Culling the herding:
Using real world randomized experiments to measure social bias with known costly goods

Miguel Godinho de Matos[1,3*‡],Pedro Ferreira[1,2‡],Michael Smith[1],Rahul Telang[1]

[1] Heinz College, Carnegie Mellon University
[2] Department of Engineering and Public Policy, Carnegie Mellon University
[3] Católica Lisbon School of Business and Economics, Universidade Católica Portuguesa

‡Correspondence should be addressed to miguelgodinhomatos@cmu.edu and pedrof@cmu.edu
*This work is part of Godinho de Matos' PhD dissertation. Other authors are ordered alphabetically.

July 28, 2015

Abstract

Peer-ratings have become increasingly important sources of product information, particularly in markets for "information goods." However, in spite of the increasing prevalence of this information, there are relatively few academic studies that analyze the impact of peer-ratings on consumers transacting in "real world" marketplaces. In this paper, we partner with a major cable company to analyze the impact of peer-ratings in a real-world Video-on-Demand market where consumer participation is organic and where movies are costly and well-known to consumers. After experimentally manipulating the initial conditions of product information displayed to consumers, we find that, consistent with the prior literature, peer-ratings influence consumer behavior independently from underlying product quality. However, we also find that, in contrast to the prior literature, at least in our setting there is little evidence of long-term bias due to herding effects. Specifically, when movies are artificially promoted or demoted in peer-rating lists, subsequent reviews cause them to return to their true quality position relatively quickly. One explanation for this difference is that consumers in our empirical setting likely had more outside

1

information about the true quality of the products they were evaluating than did consumers in the studies reported in prior literature. While tentative, this explanation suggests that in real-world marketplaces where consumers have sufficient access to outside information about true product quality, peer-ratings may be more robust to herding effects and thus provide more reliable signals of true product quality, than previously thought.

Keywords. Herding Behavior, Randomized Experiment, Likes, Video-on-Demand

1   Introduction

Consumers constantly acquire information to make purchase decisions. This is particularly true in the context of "experience goods" where the quality and fit of the product are important to the consumer but difficult to evaluate *ex ante* [Nelson, 1970]. In this context, the Internet makes it easier for users to both acquire and provide product information and feedback. As a result, ratings, rankings, and review information have all become more prevalent online, particularly for "experience goods" — and the literature suggests that their increasing prevalence may result in increased market efficiency and consumer surplus if the reported product quality accurately reflects the underlying sentiment of the community [Brynjolfsson et al., 2003].

However, in spite of the increasing prevalence of peer-review systems, and their potential importance for market efficiency, there are relatively few studies that empirically analyze consumer behavior in the context of user-generated product information, and even fewer papers that analyze the impact of user-generated reviews in markets with real consumers, real products, and real monetary transactions. The scarcity of academic studies in "real-world" marketplaces is of course understandable: the simultaneity between user ratings and product quality makes it nearly impossible to make causal statements about

2

the impact of user-generated information from observational data [Aral et al., 2009], and running a randomized field experiment in a "real-world" marketplace is extremely challenging because of the difficulty in obtaining merchant cooperation to conduct experiments for academic research.

Our research seeks to contribute to the literature by running just such a real-world field experiment through a unique collaboration with a prominent telecommunications provider – hereinafter called OurTelco – who conducted an experiment using their Video on Demand (VoD) system. During this experiment a new menu was introduced in this VoD system that showed the "most popular movies during the past weeks". OurTelco randomly changed the order of movies (and number of reported "likes") displayed in this menu as a way of experimentally identifying their impact on consumption decisions. This experiment ran for 24 weeks starting in July 2012. During this time 22,043 users leased movies from this new menu at prices ranging from $1.30 to $5.20.

The experimental approach described above allows for separating the role played by the number of "likes" from the underlying (unobserved) quality of the product. Consistent with the existing literature, we find that an increase in the number of likes of a movie (that is, user-generated promotion) increases its sales independently of its underlying quality. However, in contrast to the existing literature, our results reveal little evidence of herding effects, that is, manipulated initial conditions do not lead to significantly different market outcomes in the long-run. We find that, on average, weekly movies sales increase by 4% when a movie is promoted one rank. However, across all manipulations introduced during this experiment, a movie promoted to a better rank receives, on average, 15.9% fewer sales and 33.1% fewer likes than a movie naturally placed at that rank would receive. Likewise, a movie demoted to a worse rank receives 27.7% more sales and 30.1% more likes than a movie naturally placed at that rank would receive. We further find that these differences

are moderated by how well-know the manipulated movie is (as measured by number of IMDb votes), in particular, better-known movies are less sensitive to manipulations than other movies are. For example, the effect of a rank manipulation per rank manipulated is 84% lower for more popular movies – measured as being in the top quartile of the distribution of the IMDb votes – relative to movies in the other quartiles. Thus, our results suggest that promotion through user-generated "likes" can increase the sales of content, but that, at least in our context, the wisdom of the crowds can correct both positive and negative manipulations, neutralizing the impact of potential herding effects.

One possible explanation for the difference between our results and those reported in the prior literature is that in our case consumers likely had more outside information about the products being evaluated (i.e., widely promoted movies) than did consumers in existing studies (e.g., obscure songs in [Salganik et al., 2006], wedding vendors in [Tucker and Zhang, 2011] and breaking news in [Muchnik et al., 2013]). While tentative, and perhaps context specific, this explanation raises the possibility that an increase in the availability of outside information to consumers, combined with the "wisdom of the crowds" present in user-generated reviews, may neutralize rating bias from herding effects in "real-world" marketplace settings.

2   Related Work

Our research is most closely related to the academic literature analyzing the impact of social cues on user behavior, and particularly the impact of herding behavior on peer evaluations of quality. Many papers have noted the difficulty in separating social influence from actual product quality given the natural relationship between the underlying "quality" of the product and the resulting social cues (see for example [Manski, 1993],

[Aral et al., 2009], [Aral, 2010], [Aral and Walker, 2011], [Shalizi and Thomas, 2011], and [Bapna and Umyarov, 2014]). Several papers have overcome these challenges using randomized experiments. For example, in a seminal paper [Salganik et al., 2006] created two virtual markets for songs from unknown bands and recruited a group of subjects on a teen interest website. Each subject was randomly assigned to one of these markets. Songs were ordered randomly in one of the markets and ordered according to the number of downloads in the other market. Subjects were asked to chose songs to listen to, to rate these songs, and then to download them for free if they wished. The authors found that reported popularity was self-reinforcing for all but the very best or worst songs.

In a follow-up study [Salganik and Watts, 2008] ran a similar experiment using similar songs and a similar pool of subjects. In the setup phase they asked participants to listen to the songs and rate them. Then they order songs according to these ratings so that better songs would come last and thus seem worse. In this setting, they observed that, over time, all songs (good or bad) tended to converge to their true download rank. Taken together, these studies show that self-fulfilling prophecies in these markets are constrained by the individuals' private preferences.

[Tucker and Zhang, 2011] developed a similar experiment using an online hub for wedding service vendors to explore the impact of popularity on the number of clicks that each vendor obtained. This experiment was conducted in a real world setting with organic consumers, which obviates some of the concerns with the lab experiments described in [Salganik et al., 2006] and [Salganik and Watts, 2008]. During their experiment, the authors displayed vendors in three different categories. In one category vendors were sorted in decreasing order of the number of clicks received. In another category vendors were sorted in increasing order of the number of clicks received. In both cases, vendors were listed along with the number of clicks received. In the last category vendors were

sorted alphabetically and no information was displayed regarding the number of clicks the vendor received. The authors compared vendors across different categories, before and during their experiment, to determine the impact of popularity (measured by the number of clicks received) on future clicks. They conclude that popularity is self reinforcing and that vendors that operate in narrower markets benefit the most from this dynamic.

More recently, [Muchnik et al., 2013] examine the behavior of consumers in a social news aggregator website. The authors ran a large scale organic experiment, which obviates most of the concerns with the lab experiments that [Salganik et al., 2006] and [Salganik and Watts, 2008] run into, and observed user feedback, as apposed to just clicks received as in [Tucker and Zhang, 2011]. During this experiment, users were put in three conditions: in one treatment condition users saw a positive vote on a comment; in another treatment users saw a negative vote, and the control group was not manipulated. The authors found that the probability of an "up-vote" was much higher when users initially saw a positive vote than in the control condition, which suggests a herding effect, but they found no significant effect when users saw a negative vote. They conclude that herding effects are asymmetric: users seemed to exhibit a desire to correct negative bias but otherwise follow the herd. Positive herding was also different for different topics and shaped by whether users were commenting on friends' or enemies' opinions. The authors conclude that herding increased the likelihood of positive ratings by 32% resulting in a 25% increase, on average, in ratings after 5 months, a large bias driven solely by initial conditions.

Our paper differs from these studies along several important dimensions. In our setting the goods provided are not free. Subscribers have to make explicit decisions that involve financial risks. The price to lease movies in our VoD setting varied between $1.30 and $5.20. In [Salganik et al., 2006] and [Salganik and Watts, 2008] songs could be downloaded for free, and as such subjects did not incur any financial risk in either listening to or

downloading a song. Likewise, [Tucker and Zhang, 2011] observe click through rates on websites but do not observe actual purchase decisions. In addition,consumers in our setting (i.e., widely promoted movies) may have relatively more outside information about the quality of the products than did consumers in the studies outlined above (e.g., obscure songs, wedding vendors, news stories). This may mediate the effects observed. Finally, [Salganik et al., 2006] and [Tucker and Zhang, 2011] measure the impact of popularity on sales, versus our setting in which we measure the impact of user feedback - likes - on sales. It is possible that likes reflect subscribers' tastes and their assessment of quality differently.

In addition to these papers, our research relates to the literature analyzing the impact of quality signals on user behavior. In the context of the movie industry, most papers have used observational data (as opposed to experimental results) for their analysis. For example, [Litman, 1983] analyzed 125 movies released in the US between 1972-78 and [Wallace et al., 1993] analyzed 1687 movies released in the US between 1956-88. Both papers report a positive correlation between box office sales and reviews by movie critics. However, in a later paper [Eliashberg and Shugan, 1997] found that ratings from movie critics are not good predictors of sales during the opening week, arguing that despite being correlated with cumulative movie sales, critic's ratings do not causally influence sales.

Likewise, [Godes and Mayzlin, 2004] studied 44 TV shows released in the US between 1999 and 2000. They found that the dispersion in Word-of-Mouth (WoM) about these shows across distinct groups in Usenet groups was positively correlated to their ratings. However, they were unable to establish a link between WoM, measured by number of conversations about a show, and future rankings, which correlate to sales. Similarly, [Liu, 2006] studied data from message boards at Yahoo Movies! for about 40 movies released between May and September 2002 in the US, finding that the volume of WoM

was positively correlated with box office sales but that there was no statistically significant relationship between the sentiment implied in the messages (positive/negative comments) and sales.

These findings suggest a potential correlation between unobserved quality and ratings that might be driving any observed impact of reviews on sales, and some papers have attempted to disentangle this correlation. For example, [Reinstein and Snyder, 2005] applied a difference in difference model on a sample of more than 600 movies rated by two influential movie critics to identify the marginal impact of reviews on sales. Using the fact that some movie reviews were issued prior to the release of the movie while others were issued after the opening week, they showed that—in contrast to [Eliashberg and Shugan, 1997]—ratings from movie critics were positively correlated with sales and influenced box office sales during the opening week.

