

Online Exploration, Content Choice & Filter Bubbles: An Experiment

Preliminary – Please do not cite or circulate

Sagit Bar-Gill (*MIT*)

Neil Gandal (*Tel Aviv University and CEPR*)

September 16, 2016

Abstract: In this experiment, we create an online search environment where users can control topical search and popularity considerations separately. Users explore the TED Talks collection in our simplified environment, and choose a talk to watch. Users' clickstreams are recorded, such that their complete search sequence and subsequent choice are fully accounted for. We are interested in whether there are filter bubbles. In our setting, (I) relatively little exploration in the search process and (II) sorting talks by popularity (before choosing a talk to view) represent filter bubbles. We find that high levels of sociability and experience with the TED environment are associated with filter bubbles. Opinion leadership, on the other hand, is associated with more exploration; such users are also less likely to sort by popularity before choosing the talk to view.

Keywords: Content exploration, online search, filter bubbles.

We are grateful to the Foerder Institute for Economic Research for financial support. We are especially grateful to Ori Heffetz for comments and suggestions that significantly improved the paper. We also thank seminar participants at Tel Aviv University, the "Israel IO Day" conference at Hebrew University and the Conference in Toulouse for helpful comments and suggestions.

1 Introduction

Online search for content has become central in our everyday lives. Content search and exploration increases knowledge and shapes our opinions; furthermore, our online exploration processes affect our entertainment choices as well as our consumption patterns, both on and offline. These effects strengthen over time, as the Internet is becoming an inseparable part of our daily lives. As we become more dependent on the Internet, online search algorithms continue to evolve – constantly tweaked and refined to provide a higher degree of personalization, i.e., search results that match perceived individual tastes.¹

Increasing personalization of our online environments has led to growing concerns of *filter bubbles*. The term, coined by Eli Pariser and discussed in his 2011 book, describes how personalization results in biased exposure to content, such that online searches yield information and opinions that are in line with users’ current viewpoints, rather than providing balanced and objective information.²

In this paper, we study online exploration and choice in a search environment that allows us to analyze differences in use of topic and popularity information across individuals. We are primarily interested in whether there are filter bubbles. In our setting, (I) relatively little exploration in the search process and (II) choosing to sort talks by popularity represent filter bubbles.

We are interested in what user characteristics may lead to filter bubbles. Hence, we examine the relationship between users’ personal and social characteristics, exploration patterns, and subsequent content choices. We additionally examine whether opinion leaders or influencers exhibited different search patterns compared to individuals who are not opinion leaders.

¹ Google first introduced personalized search as a beta test in 2004, and as a non-beta service in 2005, for users signed in to their Google account. Since 2009, all Google searches are personalized, even when users do not log in to their accounts.

² These ideas relate to earlier theoretical work by Brynjolfsson and Van Alstyne (2005), which highlights the role of individual preferences for broad vs. specialized knowledge in determining integration levels.

1.1 The Experiment

We address these issues by designing an Internet application and using it to conduct an online experiment.³ We developed a unique and straightforward online search environment, where topic and popularity sorting are separately controlled.⁴ Users in our experiment explore the TED⁵ Talks collection of short videos using two buttons - *Category* and *Popularity*, with the stated goal of finding a talk they would like to watch.⁶

The *Category* button represents topical search, and produces a list of talks in the chosen category, in RANDOM order. Hitting the *Popularity* button sorts the displayed search results by popularity (or for all talks when no category is chosen) from most to least popular. Users may click each of the buttons as many times as they like, and these clicks constitute a *search sequence* with individual weights on topic and popularity (further details are provided in section 3).

We record each user's click-stream in our environment, and therefore each user's search sequence and subsequent choice are fully accounted for. Variables of interest constructed based on users' clickstream include whether or not the user chose a talk from a sorted or unsorted list, within a specific category or not, and the scroll depth required to reach the chosen talk, to name but a few. Users were further asked to report some demographic information, answered a few self-report items regarding general and online sociability, and completed an opinion leadership questionnaire (developed by Flynn et al., 1996).

An interesting feature of the environment is random assignment of users to one of two groups: (1) Popularity information (in terms of number of views) does NOT appear alongside the results

³ We ran the experiment using the Amazon Mechanical Turk (AMT) platform. AMT workers have been shown to produce similar results as both laboratory subjects and subjects in other online domains (see survey by Mason and Suri 2012). This means that AMT offers the possibility of running affordable high-quality experiments. Several other such platforms now exist. Peer et al. (2015) identified two additional crowdfunding platforms, CrowdFlower and MicroWorkers that may provide alternatives to AMT.

⁴ It took nearly two years to design and build the application and to get all of the bugs out. For example, there is not perfect compatibility across browsers. Now that the application is working, there are additional research projects we plan to undertake.

⁵ TED is a non-profit organization devoted to spreading ideas, usually in the form of short talks (18 minutes or less). TED stands for Technology, Entertainment and Design, though TED talks today may cover any topic (more at <https://www.ted.com/about/>).

⁶ Users had to watch at least five minutes of the lecture in order to receive payment for participation.

following each category click; (2) Popularity information appears alongside the results following each category click (results are shown in random order for both groups following each click on category).

1.2 Findings

As discussed, we are especially interested in whether there are filter bubbles – and if so – what are the user characteristics associated with filter bubbles. Our analysis is centered around the effects of sociability, opinion leadership, TED experience, and popularity information provision on exploration and choice characteristics.

We conjecture that highly social individuals will tend to “follow the herd” in their exploration and choice by relying more on popularity considerations than topic-based search, compared to those who report lower sociability. We expect opinion leaders to exhibit a stronger topic preference in both exploration and choice, and invest more effort in search and content selection. We also expect users with TED experience to rely more on popularity considerations. Thus we expect that highly social individuals and users with TED experience will be more susceptible to filter bubbles, while opinion leaders will be less susceptible to filter bubbles. By and large, our results are consistent with these hypotheses, although there are several unexpected findings.

