voting consists only of a numerical score: Users do not provide any other review or feedback to Threadless. For example, there is no verbal review of a design. At the end of voting, Threadless selects the designs that it wishes to retail. Threadless is not bound to a particular selection criteria or process when selecting designs (below we provide evidence of this from our data). The disaggregate votes are private and never revealed by Threadless. Threadless reveals only the mean vote and the number of votes cast for a submitted design after the voting process is complete. Each user with a design selected for retail is given a modest monetary reward (US$2000 in 2010).

We collect the largest dataset on crowdsourcing (see Table 1 for a comparison of the data used in prior empirical studies on crowdsourcing). Specifically, our data describe all votes (146,118, 048 votes submitted by 635,401 users), all submissions (150,093 designs), and all revenues from selected designs on Threadless from January 1, 2004 to July 31, 2010. From these, we drop data on 62 designs (less than 0.05% of the total data), where the identity of the user who submitted the design is missing, and 1 design (less than 0.01% of the total data) where the date of the submission is missing. Thus, our final dataset tracks 150,030 designs submitted by 48,556 users.

We summarize the information content in crowd-voting via the following variables:

1. Positive Votes = count of “4” and “5” (positive) votes received by design *d* (henceforth denoted Pos_d).
2. Neutral Votes = count of “2” and “3” (neutral) votes received by design *d* (henceforth denoted Neu_d); and

3. Negative Votes = count of “0” and “1” (negative) votes received by design *d* (henceforth denoted Neg_d);

In addition, to account for the competitive environment, we construct the following control variables (for similar practice, see Elberse and Eliashberg, 2003; Liu, 2006). To summarize competition in the voting phase, we include:

(1) Des_Vote_d: the average number of designs voted on by the community each day, over the week design *d* is voted on; and

(2) $T\text{-}S\_Vote_d$: the average number of t-shirts released to retail each day, over the four weeks prior to the start of voting on design *d*.

To summarize competition for the t-shirt based on design *d*, we include:

(3) $Des\_Release_d$: the average number of designs voted on by the community each day, over the four weeks prior to when the t-shirt based on design *d* is released to retail; and

(4) $T\text{-}S\_Release_d$: the average number of t-shirts released each day, over the four weeks prior to when the t-shirt based on design *d* is released to retail.

Table 2 summarizes the operationalization of variables, and Table 3A provides the descriptive statistics of the independent variables. We find that the distribution of voting is skewed: the count of positive votes has a mean of 178 and a median of 111; the count of neutral votes has a mean of 339 and a median of 253; and the count of negative votes has a mean of 396 and a median of 305. Table 3B provides a detailed breakdown of crowd-voting by the year. We find that (a) the voting patterns remained relatively stable over the years, suggesting that the platform had reached maturity prior to our observation period, and (b) that in all years, users cast more negative than positive votes. The finding that users on Threadless are more critical than complimentary of submitted designs is dissimilar to consumer reviews (for e.g. the mean score is reported to be above 4 on a scale of 1 to 5 in Chevalier and Mayzlin 2006 and Forman et al. 2008). This lends credence to the idea that crowd-voting is a distinct phenomenon from consumer reviews.

--- INSERT TABLES 2, 3A, AND 3B ABOUT HERE ---

The competitive environment in both voting and retail varies substantially across our data. For example, the average number of designs voted on each day over the 7-day voting period of a design has a mean of 454, with a standard deviation of 152, while the average number of designs voted on over the four weeks prior to when a t-shirt is released to retail, has a mean of 426, with a standard deviation of 122.

Figure 2 depicts the co-dependence of positive, neutral and negative votes. The figure is composed of three heat-maps. A point on the heat-map is a unique design, plotted by the count of positive, neutral, or negative votes. The density of points centered on the diagonal corresponds to concordant votes, while off-diagonal points correspond to discordant votes. Interestingly, we find that neutral and negative votes are much more closely related than positive and neutral votes, and positive and negative votes. In sum, the figure suggests that it is crucial to simultaneously consider all three types of votes.

--- INSERT FIGURE 2 ABOUT HERE ---

Selection and Revenues

Figure 3 includes three boxplots corresponding to the positive, neutral, and negative votes of the crowd, plotted for designs that were selected vs. not selected by Threadless. Recall that Threadless is not bound to a particular selection criteria. The evidence shows that positive votes have the most influence on Threadless' selections—selected designs have more positive votes than unselected designs—while neutral and negative votes seem to have limited influence on selection.

--- INSERT FIGURE 3 ABOUT HERE ---

Our research question relates to the role of crowd-voting in Threadless' selection decision. Specifically, we seek to quantify crowd-voting as a source of information for crowdsourcing portals. Therefore, we measure how crowd-voting relates to the revenue of a product based on the crowdsourced design. Consistent with the prior literature (for e.g., Kornish and Ulrich 2014; Allen et al. 2017), we focus on revenues at release, i.e. in the first three months after the t-shirt (based on a submitted design) was released on retail. We also consider long-run revenues by summing up revenues in the next 21 months, after the first three months. Note that as the cost of manufacturing t-shirts is relatively uniform across t-shirts, revenues track profits on Threadless.

Figure 4 is a histogram of (total) product revenues. We find that product revenues vary considerably across products. Therefore, the appropriate selection of designs is of vital commercial importance to Threadless. For each t-shirt based on design *d*, there are three groups of individuals:

1. "Voters": Individuals who are members of the Threadless community who voted on the design;
2. "Non-Voters": Individuals who are members of the Threadless community who did not vote on the design; and
3. "Outsiders": Individuals who are not members of the Threadless community and did not vote on the design.[4]

[4] Threadless requires individuals to be a member of the community to vote on designs but not to purchase designs.

--- INSERT FIGURE 4 ABOUT HERE ---

Table 4A provides the descriptive statistics of the overall revenue, and Table 4B provides the revenue from different users, over time. Sales to outsiders is the largest component of revenue. Sales to voters are the smallest component of net revenues, but (given the relatively modest number of votes cast for a design) represents greater purchase participation by those users than the non-voters. This may be either because these users are disproportionately more active on Threadless (they both tend to vote and buy more often), or because they voted on the design, they are also more likely to purchase the design if it is made into a product.

--- INSERT TABLE 4A and 4B ABOUT HERE ---

Figure 5 describes revenues from three different groups of individuals. Similar to the heat-maps for the elements of the votes of the crowd, a density on the diagonal indicates that two groups of individuals purchase similar products. Conversely, off-diagonal elements correspond to different purchasing behavior.

--- INSERT FIGURE 5 ABOUT HERE ---

Note that while the purchase patterns of non-voters and outsiders are relatively similar, the purchase patterns of voters do not closely correspond to those of non-voters and outsiders. To the extent that a (dis)similarity in purchasing patterns implies a (dis)similarity in tastes, this suggests that crowd-voting may not be informative of purchases by individuals who did not vote on the design. Therefore, there is a reason to doubt if the votes of the crowd are likely to be informative of revenues from individuals who did not vote on the design.

Finally, we examine the seasonality of submissions and purchases on Threadless. As we describe in the next section, we specify a copula selection model that is very flexible in its functional form. This model is identified without instruments. However, the non-parametric identification of selection models, including the copula selection model, requires a variable that enters the selection equation but not the revenue equation (i.e. an exclusion restriction). In our context, this requirement is met by the conjunction of (a) the time between voting and retail of a design, due to the need to manufacture selected designs, and (b) the differing seasonality of submissions and purchases. Specifically, prior research in crowd activities shows that crowd submissions peak when participants have slack time (Agarwal et al. 2016). This leads to a seasonality in submissions. On the other hand, the demand for fashion products, particularly t-shirts, is seasonal (Namin et al. 2017). This leads to a seasonality in purchases. Importantly, the two seasonal patterns do not coincide in our application (see Figure 6). Thus, the time required for manufacturing and the difference in seasonality patterns between submission and purchase, identifies our empirical model non-parametrically.

--- INSERT FIGURE 6 ABOUT HERE ---

MODELING FRAMEWORK

The purpose of crowd-voting (in crowdsourcing) is to help firms infer the extent to which a submission is aligned with the community's preferences. Specifically, in our context, crowd-voting variables discussed in the prior section allow the crowdsourcing platform to predict the revenue potential of a design. Accordingly, we build an empirical model where we model revenues as having two components: a component predicted by crowd-voting and a stochastic component orthogonal to the predicted component.

To ensure a fair vote, Threadless randomizes the order in which users encounter designs on the voting page in the voting period. In addition, Threadless does not inform users (including the user who submitted a design) of what votes were submitted during the voting process. These are deliberate design decisions for the following reason. If users are informed of, or are able to infer, the nature of the prior voting on a design, they may shade their own vote to match the perceived social norm, leading to the bandwagon effect. This would undermine the aim of crowd-voting by reducing the truthfulness of individuals' - and by extension, the crowd's - votes. As such, the exogenous assignment of designs to users and the lack of information regarding the nature of prior votes, considerably simplifies our econometric model, as the submitted votes are therefore unlikely to be influenced by the (orthogonal) stochastic component.

We begin by describing our revenue model for a design that was selected by Threadless. As both revenues and votes are skewed in the data, we follow the common practice to log-transform the variables. This has two benefits. First, the log-transformation accounts for any potential non-linearities from these values being able to take unbounded positive values (c.f. Liu 2006; Gopinath et al. 2013). Second, the coefficients of the model are elasticities. This makes our findings comparable across models and across contexts. As such, we model revenues as:

$$(1)\ \log(\text{Revenue}_{dk}) = \alpha_{0k} + \alpha_{1ky} \log(\text{Pos}_d) + \alpha_{2ky} \log(\text{Neu}_d) + \alpha_{3ky} \log(\text{Neg}_d) + \alpha_{4k} \log(\text{Des\_Vote}_d) + \alpha_{5k} \log(\text{T-S\_Vote}_d) + \sum_{q=2}^{Q} \gamma_{qk}\, \text{Quarter}_{rqd} + \varepsilon_{dk},$$

where Revenue_{dk} is the revenue from individuals of group k (i.e. voters, non-voters, or outsiders, or from all three groups of individuals), for the t-shirt based on design d. y is the year that design d was released to retail; $\{\alpha_{1ky}, \alpha_{2ky}, \alpha_{3ky}\}$ are year-specific coefficients that correspond to the year-specific sales elasticities of positive, neutral, and negative votes, respectively. Q is the total number of quarters in our data (arranged sequentially from the first to the last quarter). Quarter_{rqd} is an indicator variable that is 1 if design d was released to retail in quarter q, and 0 otherwise. The inclusion of a quarter-specific fixed effect non-parametrically accounts for both any trend in revenues from changes in the platform and the (earlier discussed) seasonality in revenues. ε_{dk} is the stochastic component of the model.

We estimate the revenue model on purchases by different groups of users as dependent variables. This allows us to uncover the relationship of the crowd's votes with revenue from these users. Furthermore, we allow for a specification where the relationship of the submitted votes with revenue, may vary over time. Therefore, we are able to test the extent to which a variable is a consistent versus an inconsistent predictor of revenues.

We model selection of a design by Threadless as:

$$\text{(2)}\quad \text{Select}^*_d = \beta_{0k} + \beta_{1ky}\log(\text{Pos}_d) + \beta_{2ky}\log(\text{Neu}_d) + \beta_{3ky}\log(\text{Neg}_d)$$

$$+ \beta_{4k}\log(\text{Des\_Vote}_d) + \beta_{5k}\log(\text{T-S\_Vote}_d) + \sum_{q=2}^{Q}\rho_{qk}\,\text{Quarter}_{rqd} + \xi_d,$$

where design d is selected by Threadless if $\text{Select}^*_d > 0$, and not selected otherwise. y is the year that design d was submitted for voting; $\{\beta_{1ky}, \beta_{2ky}, \beta_{3ky}\}$ are the year-specific coefficients of positive, neutral, and negative votes, respectively. Q is the total number of quarters in our data (arranged sequentially from the first to the last quarter). Quarter_{sqd} is an indicator variable that is 1 if design d was submitted in quarter q and 0 otherwise. The inclusion of a quarter-specific fixed effect non-parametrically accounts for both any trend in submissions and seasonality in submissions. ξ_d is the stochastic component, which may be correlated with the stochastic component (ε_{dk}) of equation (1).

An important consideration is the statistical dependence between the equations. Specifically, if the two error terms are modeled as being from a bivariate normal distribution, one obtains the well-known Heckman model (Heckman 1979). The bivariate normal distributional assumption in the Heckman model, however, is known to be an important restriction on model flexibility (Manski 1989; Newey 1999).

