The Impact of User-Generated Content on Sales: A Randomized Field Experiment

Jörn Grahl, Franz Rothlauf
Information Systems and Business Administration
Johannes Gutenberg-University Mainz
grahl, rothlauf@uni-mainz.de

Oliver Hinz
Electronic Markets
TU Darmstadt
hinz@wi.tu-darmstadt.de

Abstract

This study examines the causal relationship between popularity information and purchasing behavior in an online store. In a randomized field experiment we exogenously manipulated the visibility of user-generated content similar to Google's +1s or Facebook's Likes. Displaying the number of people who "Like" a product caused a +12.97% sales increase (13,883.74 EUR in the treatment group; 12,289.46 EUR in the control group). We find that popularity information influences shopping behavior significantly if it is displayed in the consumers' leisure time. This result is consistent with observational learning. For well-planned and goal-oriented purchases, knowing the preferences of others is of little importance. This information is more valuable on not so goal-oriented and, hence, more time-consuming shopping trips where consumers are searching for interesting, new products. The results also suggest that Likes have a significant monetary value, but without orthogonal variation, the valuation of Likes can easily be overestimated (by a factor of 2.26 in our sample).

1 Introduction

In February 2013, San Francisco-based startup Pinterest raised $200 million of funding, bringing its company value to $2.5 billion (Mitroff 2013). Pinterest maintains a website where users can "pin" movies, services and all kinds of products to their board, so that these preferences and tastes become visible to the public or to friends. This simple, but accessible idea has allowed Pinterest to grow rapidly: Despite only opening in 2010, the company is currently ranked among the top 40 websites with respect to site visits and was selected as one the 50 best websites by TIME Magazine (McCracken, 2011). Not far from Pinterest's headquarters resides Facebook, the world's leading online social network site. Over one billion users have registered for the online service, which allows them to manage social contacts, post and share content, and communicate their tastes by clicking on the "Like" buttons that are nowadays available on many websites. Founded in 2004, Facebook is valued at $172 billion as of July 2014—an impressive market valuation, especially for a relatively young firm.

The market capitalization of these firms cannot be justified only by its tangible assets like machines or inventories but mainly depends on the unique consumer data possessed by these firms. The value is evident in their users, who leave digital footprints, reveal their preferences, or make social recommendations. Especially these social recommendations seem to be interesting for businesses. Most online stores not only display detailed product information like price, availability, and technical specifications. They also show user-generated content like customer reviews, ratings, or simply the number of people who "like this product", making the data collected by companies like Pinterest and Facebook valuable for e-business. Surveys find that consumers trust online recommendations or opinions far more than they trust paid ads (e.g. The Nielsen Company, 2007).

It is therefore not surprising that a large body of scientific literature has tried to uncover the impact of user-generated content on demand or sales (e.g. Chevalier and Mayzlin, 2006; Dellarocas et al., 2007; Dhar and Chang, 2009; Trusov et al., 2009). A major challenge for researchers in this area is identifying the true causal effects of user-generated content. A common problem is potential endogeneity of user-generated content due to unobserved causes like product quality. To illustrate, consider the stylized case where a customer buys a product that has received many recommendations. She could make this purchase for (at least) two reasons: First, she might choose the product because of its high quality, and the good recommendations could reflect this quality. In this case, products are bought due to their high quality and user-generated content only measures the underlying product quality. Put differently, product quality may be the underlying cause of both the recommendations and the purchase; the customer, however, ultimately chooses the product for its quality, not because of the recommendations. Second, and on the contrary, she might buy the product because of the recommendations (i.e., the user-generated content). This effect is often denoted as herd behavior (e.g. Banerjee, 1992; Bikhchandani et al., 1992). The two purchases look identical in an observational dataset of sales and user-generated content, but it is important to distinguish between them. Purchases caused by product quality can occur without user-generated content, whereas herding sales are caused by the provision of user-generated content. It is econometrically challenging to distinguish between the two cases, but it is nonetheless important that online marketers understand whether user-generated content can cause herding, if this effect is economically significant, and which contextual factors moderate its strength.

The goal of this work is to examine the causal relationship between user-generated content and shopping behavior. To this end, we conducted a randomized field experiment where we exogenously manipulated the visibility of the user-generated content. This allows us to make causal inferences (Shadish et al., 2002). We restrict our study to one form of user-generated content: the rather simple but prevalent popularity information "Like". Likes and other popularity information aggregate the decisions of subjects into a simple summary statistic. It is a quantitative alternative to other methods of expressing reactions to products, like writing a text review. In comparison to dyadic information (e.g. "Your friend Bob has just purchased product XY", see, e.g., Aral and Walker, 2012; Bapna and Umyarov, 2013), which gives users information about the preferences of their peers in a social network, "Likes" are collected from a complete population. Although users usually give "Likes" only to products which they appreciate for their quality, there are also other reasons for giving "Likes". For example, due to social influence users may assign "Likes" to products that are already popular (Muchnik et al., 2013). For our experiment, we use real Likes data generated by users of an online store and ignore the users' motivations for assigning Likes to products. "Likes" are widely collected on the Web, largely because they do not cause high frictional costs and their plugins are easy to integrate. As a result, more than 20% of the top 1 million web sites collect and display Facebook Likes, while more than 14% use Google's equivalent "+1" (Web Technology Surveys, 2013).

In our experimental study, we show real Likes data to visitors of an online store and examine the economic consequences. We seek to answer the following questions:

1. Does the visibility of Likes influence sales in Euros?

2. How is the influence of Likes moderated by contextual factors?

We also examine how biased our results would be if we had obtained them from purely observational data. Furthermore, as the impact of popularity information may differ across blockbuster and niche products sales, we differentiate between these categories and investigate how Likes shift demand.

2 Literature Review

User-generated content is mostly about what others do, think or like. An important theoretical foundation for studying the influence of such information on the behavior of others was presented by Banerjee (1992) and Bikhchandani et al. (1992). They demonstrated analytically that it can be optimal for rational decision makers to observe the previous decisions of others and do "what others are doing, rather than using [their own] information." On this basis, a substantial body of empirical work has sought to understand how consumers react to other consumers' preferences or previous decisions by linking the characteristics of user-generated content to economic outcomes. Table 1 summarizes these studies and indicates how ours differs. The order in which we discuss the studies is based on the causal identification strategy employed: We start with observational studies (where no causal inference is possible), continue to studies that apply difference-in-difference estimators or instrumental variables, and finally consider experiments with random assignment, the latter of which are closest to our study.