For content exploration, we find that highly social individuals (both male and female) show a weaker preference for exploration by topic, and rely more heavily on popularity sorting, compared to individuals who report lower sociability.⁷

Male users with TED experience are also more likely to rely more heavily on popularity sorting, but this effect does not exist for women, suggesting that filter bubbles based on experience are less likely for women.

Opinion leadership is found to affect exploration patterns, but only for men in our sample. Specifically, male opinion leaders exhibit a stronger preference for exploration by topic of interest, and invest more effort in search, performing more search clicks than non-opinion leaders.

⁷ This effect is found for men’s reported sociability, and for women’s reported number of Facebook friends, which serves as another proxy for sociability in our setting.

For content choice, we find that highly social individuals are less likely to choose from unsorted category-specific results. We also find that users with TED experience are more likely to choose talks based on popularity; this suggests that familiarity with the process and content makes it less likely that users explore and choose content that intrinsically interests them. Hence, sociability and familiarity with the process and content leads to filter bubbles.

We further find that male opinion leaders' choice is characterized by a stronger topic preference, as represented by choice from unsorted topical results. The effect of popularity information provision is statistically significant only for male opinion leaders, who are very responsive to the added information on number of views, and tend to complement it by more sorting. Interestingly, providing this information did not affect the choice of non-opinion leaders.

2 Literature Review

We relate to the nascent literature on online exploration behavior, to the herding and observational-learning literature, and to the literatures on opinion leaders and influence.

Early studies of online exploration examined online shopping behavior, distinguishing between planned purchasers who conduct directed search for a product, and hedonic browsers who explore the online product space (e.g., Moe 2003; Moe and Fader 2004). Subsequent research on online exploration has examined the effects of reviews, recommendation algorithms, and user generated content on content exploration efficiency and success (e.g., Goldenberg, Oestreicher-Singer, and Reichman 2012; Wang, Goh, and Lu 2013). We add to these works by decomposing exploration into popularity and topic-based components, and examining heterogeneity in the use of popularity and topical information in content exploration.

The notion of employing popularity information in individual decision making relates to early models of observational learning and herding behavior (Banerjee 1992; Bikhchandani, Hirshleifer, and Welch 1992). More recently, evidence of the effects of popularity rankings on choices and market outcomes have been demonstrated in several online and offline settings (Salganik, Dodds, and Watts 2006; Salganik and Watts 2008; Cai, Chen, and Fang 2009; Tucker and Zhang 2011; Muchnik, Aral and Taylor 2013), and heterogeneity in herding behavior has been shown for investors' trading decisions (e.g., Merli and Roger 2012, Pentland 2013).

In our experiment, opinion leaders exhibit different exploration and choice patterns compared to non-leaders. Opinion leadership - the tendency of certain individuals to influence others' opinions or choices - has been studied in the psychology, sociology and marketing literatures since the early works of Lazarsfeld et al. (1948), Katz (1957), and Rogers (1962). More recently, work in marketing and social network research has examined the role of opinion leaders and influencers in new product diffusion and other processes of social contagion (e.g., Weimann 1994; Van den Bulte and Joshi 2007; Doumit et al. 2007; Tucker 2008; Goldenberg et al. 2009; Godes and Mayzlin 2009; Kratzer and Lettl 2009; Iyengar, Van den Bulte, and Valente 2011).

Related to this, a growing literature has focused on identifying social influence in online networks, in several behavioral contexts, and measuring its strength (e.g., Aral, Muchnik and Sundararajan 2009; Bakshy et al. 2012; Aral and Walker 2014; Bapna and Umyarov 2014). Recent papers distinguish between influence and susceptibility, demonstrating their differential effects in diffusion processes (e.g., Watts and Dodds 2007; Aral and Walker 2012). The literature has thus focused on the role of influencers in the propagation of behaviors, and on identifying influence in networks. We study a related yet distinct question – whether influencers and followers exhibit different content exploration and choice patterns, further examining the effects of sociability and information provision on these processes.

3 The Experimental Design

3.1 The Online Search Environment

Our search environment, named *TED-it*,⁸ allows users to browse the collection of TED-talks (roughly 1600 short videos) using two buttons – *Category* and *Popularity*.⁹ Users are instructed to search talks using these two buttons until they find a talk they would like to watch. They are further instructed to watch the chosen talk for at least five minutes, after which a *Sign Out* button becomes active, and must be clicked to receive payment for participation. The requirement to watch a video

⁸ The URL for TED-it is <http://ted-it.tau.ac.il/tedit/turk.php>.

⁹ The location of the buttons is randomized, such that Popularity appears on the right only for 50% of the users. This is to rule out possible location effects.

for at least five minutes has been shown to motivate non-trivial search activity in an early stage pilot.

Users are provided with explanations regarding the buttons' functionality, as follows. A click on *Popularity* sorts any list of search results according to their number of views on Youtube.com. If the first click is *Popularity*, this click produces a sorted list of all talks.

A click on *Category* produces a dropdown menu with 15 categories, from which the user may choose one. A choice of category produces a screen with search results, where talks appear in random order. An interesting feature of the environment is random assignment of users to one of two groups: (1) Popularity information (in terms of number of views) is NOT provided alongside the results following each category click; (2) Popularity information is provided alongside the results following each category click (results are shown in random order for both groups following each click on category).

Figure 1 presents a screenshot of a new search screen (prior to any clicking), and figure 2 presents a screenshot with search results appearing after a user randomly assigned to group (2) had chosen to search TED talks in the Entertainment category.