In a similar context, [Zhang and Dellarocas, 2006] developed a structural model to study the impact of consumer and movie critic ratings on sales. They showed that good reviews drove movies sales but that the volume and dispersion of reviews did not. Likewise, [Dellarocas et al., 2007] developed a predictive model for movie sales showing that the volume, valence, and dispersion of reviews were all positive and statistically significant predictors of box office sales. Finally, [Duan et al., 2008] proposed a model with simultaneous equations to estimate both user movie ratings and movie box office sales. They conclude that WoM is a strong driver of box office sales, in contrast to the findings in [Zhang and Dellarocas, 2006].

3 Experimental Context

3.1 The Company and its Dataset

The experiment analyzed in this paper was performed using a real world VoD system from a major telecommunications provider, hereafter called OurTelco. OurTelco offers TV, Internet, telephone and mobile phone service and is the market leader for Pay-TV services in the country where it operates. It serves over 1 million households, 69% of which purchase triple play bundles that include TV, Internet and fixed telephony.

We had access to OurTelco's VoD database between February 2009 and December 2012. 623,516 of the active subscribers subscribe to services that include VoD. Overall, 681,036 subscribers watched VoD content at least once and 465,059 subscribers paid for VoD content at least once during this 41-month period. The remaining subscribers with VoD capabilities never used the service. We also had access to all (paid and free) VoD transactions. During this period we observe 89,074,657 transactions, of which 6,293,557 correspond to paid leases.

We have the anonymized identifier of the subscriber requesting each transaction as well as the anonymized identifier for the MAC address of the specific Set-Top Box (STB) that did so. For each transaction we have a timestamp, the price, and the identifier of the leased movie. For each movie in OurTelco's database we have the title, director, studio, play length, synopses, cast, genres, and asset type (movie, music, documentary, etc). We also have information on the daily layout of the TV screen that subscribers saw when they logged into the VoD system between. This information includes the tree of menus displayed as well as the rank order in which movies were displayed under each menu on the screen from left to right. Menus are associated with different editorial lines as described in the next section. We also have daily information on the number of *likes* issued to each

movie.

3.2 VoD Service and Interface

OurTelco provides service over wired and satellite infrastructure. However, satellite subscribers cannot subscribe to VoD. Wired subscribers can obtain one of three types of services: basic, standard or premium. All subscribers, regardless of type, can watch TV and subscribe to specific channels such as movies, sports, children's and adult content. Figure 1 summarizes the services available to each subscriber. For our setting it is important to note that only standard and premium subscribers can use the VoD service and only premium subscribers can issue *likes* for VoD movies and TV programs. As such, in this paper, we will focus only on standard and premium subscribers (respectively 84% and 16% of OurTelco's DoD-enabled subscribers in January 2012).

Feature	BASIC/LEGACY	STANDARD	PREMIUM
Watch TV	YES	YES	YES
Subscribe to Thematic Channels	YES	YES	YES
Video on Demand	NO	YES	YES
DVR Capabilities	NO	YES	YES
Restart TV Features	NO	NO	YES
Like Ability & Facebook Link	NO	NO	YES
Mandatory Internet Connection Service	NO	NO	YES

Figure 1: Summary of the main features available to subscribers of OurTelco.

The look and feel of the VoD screen for standard and premium subscribers is different, but the organization of the content into menus is hierarchically similar. Both standard and premium subscribers can access the VoD system using a hot-key in their STB remote

control. Pressing this hot-key displays the entry screen of the VoD system. This screen, called the *Highlights Section*, contains a set of menus filled with movies, chosen by an editorial team, which are easy to access based on their location on the screen. Movies are organized into menus such as Promotions, Suggestions, and Newest Releases. Each menu has a header with a name that clearly identifies the type of movies it contains. Menus are horizontal lines on the screen. Different menus are stacked vertically. Two menus fit the screen at each time. A cursor highlights a single movie cover at a time. Users can scroll across menus. The natural scrolling direction is down, though premium consumers can also scroll up. Figure 2 depicts the look and feel of the main screen of the highlights section for premium consumers.

Upon scrolling to a new menu, a number of movie covers are visible under that menu and the cursor highlights the movie farthest to the left. Premium users see 8 movie covers under each menu immediately. Standard subscribers see either 4 or 8 movie covers under each menu depending on whether they connect their TV to their set-top-box using a SCART or a HDMI cable. Unfortunately, we do not know the type of cable used by each standard subscriber. Users can also scroll right and left within the same menu. All users can scroll past the last movie cover on the screen to unveil hidden movies under the same menu. OurTelco displays at most 15 movies per menu. Premium and standard subscribers see the same number of movies under the same menu.

The title of the movie highlighted by the cursor is shown on the screen. Premium consumers also see the number of likes for this movie. Clicking on the cover of a movie leads to a new screen with the year of release, play length, cast and synopsis. A number of actions are then available including the ability to lease the movie, use a promotional coupon or watch the movie trailer (if one is available). Figure 3 provides an illustration of the title page of a movie.

Finally, subscribers can leave the *Highlights Section* of the VoD interface and search for movies in the complete *Catalog.* The catalog is hierarchically organized into content categories such as movies, music, TV-shows, and documentaries. Within each of these categories, screens are organized as described above with menus for genres. In addition to browsing through the entire catalog, subscribers can use a keyword search function to look for specific content based on movie titles, movie directors and actors' names.

The number of *likes* shown along with movie covers is cumulative from the movie's debut at OurTelco's VoD system. Subscribers do not know who liked a particular movie or how many people leased a particular movie.
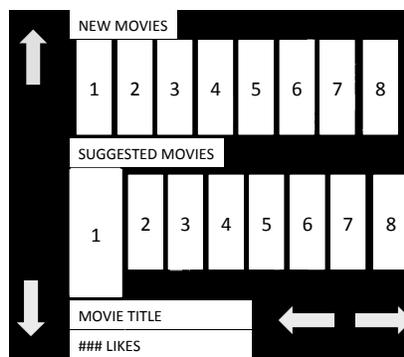

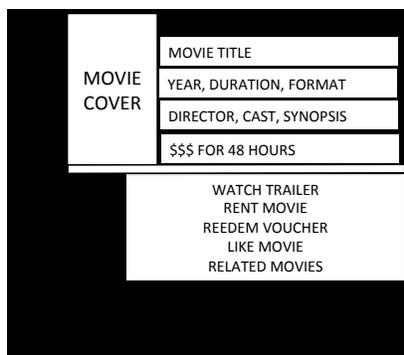Figure 2: Example of the interface look and feel in the highlights section

Figure 3: Example of the movie page in the video-on-demand system

4 Experimental Design

The experiment discussed here is based on a new menu, named "The Most Popular Movies During the Past Weeks," that was introduced in the highlights Section OurTelco's VoD system on July 3rd 2012. This menu was available for both standard and premium subscribers and included the 15 movies that obtained the highest number of *likes* in the last few weeks. These movies were shown under this menu in decreasing order of number of *likes* from left to right. The experiment ran in 1-month cycles for a period of 6 consecutive months. Each cycle was further split into mini-cycles of 1 week each. Each week started at a time of low VoD usage, namely Tuesdays at around 2pm. Weeks were named true or false. During a true week all movies under this menu were shown in their true rank. The true number of *likes* they obtained in the recent past was shown to premium consumers. During a false week a randomized procedure was used to swap some movies under this menu.

Formally, the experiment ran as follows: Let t_i represent the time at which cycle i began, with $i \in \{1, 2, 3, 4, 5, 6\}$. Let x represent a week's time. At time t_i, all movies in OurTelco's VoD system were sorted according to the number of *likes* received between $t_i - 2x$ and t_i. All movies that OurTelco decided to use in other menus under the highlights section were removed from this sorted list. For example, some movies were displayed under Promotions or Newest Releases. Cleaning these movies from the above list avoided displaying them under more than one menu in the highlights section. OurTelco's log system does not allow for identifying the menu under the highlights section from which a lease originates and thus this cleaning procedure ensures that leases of movies under the new menu came only from this menu. The 45 movies at the head of the resulting list were used in the experiment, which we call list L. Movies in list L had on average, more likes

than the movies displayed in other menus ($p-val < 0.001$) of the highlights section.

After the setup phase described above, which took place at the beginning of each cycle, a true week ensued to adjust the subscribers' expectations to the true quality of the movies show under the new menu. The nature of each of the other 3 weeks within every cycle was determined at random. This allowed for preventing a static pattern of true/false weeks that subscribers could perceive. Table 1 shows the order of true and false weeks during the experiment.

Table 1: Cycles and the nature of sub-cycles during our experiment

Cycle 1	t_1: True	$t_1 + x$: True	$t_1 + 2x$: False	$t_1 + 3x$: False
Cycle 2	t_2: True	$t_2 + x$: False	$t_2 + 2x$: True	$t_2 + 3x$: False
Cycle 3	t_3: True	$t_3 + x$: False	$t_3 + 2x$: False	$t_3 + 3x$: True
Cycle 4	t_4: True	$t_4 + x$: False	$t_4 + 2x$: False	$t_4 + 3x$: False
Cycle 5	t_5: True	$t_5 + x$: False	$t_5 + 2x$: False	$t_5 + 3x$: False
Cycle 6	t_6: True	$t_6 + x$: False	$t_6 + 2x$: True	$t_6 + 3x$: False

At the beginning of every week all movies in L were sorted according to the number of *likes* that they obtained between $t_i - 2x$ and $t_i + nx$ with $n \in \{1, 2, 3\}$ indicating how many weeks elapsed since the start of the current cycle. In a true week, the first 15 movies in L were displayed under the new menu from left to right on the TV screen. In false weeks list L was partitioned into 3 sub-lists. List L_1 comprised the 15 movies at the head of list L. List L_2 included the movies between ranks 16 and 30 in list L. Finally, list L_3 contained the movies positioned between ranks 31 and 45 in list L. Then, the following swaps were performed:

- *Within Swap:* y_i and y_j were selected randomly from $\{1, ..., 15\}$ such that $y_i \neq y_j$ and the number of *likes* associated with the y_i^{th} and y_j^{th} movies in list L_1 were swapped;

- *Between Swap:* z_i was selected randomly from $\{1, ..., 15\}$ such that $z_i \neq y_i$ and

$z_i \neq y_j$ and z_j was selected randomly from $\{1, ..., 15\}$. Then, the number of *likes* of the z_i^{th} movie in list L_1 was swapped with the number of *likes* of the z_j^{th} movie in either list $L2$ or list $L3$.

Two *Within Swaps* and one *Between Swap* were performed at each false week during the first three cycles of the experiment. The latter alternated between lists $L2$ and $L3$. Three *Within Swaps* and two *Between Swaps* were performed at each false week, one involving $L2$ and one involving $L3$, during the last three cycles of the experiment. The frequency of swaps in the final three cycles of the experiment was increased to obtain more treated observations.