Table 1: Literature on user-generated content

Study	Inference	Data & Methods	Type of UGC	Outcome	Moderators
Godes and Mayzlin (2004)	-	Observational data	Usenet conversations	Nielsen Ratings	-
Liu (2006)	-	Observational data	Messages on Yahoo!	Box office revenues	-
Dellarocas et al. (2007)	-	Observational data	Movie reviews	Box office revenues	-
Dhar and Chang (2009)	-	Observational data	Album reviews	Sales rank	-
Moe and Trusov (2011)	-	Observational data	Ratings	Dynamics of ratings and sales	-
Tirunillai and Tellis (2012)	-	Observational data	Reviews	Stock market performance	-
Chevalier and Mayzlin (2006)	Limited	Observational data + DiD	Book reviews	Sales rank	-
Forman et al. (2008)	Limited	Observational data + DiD	Book reviews	Sales rank, Inf. disclosure	-
Chintagunta et al. (2010)	Limited	Observational data + IV	Movie reviews	Box office revenues	-
Zhu and Zhang (2010)	Limited	Observational data + DiD	Game reviews	Influence of reviews	Popularity, experience
Salganik et al. (2006)	Yes	Laboratory experiment	Number of downloads	Downloads	-
Cai et al. (2009)	Yes	Field experiment	Top 5 dishes	Number of orders	Sporadic visits
Tucker and Zhang (2011)	Yes	Field experiment	Previous clicks	Click rates	Breadth of appeal
Muchnik et al. (2013)	Yes	Field experiment	Ratings	Future ratings	Topics, positivity
Matos et al. (2013)	Yes	Field experiment	Rank	Number of leases	-
This study	Yes	Field experiment	Number of Likes	Sales in Euros	Time

Godes and Mayzlin (2004) provide a seminal work based on observational data. Relating discussions about TV shows on the Usenet to their Nielsen ratings, they showed that the dispersion of conversations across different discussion groups has explanatory power for a movie rating. Several subsequent studies have explored the predictive value of user-generated content in different contexts and models. Liu (2006) and Dellarocas et al. (2007) studied the explanatory power of movie reviews on box office revenues. Dhar and Chang (2009) analyzed how online and offline music album reviews relate to their Amazon.com sales rank. Moe and Trusov (2011) assessed the dynamics of ratings for beauty products and how they relate to future ratings and sales. In a recent contribution, Tirunillai and Tellis (2012) related reviews to the dynamics of a firm's stock market performance.

As Godes and Mayzlin (2004) said, it is "difficult to draw clean inferences of causality with traditional econometrics." Thus, the following works study observational data with more sophisticated methods, like difference-in-differences (DiD) or instrumental variables. Chevalier and Mayzlin (2006) used DiD to study book sales at Amazon.com and Barnesandnobles.com, noting that improvements in a book's reviews increase the relative sales of the book at that site. Forman et al. (2008), also using DiD, provided evidence that disclosing the identity of reviewers is associated with an increase in future sales. In their study of movie box office revenues, Chintagunta et al. (2010) used exogenous data from markets where movies have previously been released as instruments for subsequent markets. They identify the valence of reviews as a main driver for box office performance. Zhu and Zhang (2010) utilized DiD to study the moderating role of product and consumer characteristics. Using data from Gamespot.com, they showed that, for the adoption of video games, reviews are more influential for less well-known titles and among players who are more experienced with using the Internet.

Similar to our study, several other researchers have conducted randomized experiments to understand how popularity influences choices. The analytical results by Banerjee (1992) were first confirmed in stylized laboratory experiments (e.g. Anderson and Holt, 1997), which suggested that visible popularity information can indeed cause decision makers to choose what is already popular and thereby spur herd behavior. Building on these findings, Salganik et al. (2006) developed an artificial electronic market for MP3s where visitors can choose to download unknown songs from unknown bands for free. If the number of downloads is visible, "popular songs are more popular and unpopular songs are less popular." This suggests that popularity information can trigger herd behavior and influence the distribution of demand in a realistic setting. In a randomized natural field experiment, Cai et al. (2009) showed guests in a restaurant the top choices of other guests and estimated the influence of this treatment on orders. They rule out salience effects as alternative explanations for changes in orders and highlight that the popularity of a dish can be self-enforcing. They attribute the changes in orders to observational learning: customers process the popularity information and extract quality information contained therein before choosing a meal. Muchnik et al. (2013) show that prior ratings on a social news aggregation website biased future rating behavior. In a study of the influence of movie ranks on the number of paid movie leases, Matos et al. (2013) showed that popularity may be self-enforcing: Randomly manipulating the position of a movie in a ranking influences its demand in the short run.

Popularity effects may be moderated by customer or product characteristics, and may not endure forever. In the restaurant setting, Cai et al. (2009) found that sporadic visitors are particularly susceptible to the social influence exerted by popularity information. Further, in an online field experiment based on click-stream data, Tucker and Zhang (2011) demonstrated that popularity benefits niche products, causing more additional clicks at products with narrow appeal. Muchnik et al. (2013) show that rating biases depend on the topic and whether one views the opinions of "friends or enemies". Finally, Matos et al. (2013) find that manipulating a product's rank has a short-term influence, on the long run, however, movies "tend to move back to their true slot over time."

Our study is similar to previous works in that it relates user-generated content to economic outcomes. Like the experimental works mentioned above, we also focus on popularity information and try to identify causal relationships between this information and consumer choices. However, our study is different in several aspects. Previous experiments considered free goods or did not directly model sales. The works that did model sales could not identify causal relationships between sales and user-generated content. Hence, the monetary impact of showing popularity information is unclear. To the best of our knowledge, we present the first randomized controlled trial that examines the impact of user-generated content on real sales data. Furthermore, some previous studies used popularity information to rank items and study the influence of rank. In contrast, we directly examine how the number of Likes influences sales and do not examine the influence of rankings based on Likes. Unlike Muchnik et al. (2013) we do not explain the number of Likes and do not decompose it into several components. We regard the number of Likes as given and study the influence of displaying these data on customer behavior.