Therefore, we turn to copulas to model the statistical dependence across the two stochastic components in the two equations. Copulas were first introduced in marketing by Danaher and Smith (2011). A seminal result by Sklar shows that any multivariate distribution function has a copula representation: the distribution function can be expressed as its marginal distributions and a copula (Sklar 1959). This is the fundamental building block for the use of copulas in applied (empirical) contexts, as Sklar's theorem implies that the statistical link between any set of events can be modeled using a diverse set of copulas, with likelihood based testing used to resolve which copula fits the data best, and a convergence across copulas to the true distribution function. Formally, the joint multivariate cumulative distribution is expressed as

$$F(y_1, y_2, \dots) = C\{F_1(y_1), F_2(y_2), \dots; \theta\} = C(\varphi_1, \varphi_{2,\dots}; \tau)$$

where $F(y_1, y_2, \dots)$ is the multivariate cumulative distribution function, $\varphi_i = F_i(y_i)$ is the marginal cumulative distribution function of y_i for i = 1, 2, …, and $C\{\varphi_1, \varphi_2, \dots;\ \theta\}$ is a copula function. Accordingly, we model ξ_d and ε_{dk} as being distributed as described by various copulas.

We investigate nine distinct copulas: Ali-Mikhail-Haq (AMH), Clayton, Farlie-Gumbel-Morgenstern (FGM), Frank, Gaussian, Gumbel, Joe, Plackett, and Product (these copulas are described in detail in web appendix A). Importantly, the copulas have different statistical dependence properties. For example, amongst the nine listed copulas, the product copula assumes that the two stochastic components are independent, the FGM, Frank, Gaussian, and Plackett copulas assume that the upper and lower tails of the stochastic components are symmetrically dependent, while the AMH, Clayton, Gumbel and Joe copulas allow for asymmetric dependence between the upper and lower tail[5]. Note that similar to the traditional

[5] A detailed discussion of copula theory is beyond the scope of the paper. Please see Durante and Sempi (2010) for a review of the alternate classifications of these copulas and a detailed discussion of their properties.

selection model, the use of a copula to model the link between the equations implies the existence of a likelihood function. Therefore, we estimate the model by Maximum Likelihood Estimation. We select the copula that best empirically fits the data.

Our model specification reflects two important design decisions. First, we chose not to use a "machine learning model" (for e.g., Support Vector Machines or Gradient Boosted Models, c.f. Rafieian and Yoganarasimhan 2017) to model revenues. Machine learning models are algorithms that learn from "experience": given a set of data, a set of tasks, and a set of performance measures, the algorithms learn from the data such that they improve their performance on the tasks, as rated by performance measures.

While machine learning models are well suited to prediction tasks, they are not well suited to explanatory tasks where we seek to identify the informational properties of variables. This is because machine learning models are deterministic, rather than statistical, models that are defined via a set of heuristics (for e.g., the loss function defines the objective function when training a machine learning model). The parameters of a fitted machine learning model are not parsimonious, do not have a statistical (Bayesian or frequentist) interpretation, and only have meaning in the context of the data features used to train the model. Thus, machine learning models do not allow for inference across different data features and applications. In contrast, our econometric model yields sales elasticities, which have a clear economic interpretation. Importantly, this allows us to compare our results across various model specifications and develop empirical generalizations for contexts and applications beyond Threadless.

Second, as is common in empirical studies of products based on crowdsourced designs (for e.g. see Kornish and Ulrich 2014; Allen et al. 2017), equations (1) and (2) do not include fixed effects for each submitting user. This reflects the information structure of crowdsourcing.

Conceptually, at the point of selecting designs, a crowdsourcing firm or portal does not know the fixed effect corresponding to a user – the extent of revenue that (post-hoc) can be attributable to the user. Instead, the firm bases its decisions only on the observed votes. Accordingly, our empirical model examines the unconditional (i.e. without conditioning on a user's identity) voting elasticities. In addition, estimating a model with user-specific fixed effects requires dropping data from users who have less than two designs selected. This accounts for the vast majority of our data (as is the norm in crowdsourcing where only a small percentage of users have more than one selected idea or design; see Bayus 2013). Thus, the empirical specification reflects both the information structure of crowdsourcing and the predominantly cross-sectional structure of crowdsourcing data. Furthermore, we cluster standard errors by user. This ensures our inference is robust to both intra-panel correlation and conditional heteroscedasticity in the error terms.

RESULTS

Table 5 summarizes[6] results from a model on total revenue (i.e., the sum of revenue from voters, non-voters, and outsiders). Positive votes are associated with higher revenues. Conversely, neutral and negative votes are associated with lower revenues. This is consistent with the theorizing of Clemons et al. (2015), who suggest that consumers tend to only buy what they really love and not what they only like. We estimate models corresponding to all nine copulas listed in web appendix A. The model based on the Plackett copula best fits our data (by both the

[6] To conserve space, we suppress the discussion of our estimates of equation (2). Tables B1, B2, and B3 in web appendix B describe the estimates of equation (2) for the models described in tables 5, 6 and 7 respectively. Importantly, in all cases, the quarter fixed effects are jointly significant indicating the models are both non-parametrically and parametrically identified.

Akaike Information Criteria and the Bayesian Information Criteria). Therefore, we base further analysis on the Plackett copula.

--- INSERT TABLE 5 ABOUT HERE ---

Revenues from voters, non-voters, and outsiders

Table 6 summarizes our results for the main model on revenues from the three different user groups. We find that more positive votes consistently predict higher revenues for all three user groups: the average sales elasticity for positive votes is 1.74, 1.26, and 1.27 to voters, non-voters, and outsiders respectively. In addition, more neutral votes consistently predict lower revenues for all three user groups: the average sales elasticity for neutral votes is -1.42, -0.95, and -0.80 to voters, non-voters, and outsiders respectively.

--- INSERT TABLE 6 ABOUT HERE ---

In contrast, the effect of negative votes is nuanced. In some years, negative votes predict higher sales from users who voted on the design (average sales elasticity is 0.57). This is consistent with users voting strategically to sabotage the chances of their best competitors. However, negative votes do not predict revenues from non-voters, and predict lower revenues from outsiders (average sales elasticity is 0.57). These effects are consistent with negative votes also being a signal of a less attractive design. Perhaps due to the inherent tension in these two signals – sabotage suggesting higher revenues and a (true) negative opinion suggesting lower revenues – on net, negative votes are an inconsistent predictor of sales.

The competitive environment (summarized by Des_Release and T-S_Release) is insignificant. This is likely due to (a) the t-shirt designs being highly differentiated, and (b) a strong demand for variety in t-shirt designs, reducing the substitution across designs. Finally, the test statistics of the Wald (χ^2) test of independence are 9.22, 19.07, and 19.49 for the three groups of consumers respectively. In all cases, the test rejects the null, suggesting that the selection and revenue equations are statistically inter-linked.

Long-run revenues

The above results focus on revenue from the first three months. Do these findings extend to the long run? To answer this question, we estimate our empirical models on revenues from the next 21 months after the first 3 months (i.e. from month 4 to 24). In Table 7, we summarize our findings for our focal crowd-voting variables. The effects of positive and neutral votes are similar between the short run and the long run. Negative votes predict higher sales (in some years) to voters. However, perhaps reflecting the distortion caused by sabotage, negative votes are not predictive of long-run revenues from non-voters and outsiders.

In sum, we find that positive votes are the most consistent, and negative votes are the least consistent, predictor of sales, across both individuals and time. In addition, more neutral votes consistently predict lower sales, while the effect of negative votes varies by user and by time. Therefore, different votes have profoundly different implications for sales. These findings lend support to the view of Babić Rosario et al. (2016) that ratings are best represented via composite volume metrics.

--- INSERT TABLE 7 ABOUT HERE ---

Sensitivity Analyses

To examine the sensitivity of our results, we conduct the following analyses. We report our estimates in Tables C1-C4 in web appendix C. Our substantive conclusions remain unchanged across these analyses.

1. To ensure our results are robust to the removal of outliers, we sequentially remove designs that match the following criteria, and re-estimate the model:

   a. The vote is unequivocally negative: the mean vote is less than 1 (see Table C1a).
   b. The vote is unequivocally positive: the mean vote is greater than 4 (see Table C1b).
   c. The design received a few votes: the volume of votes is less than 100 (see Table C1c).
2. To account for information inherent in a more diverse set of ratings, we include the standard deviation of the crowd's votes in both equations (see Table C2).
3. There may be a loss of attraction between when the design was voted on and when it is available for sale. Therefore, we add the length of time between when a design was voted on and when it was released to retail in the revenue equation (see Table C3).
4. Users who have been on the platform for a longer time may be more/less successful than users who recently joined the platform. Therefore, we add the length of time a submitting user has been registered on the platform to both equations (see Table C4).

CONCLUSION

We contribute to the literature in being the first to empirically investigate how crowd-voting relates to revenues on Threadless, a pioneer and leader in crowdsourcing. Our study differs from prior work in examining how ratings predict demand from (a) those who voted on the design and are members of the community, (b) those who did not vote on the design but are members of the community, and (c) those who are outsiders to the community, and how these relationships change over time. Hence, our study provides a valuable and timely addition to the empirical knowledge base by examining the informativeness of the crowd's opinion in an understudied and fast-growing context (crowdsourcing), and in an under-studied product category (fashion/apparel).[7]

Our study provides several important and interesting insights to both users and decision makers at crowdsourcing portals. Our finding that more positive votes are strongly associated with higher revenues, is consistent with findings in the marketing literature. Our finding that more neutral votes are associated with lower revenue, adds new insights to the literature. Our finding that the negative votes are associated with higher revenues from the group that voted on the design but lower revenues from the group that did not vote on the design, is consistent with the presence of sabotage. Beyond crowdsourcing, these findings are of importance to marketers considering crowd-voting for pre-launch analytics (Lakhani 2016). For example, ModCloth, an e-retailer of vintage clothing, crowdsources its assortment in the product line based on the product concepts proposed in crowd-voting, and Lego Ideas requires a submitted idea to gain support from 10,000 users before it is considered for production.

Our study suffers from some limitations that may provide opportunities for future research. We examine a context in which products are only horizontally differentiated. It would be interesting to more broadly consider crowd-voting in crowdsourcing contexts where products are both horizontally and vertically differentiated. Further, we focus on a relatively understudied category, apparel. It also would be important to measure how our findings compare across crowd-voting in other crowdsourcing categories, given that the impact may vary with product-related factors. Finally, we focus on a portal that is based in the United States, where cultural entrepreneurism is encouraged. It would be necessary to examine the extent to which such

[7] For e.g., none of the 40 different platforms described in 96 empirical studies analyzed in Babić Rosario et al. (2016) relate to crowdsourcing and none of the 18 different categories in the 51 empirical studies analyzed in You et al. (2015) relate to fashion or apparel.

findings may vary across countries and to relate findings to cultural differences. Given the potential of crowd-voting in crowdsourcing, we hope that our study paves the way for further research.

REFERENCES

Aaker, David A., V. Kumar, Robert Leone, and George S. Day (2013), Marketing Research (Eleventh Edition), John Wiley & Sons.

Agrawal, Ajay, Christian Catalini, and Avi Goldfarb (2015), "Slack Time and Innovation," National Bureau of Economic Research. Rotman School of Management Working Paper No. 2599004; MIT Sloan Research Paper Forthcoming. Available at SSRN: https://ssrn.com/abstract=2599004 or http://dx.doi.org/10.2139/ssrn.2599004.

Ajzen, Icek, and Martin Fishbein (1977), "Attitude-behavior Relations: A Theoretical Analysis and Review of Empirical Research," *Psychological Bulletin* 84(5), 888.

Allen, B.J., Deepa Chandrasekaran, and Suman Basuroy (2017), "Design Crowdsourcing: The Impact on New Product Performance of Sourcing Design Solutions from the Crowd," *Journal of Marketing.* https://doi.org/10.1509/jm.15.0481.

Anderson, Eugene W., Claes Fornell, and Sanal K. Mazvancheryl (2004), "Customer Satisfaction and Shareholder Value," *Journal of Marketing*, 68 (4), 172-85.

Aral, Sinan (2014), "The Problem with Online Ratings," *MIT Sloan Management Review*, 55 (2), 47.

Åstebro, Thomas, and John L. Michela (2005), "Predictors of the Survival of Innovations," *Journal of Product Innovation Management,* 22(4), 322-35.

Babić Rosario, Ana, Francesca Sotgiu, Kristine De Valck, and Tammo HA Bijmolt (2016), "The Effect of Electronic Word of Mouth on Sales: A Meta-analytic Review of Platform, Product, and Metric Factors," *Journal of Marketing Research*, 53 (3), 297-318.

Baker, Andrew M., Naveen Donthu, and V. Kumar (2016), "Investigating How Word-of-Mouth Conversations about Brands Influence Purchase and Retransmission Intentions," *Journal of Marketing Research* 53 (2), 225-239.

Bearden, William O., and Richard G. Netemeyer. *Handbook of marketing scales: Multi-item measures for marketing and consumer behavior research*. Sage, 1999.

Bauer, Julia, Nikolaus Franke & Philipp Tuertscher (2016), "Intellectual Property Norms in Online Communities: How User-organized Intellectual Property Regulation Supports Innovation," *Information Systems Research*, 27(4), 724-750.

Bayus, Barry L. (2013), "Crowdsourcing New Product Ideas over Time: An Analysis of the Dell IdeaStorm Community," *Management Science*, 59 (1), 226-44.