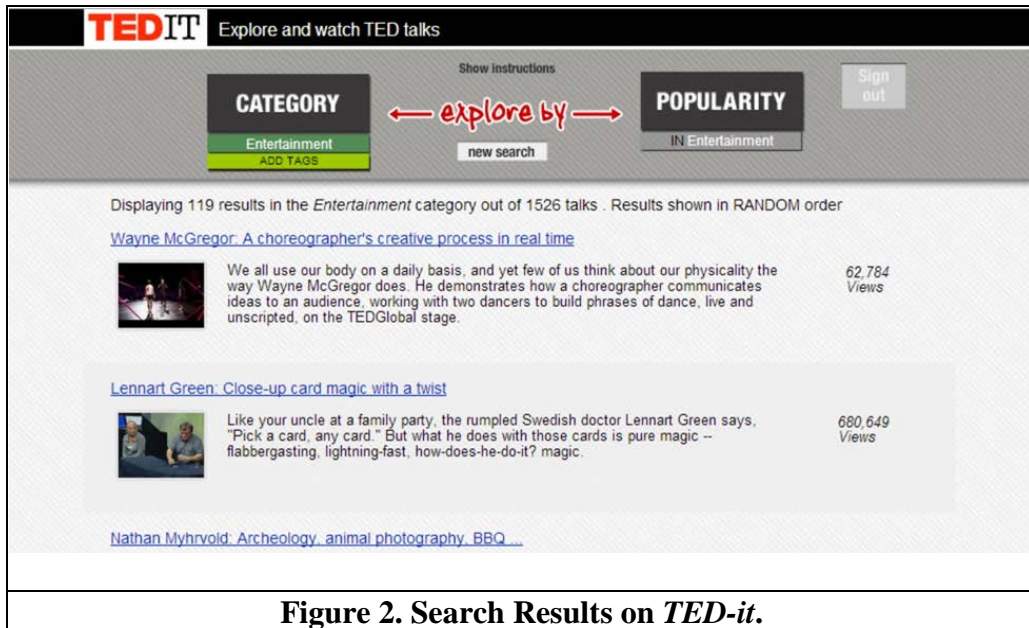


Figure 2. Search Results on TED-it.

Adding tags is possible after a category is chosen, and provides another layer of topical search. The option of adding tags appears as an extension of the *Category* button with the text *Add Tags*, after a category has been chosen. Clicking *Add Tags* produces a small pop-up window with roughly 20-40 tags (depending on the category), from which the user may select up to 3 tags. This allows the user to explore deeper within his chosen category.

Our experimental search environment will be replicated as a Facebook application, to allow us to examine the effects of a social media setting on content exploration and choice, in a planned extension of this work. The main difference in the Facebook version is that users on Facebook are *not required* to view a talk for any length of time, but rather to use the application as they wish. Therefore, a *Sign Out* button is not needed. The *Sign Out* button (irrelevant on Facebook) is replaced by a *Favorites* button, and talks are added to favorites by clicking an *Add to Favorites* button when viewing a talk. The TED-it Facebook application thus allows personal archiving of a user's favorite TED talks. All other functionalities and appearances are exactly the same except for the Facebook frame (which is out of our control), surrounding the TED-it environment. Figure 3 shows a search results screen in the Facebook application. (We do not have results regarding the use of the TED-it Facebook application at this stage.)



Figure 3. Search Results – TED-it on Facebook.

3.2 The Data Collected

The experiment was run on Amazon Mechanical Turk (AMT). The recruitment statement informed AMT workers that they were invited to participate in an experiment studying search behavior, and that payment for participation was 1.50 USD, paid via AMT.¹⁰ Consenting workers followed a link to the *TED-it* website from AMT, completed the task, and then received a code to paste back into AMT, to receive compensation.

On the *TED-it* website, workers answered several short questionnaires, either before or after the content exploration task (the timing of the questionnaires is randomly chosen for each user). The data collected includes demographics (gender, age, country, education level attained), as well as self-reported number of Facebook friends¹¹ and subjective level of sociability on a 1-5 scale (in response to “Do you consider yourself a social person? (1-not social, 5-very social)). Users further report whether or not they have previously watched a TED talk, and if their reply is positive, are asked to report roughly how many talks they’ve watched (by marking one of three options: 1-3 / 4-6/ 7 or more).

¹⁰ As noted, AMT workers have been shown to produce similar results as both laboratory subjects and subjects in other online domains (see survey by Mason and Suri 2012).

¹¹ Users without a Facebook account report 0.

In addition, users respond to a six-item Opinion Leadership questionnaire (adapted by Goldsmith et al. 2003 from Flynn et al. 1996). Responses to each item are on a 1-7 Likert scale, and their summation yields an opinion leadership score between 6 and 42 (see 7.2 in the appendix).

Exploration behavior is collected based on users' click-stream. Each and every user click on *TED-it* buttons is recorded and saved to our data base. A search sequence is a string specifying the buttons clicked on by the user. For example, the search sequence "c,c,p" means that the first two clicks were on two different categories, followed by a *Popularity* click to sort the list of talks in the second category by popularity, before making his/her choice (from within that category). In such a case, the user chose a video from a sorted category. On the other hand, a clickstream pattern of "c,p,c" means the user explored one category, sorted the TED talks in that category by popularity, and then explored an additional category, from which he/she chose a video to watch. In this case, the user chose a video from an unsorted list in the second category he/she explored.¹²

Exploration activity is organized by user and session, and includes: the user's sequence of clicks, identifiers of the talks viewed, number of seconds viewed for each talk, the scroll depth required to reach the chosen talk (i.e., its location in the list of search results), and more.

The result is a rich dataset, where we have for each user: (1) Demographics; (2) Social characteristics; (3) Detailed exploration and choice pattern in our environment.

3.3 Variables Used in the Analysis

The variables used in the analysis can be broadly classified into four groups. We discuss these variables by group; descriptive statistics and correlations among the variables are in Appendix 7.1.

1. Demographics, experience with TED talks, and popularity information provision:
 - (1) *Age* –users must be 18 or older to participate;
 - (2) *Gender*
 - (3) *HigherEd* – Dummy variable that takes on the value one if the user has at least some college education, and zero otherwise.

¹² For those who clicked *only* on popularity, some clicked several times, possibly due to congestion on our server that slowed down our website. Since this does not change the search results, it makes sense to regard redundant popularity clicks as a single click. That is, the sequence "p,p,p" is treated the same as "p".

(4) *PreviousTed* – Dummy variable that takes on the value one if the user has previously watched a TED video, and zero otherwise.

(5) *Information* - Dummy variable that takes on the value one for users who were randomly assigned to see popularity information when clicking on category, and zero otherwise.¹³

2. Social characteristics:

(1) *Social* – Response on a 1-5 scale to the subjective question “Do you consider yourself a social person? (1-not social, 5-very social)”.