While previous work has uncovered the moderating role of consumer and product characteristics, the moderating role of the time of purchase has not yet been investigated. Considering the time of purchase as a contextual factor may be important because consumers shop for different reasons (Tauber, 1972). The main distinction is between "hedonic and utilitarian shopping" (Babin et al., 1994), that is, between goal-directed, well-planned purchases (utilitarian) and purchases made more for personal pleasure (hedonic). Goal-directed purchases are made "fast and easy," so that a customer can "get exactly what [she wants], in the least amount of time," but consumers shopping for "Inspiration" want to "learn about new products" or perhaps "discover products that are new" to them (Wagner and Rudolph, 2010). With well-planned purchases, the decision is already stable, so the consumer may not need to seek the opinions of others. However, looking for a new, interesting product that matches one's taste can take much longer; and the consumer might find it useful to consult other people's previous choices. On top of this, consumer time budgets are typically distributed unequally over a day and a week (more spare time for extensive online shopping trips is usually available in the evening hours or during the weekend). Hence, the time of purchase may constitute an important contextual factor that could moderate the strength of the influence of user-generated content.

3 Field Experiment

We describe the experimental setup (Section 3.1), compare treatment and control groups (Section 3.1), and provide descriptive statistics for the data collected during the experiment (Section 3.3).

3.1 Setup

We conducted the experiment in cooperation with a German online store selling toys and board games. The online store was founded in 2002 and is an established player in the market. It provides a product information page for every product. The product pages are similar to those of Amazon.com and display price, shipping information, some photos of the product, product information, and user-generated content. In addition to the online store, the company has been hosting a large ($n > 6,500$) and active online social network (OSN) since 2010. The OSN is similar to Facebook or LinkedIn in terms of functionality: A user has a profile page that she can fill with personal information (including a picture), and she can connect with friends, post and share content, or engage in a discussion. Membership in the OSN is not tied to being a customer at the store. Users of the OSN can "Like" products that are sold through the online store.

The user-generated content shown on a product page includes customer reviews, ratings, and a box displaying the number of people who "Like" the product. The Likes display is similar to that of

Figure 1: Screenshot of Treatment Condition (left) and Control Condition (right)

Facebook's Like plug-in. In addition to the number of Likes, the box also displays miniature profile pictures of randomly selected users who like the product.

In our experiment, we used first-time visitors as subjects. We focus on first-time visitors for two reasons: First, this limits the potential financial loss for the online store. The policy of the online store was to show all user-generated content (including the number of Likes) to all customers. Performing an experiment like this, where some user-generated content is not shown to all customers, may lead to reduced sales for the store. Second, using first-time customers eliminates problems with returning customers, who might be confused if some user-generated content is missing on a return visit to the online store. This might lead to confounding effects like the Hawthorne effect (see, e.g. Parsons, 1974) as well as other unwanted effects.

Like most commercial websites, the online store sets cookies with a unique customer ID to identify web site visitors. We assume a customer to be a first-time visitor if the web server does not detect an existing cookie. If there was a cookie from a previous visit or if cookies were disabled in the browser, the corresponding visitor did not participate in the experiment.

All first-time customers were randomly assigned with a 50:50 chance to either the treatment or control group. On each product page, the members of the treatment group could see the number of people "who like the product" and a random selection of miniature profile pictures of people who liked the product. The members of the control group did not see this user-generated content. All other information remained unchanged. Figure 1 depicts screenshots of the treatment and control conditions.

We conducted the experiment for about three weeks (23 days) in February 2013, tracking the browsing and purchasing behavior of the subjects. When a subject visited a product page, we stored the corresponding product information and also whether the customer purchased the product.

3.2 Similarity of Treatment and Control Group

The treatment and the control group need to be structurally similar in order to make causal claims about the influence of the treatment. To check for similarity, we first compared the number of treated and non-treated subjects. Overall, $n = 72,849$ first-time visitors were assigned to either our treatment or control group. Out of these, 36,454 (50.04%) subjects received the treatment and 36,395 (49.96%) did not. This allocation does not deviate significantly from an "optimal" allocation that assigns 50% of the subjects to each group ($p = .8270$).

In the next step, we tested whether subjects were similar across observable characteristics and thereby checked whether the experimental split was successful. This task is challenging in an online context: Visitors who do not buy a product leave little personal information and most of their observable behavior may be influenced by the treatment. To properly address this challenge, we compared several exogenous characteristics that were observable at the beginning of and during the experiment. We then compared shopping behavior after the experiment ended.

Table 2: Comparison of treatment and control group

Characteristics of first product	Group	Min.	Mean	Max	Std. Err.	t	p
Product price in Euros	Control	.79	25.86	449	.129	-.6252	.5318
	Treatment	.79	25.97	479.99	.130		
Number of Likes	Control	0	30.83	480	.306	-.0421	.9664
	Treatment	0	30.85	480	.309		
Day	Control	1	11.70	23	.034	-13.924	.1638
	Treatment	1	11.77	23	.034		
Technology choice: Browser	**Group**	**Min.**	**Mean**	**Max.**	**Std. Err.**	z	p
IE/Mozilla-compatible (0,1)	Control	0	.728	1	.002	-1485	.1375
	Treatment	0	.733	1	.002		
Mozilla4-compatible (0,1)	Control	0	.125	1	.002	11169	.2640
	Treatment	0	.122	1	.002		
Opera 9.8 (0,1)	Control	0	.021	1	.00076	.7282	.4665
	Treatment	0	.020	1	.00074		
Technology choice: Platform	**Group**	**Min.**	**Mean**	**Max.**	**Std. Err.**	z	p
Windows (0,1)	Control	0	.442	1	.002	-0.2976	.7660
	Treatment	0	.443	1	.002		
IPad (0,1)	Control	0	.046	1	.001	-0.3203	.7487
	Treatment	0	.047	1	.001		
Linux (0,1)	Control	0	.080	1	.001	.9978	.3184
	Treatment	0	.078	1	.001		
Shopping behavior after experiment	**Group**	**Min.**	**Mean**	**Max.**	**Std. Err.**	z, t	p
ConversionPostExp (0,1)	Control	0	.021	1	.011	-.6411	.5215
	Treatment	0	.032	1	.012		
OrdersPostExp	Control	0	.090	9	.055	-.5895	.5559
	Treatment	0	.139	7	.060		
SalesPostExp	Control	0	8.38	1,088.73	6.07	.1888	.8504
	Treatment	0	7.06	438.32	3.46		
Socio-demographics	**Group**	**Min.**	**Mean**	**Max.**	**Std. Err.**	z	p
Gender (0,1)	Control	0	.609	1	.035	-.8573	.3913
	Treatment	0	.652	1	.035		

The first product inspected by a customer is exogenous to the treatment. This is because treated customers see the number of Likes only on the product detail page and not before they have clicked on the product. This allowed us to compare the characteristics of the first inspected product: Specifically, we examined the price and product popularity (number of Likes) as well as on which day of the experiment the first product view took place (variable *Day* in Table 2 is 1 on the first day of the experiment). All subsequent product views and purchase decisions are endogenous to the treatment.