Blohm, Ivo, Christoph Riedl, Johann Füller, and Jan Marco Leimeister (2016), "Rate or Trade? Identifying Winning Ideas in Open Idea Sourcing," *Information Systems Research*, 27 (1), 27-48.

Bockstedt, Jesse, Cheryl Druehl, and Anant Mishra (2015), "Problem-Solving Effort and Success in Innovation Contests: The Role of National Wealth and National Culture," *Journal of Operations Management*, 36, 187-200.

Bolton, Ruth N. (1998), "A Dynamic Model of The Duration of The Customer's Relationship with A Continuous Service Provider: The Role of Satisfaction," *Marketing Science*, 17 (1), 45-65.

Boudreau, Kevin J. and Karim R. Lakhani (2011), "Field Experimental Evidence on Sorting, Incentives and Creative Worker Performance," Harvard Business School, 11-107. Available at SSRN: http://ssrn.com/abstract=1815370.

Butler, Alexander W., Jess Cornaggia, and Umit G. Gurun (2016), "Do Local Capital Market Conditions Affect Consumers' Borrowing Decisions?," *Management Science*. 1-13. http://dx.doi.org/10.1287/mnsc.2016.2560

Chandon, Pierre, Vicki G. Morwitz, and Werner J Reinartz (2005), "Do Intentions Really Predict Behavior? Self-generated Validity Effects in Survey Research," *Journal of Marketing*, 69 (2), 1-14.

Chen, Liang, Pei Xu, and De Liu, (2016), "Experts Versus the Crowd: A Comparison of Selection Mechanisms in Crowdsourcing Contests," Available at SSRN: https://ssrn.com/abstract=2542857 or http://dx.doi.org/10.2139/ssrn.2542857.

Chevalier, Judith A. and Dina Mayzlin (2006), "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, 43 (3), 345-54.

Chowdhury, Subhasish M., and Oliver Gürtler, (2015), "Sabotage in Contests: A Survey," *Public Choice,* 164(1-2), 135-55.

Chua, Roy Y.J., Yannig Roth, and Jean-François Lemoine (2015), "The Impact of Culture on Creativity: How Cultural Tightness and Cultural Distance Affect Global Innovation Crowdsourcing Work," *Administrative Science Quarterly*, 60 (2), 189-227.

Cialdini, Robert B., Carl A. Kallgren, and Raymond R. Reno (1991), "A Focus Theory of Normative Conduct: A Theoretical Refinement and Reevaluation of The Role of Norms in Human Behavior," *Advances in Experimental Social Psychology*, 24, 201-34.

Clemons, Eric K., Guodong Gordon Gao, and Lorin M. Hitt (2015), "When Online Reviews Meet Hyperdifferentiation: A Study of The Craft Beer Industry," *Journal of Management Information Systems*, 23(2), 149-71.

Danaher, Peter J. and Michael S. Smith (2011), "Modeling Multivariate Distributions Using Copulas: Applications in Marketing," *Marketing Science*, 30 (1), 4-21.

Dissanayake, Indika, Jie Zhang, and Bin Gu (2015), "Task Division for Team Success in Crowdsourcing Contests: Resource Allocation and Alignment Effects," *Journal of Management Information Systems*, 32 (2), 8-39.

Doh, Sun-Jae and Jang-Sun Hwang (2009), "How Consumers Evaluate eWOM (Electronic Word-of-Mouth) Messages," *CyberPsychology & Behavior*, 12 (2), 193-97.

Durante, Fabrizio and Carlo Sempi (2010), "Copula Theory: An Introduction," in Copula Theory and Its Applications, P. Jaworki and F. Durante and W. Härdle and T. Rychlik, eds. New York: Springer.

Elberse, Anita and Jehoshua Eliashberg (2003), "Demand and Supply Dynamics for Sequentiallly Released Products in International Markets: The Case o f Motion Pictures," *Marketing Science*, 22 (3): 329-354.

Fazio, Russell H (1990), "Multiple Processes by Which Attitudes Guide Behavior: The MODE Model as An Integrative Framework," *Advances in Experimental Social psychology*, 23, 75-109.

Floyd, Kristopher, Ryan Freling, Saad Alhoqail, Hyun Young Cho, and Traci Freling, (2014), "How Online Product Reviews Affect Retail Sales: A Meta-Analysis," *Journal of Retailing,* 90(2), 217-32.

Forman, Chris, Anindya Ghose, and Batia Wiesenfeld (2008), "Examining the Relationship between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets," *Information Systems Research*, 19(3), 291-313.

Gopinath, Shyam, Pradeep K Chintagunta, and Sriram Venkataraman (2013), "Blogs, Advertising, and Local-Market Movie Box Office Performance," *Management Science*, 59 (12), 2635-54.

Haas, Martine R, Paola Criscuolo, and Gerard George (2015), "Which Problems to Solve? Online Knowledge Sharing and Attention Allocation in Organizations," *Academy of Management Journal*, 58 (3), 680-711.

Heckman, James J. (1979), "Sample Selection Bias as a Specification Error," *Econometrica,* 47(1), 153-162.

Ho-Dac, Nga N, Stephen J Carson, and William L Moore (2013), "The effects of positive and negative online customer reviews: do brand strength and category maturity matter?," *Journal of Marketing*, 77 (6), 37-53.

Ho, Yi-Chun, Junjie Wu, and Yong Tan (2017), "Disconfirmation Effect on Online Rating Behavior: A Structural Model," *Information Systems Research*, 28 (3), 626-42.

Howe, Jeff (2006), "The Rise of Crowdsourcing," *Wired*, 14.

Huang, Yan, Param Vir Singh, and Kannan Srinivasan (2014), "Crowdsourcing New Product Ideas under Consumer Learning," *Management Science*, 60 (9), 2138-59.

Hutter, Katja, Johann Füller, Julia Hautz, Volker Bilgram, and Kurt Matzler (2015), "Machiavellianism or Morality: Which Behavior Pays Off in Online Innovation Contests?," *Journal of Management Information Systems*, 32 (3), 197-228.

Jame, Russell, Rick Johnston, Stanimir Markov, and Michael C Wolfe (2016), "The Value of Crowdsourced Earnings Forecasts," *Journal of Accounting Research*, 54 (4), 1077-110.

Jamieson, Susan (2004), "Likert scales: how to (ab) use them," *Medical education* 38(12): 1217-1218.

Kornish, Laura J. and Karl T. Ulrich (2014), "The Importance of The Raw Idea in Innovation: Testing Ihe Sow's Ear Hypothesis," *Journal of Marketing Research*, 51 (1), 14-26..

Kumar, Viswanathan, Andrew J. Petersen, and Robert P. Leone (2007), "How Valuable is Word of Mouth?," *Harvard business review*, 85 (10), 139.

Kuppuswamy, Venkat and Barry L. Bayus (2015), "A Review of Crowdfunding Research and Findings," *Handbook of New Product Development Research*, Forthcoming.

Lakhani, Karim R. (2016), "The Antidote to HiPPOs: Crowd Voting," *Harvard Business Review*, February 02, 2016. Available at HBR: [https://hbr.org/2016/02/the-antidote-to-hippos-crowd-voting].

Lakhani, Karim R., David A. Garvin, and Eric Lonstein (2010), "Topcoder (a): Developing Software through Crowdsourcing," Harvard Business School General Management Unit Case No. 610-032. Available at SSRN: [https://ssrn.com/abstract=2002884].

Lamberton, Cait and Andrew T. Stephen (2016), "A Thematic Exploration of Digital, Social Media, and Mobile Marketing: Research Evolution from 2000 to 2015 and an Agenda for Future Inquiry," *Journal of Marketing*, 80 (6), 146-72.

Lee, Young-Jin, Kartik Hosanagar, and Yong Tan (2015), “Do I Follow My Friends or The Crowd? Information Cascades in Online Movie Ratings,” *Management Science*, 61 (9), 2241-58.

Li, Xinxin and Lorin M. Hitt (2008), “Self-selection and Information Role of Online Product Reviews,” *Information Systems Research*, 19 (4), 456-74.

Liu, Tracy X., Jiang Yang, Lada A. Adamic, and Yan Chen (2014), “Crowdsourcing with All-Pay Auctions: a Field Experiment on Taskcn,” *Management Science*, 60 (8), 2020-37.

Liu, Yong (2006), “Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue,” *Journal of Marketing*, 70 (3), 74-89.

Lovett, Mitchell J., Renana Peres, and Ron Shachar (2013), “On brands and word of mouth,” *Journal of Marketing Research*, 50 (4), 427-44.

Luca, Michael and Georgios Zervas (2016), “Fake It Till You Make It: Reputation, Competition, and Yelp Review Fraud,” *Management Science*, 62 (12), 3412-27.

Manceau, Delphine, Jehoshua Eliashberg, Vithala R. Rao, and Meng Su (2014), “A Diffusion Model for Preannounced Products,” *Customer Needs and Solutions,* 1(1), 77-89.

Manski, C. F. (1989), “Anatomy of the Selection Problem,” *Journal of Human Resources*, 24, 343–360.

Mittal, Vikas and Wagner A. Kamakura (2001), “Satisfaction, Repurchase Intent, and RepurchaseBehavior: Investigating the Moderating Effect of Customer Characteristics,” *Journal of Marketing Research*, 38 (1), 131-42.

Namin, Aidin, Brian T. Ratchford, and Gonca P. Soysal (2017), “An Empirical Analysis of Demand Variations and Markdown Policies for Fashion Retailers,” *Journal of Retailing and Consumer Services*, 38, 126-36.

Newey, W. K., (1999), “Consistency of the Two-Step Sample Selection Estimators Despite Misspecification of Distribution,” *Economics Letters*, 63, 129–132.

Nishikawa, Hidehiko, Martin Schreier, Christoph Fuchs, and Susumu Ogawa (2017), “The Value of Marketing Crowdsourced New Products as Such: Evidence from Two Randomized Field Experiments,” *Journal of Marketing Research*, 54(4), 525-39.

Ogawa, Susumu and Frank T. Piller (2006), “Reducing the Risks of New Product Development,” *MIT Sloan Management review*, 47(2), 65.

Onishi, Hiroshi and Puneet Manchanda (2012), "Marketing Activity, Blogging and Sales," International Journal of Research in Marketing, 29 (3), 221-34.

Piezunka, Henning and Linus Dahlander (2015), “Distant Search, Narrow Attention: How Crowding Alters Organizations’ Filtering of Suggestions in Crowdsourcing,” *Academy of Management Journal*, 58 (3), 856-80.

Poetz, Marion K. and Martin Schreier (2012), “The Value of Crowdsourcing: Can Users Really Compete With Professionals in Generating New Product Ideas?,” *Journal of Product Innovation Management*, 29 (2), 245-56.

Rafieian, Omid and Hema Yoganarasimhan (2017), “Targeting and Privacy in Mobile Advertising,” [available at http://faculty.washington.edu/hemay/ad-ctr.pdf].

Richins, Marsha L. (1983), “Negative Word-of-Mouth by Dissatisfied Consumers: A Pilot Study,” *Journal of Marketing,* 47(1), 68-78.

Schemmann, Brita, Andrea M Herrmann, Maryse M.H. Chappin, and Gaston J. Heimeriks (2016), "Crowdsourcing Ideas: Involving Ordinary Users in the Ideation Phase of New Product Development," *Research Policy*, 45 (6), 1145-54.

Seiders, Kathleen, Glenn B. Voss, Dhruv Grewal and Andrea L. Godfrey (2005), "Do Satisfied Customers Buy More? Examining Moderating Influences in aRetailing Context," *Journal of Marketing,* 69 (4), 26–43.

Sklar, A. (1959), "Fonctions de répartition à n dimensions et leurs marges," *Publ. Inst. Statist. Univ. Paris*, 8: 229–231.

Stephen, Andrew T, Peter Pal Zubcsek, and Jacob Goldenberg (2016), "Lower Connectivity Is Better: The Effects of Network Structure on Redundancy of Ideas and Customer Innovativeness in Interdependent Ideation Tasks," *Journal of Marketing Research.* 53(2), 263-279.

Sun, Monic (2012), "How Does The Variance of Product Ratings Matter?," *Management Science*, 58 (4), 696-707.

Sutton, Stephen (1998), "Predicting and Explaining Intentions and Behavior: How Well Are We Doing?," *Journal of Applied Social Psychology*, 28 (15), 1317-38.

Toubia, Olivier and Oded Netzer (2016), "Idea Generation, Creativity, and Prototypicality," *Marketing Science*, 36 (1), 1-20.

Wicker, Allan W. (1969), "Attitudes versus Actions: The Relationship of Verbal and Overt Behavioral Responses to Attitude Objects," *Journal of Social Issues*, 25 (4), 41-78.

Weijters, Bert, Elke Cabooter, and Niels Schillewaert (2010), "The effect of rating scale format on response styles: The number of response categories and response category labels," *International Journal of Research in Marketing* 27(3): 236-247.

Wirtz, Jochen, Ping Xiao, Jeongwen Chiang, and Naresh Malhotra (2014), "Contrasting The Drivers of Switching Intent and Switching Behavior in Contractual Service Settings," *Journal of Retailing*, 90 (4), 463-80.