The movies in list L_1 were then displayed under the new menu from left to right on the TV screen. The two types of random swaps introduced with this experiment were aimed at capturing the particular characteristics of the look and feel of OurTelco's VoD interface. *Within Swaps* allow for determining whether changes in ranks across the movies displayed under the new menu have an impact of sales. *Between Swaps* allow for determining the impact of bringing movies from the catalog into the new menu and of removing movies from the new menu into the catalog. A *Within swap* changes the search cost of the swapped movies only slightly but a *Between Swap* reduces substantially the search costs for a movie that is promoted from the catalog to the new menu and increases substantially the search costs for a movie that is demoted from the new menu into the catalog.

5 Empirical Model

5.1 Movie Level Specification

We use the following reduced form equation to study the effect of rank on sales:

$$y_{it} = \alpha + x_{it}\beta + w_{it}\gamma + m_i + u_{it}, t = 1, \ldots T \tag{1}$$

where y_{it} represents the sales of movie i during week t, x_{it} includes time varying observed movie characteristics such as age, rank and the number of distinct menus where the movies is displayed [1] w_{it} is the vector of exogenous random treatments, m_i are time constant movie fixed effects such as price, and u_{it} is an idiosyncratic error term. This equation represents the classic fixed effects specification, which we estimate using first differences with robust standard errors. Therefore, we estimate the following model:

$$\delta y_{it} = \epsilon + \Delta x_{it}\beta + \Delta w_{it}\gamma + \Delta u_{it}, t = 2, \ldots T \tag{2}$$

Note that the time constant movie fixed effects in z_i drop despite being observed. In particular, the retail price drops from our regression. Price includes a fixed margin on the top of the royalty fee and the latter did not change during our experiment. Furthermore, prices do not change in response to changes in demand in our setting as OurTelco does not engage in dynamic pricing.

---

[1]The same movie can be labelled under several genres and therefore appear several menus in the VoD catalog. We control for how many times the movie appears in all menus because this affects exposition to consumers.

5.2 The Magnitude of Treatment

Consider movies A and B under the new menu in ranks a and b, respectively, at time $t_i + nx$, with $n < 3$. When these movies are swapped their new ranks in list L are, momentarily, b and a, respectively. At time $t_i + (n+1)x$, movies in this list are reordered according to number of *likes* as described before. As a result, assume that the movie at rank a shifts to rank a' and the movie at rank b shifts to rank b'. Subscribers see only the cumulative effect of swaps and sorting. Thus, from their perspective, movie A moved from rank a to rank b' (a change of $b' - a$ ranks) and movie B moved from rank b to rank a' (a change of $a' - b$ ranks).

If the swap did not occur, subscribers would have seen that movie A moved from rank a to rank a' and movie B moved from rank b to rank b'. Therefore, the effect of the random exogenous swap is given by $(b' - a) - (a' - a) = b' - a'$ for movie A and by $(a' - b) - (b' - b) = a' - b'$ for movie B. Note that this difference is zero for control movies. We introduce this difference, which hereinafter we call $rank\_manipulation$, into Δw_{it} in equation 2. We code it so that it is positive when a movie is promoted and negative when a movie is demoted. Also, a' and b' are the true ranks for movie A and B, respectively, which hereafter we call $TrueRank$. Therefore, we have $Rank = TrueRank - RankManipulation$.

5.3 Identification and Exogeneity

Identification is obtained by design in our setting. In equation 2, Δw_{it} is not correlated to Δu_{it} because movie swaps are randomly and exogenously determined. The two movies involved in a swap are randomly selected and the timing of the swap is also random. Therefore, movies are treated at random and the magnitude of the treatment is also random. Random assignment of the treatment also ensures that Δw_{it} is uncorrelated to

Δx_{it}. Therefore, our estimates for γ in Equation 2 obtained with OLS are unbiased.

5.4 Rank Level Specification

Movies are reordered according to the number of *likes* at the beginning of each week. This establishes a truthful relationship between rank and perceived quality for control movies in the eyes of OurTelco's subscribers. Therefore, we can compare the sales obtained by control and treated movies at each rank and determine whether promoted or demoted movies sell differently than true movies placed at that rank. A true movie at a rank is a movie that was placed at this rank as a result of the number of *likes* obtained from subscribers, as opposed to being manipulated. If sales at a rank are not different when the movie at that rank is manipulated, then rank alone determines movie sales. We test this hypothesis with the following model:

$$y_{rt} = \alpha + x_{rt}\beta + w_{rt}\gamma + m_r + u_{rt}, t = 1, \dots T \tag{3}$$

where y_{rt} represents the sales of the movie at rank r during week t, x_{rt} includes observed characteristics of the movie at rank r in week t, such as age, price, IMDb rating, and the number of distinct menus where the movie appears, m_r is the intrinsic perceived quality of rank r, and u_{rt} is an idiosyncratic error term. w_{rt} is a vector of exogenous random treatments indicating whether the movie at rank r in week t was promoted, demoted or neither. A promoted movie should have, on average, lower quality than the movie it displaces and as a result may possibly sell less. Conversely, a demoted movie should have, on average, higher quality than the movie it displaces and may possibly sell more. We estimate Equation 3 using a dummy variable for each rank.

6 Results and Discussion
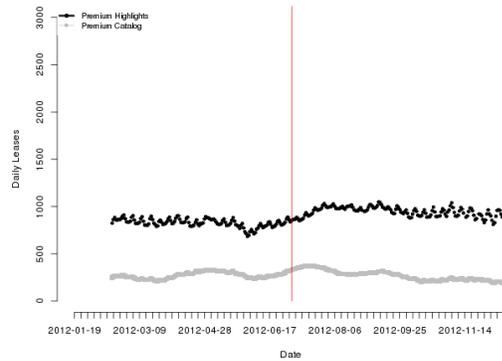
6.1 Descriptive Statistics

The combined stock of standard and premium consumers at OurTelco grew from 607,000 in January 2012 to 633,000 in December 2012[2]. Figure 4 shows the 30-day moving average of daily sales in the highlights section and in the catalog for premium and standard subscribers. Most sales came from standard subscribers though this gap reduced significantly towards the end of the year as the number of standard subscribers decreased and they started leasing fewer movies. Sales increased both in the highlights section and in the catalog around the time the experiment started. Figure 5 shows the 30-day moving average of daily sales for menus under the highlights section (expect for "The most seen during the last week" and "Newest Releases", which sold more that the menus shown in this figure). This figure shows that the new menu was well received by consumers and started selling well. The new menu sold more than any other menu shown in this figure during the first 10 weeks of the experiment. At week 10, OurTelco introduced two new menus into the highlights section, called "The most seen of all times" and "The most popular on IMDb." These menus competed with the menu used for the experiment both in terms of consumer attention and movies. In fact, when a movie under "The most popular during the past weeks" was also among "The most seen of all times" or "The most popular on IMDb" it would be pulled into the latter two menus and deleted from the former to avoid

---

[2]The share of VoD-enabled premium subscribers increased from 16% to 34% during the same period. In the first half of 2012 premium users leased an average of 1,100 movies per day. This increased to 1,200 during the second half of the year. These statistics were 2,300 and 1,700, respectively, for standard users. Yet, the average number of leases per subscriber decreased from 3.2 to 1.1 from the first to the second half of the year for premium subscribers. These statistics were 1.8 and 1.5 for standard subscribers, respectively. During the first half of 2012, 75% of the leases from premium users originated in the highlights section. This statistic increased to 79% in the second half of the year. These statistics were 64% and 68% for standard users, respectively.

duplication.
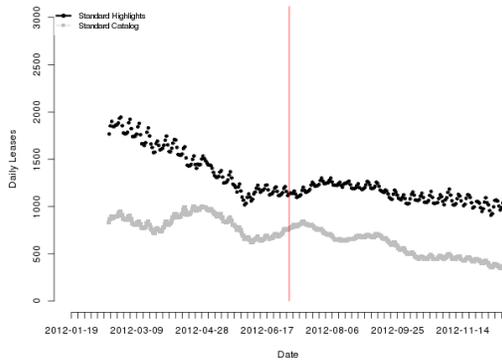
(a) Leases from Standard Subscribers   (b) Leases from Premium Subscribers

Figure 4: 30-Day Moving Average of Daily Leases in Highlights and Catalog in OurTelco's VoD.

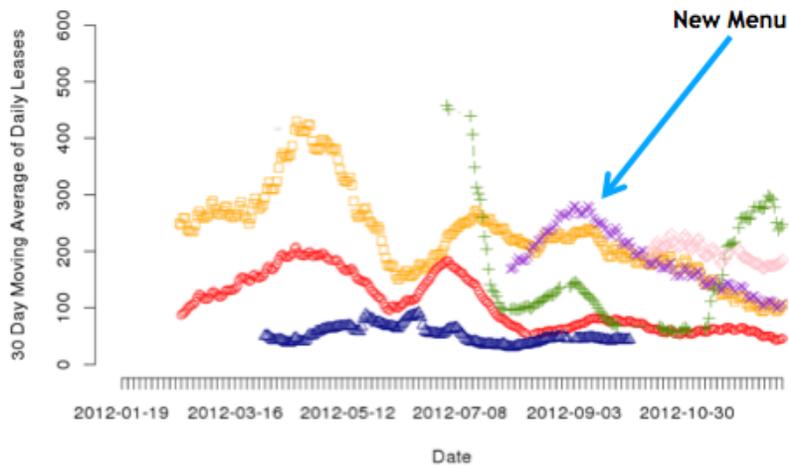Figure 5: 30-Day Moving Average of Daily Leases in Highlights per Menu in OurTelco's VoD.

Figure 6 shows the weekly sales in the new menu over time. Unlike overall VoD consumption, the majority of sales under this menu came from premium subscribers. The new menu was visible in the entry screen of the VoD system for premium subscribers and reachable with 1 click up, whereas standard subscribers did not see this menu immediately

20

when they entered the highlights section and needed to click 10 times down to reach it. In addition, standard subscribers do not see the number of likes, which might have rendered this menu less meaningful to them.
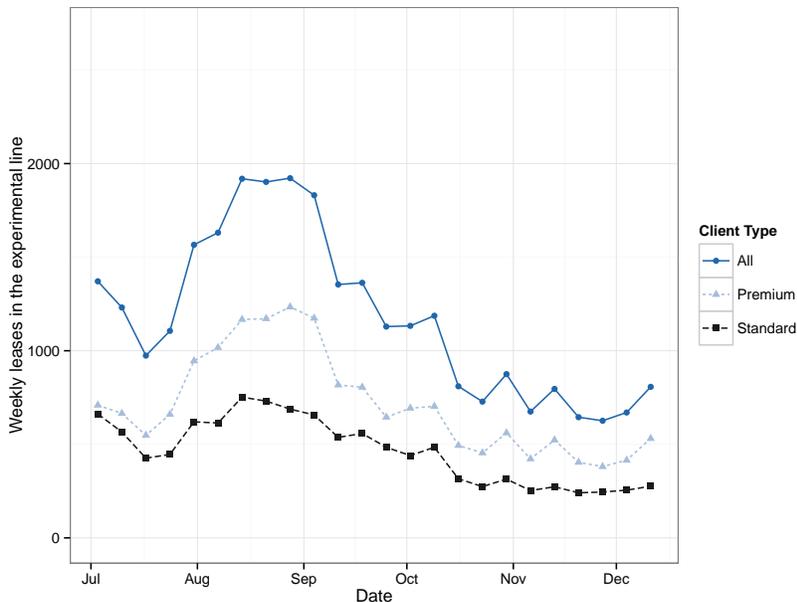

Figure 6: Sales in the New Menu During the Experiment.