Further, the subjects' choices of Internet technology are exogenous as they were made before the beginning of the experiment. The online store logs which web browser and operating system a subject uses. There are many browsers and operating systems in our sample. We present a comparison of three popular browsers and operating systems. The variables are 1 if a user uses this technology, and it is 0 otherwise. The results for the remaining browsers and OSes are qualitatively similar.

Finally, we studied the shopping behavior after the end of the experiment. We conducted the experiment in February 2013. Since then the store has operated in "normal mode", showing the Likes to all visitors. If the treatment and control groups are similar, their shopping behavior should converge after the experiment—that is, the experiment should not influence what people buy several months later. To this end, from the set of 791 customers having bought at least one product during the 23 days of the experiment we randomly selected 187 non-treated and 187 treated customers. All selected customers visited the online shop again at least once in the time after the experiment (between March and August 2013). We compared the conversion rates of the returning customers between March and August 2013 (the variable *ConversionPostExp (0,1)* is 1 if a customer has placed at least one order and 0 otherwise), the overall number of orders per returning customer between March and Angust 2013 (variable *OrdersPostExp*) and the overall Sales per returning customer between March and August 2013 in Euros (variable *SalesPostExp*). We also received the first names of these returning customers, which allowed use to deduce their gender and compare the ratio of males to females in the two groups (variable *Gender (0,1)* is 1 if a returning customer is male and 0 otherwise.)

Table 2 contains the descriptive statistics and a comparison between the treatment and control groups. We compare continuous characteristics with t-tests. Characteristics like technology choice or gender are modeled as indicator variables and we compare the groups with Z-tests for proportions. The results show that the treatment group is similar to the control group across observable characteristics and we hence conclude that the two groups are practically identical.

3.3 Description of the Dataset

During the 23 days of the experiment, 72,849 first-time visitors viewed 112,362 product detail pages, and 791 of the visitors placed at least one order and became customers. In total, those 791 customers contributed 26,173.20 EUR of sales. Table 3 describes the dataset in more detail: It shows descriptive statistics for the outcomes of our statistical models and describes the purchases made by the treatment and the control groups. As opposed to the terms "subject" and "visitor," which we use interchangeably for all participants in the experiment, we refer to subjects who have bought at least one product as "customers". Table 3 also shows the distribution of the number of Likes and product Prices across the 9,559 products in our dataset. The number of Likes range from 0 to 477 with a mean value of about 7, while the price range is between .79 EUR and 480 EUR with a standard deviation of about 20.

4 Results

We analyze the influence of the treatment on sales per visitor in Euros (Section 4.1), disentangle herd behavior from how the number of Likes measures popularity (Section 4.2), and examine the influence of the treatment on the distribution of sales (Section 4.3).

Table 3: Descriptives statistics for outcomes and covariates

Outcome	Group	Min.	Mean	Max.	S.D.
Page requests	Control	1	1.522	65	2.070
	Treatment	1	1.563	145	2.405
Conversion rate	Control	0	.0104	1	.101
	Treatment	0	.0113	1	.106
Sales per visitor in Euros	Control	0	.360	254	4.30
	Treatment	0	.402	170	4.40
Sales per order in Euros	Control	1.41	32.51	254	24.99
	Treatment	1.41	33.62	170	22.26
Average price of ordered items in Euros	Control	1.09	21.09	94.90	15.71
	Treatment	.99	21.76	110	15.50
Number of items per order	Control	1	2.074	19	2.028
	Treatment	1	2.015	30	2.038
Covariate		Min.	Mean	Max.	S.D.
Likes		0	6.891	477	20.94
Price in Euros		.790	18.46	480	20.02

4.1 Impact of Popularity Information on Sales

The treatment group appears to have shopped differently than the control group (cf. descriptive statistics in Table 3). The treated visitors were more likely to purchase (higher conversion rate), they spent more money (higher sales per visitor), and they looked at more products (higher number of page requests). Total sales generated by the treatment group (13,883.74 EUR) were +12.97% higher than sales generated by the control group (12,289.46 EUR). This is a substantial increase and reveals the economic importance of user-generated content for online retailers.

As discussed in Section 2, besides user-generated content there are other variables that may have an influence on purchase decisions. We want to distinguish between goal-oriented buyers and impulse shoppers with less planned behavior (Tauber, 1972). It seems likely that the latter type of consumer is more susceptible to popularity information than goal-oriented ones who know exactly what they want when, for example, working off a shopping list. However, in an online context, we usually do not have much information on the visitors' or buyers' traits that would allow a segmentation or clustering of the visitors in goal-oriented and impulse-oriented shoppers. Therefore, if we want to address the problem of large variance, heterogeneity and noise, we have to focus on exogenous factors that are likely to impact the visitors' behavior. For this purpose, we use the time of visit as exogenous factor which allows us to distinguish between hedonic and utilitarian shopping. We assume that on average, visitors during regular working hours are more goal-oriented and have less time to browse and to engage in time-intensive shopping behavior. During regular working hours, subjects more likely visit the shop, complete their shopping list (e.g., buy a present for a friend that they need in the next few days), leave the shop, and continue with their work. The situation is different in the evening hours and during the weekends. Then, purchases are more frequently made for personal pleasure and subjects are more engaged in "hedonic shopping". In the evening or during weekends, visitors have on average more control over their time and are more free to spend time in an online shop making them more susceptible to popularity information.