Yadav, Manjit S. and Paul A. Pavlou (2014), "Marketing in Computer-Mediated Environments: Research Synthesis and New Directions," *Journal of Marketing*, 78 (1), 20-40.

You, Ya, Gautham G. Vadakkepatt, and Amit M Joshi (2015), "A Meta-Analysis of Electronic Word-of-Mouth Elasticity," *Journal of Marketing*, 79 (2), 19-39.

Zhang, Shunyuan, Param Vir Singh, and Anindya Ghose, (2016), "A Structural Analysis of the Role of Superstars in Crowdsourcing Contests," Available at SSRN: https://ssrn.com/abstract=2764553.

Table 1: Prior Empirical Literature on Crowdsourcing

Paper	Journal	Dependent Variable	Data Description
Bayus (2013)	Management Science	Was the submission selected?	8,801 ideas submitted by 4,285 users on Dell IdeaStorm from Feb 2007 to Feb 2009.
Huang et al. (2013)	Management Science	Did a user submit an idea?	Ideas submitted on Dell IdeaStorm from Feb 2007 to Sep 2008.
Kornish and Ulrich (2014)	Journal of Marketing Research	Sales of a product.	249 raw ideas submitted on Quirky from May 2011 to Mar 2013. Sales information for 160 products developed from 149 raw ideas.
Chua et al. (2015)	Administrative Science Quarterly	Did a user submit an idea? Was the submission selected?	99 of 600 creative contests (e.g. an advertising campaign) on an anonymous crowdsourcing portal from Jan 2010 to Dec 2011.
Liu et al. (2014)	Management Science	Number of submissions. What is the quality of the submissions?	Field experiment on Taskcn.
Bockstedt et al. (2015)	Journal of Operations Management	Number of submissions. Was a submission ranked in the top three?	1024 contests and 2,626 contestants on logomyway from August 1, 2010 to February 12, 2011.
Dissanayake et al. (2015)	Journal of Management Information Systems	Rank of a team in a contest.	732 teams in 52 problem-solving contests on Kaggle from April 2010 to July 2012.
Hutter et al. (2015)	Journal of Management Information Systems	Number of submissions. Number of comments.	61 survey observations and 2,233 comments from a large innovation contest in the field of public transportation.
Piezunka and Dahlander (2015)	Academy of Management Journal	Did a suggestion receive attention from the government?	1,077,669 suggestions submitted to a large number of from November 2007 to June 2011.
Bauer et al. (2016)	Information System Research	Number of submissions.	Netnography, a survey, and a field experiment.
Blohm et al. (2016)	Information System Research	Users' evaluation quality.	Lab experiments.
Jame et al. (2016)	Journal of Accounting Research	Accuracy of user's earnings forecasts.	51,012 forecasts on Estimize from 2012 to 2013
Schemmann et al. (2016)	Research Policy	Was an idea evaluated by the firm? Will the idea be implemented?	90,000 ideas on a portal run by an international beverage producer and retailer from the start to spring 2012.
Stephen et al. (2016)	Journal of Marketing Research	Innovativeness of submitted ideas.	Lab experiments.
Allen et al. (2017)	Journal of Marketing	Total number of units sold.	The raw product idea and monthly unit sales of 86 different products from Quirky.
Nishikawa et al. (2017)	Journal of Marketing Research	Lift in sales of products from labeling "crowdsourced from consumers".	Field experiments in Muji stores across Japan.

Table 2: Operationalization of Variables

Variable	Definition
Pos_d	Count of “4” and “5” (positive) votes received by design *d.*
Neu_d	Count of “2” and “3” (neutral) votes received by design *d.*
Neg_d	Count of “0” and “1” (negative) votes received by design *d.*
Des_Vote_d	The average number of designs voted on by the community each day, over the week design *d* is voted on.
$T\text{-}S\_Vote_d$	The average number of t-shirts released to retail each day, over the four weeks prior to the start of voting on design *d.*
$Des\_Release_d$	The average number of designs voted on by the community each day, over the four weeks prior to when the t-shirt based on design *d* is released to retail.
$T\text{-}S\_Release_d$	The average number of t-shirts released each day, over the four weeks prior to when the t-shirt based on design *d* is released to retail.
$Revenue_{kd}$	Revenue from group of individuals *k* (either voters, non-voters, or outsiders) of the t-shirt (based on design *d*) in the first three months after release.

Table 3A: Descriptive Statistics of Independent Variables [1]

	N	Mean	Median	Std. Dev.
Pos	150,030	178	111	184
Neu	150,030	339	253	259
Neg	150,030	396	305	284
Des_Release	1,790	426.30	397.32	122.84
T-S_Release	1,790	25.11	24	8.45
Des_Vote	150,030	454.13	421	152.38
T-S_Vote	150,030	21.88	23	8.81

Table 3B: Crowd-Voting, by Year [1]

	Votes	N	Mean	Median	Std. Dev.
2004	Pos	9,436	63	48	52
	Neu	9,436	125	118	57
	Neg	9,436	205	193	83
2005	Pos	23,268	139	106	136
	Neu	23,268	229	190	173
	Neg	23,268	300	289	191
2006	Pos	20,588	322	327	250
	Neu	20,588	588	701	295
	Neg	20,588	727	761	284
2007	Pos	23,047	279	269	234
	Neu	23,047	565	642	303
	Neg	23,047	673	680	273
2008	Pos	25,613	174	136	168
	Neu	25,613	337	317	222
	Neg	25,613	396	342	209
2009	Pos	27,545	115	99	89
	Neu	27,545	225	243	86
	Neg	27,545	230	230	93
2010	Pos	20,533	111	102	86
	Neu	20,533	216	239	93
	Neg	20,533	175	176	72

Notes:
1. Std. Dev. = Standard Deviation.

Table 4A: Descriptive Statistics of Revenues [1]

	N	Mean	Median	Std. Dev.
Voters	1,790	417	433	328
Non-Voters	1,790	7,748	5,119	6,609
Outsiders	1,790	11,023	9,028	9,442
Total	1,790	19,188	16,560	13,855

Table 4B: Revenues, by Group, by Year [1]

		N	Mean	Median	Std. Dev.
2004	Voters	52	205	193	83
	Non-Voters	52	125	118	57
	Outsiders	52	63	48	52
2005	Voters	171	300	289	191
	Non-Voters	171	229	190	173
	Outsiders	171	139	106	136
2006	Voters	256	727	761	284
	Non-Voters	256	588	701	295
	Outsiders	256	322	327	250
2007	Voters	348	673	680	273
	Non-Voters	348	565	642	303
	Outsiders	348	279	269	234
2008	Voters	331	396	342	209
	Non-Voters	331	337	317	222
	Outsiders	331	174	136	168
2009	Voters	347	230	230	93
	Non-Voters	347	225	243	86
	Outsiders	347	115	99	89
2010	Voters	285	175	176	72
	Non-Voters	285	216	239	93
	Outsiders	285	111	102	86

Notes:
1. Std. Dev. = Standard Deviation.

Table 5: Determinants of Revenues [1]

Variable [2]	Coeff.	T-stat.
Pos: 2004	1.33 ***	8.20
Pos: 2005	1.07 ***	8.16
Pos: 2006	1.10 ***	8.26
Pos: 2007	1.36 ***	9.05
Pos: 2008	1.46 ***	10.42
Pos: 2009	1.20 ***	13.87
Pos: 2010	1.46 ***	12.11
Neu: 2004	-1.20 ***	-6.06
Neu: 2005	-.52 **	-3.11
Neu: 2006	-.62 **	-3.21
Neu: 2007	-.89 ***	-5.64
Neu: 2008	-1.22 ***	-5.46
Neu: 2009	-1.10 ***	-6.03
Neu: 2010	-.55 **	-2.76
Neg: 2004	.37	1.94
Neg: 2005	-.38 **	-2.92
Neg: 2006	-.45 **	-3.12
Neg: 2007	-.25	-1.86
Neg: 2008	-.31	-1.66
Neg: 2009	-.16	-1.32
Neg: 2010	-.23	-1.89
Des_Release	-.09	-0.81
T-S_Release	-.02	-0.41
σ_{du} [3]	.65 ***	
$\chi^2(\varepsilon_{du}, \xi_{du})$ [4]	18.45 ***	
Log-likelihood	-6066.56	

Notes:
1. Coeff. = Coefficient; T-stat. = T-statistic. Standard errors clustered by user. All tests two sided. * = $p < .05$; ** = $p < .01$; *** = $p < .001$.
2. Pos:*y*, Neu:*y*, and Neg:*y* are year-specific coefficients associated with the positive, neutral, and negative votes in year *y* respectively.
3. σ_{du} = Standard Deviation of ε_{du}.
4. $\chi^2(\varepsilon_{du}, \xi_{du})$ = Chisq Statistic for independence between ε_{du} and ξ_{du}.

Table 6: Determinants of Revenues from Different Individuals [1]

Variable [2]	Voters		Non-Voters		Outsiders	
	Coeff.	T-stat.	Coeff.	T-stat.	Coeff.	T-stat.
Pos: 2004	1.46 ***	6.41	1.26 ***	7.11	1.39 ***	8.47
Pos: 2005	1.59 ***	8.69	1.04 ***	8.15	1.06 ***	7.68
Pos: 2006	1.73 ***	9.88	1.05 ***	8.11	1.03 ***	7.55
Pos: 2007	1.97 ***	10.10	1.41 ***	9.22	1.28 ***	8.04
Pos: 2008	1.89 ***	8.96	1.50 ***	10.49	1.42 ***	9.67
Pos: 2009	1.63 ***	7.65	1.16 ***	13.06	1.21 ***	13.40
Pos: 2010	1.91 ***	9.27	1.41 ***	11.82	1.48 ***	11.71
Neu: 2004	-.98 **	-3.21	-1.14 ***	-5.17	-1.29 ***	-6.19
Neu: 2005	-.88 ***	-3.59	-.56 **	-3.11	-.54 **	-3.09
Neu: 2006	-1.05 ***	-4.74	-.67 ***	-3.37	-.47 *	-2.33
Neu: 2007	-1.35 ***	-5.71	-1.03 ***	-6.89	-.72 ***	-4.15
Neu: 2008	-2.05 ***	-7.73	-1.37 ***	-6.37	-1.07 ***	-4.63
Neu: 2009	-2.14 ***	-7.15	-1.14 ***	-6.35	-1.04 ***	-5.90
Neu: 2010	-1.47 ***	-4.01	-.72 ***	-3.51	-.46 *	-2.20
Neg: 2004	.53	1.93	.32	1.66	.44 *	2.12
Neg: 2005	-.04	-.20	-.31 *	-2.21	-.39 **	-3.22
Neg: 2006	.32	1.75	-.40 **	-2.65	-.58 ***	-3.84
Neg: 2007	.72 ***	4.16	-.11	-.87	-.43 **	-2.97
Neg: 2008	.80 ***	5.13	-.05	-.28	-.50 **	-2.60
Neg: 2009	.83 ***	4.22	-.06	-.47	-.24 *	-2.03
Neg: 2010	.84 ***	4.03	-.09	-.67	-.33 **	-2.61
Des_Release	-.03	-.21	-.03	-.29	-.14	-1.22
T-S_Release	-.09	-1.55	.04	-.84	-.01	-.21
σ_{du} [3]	.86 ***		.81 ***		.83 ***	
$\chi^2(\varepsilon_{du}, \xi_{du})$ [4]	9.22 ***		19.07 ***		19.40 ***	
Log-likelihood	-6598.13		-6118.82		-6129.34	

Notes:
1. Coeff. = Coefficient; T-stat. = T-statistic. Standard errors clustered by user. All tests two sided. * = $p < .05$; ** = $p < .01$; *** = $p < .001$.
2. Pos:y, Neu:y, and Neg:y are year-specific coefficients associated with the positive, neutral, and negative votes in year y respectively.
3. σ_{du} = Standard Deviation of ε_{du}.
4. $\chi^2(\varepsilon_{du}, \xi_{du})$ = Chisq Statistics for independence between ε_{du} and ξ_{du}.