(2) *FBfriends* – Self-reported number of Facebook friends (users with no Facebook account report 0).

(3) *OpinionLeader* – Dummy variable that takes on the value one, for subjects whose opinion leadership score is in the top quartile (see appendix 7.2 for details on the calculation of the opinion leadership score).

¹³ Users either receive or do not receive view count information following their first click on category. Those who receive this information following their first category click continue to receive it with every subsequent category click. Those who do not receive popularity information following their first click on category do not receive popularity information following subsequent category clicks.

3. Exploration variables:

- (1) *CatFirst* – Dummy variable that takes on the value one if the search sequence starts with a category click, and zero otherwise.
- (2) *Clicks* – Number of search clicks, where search clicks are clicks on *Popularity*, *Category*, *Tags*, or sampling a video within exploration, recorded as a *V* click (also referred to as *Total Number of Clicks*).
- (3) *percentC* – Share of *Category* clicks out of *Clicks* (also referred to as *Share of Category Clicks*).

4. Choice characteristics:

- (1) *ScrollDepth* – Location of the chosen talk within the list of search results, where 1 is the first entry at the top of the page. Note that the number of talks visible without scrolling changes according to screen size and is out of our control. We do not break down search results into pages, so, one could scroll down to the 100th talk or more (within the list of all talks, or in a large category).
- (2) *ChoiceUnsortedCat* – Dummy variable that takes on the value one if the user chose a talk from unsorted search results within a specific category, and zero otherwise.
- (3) *ChoiceSortedCat* – Dummy variable that takes on the value one if the user chose a talk from sorted search results within a specific category, and zero otherwise.
- (4) *ChoiceNonCat* – Dummy variable that takes on the value one if the user never clicks on category, i.e., clicks once on popularity to sort the results, and then chooses a talk; it takes on the value zero otherwise.

Clicks on category represent breadth of search, while clicks on popularity might represent a filter bubble.

3.4 Summary Information

The experiment was run using 1,851 AMT workers who followed a link to the *TED-it* website. These users are predominantly Americans (93%), their average age is 34, 56% are male, and 89% have at least some college education. Their average number of Facebook friends is approximately

232. Note that the average number of Facebook friends for users in the 25-34 age group is 360.¹⁴ The average user performs 0.65 *Popularity* clicks and 1.05 *Category* clicks. Fully 34% of users clicked only on *Popularity*, thereby sorting all talks, and choosing from this list.¹⁵

3.5 Informal Evidence of Filter Bubbles in Exploration: TED exposure

Does previous exposure to “TED Talks” lead users to behave differently. In the case content exploration 63% of men who had never seen TED talk picked “category” first, while 58% of men who had seen a TED Talk picked “category” first. There was no difference in the case of women.

In the case of content choice, 62% (63%) of men (of women) who had never seen TED talk chose the talk they viewed from unsorted results within a specific category, while 52% (55%) of men (women) who had seen a TED Talk chose the talk they viewed from unsorted results within a specific category. That is, those with TED experience were more likely to choose after sorting the talks by popularity.

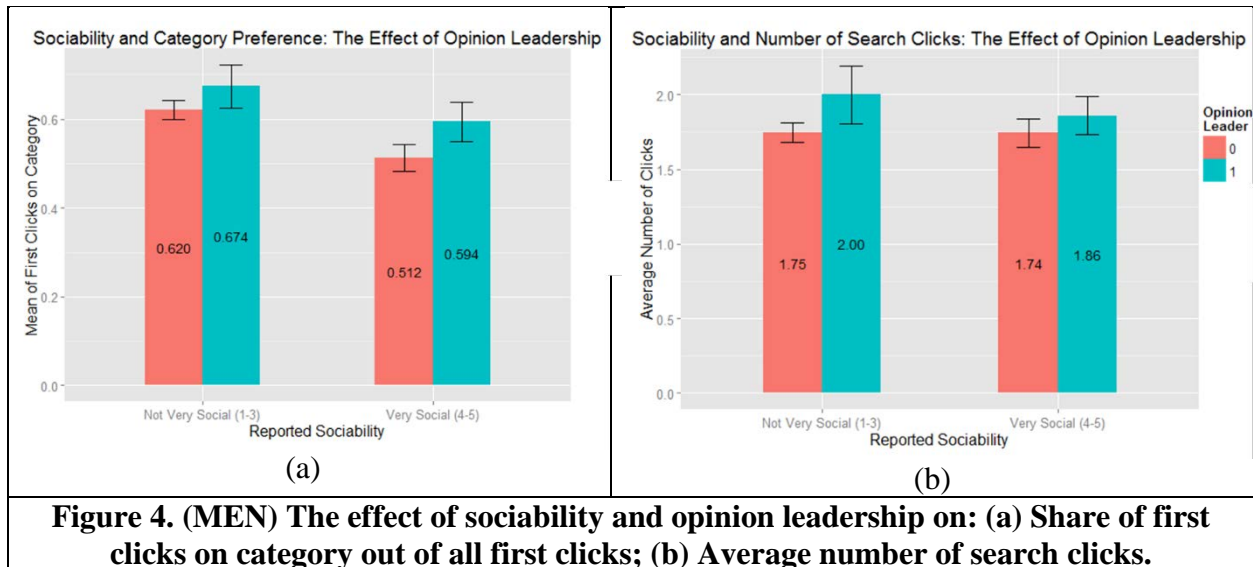
Of those who clicked on category first, 79% of males with previous TED experience who received information on popularity selected a TED-talk without sorting the talks by popularity, while 86% of the males with previous TED experience who did not receive information about popularity selected a talk without sorting them by popularity. Thus for men, information provision on popularity led more users with TED experience to sort and choose a talk by popularity. The information effect does not exist for those men without TED experience. Fully 90% of the men who chose category first selected a TED-talk without sorting the talks by popularity regardless of whether they received information about popularity. This suggests that men without previous TED experience were (I) more likely to choose a talk without sorting and (II) were not affected by popularity information. (This does not show up in the regression. The estimated coefficient on the interaction variable is not quite significant.) In the case of women, neither effect was present.