22,043 subscribers leased movies from the new menu. Figure 7 shows that roughly 80% of the subscribers leasing movies from the new menu did so only once during the experiment. 50% of the subscribers lease less than 1 movie per quarter, and roughly 20% of subscribers lease more than 1 movie per month.

Panel (a) in Figure 8 shows the number of *likes* per rank in the beginning of each week. This is a decreasing function by design and we observe a clear exponential decay in *likes* as a function of rank. Panel (b) in Figure 8 shows the number of leases during the week per rank. This function, however, is far from monotonic, which might suggest that subscribers use information besides rank and number of *likes* to decide which movies to watch.
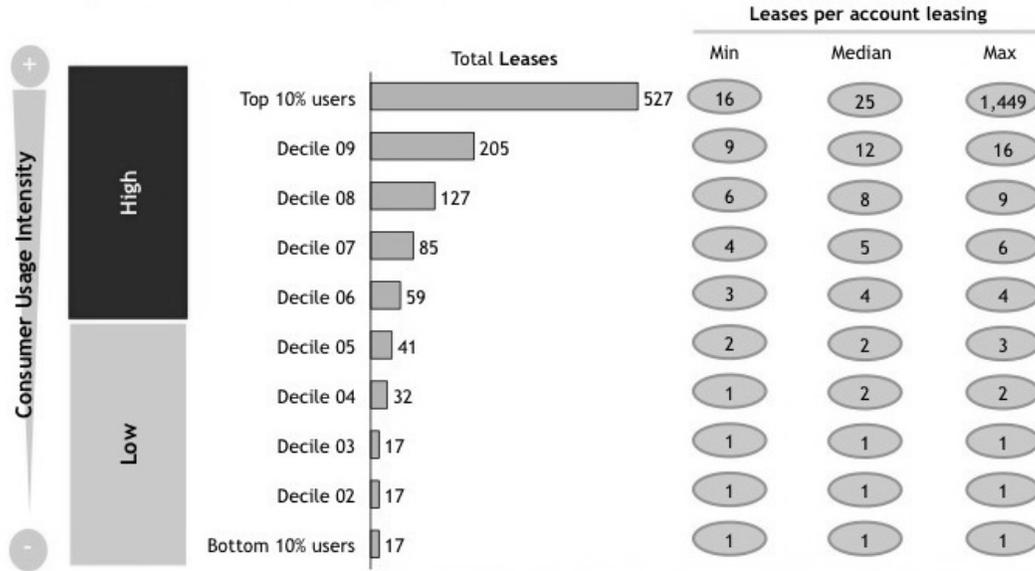
Figure 7: Statistics on VoD Consumption per Subscriber.
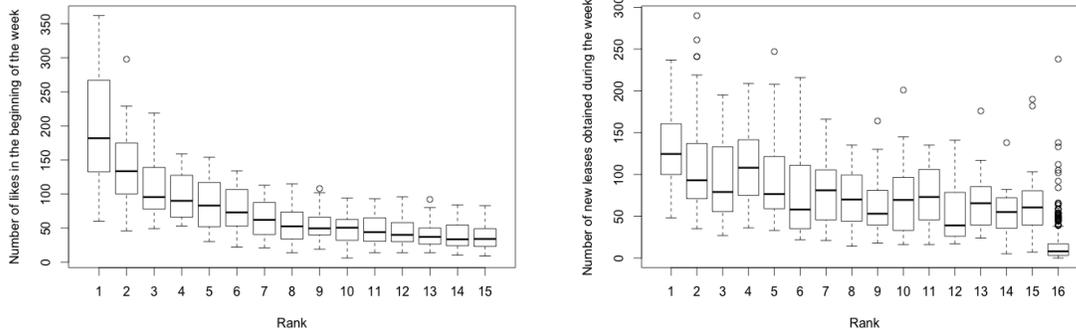
(a) Number of Likes per Rank (beginning of the week)

(b) Leases per Week as Function of Rank

Figure 8: Likes in the Beginning of the Week and Leases per Week as Function of Rank at the New Menu.

Figure 9 shows the relationship between the change in the number of leases obtained per week and the manipulation in rank introduced during the experiment. This figure provides preliminary evidence that promoted (demoted) movies sold more (less) and that the effect of between swaps dominates the effect of within swaps. Figure 10 shows that,

on average, promoted (demoted) movies tend to receive fewer (more) *likes* than untreated movies do. This provides preliminary evidence that manipulated movies might return to their true ranks when they are re-ordered in subsequent weeks according to the number of likes.
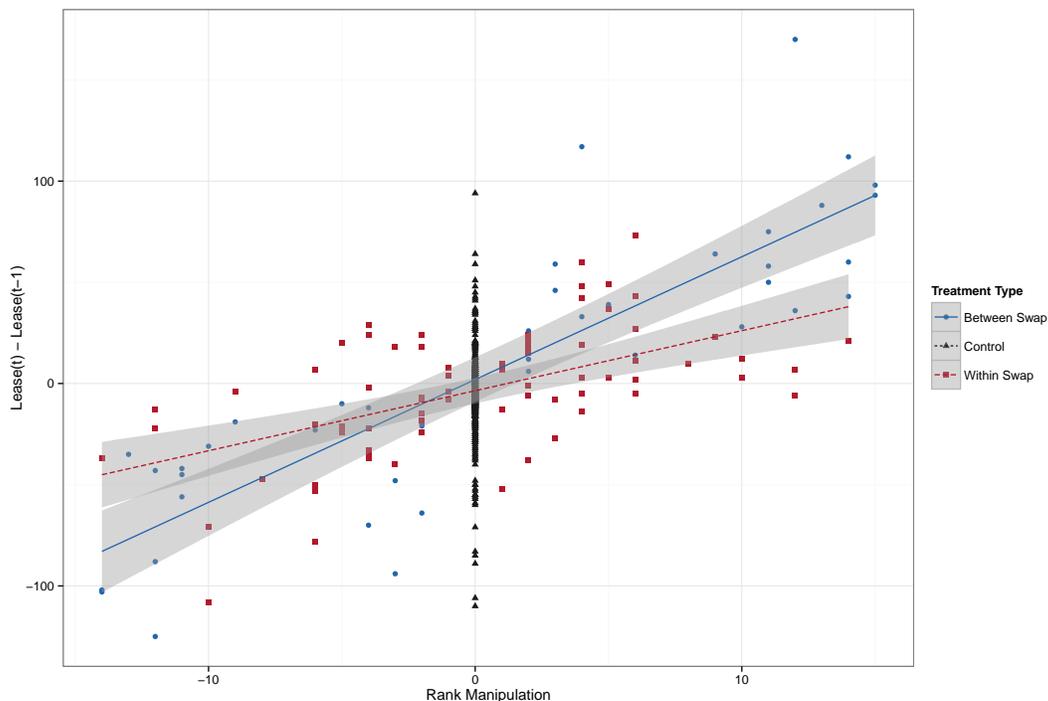
Figure 9: Relationship Between Changes in Sales and Rank Manipulation During the Experiment.

Finally, Table 2 shows descriptive statistics for the covariates used in this paper separately for all movies and for control and treated movies in the catalog and in the highlights section. T-tests to compare means between control and treated movies show they are all similar, which is as expected given our random treatment assignment. Treated movies in the catalog are movies in lists $L2$ and $L3$ that are randomly selected to be displayed in the highlights section. All other movies in these two lists are control movies in the catalog. We code the rank of all movies in these two lists as 16.
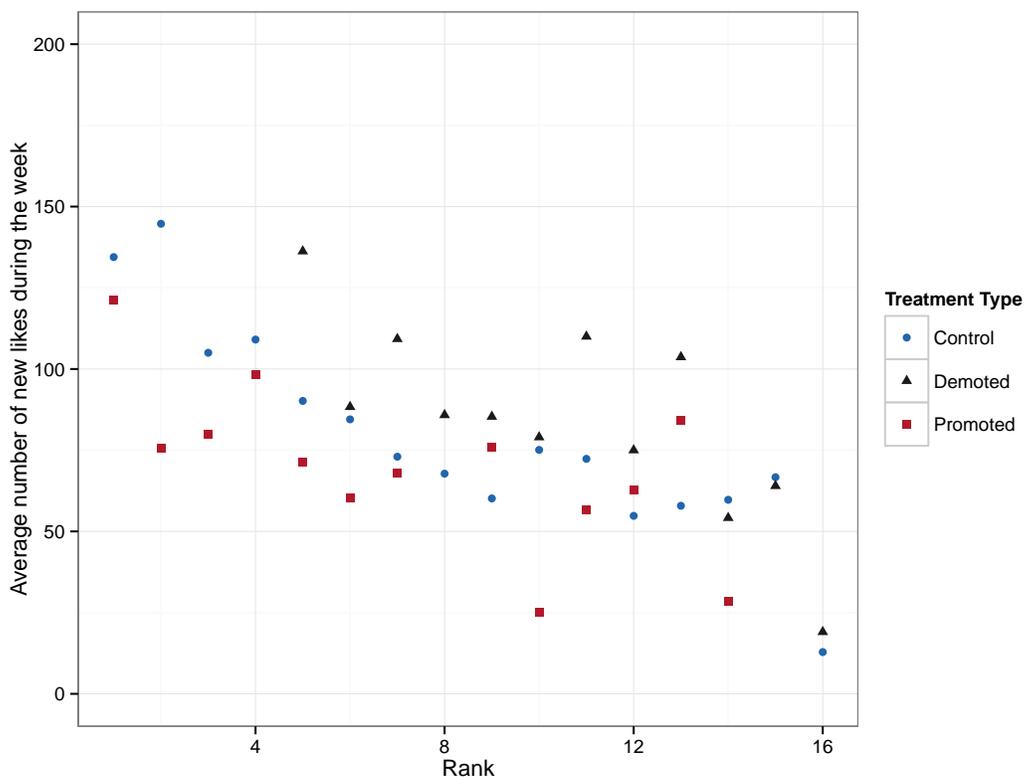
Figure 10: Likes per Week as a Function of Rank for All, Control and Treated Movies at the New Menu.

6.2 The Effect of Movie Swaps

We estimate equation 4 (which resembles equation 1) to learn the effect of rank on leases. In this regression, $TreatedWithin \times RankManipulation$ denotes the size of a rank manipulation within the top 15 ranks. $PromotedtoLine$ and $DemotedfromLine$ denote the size of rank manipulations that lead a movie to go from the catalog into the new menu or to move from the new menu into the catalog, respectively. These three types of manipulations constitute a partition of the space of possible manipulations and therefore their coefficients must be interpreted relative to our control movies. $Treated$ indicates whether a movie has been treated.

Table 2: Descriptive Statistics for All Covariates used by Catalog ($L2$ and $L3$) and Highlights ($L1$).