Table 7: Determinants of Short-Run and Long-Run Revenues [1, 2]

Variable [3]	Voters		Non-Voters		Outsiders	
	Short-run	Long-run	Short-run	Long-run	Short-run	Long-run
Pos: 2004	1.46 ***	2.23 **	1.26 ***	3.63 ***	1.39 ***	3.79 ***
Pos: 2005	1.59 ***	2.66 ***	1.04 ***	3.35 ***	1.06 ***	3.53 ***
Pos: 2006	1.73 ***	2.24 ***	1.05 ***	4.10 ***	1.03 ***	4.14 ***
Pos: 2007	1.97 ***	2.36 ***	1.41 ***	4.00 ***	1.28 ***	4.13 ***
Pos: 2008	1.89 ***	1.20 **	1.50 ***	3.25 ***	1.42 ***	3.59 ***
Neu: 2004	-.98 **	-.61	-1.14 ***	-3.51 ***	-1.29 ***	-3.59 ***
Neu: 2005	-.88 ***	-2.63 ***	-.56 **	-2.87 *	-.54 **	-2.98 *
Neu: 2006	-1.05 ***	-3.58 **	-.67 ***	-6.01 ***	-.47 *	-6.00 ***
Neu: 2007	-1.35 ***	-2.46 **	-1.03 ***	-4.20 ***	-.72 ***	-4.01 ***
Neu: 2008	-2.05 ***	-1.97 **	-1.37 ***	-3.61 ***	-1.07 ***	-3.50 ***
Neg: 2004	.53	.02	.32	.01	.44 *	-.02
Neg: 2005	-.04	1.00	-.31 *	-.20	-.39 **	-.31
Neg: 2006	.32	1.91 *	-.40 **	2.11	-.58 ***	2.10
Neg: 2007	.72 ***	2.13 **	-.11	.93	-.43 **	.54
Neg: 2008	.80 ***	.85	-.05	.28	-.50 **	-.04

Notes:
1. Standard errors clustered by user. All tests two sided. * = p < .05; ** = p < .01; *** = p < .001.
2. Short-run = revenues in the first 3 months after release; Long-run = revenues in months 4 to 24 after release.
3. Pos:*y*, Neu:*y*, and Neg:*y* are year-specific coefficients associated with the positive, neutral, and negative votes in year *y* respectively.

Figure 1: Time-line of a Design Submission

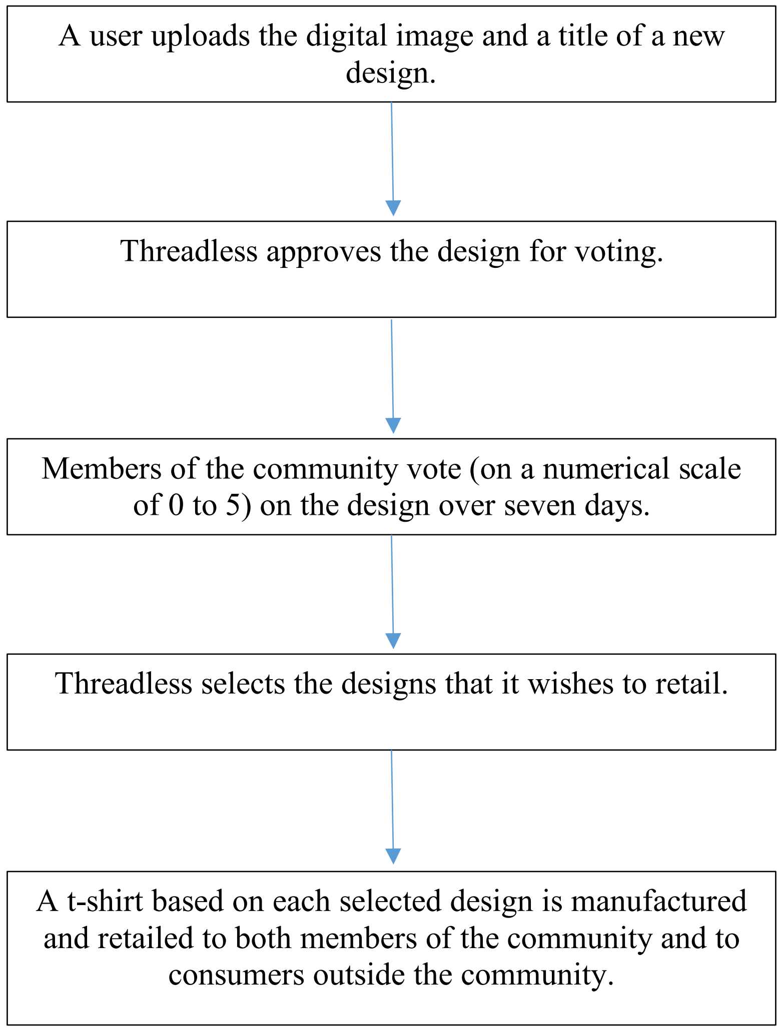
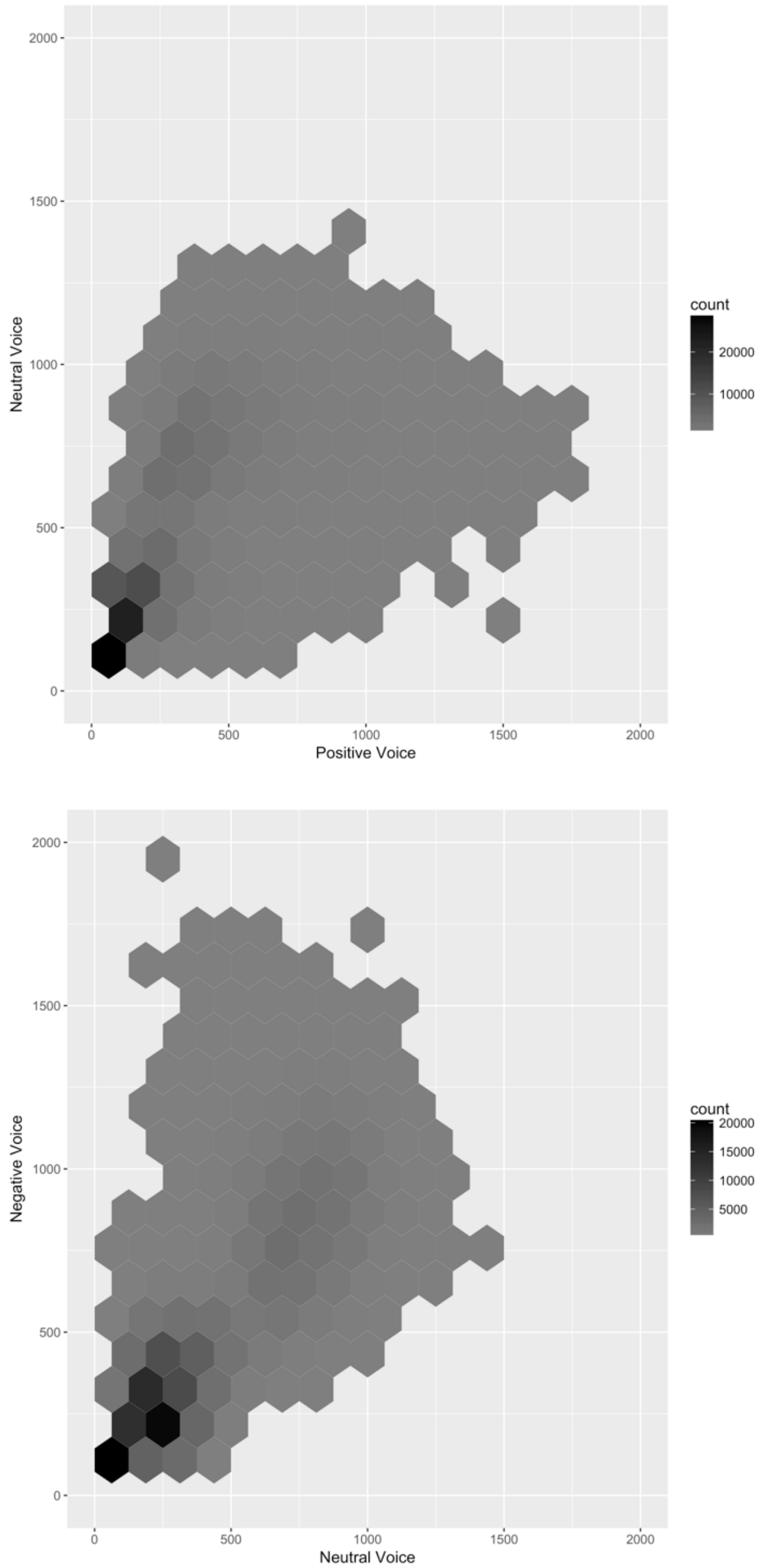
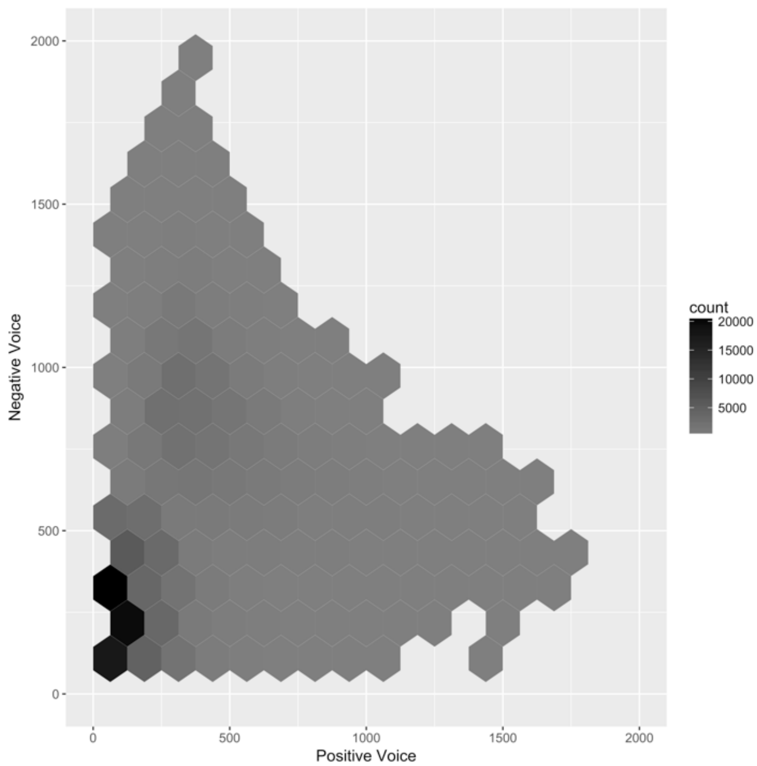


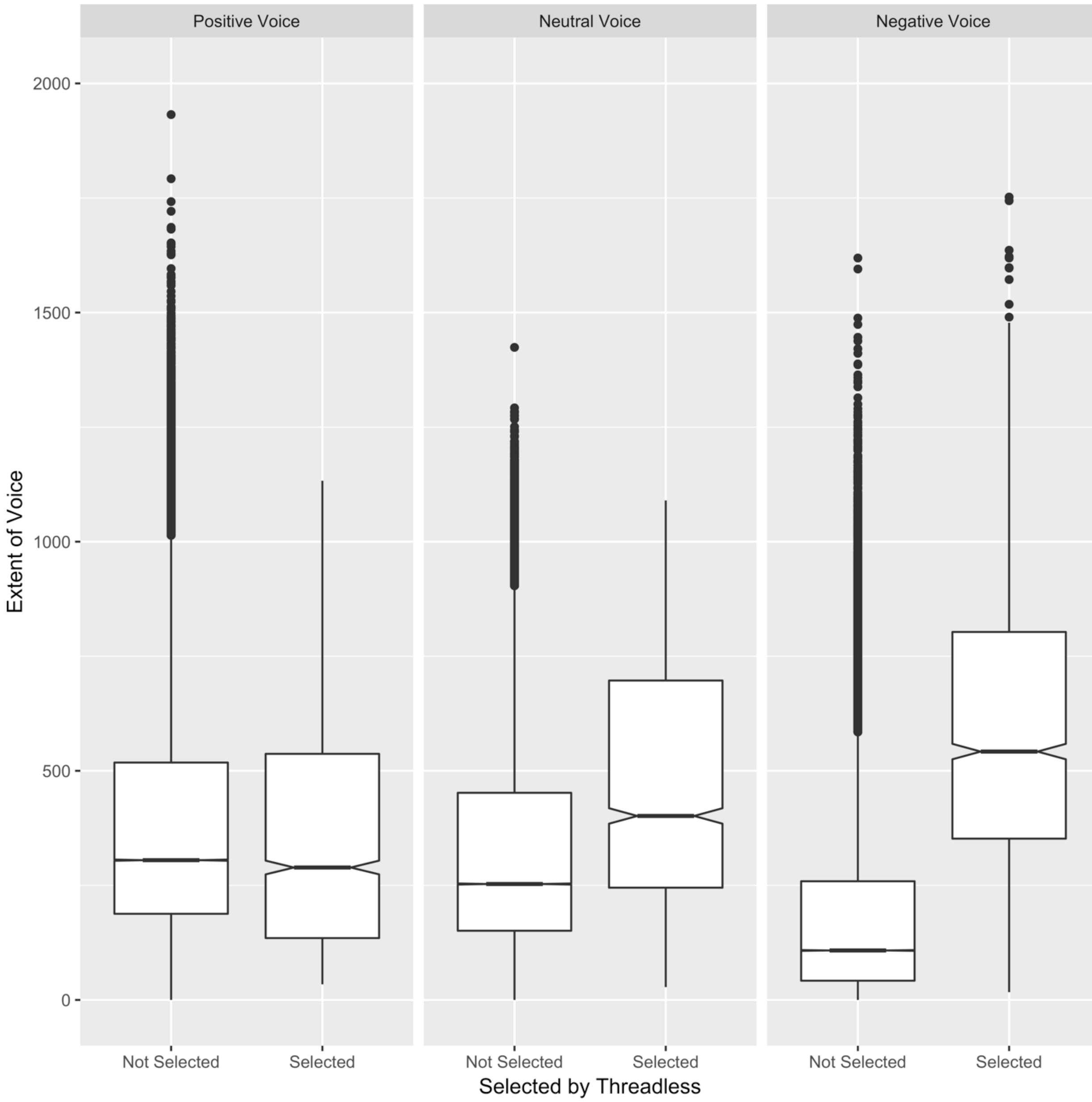
Figure 2: Heat-maps of Positive, Neutral, and Negative Votes [1]