¹⁴ According to studies by Edison Research and Triton Digital in January and February 2014.

¹⁵ That is, for 34% of the users, the dummy variable ChoiceNonCat=1.

3.6 Informal Examination of Filter Bubbles in Exploration: Sociability & Opinion Leadership

Figure 4 depicts the effects of sociability and opinion leadership on the likelihood of choosing category first for men.¹⁶ For men, Figure 4 shows a positive correlation between opinion leadership and the likelihood of clicking first on category. This suggests a stronger category reliance in opinion leaders' exploration. Figure 4 further shows a negative correlation between sociability and a preference for category search.



3.7 How Did our Users Choose?

Not surprisingly, whether the user chose from unsorted categories, sorted categories, or sorted talks (from all talks) determined to a large extent the popularity of the talk chosen. If the user chose from:

¹⁶ Figure 4 is for men, since both the effects of opinion leadership and sociability seem to be significant on content exploration. This is confirmed by the regression results in Table 1.

“ChoiceUnsortedCat,” the mean views per talk (on the TED website) was 521,007.

“ChoiceSortedCat,” the mean views per talk (on the TED website) was 2,909,773.

“ChoiceNonCat,” the mean views per talk (on the TED website) was 5,032,500.

In the analysis, we will examine what characteristics led the user to reach these decisions.

3.8 What Did our Users Choose?

The talk chosen most often by the users in our experiment was Tony Robbins: “Why we do what we do?” The talk, which is one of the most popular on the TED website, had 8,063,277 views on the TED site, and 123 users in our experiment chose this talk in the “science and technology” category. The second most popular talk for our users was David Blaine: “How I held my breath for 17 minutes.” One hundred and fifteen of our users selected this popular talk, which had 5,398,543 views on the TED website. But talks chosen frequently by our users were not necessarily popular on the TED website. Forty-two individuals in our experiment chose the Ted talk: “Are games better than life?” The talk had 97,808 views on the TED website.

4 Analysis

We now formally study the effects of opinion leadership, sociability, previous TED experience and information provision on exploration patterns (section 4.1) and subsequent content choice (section 4.2). Inherent differences between men and women imply large heterogeneity in our sample, masking some of the effects we wish to study. Thus, we present the results separately for men and women, as some effects are persistent and significant only for one group.

4.1 Content Exploration

Previous TED experience and content exploration

As discussed in the descriptive data, users without TED experience were more likely to choose category first than those with TED experience. It also seems intuitive that users without TED experience will choose more category clicks than users with TED experience.

Opinion leadership and content exploration. Opinion leaders exert some influence on others in their social circle, and are characterized by an intrinsic motivation to influence their peers. Since already popular content decreases opinion leaders’ capacity to act as thought leaders, we expect

these individuals to seek out content based on their topical interests, to create new avenues for influence. This implies that, on average, opinion leaders will invest more effort in content search and display some category preference in our environment.

Sociability and content exploration. High reported sociability is associated with a tendency to follow the crowd, and therefore with stronger reliance on popularity information in search. This is quite intuitive, as highly social individuals are likely to consider their peers' opinions, and do not stray far from the herd. We consider the number of Facebook friends as another proxy for sociability, and expect it to have a similar effect.

The above leads to the following hypotheses, which we take to the data.

H1: Those without TED experience will exhibit a stronger preference for exploration by topic

H2: Opinion leaders will: (a) exhibit a stronger preference for exploration by topic of interest, and (b) will invest more effort in search, compared to non-opinion leaders

H3: Highly social individuals will show a weaker preference for exploration by topic

We run regressions using two different dependent variables that measure various aspects of content exploration: (1) *Category First* and (2) *Share of Category Clicks*. Regressions results are reported in Table 1 below.

In the case of men, we find support for H1. The estimated coefficient for the effect of *TED* experience is negative and statistically significant in both cases (1) and (2), suggesting that those with TED experience are less likely to click on “category first” and have a smaller share of clicks on category than those without TED experience.

In the case of men, we find some support for H2. The estimated coefficient for the effect of *OpinionLeader* is positive and statistically significant in case (1) and positive although not statistically significant in the case of (2). That is, opinion leaders are more likely to start their

exploration with a category click. This suggests a stronger preference for topic-based search, and higher search effort, for opinion leaders.¹⁷

We also find support for H3 in the case of men. The estimated coefficient on the variable *Social* is negative and statistically significant in both cases. That is, highly social individuals are less likely to explore by topic, and perform fewer clicks than individuals who report lower sociability.

For women, we find support only for H3. H1 and H2 are not supported, since the estimated coefficient on the variables *PreviousTED* and *OpinionLeader* are not statistically significant in either of the regressions.¹⁸ Regarding Facebook friends as a proxy for sociability, we find support for H3 for women in both specifications (1) and (2), where the more Facebook friends (*FBfriends*) a user has, the less she explores by topic.

Hence, we find evidence of filter bubbles for men in the case of content exploration. Both TED experience and “sociability” lead to filter bubbles in the sense that these characteristics are associated with less exploration by category and more reliability on popularity. Opinion leaders are less likely to be associated with filter bubbles. Our results suggest that women are less likely to “fall into” filter bubbles (regarding content exploration.)

The effect of age is also interesting. Younger individuals, both men and women, are less likely to choose category first, and click on category less than older individuals. This suggests that younger users may be more likely to fall into filter bubbles.

¹⁷ If we restrict attention to those MEN without TED experience, the estimated coefficient on opinion leadership is also positive and statistically significant as well (coefficient=0.12, t-stat=1.87) when the dependent variable is the share of category clicks. Also, when the dependent variable is *Category First*, for those without TED experience, the estimated coefficient on opinion leadership is larger (0.39 vs. 0.17) and more statistically significant (t-stat=1.92 for no TED experience vs. t-stat=1.67 overall.)

¹⁸ If we restrict attention to those women who did not have previous TED experience, then opinion leadership is positively associated with both (I) whether the first choice is category and (II) the share of category clicks and this effect is statistically significant.