Vars	Stats	All	Catalog ($L2$ and $L3$)		Highlights ($L1$)	
			Control	Treated	Control	Treated
Number of Leases						
	mean	36.341	12.85	19.05	80.461	82.29
	sd	45.788	18.201	18.735	50.758	45.904
Number of Leases (Premium)						
	mean	19.648	4.174	6.9	48.23	51.527
	sd	29.136	6.124	7.247	33.785	30.23
Number of Leases (Standard)						
	mean	16.693	8.676	12.15	32.23	30.763
	sd	19.262	12.644	12.123	20.839	20.25
Rank						
	mean	13.311	16	16	8.531	7.151
	sd	4.487	0	0	4.197	4.369
True Rank						
	mean	13.348	16	8	8.531	9.28
	sd	4.438	0	4.472	4.197	5.247
Rank Manipulation						
	mean	0.037	0	-8	0	2.129
	sd	2.5	0	4.472	0	6.811
Number of Menus						
	mean	1.984	1.708	1.65	2.609	2.258
	sd	1.058	0.932	0.813	1.193	0.674
Price						
	mean	287.741	260.883	324	331.617	346.312
	sd	92.662	84.763	96.655	90.21	74.951
IMDb Rating						
	mean	6.328	6.31	5.98	6.427	6.253
	sd	1.242	1.215	1.485	1.261	1.304
IMDb Votes						
	mean	82434.666	87387.516	73270.75	80008.728	58978.022
	sd	114271.836	117701.293	168947.22	111504.825	76944.554
Movie Age						
	mean	250.257	291.779	266.294	166.844	187.112
	sd	380.553	415.368	429.838	277.441	314.458
Observations		1017	648	20	256	93

$$leases_{it} = \lambda + \beta_1 Log(MovieAge_{it}) + \beta_2 Number of Menus_{it} + \beta_3 Treated_{it} +$$

$$\beta_4 TrueRank_{it} + \beta_5 TreatedWithin \times RankManipulation_{it} +$$

$$\beta_6 PromotedtoLine_{it} + \beta_7 DemotedfromLine_{it} +$$

$$+WeekDummies + m_i + \epsilon_{it} \qquad (4)$$

Table 3 shows the results obtained with first-differences for all subscribers and separately for standard and premium subscribers. The coefficient on $TreatedWithin \times RankManipulation$ shows that a promoted (demoted) movie receives more (fewer) leases than other movies do. This result is statistically significant for both standard and premium subscribers, although less for the former. On average, a manipulation that increases rank by 1 leads to 2.3 (0.5) more leases from premium (standard) subscribers. This corresponds to a 4.7% (1.6%) increase in the number of leases. Promoting a movie to the new menu increases 7.2 (2.1) times the number of leases from premium (standard) subscribers, on average. This significant jump is associated with the difference in search costs between the catalog and the highlights section. Demoting movies from the new menu yields the opposite effect for premium subscribers: the number of leases reduces by 37%. The effect of demotions from the new menu is not statistically significant for standard subscribers. The new menu was much harder to reach for standard subscribers and thus standard subscribers that use this menu might already be more willing to search for good movies. We also note that, as expected given the random assignment of treatments in our experiment, the coefficient on $Treated$ is not statistically significant.

Other covariates behave as expected. The effect of age is negative, showing that movies tend to sell less as they become old. The effect of number of menus is positive, showing that movies displayed in more menus tend to sell more. The effect of true rank is negative for premium consumers, showing that better ranked movies tend to sell more. This coefficient is not statistically significant for standard subscribers, which again might indicate that these consumers are more willing to search for good movies to watch.

Table 3: The Effect of Swaps Within the New Menu and Between the New Menu and the Catalog on Sales.

	Subscribers	All	Standard	Premium
	Model	FD	FD	FD
	Variables	$leases_{it}$	$leases_{it}$	$leases_{it}$
	(Intercept)	-5.621*	-2.693	-2.928
		(3.083)	(1.637)	(1.805)
	Log(Movie Age)	-11.852**	-11.775***	-0.076
		(5.617)	(3.657)	(2.788)
	Number of Menus	12.3***	5.731***	6.569***
		(3.253)	(1.678)	(1.824)
	Treated	1.356	0.387	0.969
		(3.039)	(1.039)	(2.571)
	True Rank	-0.62	0.137	-0.756**
		(0.752)	(0.555)	(0.321)
Treated Within $\times$ Rank Manipulation		2.821***	0.509*	2.313***
		(0.488)	(0.278)	(0.333)
	Promoted to Line	36.31***	9.366***	26.945***
		(6.091)	(2.084)	(4.579)
	Demoted from Line	-23.039***	-4.848	-18.191***
		(7.091)	(3.131)	(4.646)
	Week Dummies	Yes	Yes	Yes
	Observations	817	817	817
	R-Squared	0.448	0.264	0.478
	R-Squared Adj	0.431	0.254	0.461
	F-Stat (p-value)	0	0	0

Note 1: Robust standard errors in ( );
Note 2: *** p<0.01, ** p<0.05, * p<0.1
Note 3: First Differences Estimator

Swapping the order in which movies were displayed under the new menu was accomplished by swapping their number of likes as described in Section 4. Therefore, we can also look at the effect of the number of likes on sales. In fact, and on average across all manipulations introduced during our experiment, promoting a movie by one rank was associated with an increase of 12 likes. We define $LikesManipulation$ to indicate the exogenous change in the number of likes of the manipulated movies. This covariate is zero by design for control movies. If movie A at rank a with l_a likes swaps with movie B at rank b with l_b likes, with $a > b$ and thus $l_a > l_b$, then $LikesManipulation = l_a - l_b$ for movie B and $Likes\ Manipulation = l_b - l_a$ for movie A. Table 4 shows the effect of likes

on leases. Each additional like increases weekly sales by 0.22 units.

Table 4: The Effect of Swapping Likes Within the New Menu and Between the New Menu and the Catalog on Sales.

Subscribers	All	Standard	Premium
Model	FD	FD	FD
Variables	$leases_{it}$	$leases_{it}$	$leases_{it}$
(Intercept)	-6.824**	-2.865*	-3.959**
	(3.266)	(1.683)	(1.893)
Log(Movie Age)	-10.687*	-11.92***	1.233
	(5.967)	(3.857)	(3.028)
Number of Menus	13.062***	5.649***	7.413***
	(3.283)	(1.607)	(1.89)
Treated	0.974	0.317	0.657
	(2.291)	(0.973)	(1.98)
True Likes	-0.011**	-0.002	-0.009**
	(0.005)	(0.003)	(0.004)
Treated Within × Likes Manipulation	0.219***	0.038*	0.181***
	(0.03)	(0.02)	(0.021)
Promoted to Line	35.659***	9.494***	26.166***
	(5.551)	(2.01)	(4.094)
Demoted from Line	-21.599***	-5.352**	-16.247***
	(5.296)	(2.413)	(3.5)
Week Dummies	Yes	Yes	Yes
Observations	817	817	817
R-Squared	0.472	0.266	0.513
R-Squared Adj	0.455	0.256	0.495
F-Stat (p-value)	0	0	0

Note 1: Robust standard errors in ();
Note 2: *** p<0.01, ** p<0.05, * p<0.1;
Note 3: First Differences Estimator

The effect of rank on sales reported before was assumed constant across ranks but in fact promoting a movie from rank 15 to rank 14 may be very different from promoting a movie from rank 2 to rank 1. Ideally, we would measure each of these effects separately but we do not have enough data to do so. However, we can still show that promoting a movie by 1 rank changes sales differently at different ranks if, for example, we assume a linear

trend for this effect. Such a trend can be captured by interacting $RankManipulation$ with $TrueRank$. Table 5 shows the results obtained. We estimate the average effect of rank manipulation to be 5.18 and the effect of the interaction to be -0.31. This results suggest, as expected, that promotions have a stronger effect for higher ranks than for lower ranks: a promotion from rank 2 to rank 1 increases weekly sales by 4.56, whereas a promotion from rank 8 to rank 7 increases weekly sales by 2.69 only.

Table 5: The Effect of Swaps on Sales Across Ranks Within the New Menu.

Subscribers	All	Standard	Premium
Model	FD	FD	FD
Variables	$leases_{it}$	$leases_{it}$	$leases_{it}$
(Intercept)	-5.948*	-2.854*	-3.095*
	(3.116)	(1.635)	(1.828)
log(Movie Age)	-11.291**	-11.5***	0.209
	(5.608)	(3.573)	(2.856)
Number of Menus	12.188***	5.676***	6.512***
	(3.254)	(1.664)	(1.836)
Treated	6.951*	3.136**	3.815
	(3.593)	(1.31)	(2.908)
True Rank	-0.481	0.205	-0.686**
	(0.754)	(0.556)	(0.325)
True Rank × (Rank Manipulation × Treated Within)	-0.312***	-0.153***	-0.159*
	(0.11)	(0.041)	(0.088)
Treated Within × Rank Manipulation	5.184***	1.669***	3.515***
	(0.918)	(0.504)	(0.63)
Promoted to Line	30.913***	6.713***	24.199***
	(6.382)	(2.175)	(4.833)
Demoted from Line	-28.528***	-7.545**	-20.983***
	(7.245)	(3.28)	(4.631)
Week Dummies	Yes	Yes	Yes
Observations	817	817	817
R-Squared	0.464	0.284	0.488
R-Squared Adj	0.447	0.273	0.469
F-Stat (p-value)	0	0	0

Note 1: Robust standard errors in ();
Note 2: *** p<0.01, ** p<0.05, * p<0.1;
Note 3: First Differences Estimator

Another way to see that the effect of manipulations is not constant across ranks is to consider what happens in the visible and in the hidden parts of the menu. Table 6 shows the results obtained when we interact $RankManipulation$ with whether the manipulation occurs within ranks 1-8 (visible part), within ranks 9-15 (hidden part), from ranks 1-8 to ranks 9-15 or from ranks 9-15 to 1-8. Swaps within the visible part of the menu have a significant impact on sales (2.5x more than that associated to swaps between the visible and the hidden parts of the menu) but swaps within the hidden part of the menu do not seem to change sales. Note, however, that this analysis is performed only for premium subscribers because we cannot know for sure how many movies are shown in the visible part of the menu to standard subscribers.

6.3   Robustness Check

Treated movies receive a number of likes that is different from what they otherwise would. Consequently, they might be displayed in a rank different from the rank they would otherwise occupy when at the end of a false week swaps are reverted. One approach to circumvent this problem is to ignore treated movies within the same month after they are first treated. Yet, some treated movies remain in our panel from one month to the next. Thus, a second approach to address this problem is to ignore treated movies once they are first treated until the end of the experiment. In this section we test the degree to which prior treatment may bias our results by conducting a robustness test using both approaches described above.

Table 7 shows the results obtained using these approaches for all consumers, standard consumers alone and premium consumers alone. The columns marked "Trimmed Month" refer to the former approach while the columns marked "Trimmed Experiment" refer

Table 6: The Effect of Manipulations in the Visible and Hidden Parts of the New Menu.