We operationalize the time of visit as follows: We define "leisure hours" as the time between 2pm

Table 4: Effect of Treatment on page requests per visitor, Sales per Visitor in Euros, conversion Rate, and Sales per order in Euros

Page requests per visitor		Min.	Mean	Max.	SD	Diff.	SE	t	p
Working hours		1	1.537	145	2.35	-0.0526	.0189	-2.7727	0.0056
Leisure hours		1	1.589	81	2.29				

Sales per visitor in Euros		Min.	Mean	Max.	SD	Diff.	SE	t	p
Leisure hours	Control	0	0.35	254.0	4.14	0.092	0.041	-2.25	0.025
	Treatment	0	0.44	170.0	4.66				
Working hours	Control	0	0.38	217.9	4.63	-0.064	0.057	1.12	0.26
	Treatment	0	0.32	94.9	3.77				

Conversion rate		Min.	Mean	Max.	SD	Diff.	SE	z	p
Leisure hours	Control	0	0.011	1	0.10	0.0022	0.0010	-2.15	0.031
	Treatment	0	0.013	1	0.11				
Working hours	Control	0	0.011	1	0.11	-0.0018	0.0014	1.33	0.18
	Treatment	0	0.0096	1	0.098				

Sales per order in Euros		Min.	Mean	Max.	SD	Diff.	SE	t	p
Leisure hours	Control	1.55	32.0	254.0	23.6	1.69	1.99	-0.85	0.40
	Treatment	1.41	33.7	170.0	23.2				
Working hours	Control	1.41	33.5	217.9	27.6	-0.26	3.24	0.081	0.94
	Treatment	3.49	33.2	94.9	19.8				

and midnight on workdays, plus Saturdays and Sundays. The rationale behind this is that students constitute a substantial portion of the shop's customers and school typically ends around 1pm in Germany. Furthermore, many part-time employees—which represent a large part of the working population in Germany (about 45.3% of Germany's female employees work part-time; Eurostat, 2008)—typically work in the morning hours and finigh work around lunch-time. Consequently, we define "working hours" as the remaining time, which are weekdays between midnight and 2pm. In the Appendix, we provide robustness checks where we systematically vary the time partitioning).

If visitors are more susceptible to popularity information on hedonic shopping trips and hedonic shopping is more prevalent during leisure hours, we would expect that the number of page requests is higher during leisure hours than working hours. Table 4 shows the results from comparing the shopping behavior with t-tests and z-tests. We find support for this hypothesis as the number of page views during leisure hours is significantly higher than during working hours ($p < .01$). Furthermore, the influence of the treatment on sales in the leisure hours is significant, while its influence during working hours is negligible. During leisure hours, the display of Likes increases the sales per visitor by +0.09 EUR (+25.7%, $p < .05$) and the conversion rate from 1.1% to 1.3% (+18.2%, $p < .05$). The sales per order were not influenced by the user-generated content ($p > .1$).[1]

Thus, popularity information can sigficantly increase the conversion rate and the average basket value, but only when visitors are susceptible to the influence exerted by popularity information. The time of the day is a strong exogenous moderating factor: user-generated content seems to unfold its influence only during the leisure time of the day and on the weekends. The visitors have more time

[1] Results obtained ignoring the moderating effect of time are in the Appendix.

available during leisure time, which induces more browsing behavior and ultimately more sales caused by the user-generated content.

4.2 Disentangling the Measurement Capability and the Herding Effect

The previous section studied the influence of the treatment on the level of subjects. This section focuses on the problem of endogeneity: we disentangle measurement capability (inherent product quality) and herding effect and discuss ways to estimate the value of a single Like. We model on the level of products because Likes are properties of products, not of visitors.

The dependent variable is sales per product in Euros. Table 5 summarizes the distribution of this outcome and of all covariates used in the statistical models. The sales per product varied between 0 and 389.90 EUR in the control group and 0 and 215.90 EUR in the treatment group. Average sales were higher in the treatment group (1.20 EUR compared to 1.08 EUR). The average price of the products (Price) was 20.10 EUR and the average product had about 11 Likes (Number of Likes). Furthermore, we report details on the product ratings provided by the users. All registered shop users (including members of the OSN) can rate a product on a scale between 0 and 6 (6 is best). On average, a product obtained 3.7 ratings (Volume of ratings), the average rating of a game is 4.81 (Valence of ratings), and the average variance of ratings for a product is 0.47 (Variance of ratings).

Table 5: Descriptive Statistics on Outcome Variable and Covariates

Outcome	Group	Min.	Mean	Max.	SD
Sales per product in Euros	Control	0	1.08	389.9	9.24
	Treatment	0	1.20	215.9	7.56
Covariate		Min.	Mean	Max.	SD
Price		0.79	20.1	449	17.8
Number of Likes		0	10.9	395.2	25.4
Volume of ratings		0	3.7	203	8.45
Valence of ratings		0	4.81	6	0.98
Variance of ratings		0	0.47	6.25	0.71

In our experiment, we randomized the assignment of subjects to treatment and control groups. However, we did not randomize the number of Likes. Hence, modeling on the level of products requires that we control for product-specific heterogeneity. This is supported by a Breusch-Pagan Lagrange Multiplier test which finds significant differences across products ($p < .001$). As a consequence, we use panel regressions. To decide between a fixed-effects and a random-effects specification we conducted a Hausman test. The results supported the hypothesis that a random effects model is preferable ($p > .99$).

We considered all products that were visited by members of both treatment and control group. We excluded all products, where the number of Likes was not yet stable and strongly changed during the experiment. Taking such products into account would heavily bias the results if the number of Likes are covariates. In particular, we excluded all products, where the standard deviation of the number of Likes that were shown to the users who visited the product page is equal or greater than two. After excluding these games, our final subsample contains 99.6% of the products (n=4,958 products) in a quasi-experimental setup.

We start with an illustrative case that does not exploit the experimental variation and only uses observational data. We regress the number of Likes per product on its sales:

$$Sales_i = \alpha_0 + \alpha_1 \cdot Likes_i + R + \epsilon_i, \tag{1}$$

where $Sales_i$ are the sales of product i in Euros, $Likes_i$ is the number of Likes of product i, and R is the random effect. The coefficient α_1 captures the contribution of an additional Like for product i to the sales of product i in Euros. Table 6 (Model 1 Observational) shows the estimation results. An additional Like is associated with an additional +0.03 EUR of sales ($p < .01$). It is important to note that this (observational) model cannot disentangle the measurement capability and the herding effect. It is possible that high quality products receive more Likes and higher sales of a product are due to its higher quality and not due to the visibility of the number of Likes.