Table 1: Effect of TED Experience, Opinion Leadership and Sociability on Content Exploration

	<i>Dependent variable:</i>			
	(Probit)	(OLS)	(Probit)	OLS
<i>Category First</i>	<i>Share of Category Clicks</i>	<i>Category First WOMEN</i>	<i>Share of Category Clicks WOMEN</i>	
	(1)	(2)	(1)	(2)
<i>OpinionLeader</i>	0.17* (1.67)	0.045 (1.34)	-0.31 (0.46)	0.017 (0.46)
<i>Social</i>	-0.088* (-1.71)	-0.033* (-1.88)	0.06 (1.10)	-0.026 (1.37)
<i>FBfriends</i>	-0.37 (-0.70)	-0.0047 (-0.25)	-0.14** (-2.51)	-0.50** (-2.53)
<i>PreviousTED</i>	-0.16* (-1.77)	-0.075** (-2.44)	-0.056 (-0.60)	-0.047 (-1.44)
<i>HigherEd</i>	0.12 (1.03)	0.058 (1.41)	-0.22 (-0.22)	-0.048 (-0.86)
<i>Age</i>	0.18*** (4.12)	.0068*** (4.69)	0.0073* (1.85)	.0027** (2.00)
<i>Constant</i>	0.10 (0.05)	0.44 (5.94)	0.41 (1.64)	0.60 (7.07)
Observations	1,034	1,034	812	812
R ²		0.04		0.02
Adjusted R ²		0.03		0.01

Note: * = p < 0.1; ** = p < 0.05; *** = p < 0.01 :
T-statistics in parentheses

4.2 Content Choice

We proceed to examine how opinion leadership, sociability, TED experience and information provision affect content choice.¹⁹ Recall that popularity information for search results in a chosen category (i.e., number of views for each of the randomly ordered talks) is provided for 50% of our users, randomly selected.

Outcome variables are the binary variables “choice from unsorted category” and “choice from sorted category.” We expect a connection between determinants of exploration and choice. Hence we expect that:

H4: Those with TED experience will be less likely to choose from unsorted categories.

H5: More sociable individuals will be less likely to choose from unsorted categories.

The interaction between opinion leaders and information provision

Those who assume the (informal) role of opinion leaders are likely to conduct more in-depth exploration, and to be comfortable with making choices that are solely topic based, independent of popularity considerations. In our setting, this would appear as a greater tendency to choose a talk from unsorted category results.

Providing information complicates the issue: As opinion leaders tend to search more, we expect they will be more likely to utilize information available to them in content exploration. Applied to our setting, we postulate that availability of popularity information for unsorted search results will affect opinion leaders more strongly than non-leaders. We expect that opinion leaders users who receive popularity information will exhibit an overall lower tendency to choose from unsorted results, compared to those who are not provided with this information.

Following the above discussion, we postulate that:

¹⁹ The effect of popularity information was not a significant factor in the exploration process. It does not affect whether the first click is on category, since it comes into play only after the first click. It was not significant regarding the share of clicks on category.

H6: When popularity information is not provided, opinion leaders' chosen talk is more likely to be selected from unsorted category results, compared to non-leaders.

H7: Popularity information provision may affect the chosen talk by: decreasing the tendency to choose from an unsorted list. These effects will be more pronounced for opinion leaders.

Figure 5 graphs the effects of popularity information provision and opinion leadership on the tendency to choose from unsorted category results, focusing on the subset of male subjects. The figure shows that when popularity information is not provided, opinion leaders' choice is more likely to be from an unsorted list, compared to non-opinion leaders.

The figure shows that information provision interacts with opinion leadership, such that, only for opinion leaders, popularity information leads to less scrolling and more sorting, before choosing a talk. For non-leaders, on the other hand, the effect of information provision appears to be not significant.

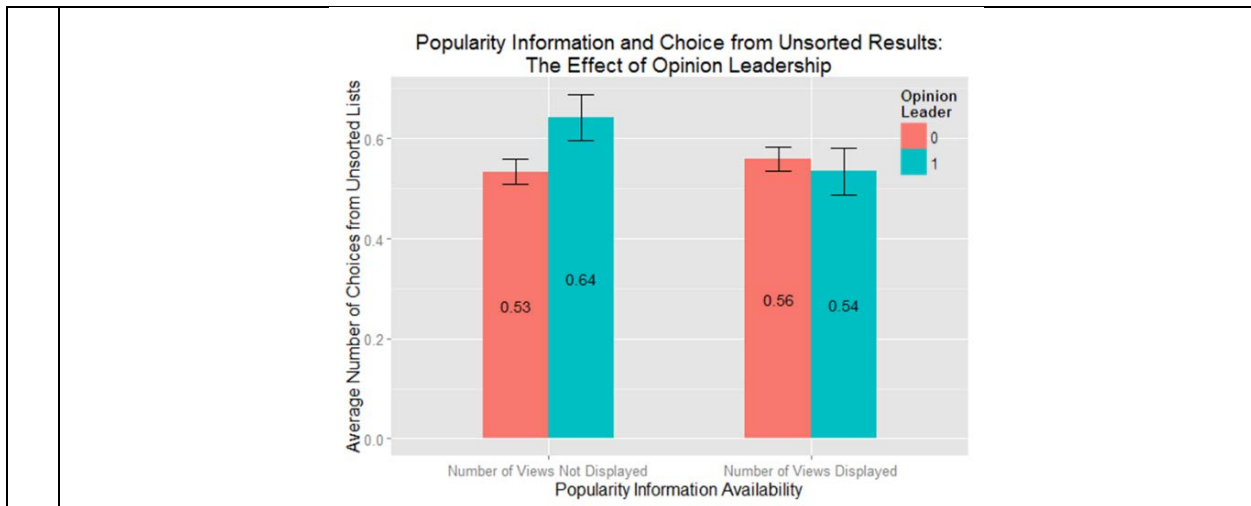


Figure 5. The effects of popularity information provision and opinion leadership on probability that user chooses from unsorted, category-specific, search results.

Regression results reported in table 2 confirm some of descriptive results from Figure 5. In table 2, we use two possible dependent variables: (1) *Choice from Unsorted Category*, and (2) *Choice from Sorted Category*.