	Dependent variable:	
	$leases_{it}$	
	(1)	(2)
Rank Manipulation	2.282***	
	(0.329)	
Rank Manipulation × To_01_08		1.762***
		(0.344)
Rank Manipulation × To_09_15		1.726**
		(0.723)
Rank Manipulation × Whithin_01_to_08		4.448***
		(0.687)
Rank Manipulation × Whithin_09_to_15		0.478
		(1.371)
Promoted to Line	31.700***	31.570***
	(3.024)	(3.046)
Demoted from Line	−19.700***	−19.590***
	(3.347)	(3.372)
Constant	−3.191**	−3.190**
	(1.589)	(1.607)
Week Dummies	Yes	Yes
Observations	817	817
R^2	0.447	0.476
Adjusted R^2	0.433	0.459
F Statistic	25.580*** (df = 25; 791)	25.570*** (df = 28; 788)

Note: *p<0.1; **p<0.05; ***p<0.01
Robust standard errors in ()
First differences estimator

to the latter. Columns marked "Original" replicate the results in Table 3 to facilitate the comparison. Our results remain unchanged when we ignore treated movies after first treatment, although this of course results in fewer observations which increases our standard errors. Therefore, it is reasonable to believe that this type of contamination does not hurt our findings.

Table 7: Movie Level Regression Results Ignoring Treated Movies After Treatment.

Vars	ALL Original $leases_{it}$	ALL Trimmed Month $leases_{it}$	ALL Trimmed Whole $leases_{it}$	Standard Original $leases_{it}$	Standard Trimmed Month $leases_{it}$	Standard Trimmed Whole $leases_{it}$	Premium Original $leases_{it}$	Premium Trimmed Month $leases_{it}$	Premium Trimmed Whole $leases_{it}$
(Intercept)	-5.621*	-5.068*	-5.4*	-2.693	-2.419	-2.567	-2.928	-2.649*	-2.833*
	(3.083)	(2.711)	(2.83)	(1.637)	(1.565)	(1.601)	(1.805)	(1.559)	(1.608)
Log(Movie Age)	-11.852**	-15.077**	-16.646***	-11.775***	-13.841***	-14.634***	-0.076	-1.235	-2.013
	(5.617)	(6.026)	(6.284)	(3.657)	(4.035)	(4.169)	(2.788)	(3.24)	(3.218)
Number of Menus	12.3***	7.009**	6.76**	5.731***	3.641**	3.643**	6.569***	3.369**	3.117**
	(3.253)	(2.773)	(2.767)	(1.678)	(1.49)	(1.494)	(1.824)	(1.606)	(1.585)
Treated	1.356	-0.386	2.68	0.387	0.407	1.542	0.969	-0.794	1.138
	(3.039)	(3.361)	(5.318)	(1.039)	(1.528)	(2.421)	(2.571)	(2.484)	(3.523)
True Rank	-0.62	0.618	0.73	0.137	0.878	0.926	-0.756**	-0.26	-0.196
	(0.752)	(1.043)	(1.142)	(0.555)	(0.881)	(0.976)	(0.321)	(0.372)	(0.383)
Treated Within × Rank Manipulation	2.821***	3.108***	3.2***	0.509*	0.478	0.491	2.313***	2.63***	2.709***
	(0.488)	(0.667)	(0.895)	(0.278)	(0.325)	(0.55)	(0.333)	(0.459)	(0.635)
Promoted to Line	36.31***	54.33***	55.661***	9.366***	15.722***	16.268***	26.945***	38.608***	39.393***
	(6.091)	(7.82)	(10.475)	(2.084)	(3.109)	(4.34)	(4.579)	(5.381)	(6.813)
Demoted from Line	-23.039***	-45.169***	-42.083***	-4.848	-14.677***	-11.093*	-18.191***	-30.492***	-30.99***
	(7.091)	(10.206)	(13.879)	(3.131)	(4.756)	(6.116)	(4.646)	(6.308)	(8.789)
Week Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	817	687	596	817	687	596	817	687	596
R-Squared	0.448	0.502	0.493	0.264	0.321	0.326	0.478	0.524	0.53
R-Squared Adj	0.431	0.48	0.469	0.254	0.307	0.31	0.461	0.501	0.504
F-Stat (p-value)	0	0	0	0	0	0	0	0	0

Note 1: *** p<0.01, ** p<0.05, * p<0.1
Note 2: Robust standard errors in ( )
Note 3: First Differences Estimator

6.4   The Role of Rank in the VoD System

During the experiment, some movies were sometimes exogenously and randomly swapped and thus shown in fake ranks. This variability allows us to study whether a movie placed in a fake rank sells differently from a true movie placed at that rank. To do so, we estimate equation 5 (which resembles equation 3):

$$leases_{rt} = \lambda + \beta_1 log(MovieAge_{rt}) + \beta_2 NumberofMenus_{rt} + \beta_3 Price_{rt} + \beta_4 IMDbRating_{rt} +$$

$$\beta_5 Promoted \times TreatedWithin_{rt} + \beta_6 Demoted \times TreatedWithin_{rt} +$$

$$\beta_7 Promoted \times TreatedBetween_{rt} + \beta_7 Demoted \times TreatedBetween_{rt} +$$

$$+WeekDummies + RankDummies + GenreDummies + YearReleaseDummies + \epsilon_{rt} \ (5)$$

This regression allows us to compare the number of leases obtained by treated and control movies at each rank. *promoted* (*demoted*) indicates a movie that was promoted (demoted) to a fake rank. $TreatedBetween$ indicates whether a rank manipulation entails moving a movie from the catalog to the new menu or vice-versa. Therefore, the four types of manipulations included in this regression constitute a partition of the space of possible manipulations and thus their coefficients should be interpreted relative to our control movies.

Table 8 shows the results obtained. The first three columns in this table show the effect of rank manipulations on the number of leases and the last column shows the effect of rank manipulations on the number of likes. A movie that is demoted to a fake rank within the new menu sells 27.7% more than a true movie at that rank. Consumers are still able to spot high quality movies even if they have been shifted to the right on the

TV screen under the new menu. This is true for both standard and premium subscribers, though the result is less statistically significant for the former. These results suggest that subscribers use additional information besides rank to decide which movies to watch. We provide more detail on this hypothesis below. Conversely, a movie that is promoted to a fake rank within the new menu sells 15.9% less than a true movie at that rank. However, this result is weaker than the effect of demotions within the new menu. Both its magnitude and its statistical significance are lower. In fact, this effect is only statistically significant for premium subscribers.

Table 8: The Effect of Promotions and Demotions on Sales Relative to Movies at True Ranks.

Subscribers	All	Leases Standard	Premium	Likes Premium
Variables	$leases_{rt}$	$leases_{rt}$	$leases_{rt}$	$likes_{rt}$
(Intercept)	63.134***	0.986	62.148***	45.615***
	(18.904)	(10.935)	(10.906)	(5.808)
Promoted × Treated Within	-12.184*	-2.789	-9.396*	-7.614***
	(7.308)	(3.221)	(5.108)	(2.584)
Demoted × Treated Within	22.348***	7.756*	14.592***	6.955***
	(7.034)	(4.501)	(4.21)	(2.511)
Promoted × Treated Between	-4.331	1.09	-5.421	-7.853**
	(8.566)	(3.391)	(5.832)	(3.156)
Demoted × Treated Between	5.686	4.35*	1.336	1.241
	(4.448)	(2.451)	(2.524)	(1.051)
Log(Movie Age)	-3.103***	-1.814***	-1.289**	-0.305
	(1.178)	(0.689)	(0.633)	(0.275)
Number of Menus	4.000*	3.891***	0.109	1.276**
	(2.106)	(1.207)	(1.063)	(0.531)
Price	-0.005	-0.006	0.002	0.01*
	(0.026)	(0.016)	(0.013)	(0.006)
IMDb Rating	3.541**	1.005	2.536**	1.112*
	(1.429)	(0.717)	(1.096)	(0.576)
Week Dummies	Yes	Yes	Yes	Yes
Rank Dummies	Yes	Yes	Yes	Yes
Genre Dummies	Yes	Yes	Yes	Yes
Year Release Dummies	Yes	Yes	Yes	Yes
Observations	1001	1001	1001	1001
R-Squared	0.759	0.631	0.775	0.777
R-Squared Adj	0.697	0.58	0.713	0.714
F-Stat (p-value)	0	0	0	0

Note 1: Robust standard errors in ( ); **Note 2:** *** p<0.01, ** p<0.05, * p<0.1

The last column in Table 8 shows the effect of promotions and demotions on the number of *likes* obtained. A movie promoted to a fake rank receives 33.1% fewer *likes* than a true movie at that rank, and a movie demoted to a fake rank receives 30.1% more *likes* than a true movie at that rank. This result suggests that, over time, manipulated movies are likely to come back to their true ranks. We explore this hypothesis in more detail below. In addition, note that the coefficients for the effects of manipulations interacted with $TreatedBetween$ are statistically insignificant, which means that search costs dominate the effect of manipulations. If anything, standard subscribers lease demoted movies more than true movies, suggesting that standard subscribers are more willing to search for movies to watch than premium subscribers are.

6.5   The Role of Outside Information

In this section we test whether outside information about the movies shown at OurTelco's VoD system mediates the effect of promoting and demoting movies. We use the number of IMDb votes as a proxy for how well movies are known to consumers in general. The number of IMDb votes indicates how may people evaluated a movie irrespective of the rating provided. Figure 11 shows that there is significant variation in the number of IMDb votes across movies in our sample. This is not surprising given the well established super star effect in the movie industry whereby a disproportionate amount of attention is concentrated on a small number of movies [Elberse and Oberholzer-Gee, 2006]. In addition, both IMDb votes and IMDb ratings are similar between control and treated movies. We hypothesize that the movies in OurTelco's VoD system that have more outside information are less sensitive to the exogenous random manipulations introduced in this experiment. This would be consistent with the findings in [Tucker and Zhang, 2011]

showing that products with broader appeal are less likely to benefit from the popularity they obtain at the specific platforms where they are sold. [Salganik and Watts, 2008] also report similar results, but their measure of appeal is endogenous to the population of subjects used in their experiment.
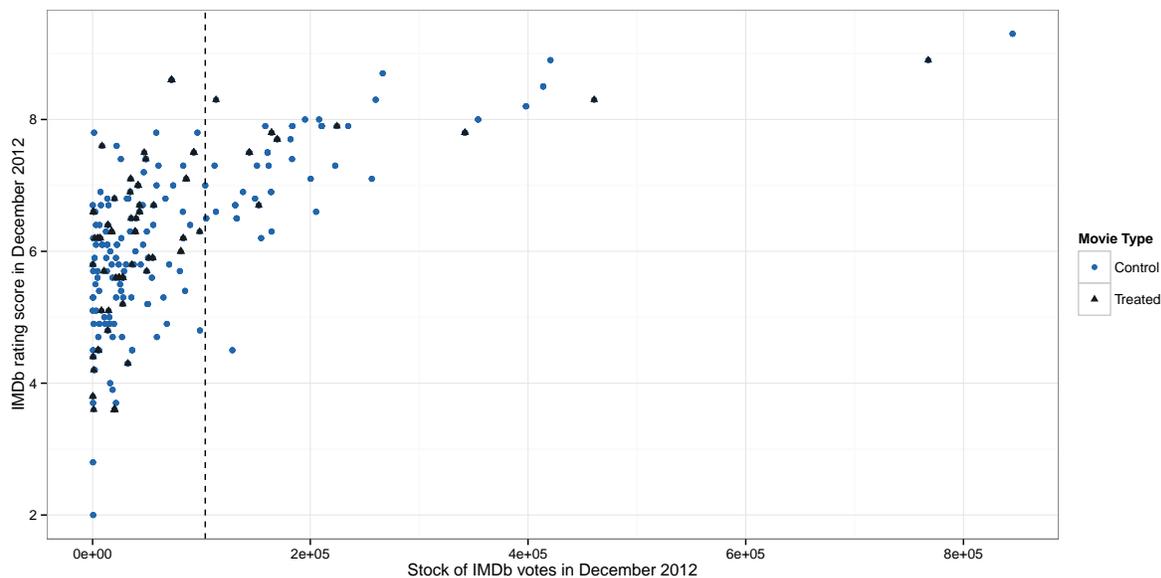

Figure 11: IMDb votes and ratings across movies in our sample.