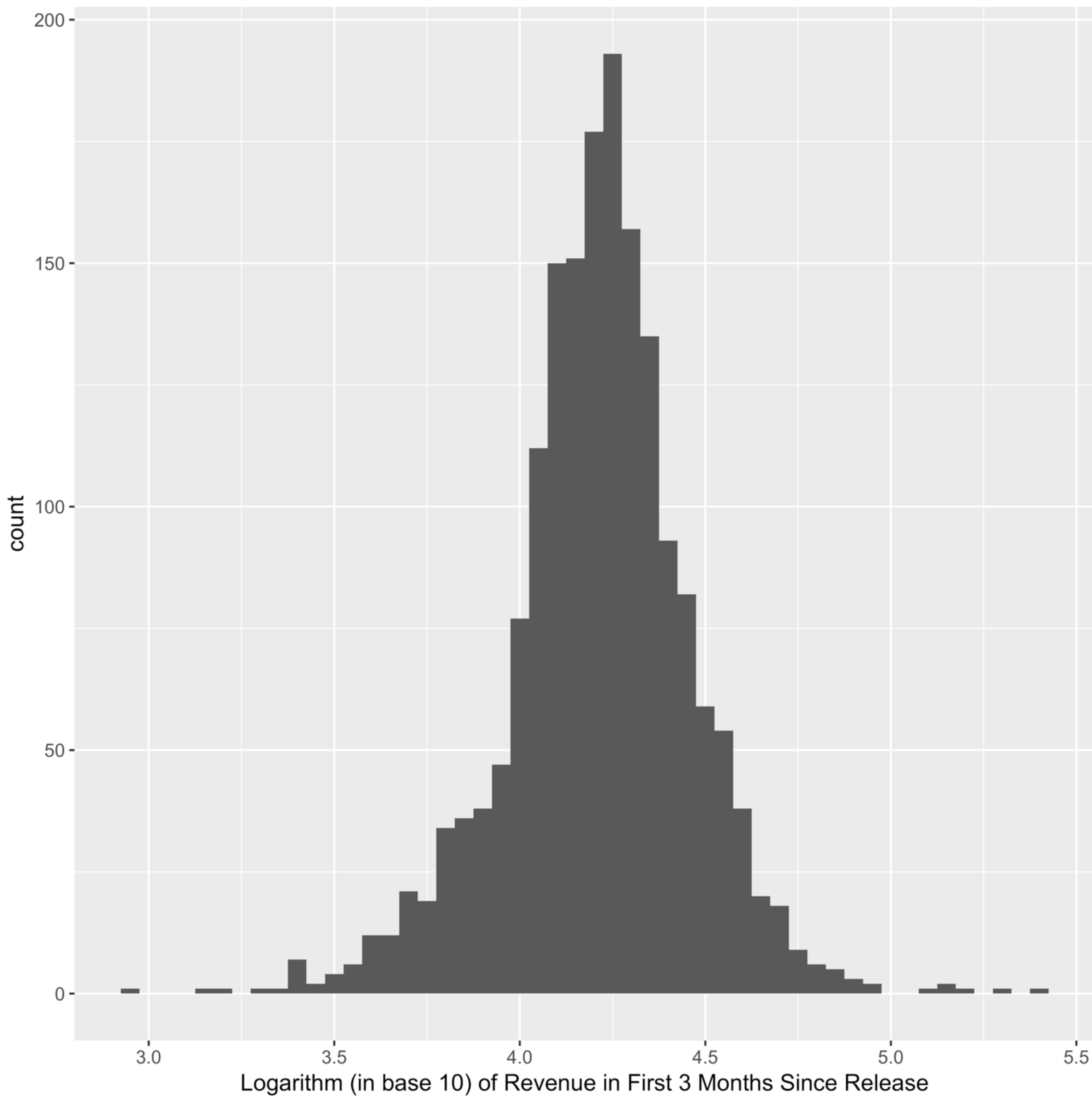
Notes:
1. The shade of each hexagonal bin corresponds to the number of designs with the corresponding positive, neutral, and negative votes.

Figure 3: Votes of the Crowd for Selected and Not Selected Designs [1]



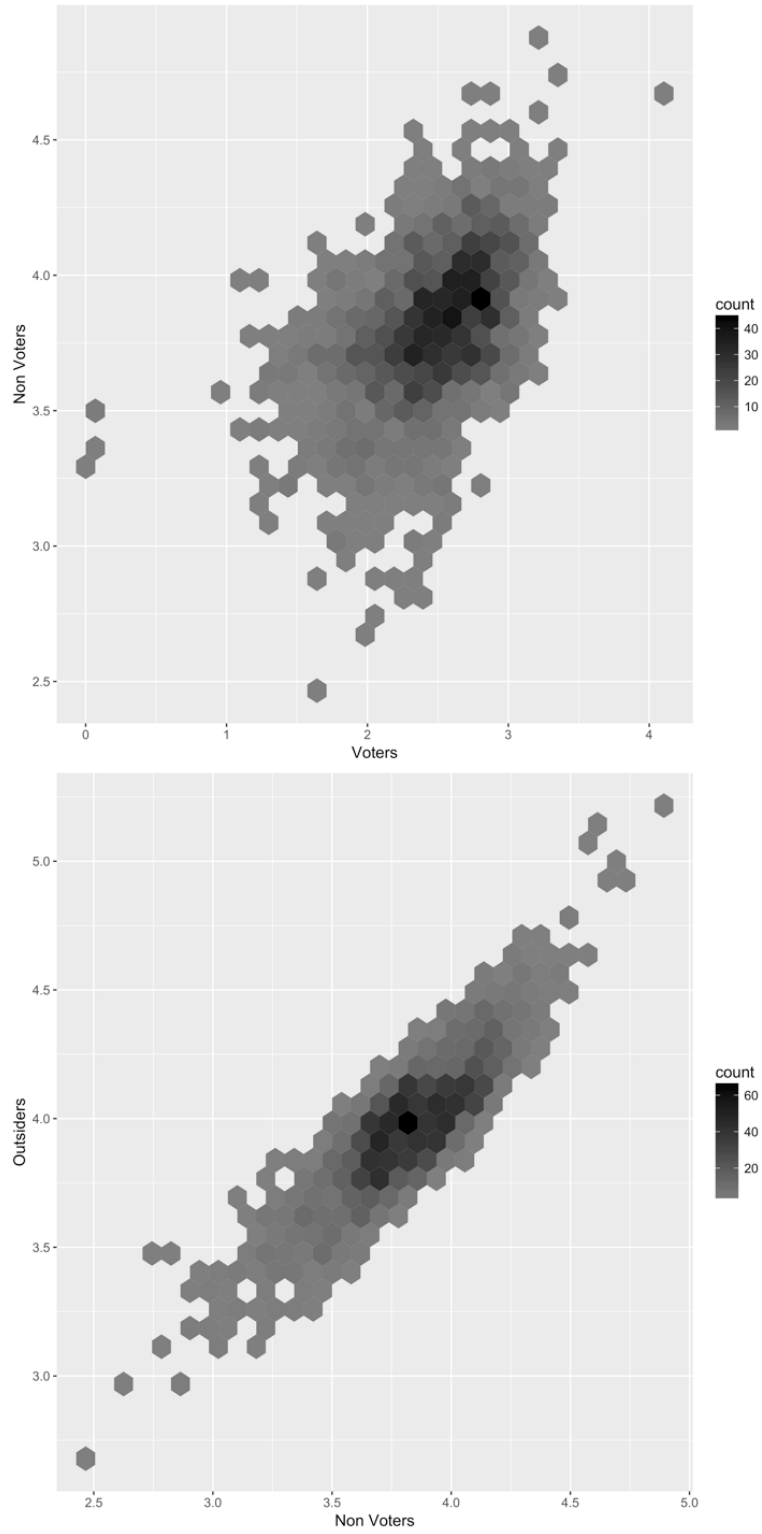
Notes:
1. A design is a point in the boxplot. The upper and lower hinge of the box correspond to the inter-quartile range (75th to the 25th percentile). The upper whisker extends to the largest value no further than 1.5 x the inter-quartile range from the upper hinge. The lower whisker extends to the smallest value no further than 1.5 x the inter-quartile range from the lower hinge. Data beyond the whiskers are plotted individually.

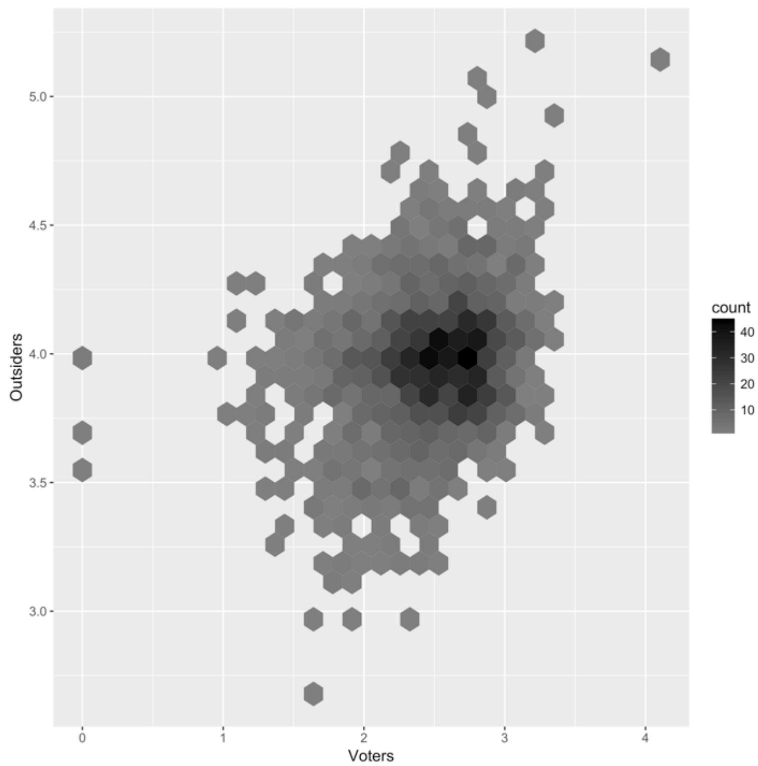
Figure 4: Histogram of Revenue [1]



Notes:
1. A selected design is a point in the histogram. For readability, the x-axis is the logarithm in base (10) of the revenue in the first 3 months.

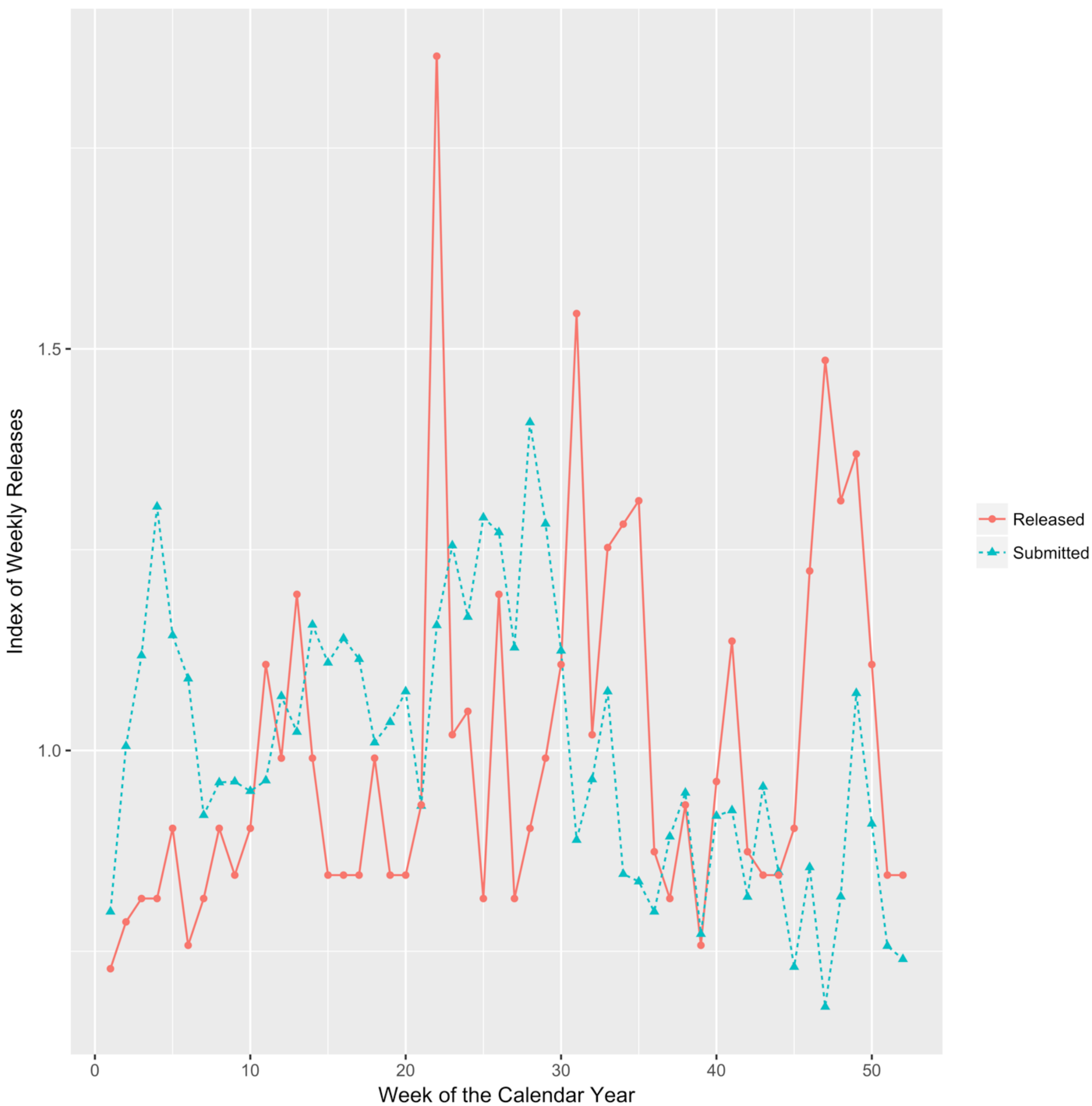
Figure 5: Revenues by Voters, Non-voters, and Outsiders [1]





Notes:
1. The shade of each hexagonal bin corresponds to the revenue of a selected design in the first 3 months since release, from the three different groups (voters, non-voters, and outsiders) of individuals.