For men, the effects *OpinionLeader* and the interaction between *Information* and *OpinionLeader* are both statistically significant in both specifications. Opinion leaders exhibit a stronger topic

preference, as represented by choice from unsorted category-specific results. This topic preference is weakened in the presence of popularity information, as opinion leaders are very responsive to information provision, and tend to complement it by more sorting. For women, we do not find any effects of information provision or opinion leadership.

In the case of reported sociability (*Social*), we find a negative and significant effect on topic-based choice for men (in specification (2)), while for women a similar effect exists for Facebook friends (*FBfriends*). This is in line with the preceding discussion on sociability and exploration, and highlights the connection between determinants of exploration and choice.

In the case of TED experience, we find that previous TED experience makes users less likely to choose from unsorted category lists. This holds for both men and women.²⁰ Younger individuals are less likely to choose from unsorted category lists for both men and women; the effect is stronger for men.

Hence, we find evidence of filter bubbles for both men and women in the case of content choice. Both TED experience and “sociability” lead to filter bubbles in the sense that these characteristics are associated with a greater likelihood to sort the talks before choosing. In the case of men, opinion leaders are less likely to be associated with filter bubbles.

²⁰ We also included an interaction variable between TED experience and information provision. Based on descriptive statistics, we might have expected that those with TED experience who received information would be less likely to choose from unsorted categories. The sign coefficient is as expected, but it is not statistically significant ($t=-1.42$.)

Table 2: Effect of TED Experience, Opinion Leadership, Sociability, and Information on Content Choice

	MEN		WOMEN	
	probit		Probit	
	<i>Choice from Unsorted Category</i>	<i>Choice from Sorted Category</i>	<i>Choice from Unsorted Category</i>	<i>Choice from Sorted Category</i>
	(1)	(2)	(1)	(2)
<i>OpinionLeader</i>	0.32** (2.30)	-0.39* (-1.72)	-0.00066 (-0.00)	-0.15 (-0.65)
<i>Information</i>	0.047 (0.52)	-0.056 (-0.46)	-0.086 (-0.84)	-0.15 (-1.03)
<i>Social</i>	-0.11** (-2.19)	0.023 (0.32)	0.069 (1.28)	0.014 (0.18)
<i>FBfriends</i>	0.015 (0.27)	-0.0070 (-0.09)	-0.12** (-2.04)	-0.013 (-0.15)
<i>PreviousTED</i>	-0.29*** (-3.19)	0.18 (1.34)	-0.17* (-1.84)	0.15 (1.06)
<i>HigherEd</i>	0.26** (2.12)	-0.37 (-0.22)	-0.21 (-1.31)	0.16 (0.62)
<i>Age</i>	0.018*** (4.22)	-0.0037 (-0.62)	0.0067* (1.71)	-0.0058 (-0.95)
<i>OpinionLeader* Information</i>	-0.32* (0.166)	0.54* (1.87)	0.12 (0.56)	0.00023 (0.00)
<i>Constant</i>	-0.21 (-0.97)	-1.33 (-4.36)	0.39 (1.54)	-1.35 (-3.61)
Observations	1,034	1,034	812	812

T-statistics in Parentheses

5 Conclusion

In this paper, we examined how TED experience, social characteristics and popularity information provision affect content exploration and choice patterns. We find that high reported sociability is associated with a strong reliance on popularity considerations in both exploration and choice. Thus,

we find that social individuals are more susceptible to filter bubbles, both in terms of content exploration and content choice. We also find that those with TED experience are more susceptible to filter bubbles, both in terms of content exploration and content choice. For men, opinion leaders are less susceptible to filter bubbles than other users. In the case of women, there is no difference between opinion leaders and other users.

Our results imply a relationship between social characteristics and online search which may be used to guide the design of search algorithms. Algorithms to date incorporate personalization methods based on users' search history and on similar individuals', or friends' previous choices. Our findings suggest that a users' sociability and opinion leadership levels may be further used in determining the leading results he is shown.

To the extent that we can identify excessive weighting of popularity considerations in search, it is possible to set policies that influence the mix of results displayed, to provide more balanced information in some environments. Such interventions are not expected to affect users' utility (at least in settings similar to ours), as we found no significant effects of sociability, opinion leadership, information, or the exploration process itself, on viewership length, which may be regarded as a proxy for enjoyment. Such policies may be used to alleviate concerns regarding biased exposure to content, or the *filter bubble*.

6 References

- Aral, S., Muchnik, L., Sundararajan, A., 2009. “Distinguishing Influence Based Contagion from Homophily Driven Diffusion in Dynamic Networks,” *Proceedings of the National Academy of Sciences (PNAS)*, Dec. 22, 106(51), pp. 21544-21549.
- Aral, S., Walker, D., 2011, “Identifying Social Influence in Networks Using Randomized Experiments,” *IEEE Intelligent Systems*, September-October, 26(5), pp. 91-96.
- Aral, S., & Walker, D., 2012, “Identifying Influential and Susceptible Members of Social Networks,” *Science*, July 20, pp. 337-341.
- Aral, S., and Walker, D., 2014, “Tie Strength, Embeddedness & Social Influence: A Large-Scale Networked Experiment,” *Management Science*, 60(6), pp. 1352-1370
- Bakshy, E., D. Eckles., R. Yan, I. Rosenn, 2012, “Social Influence in Social Advertising: Evidence from Field Experiments,” *Proceedings of the 13th ACM Conference on Electronic Commerce (EC '12)*, ACM, New York, NY, USA, pp. 146-161.
- Banerjee, A.V., 1992, “A Simple Model of Herd Behavior.” *Quarterly Journal of Economics*, 107(3), pp. 797–817.
- Bapna, R., A. Umyarov, 2015, “Do Your Online Friends Make You Pay? A Randomized Field Experiment in an Online Music Social Network,” *Management Science*, forthcoming.
- Bikhchandani, S., D. Hirshleifer, and I. Welch, 1992, “A Theory of Fads, Fashion, Custom, and Cultural Change in Informational Cascades,” *Journal of Political Economy*, 100(5), pp. 992–1026.
- Brynjolfsson, E., M. Van Alstyne, 2005, “Global Village or Cyber-Balkans? Modeling and Measuring the Integration of Electronic Communities,” *Management Science*, 51(6), pp. 851-868.
- Cai, H., Y. Chen, and H. Fang, 2009, “Observational Learning: Evidence from a Randomized Natural Field Experiment,” *American Economic Review*, 99(3), pp. 864-82.
- Doumit, G., M. Gattellari, J. Grimshaw, M. A. O’Brien. 2007. “Local opinion leaders: Effects on professional practice and healthcare outcomes,” *Cochrane Database Syst. Rev.* 1 CD000125.
- Godes, D., D. Mayzlin, 2009, “Firm-created word-of-mouth communication: Evidence from a field test,” *Marketing Science*, 28(4), pp. 721–739.
- Goldenberg, J., Oestreicher-Singer, G., and Reichman, S. 2012. “The Quest for Content: How User-Generated Links Can Facilitate Online Exploration,” *Journal of Marketing Research* (49:4), pp. 452-468.
- Goldenberg, J., S. Han, D. R. Lehmann, J. W. Hong. 2009, “The role of hubs in the adoption process,” *Journal of Marketing* 73(2), pp. 1–13.