We classify each movie in our sample according to the number of IMDb votes it received prior to December of 2012. We define a dummy variable called $Top25IMDbVotes$ to indicate whether a movie is in the top quartile of the distribution of IMDb votes in our sample. We estimate equation 4 adding an interaction term between $RankManipulation$ and this dummy variable. In this regression, the interaction term captures the difference in the effect of rank manipulations for movies in the top quartile of the distribution of IMDb votes relative to the effect on all the other movies in our sample that were also manipulated. Table 9 presents the resulting estimates. The effect of the interaction between $RankManipulation$ and $Top25IMDbVotes$ is negative and statistically significant,

confirming our hypothesis. The effect of a rank manipulation per rank manipulated is 84% lower for movies in the top quartile of the distribution of the IMDb votes relative to movies in the other quartiles, indicating that better known movies are less susceptible to manipulations.

Table 9: The role of IMDb Votes on the Effect of Rank manipulations on Leases.

	Subscribers	All	Standard	Premium
	Model	FD	FD	FD
	Variables	$leases_{it}$	$leases_{it}$	$leases_{it}$
	(Intercept)	-5.698*	-2.723*	-2.975*
		(3.092)	(1.647)	(1.802)
	Log(Movie Age)	-11.932**	-11.807***	-0.125
		(5.624)	(3.665)	(2.78)
	Number of Menus	12.346***	5.749***	6.597***
		(3.268)	(1.682)	(1.833)
	Treated	1.193	0.323	0.87
		(2.901)	(1.037)	(2.479)
	True Rank	-0.674	0.115	-0.789**
		(0.756)	(0.559)	(0.322)
Treated Within × Rank Manipulation		3.031***	0.591*	2.44***
		(0.51)	(0.304)	(0.345)
Treated Within × Rank Manipulation × Top 25 IMDb Votes		-2.547**	-1.001*	-1.546**
		(1.133)	(0.552)	(0.678)
	Promoted to Line	36.439***	9.416***	27.023***
		(6.031)	(2.087)	(4.532)
	Demoted from Line	-23.461***	-5.014*	-18.447***
		(6.736)	(3.025)	(4.439)
	Week Dummies	Yes	Yes	Yes
	Observations	817	817	817
	R-Squared	0.452	0.267	0.482
	R-Squared Adj	0.435	0.257	0.463
	F-Stat (p-value)	0	0	0

Note 1: Robust standard errors in ();
Note 2: *** p<0.01, ** p<0.05, * p<0.1;
Note 3: First Differences Estimator

6.6 The Effect of Price

While price is static in this experiment we can still observe whether it moderates the effect of manipulations. To do so, we interact price with the size of our manipulations.

We create two dummy variables: (1) $Price > \$3.99$, which identifies movies in the top half of the price distribution; (2) $Price = \$5.19$, which identifies movies in the top quartile of the price distribution. Table 10 shows the results obtained for all consumers and for standard and premium consumers separately. As expected, users seem to follow more the number of likes for more expensive movies, in particular premium consumers who can issue likes and see the number of likes in the VoD system.

Table 10: The Mediating Effect of Price on the Effect of Rank Manipulations

	Dependent variable:			
	$leases_{it}$			
	Standard	Standard	Premium	Premium
	(1)	(2)	(3)	(4)
Rank Manipulation × (Price > US3.89)	0.614*		2.195***	
	(0.357)		(0.581)	
Rank Manipulation × (Price = US5.19)		0.506		1.907***
		(0.356)		(0.585)
Rank Manipulation	0.070	0.154	0.658	0.887**
	(0.259)	(0.257)	(0.403)	(0.406)
Promoted to Line	13.480***	13.480***	31.630***	31.620***
	(1.930)	(1.931)	(3.003)	(3.002)
Demoted from Line	−8.444***	−8.430***	−19.990***	−19.950***
	(2.424)	(2.422)	(3.328)	(3.326)
Week Dummies	Yes	Yes	Yes	Yes
Observations	817	817	817	817
R^2	0.215	0.214	0.466	0.461
Adjusted R^2	0.208	0.207	0.450	0.446

Note: *p<0.1; **p<0.05; ***p<0.01
Robust standard errors in ()
First differences estimator

6.7 Convergence to True Ranks and Effect on Sales

The last column in Table 8 shows that promoted movies obtain fewer likes than the movies they displace and demoted movies obtain more likes than the movies they displace. Unfortunately, we are unable to observe long-term effects in our experiment because swaps

were reverted at the end of every false week. Yet, these results provide suggestive evidence that manipulated movies are likely to return to their true ranks as long as movies under the new menu are periodically ordered according to the number of likes received.

The panels in Figure 12 illustrate how the sales of manipulated movies converge to their pre-treatment levels over time. The horizontal axes represent time relative to the moment of treatment. The vertical axes represent, for a particular time t in the horizontal axis, the average number of weekly leases across all movies in our sample that were t weeks away from their treatment date. On the top of each panel we indicate over how many movies each average is computed.[3] All movies tend to sell less over time before treatment mostly due to aging.

The panel on the left shows that promoted movies sell significantly more immediately after treatment. However, their sales reduce significantly with time, matching the level of sales they might have had had they not been promoted in roughly 4 weeks. Counterfactual sales are obtained from the sales trends exhibited by control movies. The panel on the right shows that demoted movies sell significantly less immediately after treatment but their sales increase significantly with time to roughly match the counterfactual level of sales also in 4 weeks.

---

[3]Note that some movies were treated more than once. To avoid confounding subsequent treatments on the same movie, in this analysis we only include the first treatment of a movie and its data up to a second treatment, or up to the end of the panel if the movie was only treated once.

Figure 12: Convergence of Sales to Pre-treatment Levels for Demoted Movies (on the left) and for Promoted Movies (on the rigth).
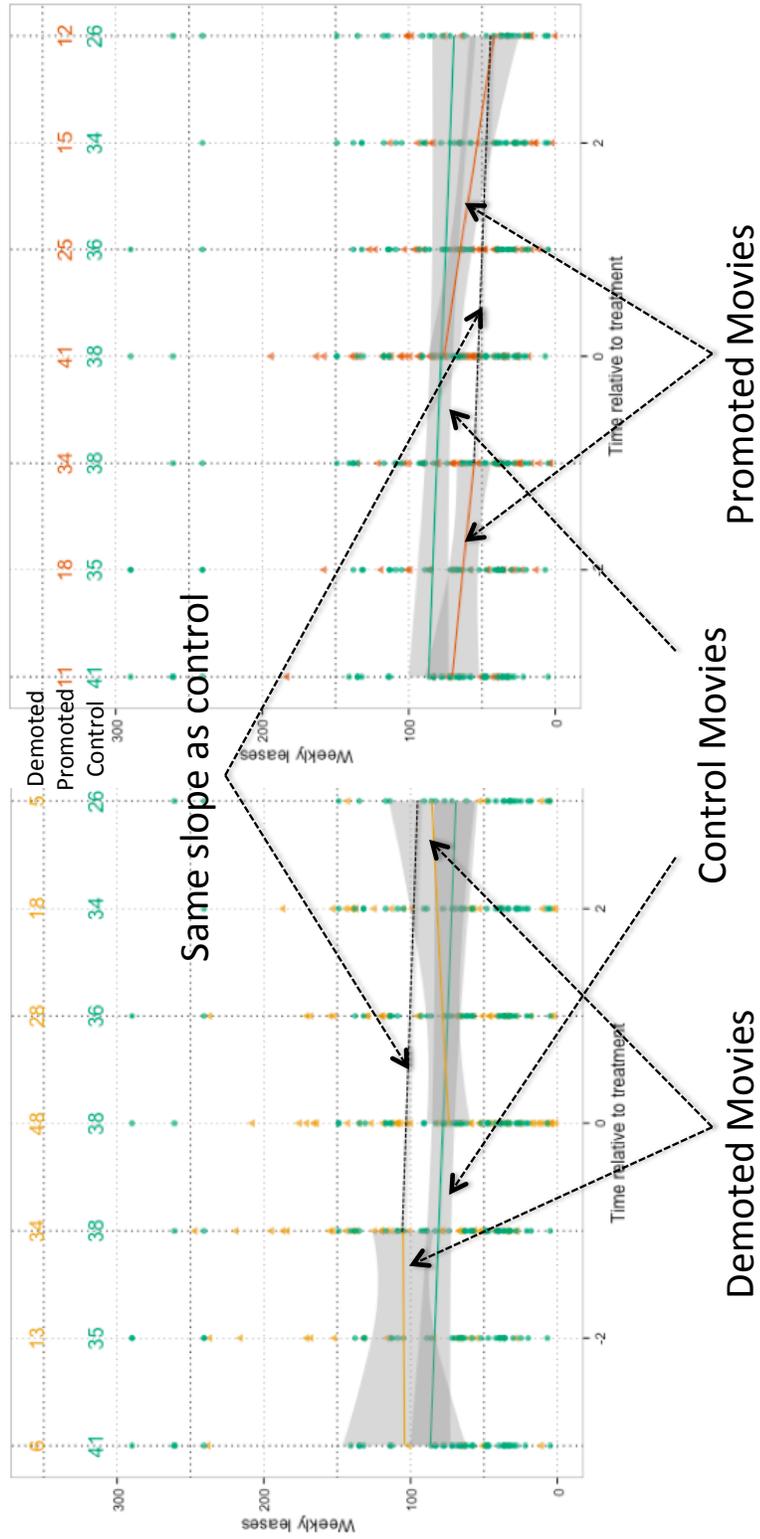
This suggests that the self-fulfilling prophecies reported in the literature do not seem to hold on in our setting. While promoted (demoted) movies initially sell (more) less, these biases seem to correct themselves over time and do not lead to the long term harding effects observed in other papers in the literature.

As discussed before, there are several differences between our study and the studies in the prior literature that may be driving this difference in results. One notable difference is that the consumers in our study likely have more outside information about the products they are purchasing (i.e., widely promoted movies) than did consumers in some of the other settings exhibiting "herding effects" reported in the literature (e.g., obscure music, wedding vendors, and breaking news stories). The increased availability of outside information in our context may cause our consumers to rely less heavily on the information about the products available in this specific marketplace (i.e. the number of likes in the VoD system) than other studies in the literature where outside information about products may have been less available to consumers. However, it is also possible that the differences in results could be driven by other differences between our setting and the settings studied in the prior literature (e.g., the nature of the user interface, the cultural or product context, the number of products under review). As we note below, these differences deserve further study.