Consequently, we use our experimental variation to model the visibility of the Likes. This decomposes $Sales_i$ into sales that were made when the Likes were *not shown*, and additional sales that were caused by the *visibility* of the Likes. While the former constitutes the measurement capability, the latter models the herding sales caused be the number of Likes:

$$Sales_i = \alpha_0 + \alpha_1 \cdot Likes_i + \alpha_2 \cdot Treatment_i + \alpha_3 \cdot Treatment_i \cdot Likes_i + R + \epsilon_i. \tag{2}$$

Model (2) is the base model. It extends (1) by adding the variable $Treatment_i$, which is 1 if the Likes of product i are visible and 0 otherwise, and an interaction term $Treatment_i \cdot Likes_i$. The coefficient α_3 captures the additional sales per product in Euros that are caused by increasing the visible number of Likes by one. This is the herding effect. Different from model (1), now the coefficient α_1 measures popularity effects that are not absorbed by the interaction term. Thus, α_1 mainly captures the effect of product quality. Finally, we introduce the coefficient α_2, which captures the monetary effect of the treatment, i.e. the effect of showing the box that contains the number of Likes. We include this term because observing the number of Likes of a product can increase the salience of the product, independently of the actual number of Likes. The coefficient α_2 captures the monetary value of this salience effect.

Model estimates are presented in Table 6 (Model 2 Base model). Because the coefficient α_3 is significantly different from zero we observe herding sales: increasing the number of Likes by one causes additional +0.015 EUR of sales ($p < .01$) per product. On the contrary, coefficient α_1 captures how the number of Likes measure popularity. Here, increasing the number of Likes by one is associated with an additional +0.026 EUR of sales ($p < .01$). We compare α_1 to α_3 and conclude that the measurement capability of Likes is dominant $\left(63.7\% \approx \frac{\alpha_1}{\alpha_1+\alpha_3} = \frac{.0258}{.0258+.0147}\right)$, which means that sales are more a result of the measurement capability than herding. Comparing α_1 from model (1) to α_3 from model (2) reveals that models based on purely observational data (like model 1) can easily overestimate the causal, economic impact of Likes (by a factor of $2.26 \approx \frac{.0332}{.0147}$ in our data). We do not observe a significant salience effect ($p > .1$ for α_2).

In the next two models we control for heterogeneity in product prices and ratings. We first integrate product prices in our model which leads to Model (3)

$$\begin{aligned} Sales_i = &\alpha_0 + \alpha_1 \cdot Likes_i + \alpha_2 \cdot Treatment_i + \alpha_3 \cdot Treatment_i \cdot Likes_i \\ &+ \alpha_4 \cdot Price_i + R + \epsilon_i, \end{aligned} \tag{3}$$

where $Price_i$ is the price of product i in Euros. The estimates are shown in Table 6 (Model 3 +Price). Finally, we control for product ratings as ratings are often claimed to absorb the effects of product quality and obtain Model (4)

$$\begin{aligned}Sales_i = &\alpha_0 + \alpha_1 \cdot Likes_i + \alpha_2 \cdot Treatment_i + \alpha_3 \cdot Treatment_i \cdot Likes_i + \alpha_4 \cdot Price_i \\ &+ \alpha_5 \cdot Volume_i + \alpha_6 \cdot Valence_i + \alpha_7 \cdot Variance_i + R + \epsilon_i, \end{aligned} \tag{4}$$

where $Volume_i$ is the number of ratings for product i, $Valence_i$ is the average rating of product i, and $Variance_i$ is the variance of product ratings for product i. Table 6 (Model 4 +Price&Ratings) lists the estimates. For model (4), we only considered products where at least one rating was available reducing the size of our subsample to 2,922. We find that controlling for product prices and ratings does not substantially change the main results obtained from our base model (2). The coefficients α_1 and α_3 are significantly different from zero, measurement effects are stronger than herding effects, and we do not observe significant salience effects (α_2 is not significantly different from zero, $p > .1$).

Table 6: Impact of Likes on Sales per Product in Euro

	(1) Observational	(2) Base model	(3) +Price	(4) +Price&Ratings
Likes (α_1)	0.0332***	0.0258***	0.0231***	0.0409***
	(0.00593)	(0.00592)	(0.00575)	(0.0111)
Treatment (α_2)		-0.0472	-0.0471	-0.129
		(0.114)	(0.114)	(0.183)
Treatment $\cdot$ *Likes* (α_3)		0.0147**	0.0147**	0.0161**
		(0.00749)	(0.00749)	(0.00782)
Price (α_4)			0.0387***	0.0516**
			(0.0122)	(0.0219)
Volume of Ratings (α_5)				-0.0929***
				(0.0260)
Valence of Ratings (α_6)				0.350**
				(0.140)
Variance of Ratings (α_7)				0.122
				(0.161)
Constant (α_0)	0.777***	0.801***	0.0551	-1.541**
	(0.0936)	(0.114)	(0.202)	(0.722)
Model statistics				
R^2 within	0.00000706	0.00220	0.00222	0.00293
R^2 between	0.0132	0.0132	0.0219	0.0291
R^2 overall	0.00998	0.0105	0.0171	0.0231
df	1	3	4	7
Wald $\chi^2(df)$	31.33	32.42	41.28	42.05
Prob$> \chi^2$	$< .0001$	$< .0001$	$< .0001$	$< .0001$
# Products (n)	4,958	4,958	4,958	2,922

Significance levels: *:$p < .1$, **:$p < .05$, ***:$p < .01$

There are several explanations for our findings. First, the theory of observational learning (Bandura, 1977) suggests that consumers extract quality signals from the number of Likes before they decide what to buy. Thus, increasing the number of Likes causes additional sales of this product. Second, observing the number of Likes of a product can increases the product's salience, which in turn leads to a purchase.

However, in our data we do not observe salience effects, which makes this explanation less likely. Third, there could be a desire for conformity where visitors try to conform with social norms. Although we can not completely rule out conformity reasons, we believe conformity plays only a minor role. First, board games are no status symbols; second, usually one instance of a board game is sufficient to play the game within a group. Thus, as soon as one member of a gaming group owns a particular game, there is no need for the other group members to also buy this game.