Figure 6: Seasonality of Weekly Releases [1]



Notes:
1. The index is calculated by dividing the number of designs submitted or released in each calendar week, by the mean number of designs submitted or released in a calendar week, averaged over the 7 years of data.

WEB APPENDIX A: COPULAS

The decision to select design *d* is given by:

$$c_d = \begin{cases} 0 \; if \; z_d\gamma + \varepsilon_{sd} \leq 0 \\ 1 \; if \; z_d\gamma + \varepsilon_{sd} > 0 \end{cases}$$

where c_d is 0 if the design is not selected and 1 if the design is selected, z_d are independent variables, and ε_{sd} is a stochastic (error) term. Revenues are observed if $c_d = 1$ and are then modeled as:

$$R_d = x_d\beta + \varepsilon_{rd},$$

where R_d is revenue from the product based on selected design *d*, x_d is a vector of independent variables, and ε_{rd} is a stochastic (error) term. The likelihood of an observation (c_d, R_d) is given by:

$$\mathcal{L}(c_d, R_d) = \begin{cases} \mathcal{L}(z_d\gamma + \varepsilon_{sd} > 0, R_d = x_d\beta + \varepsilon_{rd}) & if \; c_d = 1 \\ \mathcal{L}(z_d\gamma + \varepsilon_{sd} \leq 0) & if \; c_d = 0 \end{cases}$$

The empirical challenge is to construct a robust model of the joint distribution of the stochastic terms $(\varepsilon_{sd},\ \varepsilon_{rd})$. This is because, if the stochastic terms are statistically dependent, then solely examining the marginal distribution of ε_{rd} or inaccurately modeling the joint distribution of ε_{sd} and ε_{rd} may lead to biased estimates.

The (classic) Heckman model assumes ε_{sd} and ε_{rd} are distributed bivariate normal. However, adopting this parametric restriction induces several restrictive assumptions. For example, the bivariate standard normal distribution is symmetrical:

$$\mathrm{P}(\varepsilon_{sd} > c_1, \varepsilon_{rd} > c_2) = \mathrm{P}(\varepsilon_{sd} > c_2, \varepsilon_{rd} > c_1),$$

for any arbitrary set of values (c_1, c_2). In our model, this restriction corresponds to the probability of a draw from the right tail of ε_{sd} (being unusually lucky in selection) and a draw from the left tail of ε_{rd}(being unusually unlucky in revenues), being equivalent to a draw from the left tail of ε_{sd} (being unusually unlucky in selection) and a draw from the left tail of ε_{rd} (being unusually lucky in revenues).

To relax parametric restrictions, we turn to copulas to generate different joint distribution of the stochastic terms. A copula generates a joint distribution as follows:

$$F(y_1, y_2) = C\{F_1(y_1), F_2(y_2); \theta\} = C(\varphi_1, \varphi_2; \theta)$$

where $F(y_1, y_2)$ is the bivariate cumulative distribution, $\varphi_i = F_i(y_i)$ is the marginal cumulative distribution of y_i for i = 1, 2, and $C(\varphi_1, \varphi_2; \theta)$ is the copula. θ is the dependence parameter in the copula that governs the extent of dependence across the ingredient marginals.

Table A1 summarizes the functional form of the nine copulas we consider in our study: Ali-Mikhail-Haq, Clayton, Farlie-Gumbel-Morgenstern, Frank, Gaussian, Plackett, Product, Gumbel, and Joe. Please see Durante and Sempi (2010) for a detailed discussion of the statistical properties of these copulas. Taken together, they provide a robust capture of statistical dependence. As is common in likelihood based estimation, we estimate a model corresponding to each copula, and use log-likelihood based inference (via the Akaike Information Criteria and the Bayesian Information Criteria) to select the copula-based model that fits our data best.

Table A1: Functional Form of Various Copulas [1]

Name	Copula Function [2]
Ali-Mikhail-Haq	$\varphi_1\varphi_2\{1-\theta(1-\varphi_1)(1-\varphi_2)\}^{-1}$
Clayton	$({\varphi_1}^{-\theta}+{\varphi_2}^{-\theta}-1)^{-1/\theta}$
Farlie-Gumbel-Morgenstern	$\varphi_1\varphi_2\{1+\theta(1-\varphi_1)(1-\varphi_2)\}$
Frank	$-\theta^{-1}\log\left\{1+\frac{(e^{-\theta\varphi_1}-1)(e^{-\theta\varphi_2}-1)}{(e^{-\theta}-1)}\right\}$
Gaussian [3]	$\Phi_2\{\Phi^{-1}(\varphi_1),\Phi^{-1}(\varphi_2);\theta\}$
Joe [4]	$1-\left\{(\tilde{\varphi}_1)^\theta+(\tilde{\varphi}_2)^\theta-(\tilde{\varphi}_1\tilde{\varphi}_2)^\theta\right\}^{1/\theta}$
Plackett [5]	$\frac{r-\sqrt{r^2-4\,\phi_1\,\phi_2\theta(\theta-1)}}{2(\theta-1)}$
Product	$\phi_1\,\phi_2$

Notes:
1. $\varphi_i = F_i(y_i)$ is the marginal cumulative distribution of y_i, for i = 1, 2.
2. $C(\varphi_1,\varphi_2;\theta)$ is the copula function, with dependency parameter(s) θ.
3. Φ_2 is the cumulative distribution function of the bivariate normal distribution; Φ^{-1} is the inverse cumulative distribution function of the standard normal distribution.
4. $\tilde{\varphi}_j = 1-\varphi_j$.
5. $r = 1+(\theta-1)(\phi_1+\phi_2)$.

WEB APPENDIX B: ESTIMATES OF EQUATION (2)

This appendix describe estimates of focal crowd-voting variables of equation (2) (in the main text), corresponding to models described in tables 5, 6 and 7 in the main text.

Table B1: Estimates of Equation (2) for Table 5 and 6 [1]

Variable [2]	Total [3]	Voters [4]	Non-Voters [5]	Outsiders [6]
Pos: 2004	1.75 ***	1.84 ***	1.77 ***	1.73 ***
Pos: 2005	2.23 ***	2.19 ***	2.23 ***	2.23 ***
Pos: 2006	2.39 ***	2.41 ***	2.40 ***	2.41 ***
Pos: 2007	2.67 ***	2.68 ***	2.67 ***	2.69 ***
Pos: 2008	3.15 ***	3.10 ***	3.13 ***	3.15 ***
Pos: 2009	2.35 ***	2.36 ***	2.33 ***	2.37 ***
Pos: 2010	2.36 ***	2.33 ***	2.38 ***	2.36 ***
Neu: 2004	-1.33 ***	-1.40 ***	-1.33 ***	-1.33 ***
Neu: 2005	-1.40 ***	-1.38 ***	-1.39 ***	-1.40 ***
Neu: 2006	-1.36 ***	-1.36 ***	-1.36 ***	-1.36 ***
Neu: 2007	-1.99 ***	-2.00 ***	-1.97 ***	-2.03 ***
Neu: 2008	-2.07 ***	-2.17 ***	-2.04 ***	-2.09 ***
Neu: 2009	-1.99 ***	-2.01 ***	-1.98 ***	-2.00 ***
Neu: 2010	-1.20 ***	-1.14 ***	-1.22 ***	-1.19 ***
Neg: 2004	-.55 *	-.47	-.58 *	-.52 *
Neg: 2005	-.47 *	-.48 *	-.48 **	-.47 **
Neg: 2006	-.76 ***	-.75 ***	-.76 ***	-.76 ***
Neg: 2007	-.09	-.06	-.06	-.08
Neg: 2008	-.06	-.17	-.13	-.05
Neg: 2009	-.01	.01	-.01	.01
Neg: 2010	-.46 *	-.52 **	-.47 *	-.44 *

Notes:
1. Standard errors clustered by user. All tests two sided. * = $p < .05$; ** = $p < .01$; *** = $p < .001$.
2. Pos:*y*, Neu:*y*, and Neg:*y* are year-specific coefficients associated with the positive, neutral, and negative votes in year *y* respectively.
3. Total = Total revenue.
4. Voters = Revenue from users who voted on the design.
5. Non-Voters = Revenue from users who did not vote on the design.
6. Outsiders = Revenue from individuals who are not members of the Threadless community.

Table B2: Estimates of Equation (2) for Table 7 [1, 2]

Variable [3]	Voters [4]		Non-Voters [5]		Outsiders [6]	
	Short-run	Long-run	Short-run	Long-run	Short-run	Long-run
Pos: 2004	1.84 ***	1.77 ***	1.77 ***	1.79 ***	1.73 ***	1.78 ***
Pos: 2005	2.19 ***	2.23 ***	2.23 ***	2.15 ***	2.23 ***	2.13 ***
Pos: 2006	2.41 ***	2.43 ***	2.40 ***	2.41 ***	2.41 ***	2.43 ***
Pos: 2007	2.68 ***	2.72 ***	2.67 ***	2.74 ***	2.69 ***	2.74 ***
Pos: 2008	3.10 ***	3.22 ***	3.13 ***	3.14 ***	3.15 ***	3.12 ***
Neu: 2004	-1.40 ***	-1.31 ***	-1.33 ***	-1.35 ***	-1.33 ***	-1.35 ***
Neu: 2005	-1.38 ***	-1.41 ***	-1.39 ***	-1.32 ***	-1.40 ***	-1.31 ***
Neu: 2006	-1.36 ***	-1.38 ***	-1.36 ***	-1.29 ***	-1.36 ***	-1.30 ***
Neu: 2007	-2.00 ***	-2.11 ***	-1.97 ***	-2.08 ***	-2.03 ***	-2.07 ***
Neu: 2008	-2.17 ***	-2.21 ***	-2.04 ***	-2.14 ***	-2.09 ***	-2.11 ***
Neg: 2004	-.47	-.56 *	-.58 *	-.56 *	-.52 *	-.55 *
Neg: 2005	-.48 *	-.49 *	-.48 **	-.51 *	-.47 **	-.50 *
Neg: 2006	-.75 ***	-.77 ***	-.76 ***	-.84 ***	-.76 ***	-.84 ***
Neg: 2007	-.06	-.05	-.06	-.03	-.08	-.04
Neg: 2008	-.17	-.17	-.13	-.17	-.05	-.17

Notes:
1. Standard errors clustered by user. All tests two sided. * = $p < .05$; ** = $p < .01$; *** = $p < .001$.
2. Short-run = revenues in the first 3 months after release; Long-run = revenues in months 4 to 24 after release.
3. Pos:*y*, Neu:*y*, and Neg:*y* are year-specific coefficients associated with the positive, neutral, and negative votes in year *y* respectively.
4. Voters = Revenue from users who voted on the design.
5. Non-Voters = Revenue from users who did not vote on the design.
6. Outsiders = Revenue from individuals who are not members of the Threadless community.