6.8 Economic Impact of Within Swaps

We run an analysis at the swap level to estimate their economic impact. Each pair of ranks X and Y becomes an observation (X and Y vary between 1 and 15). In some weeks, at random, the movies in these ranks were swapped. This constitutes a treated pair of movies at these ranks. In the remaining weeks the pair of movies at these ranks is used as control.

Table 11: Effect of Swaps Within the New Menu on Sales.

	Dependent variable:		
	$leases_{it}$		
	All	Standard	Premium
	(1)	(2)	(3)
Treated Movie Pair	23.284**	9.896**	13.387**
	(9.950)	(4.817)	(6.610)
Constant	233.374***	92.099***	141.275***
	(16.230)	(7.857)	(10.782)
Week Dummies	Yes	Yes	Yes
Swap Dummies	Yes	Yes	Yes
Observations	348	348	348
R^2	0.706	0.589	0.706
Adjusted R^2	0.659	0.525	0.660
Residual Std. Error (df = 300)	46.002	22.269	30.561
F Statistic (df = 47; 300)	15.291***	9.159***	15.330***

Note: *p<0.1; **p<0.05; ***p<0.01
Includes only swaps within the new menu

We add the sales of the movies in ranks X and Y and compare them between treated and control weeks. We add both week and swap dummies to our regressions. Tables 11 and 12 show the results obtained for sales and revenues, respectively. We observe that swaps within the new menu increase both sales and revenues for both standard and premium consumers. According to Column (1) in Table 12, on average, swapping two movies increases their revenues by roughly 22.6%.

In addition, we also look at the effect of swaps on how much consumers like movies. Similarly to the approach above, in this analysis we compare the number of likes per lease obtained by movies in ranks X and Y when the movies in these ranks were and were not swapped. Table 13 shows that swaps reduce the number of likes per lease. Column (2) in this table shows that a swap within the new menu reduces this ratio by roughly 15.7%. Note that the analysis of likes per lease refers only to premium consumers, which are the ones that can issue likes and see the number of likes. These results provide suggestive

Table 12: Effect of Swaps Within the New Menu on Revenues.

	Dependent variable:		
	$log(revenue_{it})$		
	All	Standard	Premium
	(1)	(2)	(3)
Treated Movie Pair	0.226***	0.244***	0.205**
	(0.081)	(0.091)	(0.084)
Constant	7.102***	6.124***	6.563***
	(0.133)	(0.149)	(0.136)
Week Dummies	Yes	Yes	Yes
Swap Dummies	Yes	Yes	Yes
Observations	348	348	348
R^2	0.668	0.615	0.671
Adjusted R^2	0.617	0.555	0.619
Residual Std. Error (df = 300)	0.376	0.421	0.387
F Statistic (df = 47; 300)	12.869***	10.211***	13.009***

Note: *p<0.1; **p<0.05; ***p<0.01
Includes only swaps within the new menu

evidence that, in the short run, OurTelco is better off in the presence of movie swaps but consumers seem to be worse off. Therefore, and in the long run, fewer likes for the movies leased in this VoD system may turn into fewer sales as consumers lose trust in the VoD recommendations. Hence, OurTelco is unlikely to be able to use within swaps to increased revenues.

7 Conclusions

While user-generated product information has become increasingly prevalent in online markets, there are relatively few empirical studies that analyze the impact of user-generated signals of product quality in real-world settings. We attempt to fill this gap in the literature by partnering with a major telecommunications provider who implemented an online experiment to determine the role that *likes* play in Video-on-Demand (VoD)

Table 13: Effect of Swaps Within the New Menu on Likes.

	Dependent variable:	
	week likes/lease	log(week likes/lease)
	(1)	(2)
Treated Movie Pair	−0.087*	−0.157**
	(0.045)	(0.067)
Constant	0.773***	−0.275**
	(0.073)	(0.110)
Week Dummies	Yes	Yes
Swap Dummies	Yes	Yes
Observations	348	348
R^2	0.496	0.551
Adjusted R^2	0.417	0.481
Residual Std. Error (df = 300)	0.207	0.310
F Statistic (df = 47; 300)	6.284***	7.844***

Note: *p<0.1; **p<0.05; ***p<0.01
Includes only swaps within the new menu

movie sales. The VoD system of this provider was used during 24 weeks in 2012 as an experimental setting. As part of this experiment, a new menu titled "The Most Popular Movies During the Past Weeks" was introduced in the *Highlights Section* of this provider's VoD system. Movies with more *likes* were shown farthest to the left on the TV screen. During this experiment, movies were primarily placed in their true rank and shown along with their true number of *likes*. However, the position of some movies was manipulated such that a random set of movies were swapped and displayed our of order and with a fake number of *likes*. This randomization allows us to disentangle *likes* from unobserved perceived quality to identify the effect of the former on sales.

Our main finding is that sales and ratings of movies in our sample seem to be robust to artificial manipulations and do not exhibit strong "herding effects" seen in the prior literature. While, on average, promoted (demoted) movies still sell more (less) 3.5%, we find that a movie artificially promoted to a fake slot sells 15.9% less and receives 33.1%

fewer likes than a "true" movie in that slot, on average. Likewise, a movie artificially demoted to a fake slot sells 27.7% more and receives 30.1% more likes than a "true" movie in that slot, on average. Together, these results suggest that at least in our setting manipulated movies will move back to their "true" location relatively quickly (in our setting, within 1 month). We also show that better-known movies are less sensitive to manipulations (return to their "true" location more quickly) than are other movies.

One explanation for this finding is that consumers in our empirical setting have relatively more "outside information" about the quality of the products being evaluated (i.e., widely promoted movies) than did consumers in many of the other settings reported in the literature (e.g., obscure songs, breaking news stories, wedding vendors). While tentative, and perhaps context specific, this explanation would suggest that potential bias from herding effects could be muted in many "real world" marketplace settings where consumers have outside information about the true quality of the products they are evaluating. We also acknowledge that, contrary to previous studies, in our setting consumers need to make explicit decisions that involve financial risks. The price to lease movies in the VoD system used in this experiment varied between $1.30 and $5.20. It might be the case that when consumers have to pay for goods their feedback attenuates social bias more accurately. In sum, our study provides some evidence that "real world" organic markets for well known costly goods do not exhibit strong "herding effects" and identifies novel avenues for future research.

In particular, we believe that future works that attempt to measure directly the impact of financial costs and outside information on herding behavior would be of particular relevance for firms and researchers alike. If our findings stem mostly from the involment of financial costs then this suggests that product markets where social bias has the potential to be more damaging are also those that will be less affected by it. On the other hand, if

our findings originate mostly from outside information about products, a natural policy to prevent bias is to increase information dissemination efforts in such markets.

Acknowledgement

References

S. Aral. Commentary–identifying social influence: A comment on opinion leadership and social contagion in new product diffusion. *Marketing Science*, 30(2):217–223, 2010. ISSN 0732-2399 1526-548X. doi: 10.1287/mksc.1100.0596.

Sinan Aral and Dylan Walker. Creating social contagion through viral product design: A randomized trial of peer influence in networks. *Management Science*, 57(9):1623–1639, 2011.

Sinan Aral, Lev Muchnik, and Arun Sundararajan. Distinguishing influence-based conta-

gion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences*, 106(51):21544–21549, 2009.

Ravi Bapna and Akhmed Umyarov. Do your online friends make you pay? a randomized field experiment in an online music social network. *Conditionally Accepted at Management Science*, 2014.

Erik Brynjolfsson, Yu Hu, and Michael D Smith. Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers. *Management Science*, 49(11):1580–1596, 2003.

Chrysanthos Dellarocas, Xiaoquan (Michael) Zhang, and Neveen F. Awad. Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive Marketing*, 21(4):23–45, January 2007. ISSN 10949968. doi: 10.1002/dir.20087.

Wenjing Duan, Bin Gu, and Andrew B. Whinston. Do online reviews matter? an empirical investigation of panel data. *Decision Support Systems*, 45(4):1007–1016, November 2008. ISSN 0167-9236. doi: 10.1016/j.dss.2008.04.001.

Anita Elberse and Felix Oberholzer-Gee. Superstars and underdogs: An examination of the long tail phenomenon in video sales. *Harvard Business School Working Paper Series*, September 2006.

Jehoshua Eliashberg and Steven M. Shugan. Film critics: Influencers or predictors? *Journal of Marketing*, 61(2):68, April 1997. ISSN 00222429. doi: 10.2307/1251831.

David Godes and Dina Mayzlin. Using online conversations to study word-of-mouth communication. *Marketing Science*, 23(4):545–560, 2004. ISSN 0732-2399.

Barry R. Litman. Predicting success of theatrical movies: An empirical study. *The Journal of Popular Culture*, 16(4):159175, 1983. ISSN 1540-5931. doi: 10.1111/j.0022-3840.1983.1604_159.x.

Yong Liu. Word of mouth for movies: Its dynamics and impact on box office revenue. *Journal of Marketing*, 70(3):pp. 74–89, 2006. ISSN 00222429.

Charles F. Manski. Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, 60(3):531–542, 1993.

Lev Muchnik, Sinan Aral, and Sean J. Taylor. Social influence bias: A randomized experiment. *Science*, 341(6146):647–651, August 2013. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1240466. PMID: 23929980.

Phillip Nelson. Information and consumer behavior. *The Journal of Political Economy*, pages 311–329, 1970.

David A. Reinstein and Christopher M. Snyder. The influence of expert reviews on consumer demand for experience goods: A case study of movie critics. *The Journal of Industrial Economics*, 53(1):2751, 2005. ISSN 1467-6451. doi: 10.1111/j.0022-1821.2005.00244.x.

Matthew J. Salganik and Duncan J. Watts. Leading the herd astray: An experimental study of self-fulfilling prophecies in an artificial cultural market. *Social Psychology Quarterly*, 71(4):338–355, December 2008. ISSN 0190-2725, 1939-8999. doi: 10.1177/019027250807100404.

Matthew J Salganik, Peter Sheridan Dodds, and Duncan J Watts. Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, 311(5762): 854–856, February 2006. ISSN 1095-9203. doi: 10.1126/science.1121066.

C.R. Shalizi and A.C. Thomas. Homophily and contagion are generically confounded in observational social network studies. *Sociological Methods & Research*, 40(2):211, 2011. ISSN 0049-1241.

Catherine Tucker and Juanjuan Zhang. How does popularity information affect choices? a field experiment. *Management Science*, 57(5):828–842, May 2011. ISSN 0025-1909, 1526-5501. doi: 10.1287/mnsc.1110.1312.

W. Timothy Wallace, Alan Seigerman, and Morris B. Holbrook. The role of actors and actresses in the success of films: how much is a movie star worth? *Journal of Cultural Economics*, 17(1):1–27, 1993. ISSN 0885-2545. doi: 10.1007/BF00820765.

Xiaoquan Zhang and Chrysanthos Dellarocas. The lord of the ratings: Is a movie's fate is influenced by reviews? *ICIS 2006 Proceedings*, December 2006.