4.3 Impact of Popularity Information on the Distribution of Sales

Finally, we examine the effect of the treatment on product choice and the distribution of sales. This is important because profit margins can differ between blockbusters and niche products: often, retailers use blockbusters as loss leaders, whereas long-tail products offer more attractive profit margins (Elberse, 2008).

First, we tested if treated customers are more likely to buy highly recommended ("Liked") products. For this purpose, we stored all products that a subject i inspected during her visit in her consideration set $con(i)$, flagging those products in $con(i)$ that were purchased. For each product k purchased by a customer, we calculate the measure $Ratio(k)$, which is is the percentage of products from the consideration set $con(i)$ that are less or equally popular (have the same or a lower number of "Likes") than the purchased product k. $Ratio(k)$ is equal to zero if all products in the consideration set of customer i are more popular than product k; it is equal to one, if product k is the most popular product in the consideration set of customer i. We wanted to know if the treatment influences *Ratio*. Table 7 shows that the treatment has a significant influence on *Ratio* ($p < .1$).[2] Treated customers were more likely to buy products with a lower number of Likes from their consideration sets than untreated subjects. Thus, showing the number of Likes seems to motivate customers to choose products with a lower number of Likes.

Table 7: Influence of Treatment on Ratio

Ratio per customer	Min.	Mean	Max.	SD	Diff.	SE	t	p
Control	.0588	.7342	1	.2623	.0248	.0145	1.7127	0.087
Treatment	.0625	.7095	1	.2790				

To study whether the treatment changes the ratio between blockbusters and niche products, we measured the inequality among the values of the sales distribution using the Gini coefficient, which is two times the area between the line of equality and the Lorenz curve. The Lorenz curve plots the proportion of the cumulated total sales over $x\%$ of the customers. The Gini coefficient ranges from 0% to 100%; if the Gini coefficient is zero, the Lorenz curve is identical to the line of equality and each product is purchased by the same number of customers. With a higher Gini coefficient, the sales distribution becomes more unequal; if the Gini coefficient is one, customers are purchasing only one product. The Gini coefficient should be interpreted with care because an infinite number of Lorenz curves yields the same Gini coefficients (Hinz et al., 2011).

Unfortunately, dividing the rather small number of purchases against the many products in the assortment produces only descriptive results. We observe that displaying popularity information leads to an increase of the Gini coefficient from 24.54% to 27.39% ($p > .2$). Thus, the distribution of sales becomes more unequal by showing popularity information. Studying this effect more closely, we consider

[2] The results are qualitatively similar if we control for the size of the consideration sets.

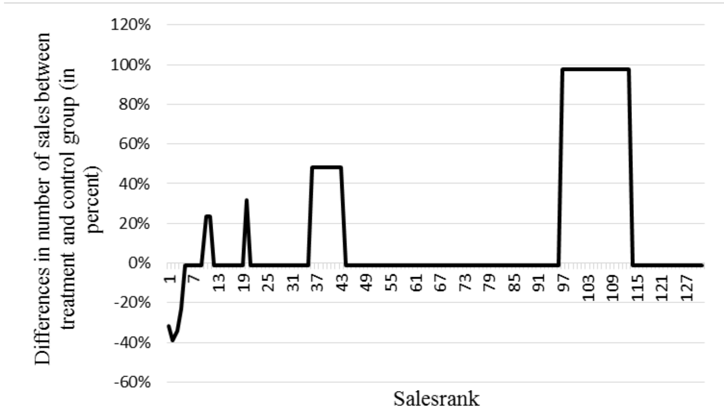

Figure 2: Influence of the Treatment of the Distribution of Sales

all products sold during the 23 days of the experiment and order them by sales rank. For each sales rank, Figure 2 plots the difference in the number of sales (in percent) between the treatment and control group. Displaying popularity information decreases the number of sold blockbusters up to 40%. Furthermore, products that are neither blockbusters nor niche products can benefit from showing popularity information. This indicates that the influence of user-generated content may be even more nuanced than previously thought. We believe that in the absence of Likes, top lists have a strong influence on consumer behavior and the display of top lists shifts demand towards blockbusters. If additional popularity information is available, then the influence of top lists is mitigated. Good quality products that match personal tastes (and are hence in the middle of the sales distribution curve) can experience additional demand. Future research could test this hypothesis.

5 Discussion and Implications

Understanding how user-generated content influences consumer behavior is vitally important to the electronic commerce industry. We conducted a randomized field experiment in cooperation with an online store. Covering a period of 23 days and considering only prospective customers, we found that displaying popularity information in the form of Likes caused a +12.97% sales increase.

Thus, our results suggest that online stores should display popularity information. However, we find that this information is effective only when visitors' time restrictions are low as the effect of showing Likes was insignificant during working hours. In contrast, during leisure hours visible Likes significantly increased the conversion rate (+18.2%) and the average shopping basket (+25.7%). Thus, user-generated content can have a strong impact on sales. A descriptive analysis revealed that presenting Likes may even change the distribution of sales by decreasing demand for blockbusters and increasing demand for products in the middle of the sales distribution.

Global ecommerce sales have reached $1 trillion (eMarketer, 2013). In our experiment, visible Likes caused almost 13% additional sales. It is difficult to assess the global impact of user-generated content on sales, but the magnitude of these numbers suggests that herding effects generated by Likes and other user-generated content are also of macroeconomic relevance.

Our study also makes contributions to theory. The results suggest that not all customers are equally susceptible to the influence exerted by user-generated content. Previous research in the area of user-generated content and decision aids did not differentiate between goal-directed, well-planned purchases (utilitarian shopping) and purchases made for personal pleasure (hedonic shopping), where visitors more often browse among products that they did not know about before. On this point, future research could study the moderating effects of the shopping motive on online shopping behavior in more detail. It would be interesting to know how different motives are to be integrated in models of consumer search (Zhang et al., 2007) and whether they should be matched with different online store designs or functionalities, which might serve customers better than a standardized version.

We have conducted an experiment to disentangle the herding effect of a Like from measurement effects. From our findings, observational studies should be careful when interpreting their estimates. It might be difficult to quantify the causal influence of user-generated content correctly without experimental manipulation. Researchers and practitioners are often more interested in the causal value of user-generated content than its measurement capability. However, our results suggest that measurement effects are much stronger than herding effects. Identifying herding effects correctly might be a challenge, even after controlling for important covariates.