WEB APPENDIX C: SENSITIVITY ANALYSES

Table C1a: Drop Designs with Average Rating less than 1 [1]

Variable [2]	Voters [3]	Non-Voters [4]	Outsiders [5]
Pos: 2004	1.46 ***	1.26 ***	1.39 ***
Pos: 2005	1.59 ***	1.04 ***	1.06 ***
Pos: 2006	1.73 ***	1.05 ***	1.03 ***
Pos: 2007	1.97 ***	1.41 ***	1.28 ***
Pos: 2008	1.89 ***	1.50 ***	1.42 ***
Pos: 2009	1.63 ***	1.16 ***	1.21 ***
Pos: 2010	1.91 ***	1.41 ***	1.48 ***
Neu: 2004	-0.98 **	-1.14 ***	-1.29 ***
Neu: 2005	-0.88 ***	-0.56 **	-0.54 **
Neu: 2006	-1.05 ***	-0.67 ***	-0.47 *
Neu: 2007	-1.35 ***	-1.03 ***	-0.72 ***
Neu: 2008	-2.05 ***	-1.37 ***	-1.07 ***
Neu: 2009	-2.14 ***	-1.14 ***	-1.04 ***
Neu: 2010	-1.47 ***	-0.72 ***	-0.46 *
Neg: 2004	0.53	0.32	0.44 *
Neg: 2005	-0.04	-0.31 *	-0.39 **
Neg: 2006	0.32	-0.40 **	-0.58 ***
Neg: 2007	0.72 ***	-0.11	-0.43 **
Neg: 2008	0.80 ***	-0.05	-0.50 **
Neg: 2009	0.83 ***	-0.06	-0.24 *
Neg: 2010	0.84 ***	-0.09	-0.33 **
Des_Release	-0.03	-0.03	-0.14
T-S_Release	-0.09	-0.04	-0.01

Notes:
1. Standard errors clustered by user. All tests two sided. * = $p < .05$; ** = $p < .01$; *** = $p < .001$.
2. Pos:*y*, Neu:*y*, and Neg:*y* are year-specific coefficients associated with the positive, neutral, and negative votes in year *y* respectively.
3. Voters = Revenue from users who voted on the design.
4. Non-Voters = Revenue from users who did not vote on the design.
5. Outsiders = Revenue from individuals who are not members of the Threadless community.

Table C1b: Drop Designs with Average Rating Larger than 4 [1]

Variable [2]	Voters [3]	Non-Voters [4]	Outsiders [5]
Pos: 2004	1.46 ***	1.26 ***	1.39 ***
Pos: 2005	1.60 ***	1.03 ***	1.06 ***
Pos: 2006	1.73 ***	1.05 ***	1.03 ***
Pos: 2007	1.97 ***	1.41 ***	1.27 ***
Pos: 2008	1.89 ***	1.50 ***	1.42 ***
Pos: 2009	1.64 ***	1.17 ***	1.18 ***
Pos: 2010	1.92 ***	1.42 ***	1.48 ***
Neu: 2004	-0.98 **	-1.14 ***	-1.28 ***
Neu: 2005	-0.88 ***	-0.55 **	-0.53 **
Neu: 2006	-1.05 ***	-0.66 ***	-0.46 *
Neu: 2007	-1.35 ***	-1.03 ***	-0.71 ***
Neu: 2008	-2.05 ***	-1.37 ***	-1.06 ***
Neu: 2009	-2.16 ***	-1.16 ***	-1.01 ***
Neu: 2010	-1.46 ***	-0.73 ***	-0.45 *
Neg: 2004	0.53	0.32	0.44 *
Neg: 2005	-0.05	-0.32 *	-0.41 ***
Neg: 2006	0.32	-0.40 **	-0.58 ***
Neg: 2007	0.72 ***	-0.11	-0.43 **
Neg: 2008	0.80 ***	-0.05	-0.50 **
Neg: 2009	0.83 ***	-0.06	-0.28 *
Neg: 2010	0.83 ***	-0.08	-0.33 *
Des_Release	-0.03	-0.03	-0.14
T-S_Release	-0.10	-0.04	-0.01

Notes:
1. Standard errors clustered by user. All tests two sided. * = $p < .05$; ** = $p < .01$; *** = $p < .001$.
2. Pos:*y*, Neu:*y*, and Neg:*y* are year-specific coefficients associated with the positive, neutral, and negative votes in year *y* respectively.
3. Voters = Revenue from users who voted on the design.
4. Non-Voters = Revenue from users who did not vote on the design.
5. Outsiders = Revenue from individuals who are not members of the Threadless community.

Table C1c: Drop Designs with Fewer than 100 Votes [1]

Variable [2]	Voters [3]	Non-Voters [4]	Outsiders [5]
Pos: 2004	1.45 ***	1.25 ***	1.37 ***
Pos: 2005	1.59 ***	1.03 ***	1.05 ***
Pos: 2006	1.72 ***	1.05 ***	1.02 ***
Pos: 2007	1.95 ***	1.40 ***	1.27 ***
Pos: 2008	1.87 ***	1.49 ***	1.42 ***
Pos: 2009	1.63 ***	1.16 ***	1.21 ***
Pos: 2010	1.90 ***	1.40 ***	1.47 ***
Neu: 2004	-0.97 **	-1.14 ***	-1.28 ***
Neu: 2005	-0.88 ***	-0.56 **	-0.53 **
Neu: 2006	-1.05 ***	-0.67 ***	-0.47 *
Neu: 2007	-1.35 ***	-1.03 ***	-0.72 ***
Neu: 2008	-2.06 ***	-1.37 ***	-1.07 ***
Neu: 2009	-2.18 ***	-1.19 ***	-1.09 ***
Neu: 2010	-1.48 ***	-0.74 ***	-0.47 *
Neg: 2004	0.51	0.29	0.41
Neg: 2005	-0.05	-0.33 *	-0.41 ***
Neg: 2006	0.31	-0.41 **	-0.59 ***
Neg: 2007	0.71 ***	-0.12	-0.43 **
Neg: 2008	0.79 ***	-0.05	-0.50 **
Neg: 2009	0.83 ***	-0.06	-0.25 *
Neg: 2010	0.83 ***	-0.09	-0.34 **
Des_Release	-0.03	-0.04	-0.15
T-S_Release	-0.09	-0.04	-0.01

Notes:
1. Standard errors clustered by user. All tests two sided. * = $p < .05$; ** = $p < .01$; *** = $p < .001$.
2. Pos:*y*, Neu:*y*, and Neg:*y* are year-specific coefficients associated with the positive, neutral, and negative votes in year *y* respectively.
3. Voters = Revenue from users who voted on the design.
4. Non-Voters = Revenue from users who did not vote on the design.
5. Outsiders = Revenue from individuals who are not members of the Threadless community.

Table C2: Add Standard Deviation of Votes [1]

Variable [2]	Voters [3]	Non-Voters [4]	Outsiders [5]
Pos: 2004	1.64 ***	1.30 ***	1.47 ***
Pos: 2005	1.68 ***	1.05 ***	1.10 ***
Pos: 2006	1.86 ***	1.08 ***	1.09 ***
Pos: 2007	2.12 ***	1.44 ***	1.34 ***
Pos: 2008	2.04 ***	1.53 ***	1.48 ***
Pos: 2009	1.66 ***	1.17 ***	1.23 ***
Pos: 2010	1.91 ***	1.41 ***	1.48 ***
Neu: 2004	-1.28 ***	-1.21 ***	-1.43 ***
Neu: 2005	-1.24 ***	-0.63 **	-0.69 ***
Neu: 2006	-1.43 ***	-0.75 ***	-0.64 **
Neu: 2007	-1.71 ***	-1.11 ***	-0.88 ***
Neu: 2008	-2.40 ***	-1.44 ***	-1.21 ***
Neu: 2009	-2.35 ***	-1.18 ***	-1.13 ***
Neu: 2010	-1.71 ***	-0.78 ***	-0.57 **
Neg: 2004	0.63 **	0.34	0.50 *
Neg: 2005	0.24	-0.26	-0.27 *
Neg: 2006	0.56 **	-0.34 *	-0.46 **
Neg: 2007	0.95 ***	-0.06	-0.32 *
Neg: 2008	1.04 ***	-0.00	-0.39 *
Neg: 2009	1.07 ***	-0.01	-0.14
Neg: 2010	1.08 ***	-0.03	-0.22
Des_Release	-0.03	-0.03	-0.14
T-S_Release	-0.09	-0.04	-0.01
Std_dev [6]	-0.98 **	-0.21	-0.43

Notes:
1. Standard errors clustered by user. All tests two sided. * = p < .05; ** = p < .01; *** = p < .001.
2. Pos:*y*, Neu:*y*, and Neg:*y* are year-specific coefficients associated with the positive, neutral, and negative votes in year *y* respectively.
3. Voters = Revenue from users who voted on the design.
4. Non-Voters = Revenue from users who did not vote on the design.
5. Outsiders = Revenue from individuals who are not members of the Threadless community.
6. Std_Dev = Standard deviation of votes.

Table C3: Add Time between Selection and Release [1]

Variable [2]	Voters [3]	Non-Voters [4]	Outsiders [5]
Pos: 2004	1.40 ***	1.25 ***	1.39 ***
Pos: 2005	1.54 ***	1.03 ***	1.06 ***
Pos: 2006	1.64 ***	1.05 ***	1.03 ***
Pos: 2007	1.92 ***	1.40 ***	1.28 ***
Pos: 2008	2.01 ***	1.50 ***	1.43 ***
Pos: 2009	1.63 ***	1.16 ***	1.21 ***
Pos: 2010	1.89 ***	1.40 ***	1.48 ***
Neu: 2004	-0.98 **	-1.14 ***	-1.28 ***
Neu: 2005	-0.91 ***	-0.56 **	-0.54 **
Neu: 2006	-1.15 ***	-0.68 ***	-0.47 *
Neu: 2007	-1.43 ***	-1.04 ***	-0.72 ***
Neu: 2008	-2.08 ***	-1.37 ***	-1.07 ***
Neu: 2009	-2.00 ***	-1.13 ***	-1.04 ***
Neu: 2010	-1.58 ***	-0.73 ***	-0.46 *
Neg: 2004	0.43	0.31	0.44 *
Neg: 2005	-0.04	-0.32 *	-0.39 **
Neg: 2006	0.28	-0.40 **	-0.58 ***
Neg: 2007	0.68 ***	-0.11	-0.43 **
Neg: 2008	0.96 ***	-0.04	-0.50 *
Neg: 2009	0.86 ***	-0.06	-0.24 *
Neg: 2010	0.92 ***	-0.08	-0.33 **
Des_Release	-0.03	-0.03	-0.14
T-S_Release	-0.09	-0.04	-0.01
Time_Between [6]	-0.98 **	-0.21	-0.43

Notes:
1. Standard errors clustered by user. All tests two sided. * = p < .05; ** = p < .01; *** = p < .001.
2. Pos:*y*, Neu:*y*, and Neg:*y* are year-specific coefficients associated with the positive, neutral, and negative votes in year *y* respectively.
3. Voters = Revenue from users who voted on the design.
4. Non-Voters = Revenue from users who did not vote on the design.
5. Outsiders = Revenue from individuals who are not members of the Threadless community.
6. Time_Between = Natural logarithm of time between when the design was submitted and released to retail.

Table C4: Add Time on Platform [1]

Variable [2]	Voters [3]		Non-Voters [4]		Outsiders [5]	
Pos: 2004	1.46	***	1.26	***	1.38	***
Pos: 2005	1.60	***	1.02	***	1.04	***
Pos: 2006	1.72	***	1.07	***	1.05	***
Pos: 2007	1.96	***	1.42	***	1.27	***
Pos: 2008	1.88	***	1.51	***	1.44	***
Pos: 2009	1.63	***	1.18	***	1.23	***
Pos: 2010	1.90	***	1.42	***	1.50	***
Neu: 2004	-0.97	**	-1.14	***	-1.29	***
Neu: 2005	-0.89	***	-0.53	**	-0.50	**
Neu: 2006	-1.04	***	-0.69	***	-0.50	*
Neu: 2007	-1.34	***	-1.03	***	-0.72	***
Neu: 2008	-2.04	***	-1.39	***	-1.10	***
Neu: 2009	-2.14	***	-1.15	***	-1.05	***
Neu: 2010	-1.48	***	-0.72	***	-0.47	*
Neg: 2004	0.53		0.30		0.43	*
Neg: 2005	-0.03		-0.34	*	-0.42	***
Neg: 2006	0.32		-0.39	**	-0.56	***
Neg: 2007	0.72	***	-0.10		-0.42	**
Neg: 2008	0.80	***	-0.05		-0.49	*
Neg: 2009	0.83	***	-0.04		-0.23	
Neg: 2010	0.84	***	-0.08		-0.33	**
Des_Release	-0.03		-0.04		-0.14	
T-S_Release	-0.09		-0.04		-0.01	
Time On Platform [6]	0.01		-0.01	*	-0.02	*

Notes:
1. Standard errors clustered by user. All tests two sided. * = $p < .05$; ** = $p < .01$; *** = $p < .001$.
2. Pos:*y*, Neu:*y*, and Neg:*y* are year-specific coefficients associated with the positive, neutral, and negative votes in year *y* respectively.
3. Voters = Revenue from users who voted on the design.
4. Non-Voters = Revenue from users who did not vote on the design.
5. Outsiders = Revenue from individuals who are not members of the Threadless community.
6. Time On Platform = Time since the user submitting the design registered an account.