Of course, our study is not without limitations which might provide avenues for future research. Maybe the most important limitation is that we study the effect of popularity information on prospective (new) customers. The magnitude of the effects might be different for existing customers. Furthermore, the magnitudes of the effects might differ across industries and online stores. Future research is needed that replicates our study in different contexts. Additionally, the design of the product detail page might be an important influence factor as well. In our case, the online store displayed the information with miniature profile pictures of randomly selected users who "Like" the product. Other designs might change the magnitude of influence. For example, Forman et al. (2008) showed that reviewer identity disclosure plays an important role in electronic markets and it would thus be very interesting to design an optimal recommendation format.

References

Anderson, L. R., C. A. Holt. 1997. Information Cascades in the Laboratory. *American Economic Review* **87**(5) 847.862.

Aral, S., D. Walker. 2012. Identifying Influential and Susceptible Members of Social Networks. *Science* **337**(6092) 337–341.

Babin, B. J., W. R. Darden, M. Griffin. 1994. Work and/or Fun: Measuring Hedonic and Utilitarian Shopping Value. *Journal of Consumer Research* **20**(4) 644–656.

Bandura, A. 1977. *Social Learning Theory.* General Learning Press, New York.

Banerjee, A. V. 1992. A Simple Model of Herd Behavior. *The Quarterly Journal of Economics* **107**(3) 797–817.

Bapna, R., A. Umyarov. 2013. Do Your Online Friends Make You Pay? A Randomized Field Experiment in an Online Music Social Network: Working Paper.

Bikhchandani, S., D. Hirshleifer, I. Welch. 1992. A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades. *Journal of Political Economy* **100**(5) 992–1026.

Cai, H., Y. Chen, H. Fang. 2009. Observational Learning: Evidence from a Randomized Natural Field Experiment. *American Economic Review* **99**(3) 864–882.

Chevalier, J. A., D. Mayzlin. 2006. The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal of Marketing Research* **43**(3) 345–354.

Chintagunta, P. K., S. Gopinath, S. Venkataraman. 2010. The Effects of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation Across Local Markets. *Marketing Science* **29**(5) 944–957.

Dellarocas, C., X. Zhang, N. F. Awad. 2007. Exploring the Value of Online Product Reviews in Forecasting Sales: The Case of Motion Pictures. *Journal of Interactive Marketing* **21**(4) 23–45.

Dhar, V., E. A. Chang. 2009. Does Chatter Matter? The Impact of User-Generated Content on Music Sales. *Journal of Interactive Marketing* **23**(4) 300–307.

Elberse, A. 2008. Should You Invest in the Long Tail? *Harvard Business Review* **86**(7/8) 88–96.

eMarketer. 2013. B2C Ecommerce Climbs Worldwide, as Emerging Markets Drive Sales Higher. URL `http://www.emarketer.com/Article/B2C-Ecommerce-Climbs-Worldwide-Emerging-Markets-Drive-Sales-Higher/1010004`.

Eurostat. 2008. Labour Force Survey: Reduced Gap in Employment Rates Between Men and Women in the EU27 - Share of Part-Time Employment for Women Four Times Higher than for Men.

Forman, C., A. Ghose, B. 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.

Godes, D., D. Mayzlin. 2004. Using Online Conversations to Study Word-of-Mouth Communication. *Marketing Science* **23**(4) 545–560.

Hinz, O., J. Eckert, B. Skiera. 2011. Drivers of the Long Tail Phenomenon: An Empirical Analysis. *Journal of Management Information Systems* **27**(4) 43–70.

Liu, Y. 2006. Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue. *Journal of Marketing Research* **70**(3) 74–89.

Matos, M. G. de, P. Ferreira, R. Telang, M. Smith. 2013. The Impact of Popularity on the Sales of Movies in Video-on-Demand: A Randomized Experiment: Working Paper.

McCracken, H. 2011. 50 Websites that make the Web great. URL `\url{http://www.time.com/time/specials/packages/article/0,28804,2087815_2088159_2088155,00.html}`.

Moe, W. W., M. Trusov. 2011. The Value of Social Dynamics in Online Product Rating Forums. *Journal of Marketing Research* **48**(3) 444–456.

Muchnik, L., S. Aral, S. J. Taylor. 2013. Social Influence Bias: A Randomized Experiment. *Science* **341**(6146) 647–651.

Parsons, H. M. 1974. What Happened at Hawthorne?: New Evidence Suggests the Hawthorne Effect Resulted from Operant Reinforcement contingencies. *Science* **183**(4128) 922–932.

Salganik, M. J., P. S. Dodds, D. J. Watts. 2006. Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market. *Science* **311**(5762) 854–856.

Shadish, W. R., T. D. Cook, D. T. Campbell. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Houghton Mifflin, Boston.

Tauber, E. M. 1972. Why Do People Shop? *The Journal of Marketing* **36**(4) 46–49.

The Nielsen Company, ed. 2007. *Through the Grapevine*. September.

Tirunillai, S., G. J. Tellis. 2012. Does Chatter Really Matter? Dynamics of User-Generated Content and Stock Market Performance. *Marketing Science* **31**(2) 198–215.

Trusov, M., R. E. Bucklin, K. Pauwels. 2009. Effects of Word-of-Mouth Versus Traditional Marketing: Findings from an Internet Social Networking Site. *Journal of Marketing* **73**(5) 90–102.

Tucker, C., J. Zhang. 2011. How Does Popularity Information Affect Choices? A Field Experiment. *Management Science* **57**(5) 828–842.

Wagner, T., T. Rudolph. 2010. Towards a Hierarchical Theory of Shopping Motivation. *Journal of Retailing and Consumer Services* **17**(5) 415–429.

Web Technology Surveys. 2013. URL `\url{http://w3techs.com/technologies/overview/social_widget/all.}`

Zhang, J., X. Fang, O. Liu Sheng. 2007. Online Consumer Search Depth: Theories and New Findings. *Journal of Management Information Systems* **23**(3) 71–95.

Zhu, F., X. Zhang. 2010. Impact of Online Consumer Reviews on Sales:The Moderating Role of Product and Consumer Characteristics. *Journal of Marketing* **74**(2) 133–148.