- Goldsmith, R.E., Flynn, L.R., Goldsmith, E.B. 2003. "Innovative Consumers and Market Mavens," *Journal of Marketing Theory and Practice* 11(4), pp. 54-64.
- Iyengar, R., C. Van den Bulte, T.W. Valente, 2011, "Opinion Leadership and Social Contagion in New Product Diffusion," *Marketing Science* 30(2), pp. 195-212.
- Katz, E., 1957, "The two-step flow of communication: An up-to-date report on a hypothesis," *Public Opinion Quarterly*, 21, pp. 61-78.
- Kratzer, J., C. Lettl, 2009, "Distinctive roles of lead users and opinion leaders in the social networks of schoolchildren," *Journal of Consumer Research*, 36(4), pp. 646-659.
- Lazarsfeld, P.F., Berelson, B. Gaudet, H., 1948, "The People's Choice: How the Voter makes up his Mind in a presidential Campaign," Columbia University Press, New York, NY.
- Mason, W. and Suri, S. 2012. "Conducting Behavioral Research on Amazon's Mechanical Turk," *Behavioral Research Methods* (44:1), pp. 1-23.
- Merli, M. and T. Roger, 2013, "What Drives the Herding Behavior of individual Investors?" *Finance*, Vol. 34(3).
- Moe, W. 2003. "Buying, Searching, or Browsing: Differentiating between Online Shoppers using In-Store Navigational Click-Stream," *Journal of Consumer Psychology* (13:1,2), pp. 29-39.
- Moe, W., P. Fader, 2004, "Dynamic Conversion Behavior at E-Commerce Sites," *Management Science* (50:3), pp. 326-335.
- Moldovan, S., S. Oreg, 2014, "Opinion Leadership as a Multi-Dimensional Trait," *Working Paper*.
- Muchnik, L., S. Aral, S. Taylor, 2013, "Social Influence Bias: A Randomized Experiment," *Science*, 341(6146), pp. 647-651.
- Netemeyer, R.G., Bearden, W.O. and Teel, J.E. 1992. "Consumer Susceptibility to Interpersonal Influence and Attributional Sensitivity," *Psychology and Marketing* (9), pp. 379-394.
- Pariser, E., 2012 "The Filter Bubble: What the Internet is hiding from you," Penguin Books.
- Pentland, A., 2013, "Beyond the echo chamber," *Harvard Business Review*, 91(11), pp. 80-86.
- Peer, E., Samat, C., Brandimarte, KL., and a. Acquisti, "Beyond the Turk: An Empirical Comparison of Alternative Platforms for Crowdsourcing Online Behavioral Research," 2015 mimeo, available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2594183.
- Racherla, P., R.A. King, 2012, "What We Know and Don't Know About Online Word-Of-Mouth: A Systematic Review and Synthesis of the Literature," *Working Paper*, <http://dx.doi.org/10.2139/ssrn.2187040>.
- Rogers, E. 1962, "Diffusion of innovations," New York: Free Press.
- Salganik, M.J., D.J. Watts, 2008, "Leading the herd astray: An experimental study of self-fulfilling prophecies in an artificial cultural market," *Social Psychology Quarterly*, 71 (4), pp. 338-355.

- Salganik, Matthew J., Peter S. Dodds, and Duncan J. Watts. 2006. "Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market." *Science*, 311 (February), pp. 854–56.
- Tucker, C., J. Zhang, 2011, "How does popularity information affect choices? A field experiment," *Management Science*, 57 (5), pp. 828-842.
- Tucker, C., 2008, "Identifying formal and informal influence in technology adoption with network externalities," *Management Science*, 54(12), pp. 2024–2038.
- Valente, T. W., P. Pumpuang, 2007, "Identifying opinion leaders to promote behavior change," *Health Education and Behavior*, 34(6), pp. 881-896.
- Van den Bulte, C., Y. V. Joshi, 2007, "New product diffusion with influentials and imitators," *Marketing Science*, 26(3), pp. 400–421.
- Wang, Q., K.Y. Goh, X. Lu (2013), "How Does User-Generated Content Influence Consumers' New Product Exploration? An Empirical Analysis of Consumer Search and Choice Behaviors," *Working paper*.
- Watts, Duncan J. and Peter Sheridan Dodds, 2007, "Influentials, Networks, and Public Opinion Formation", *Journal of Consumer Research*, 34, pp. 441 – 458.
- Weimann, Gabriel, 1994, "The influential: people who influence people," State University of New York Press, Albany, NY.

7 Appendix

7.1 Descriptive Statistics

Descriptive Statistics - All Users

Statistic	N	Mean	St. Dev.	Min	Max
OpinionLeader	1,850	24.92	7.17	6	42
Social	1,851	3.09	1.16	1	5
FBfriends	1,851	232.06	369.26	0	5,000
PreviousTED	1,851	0.68	0.47	0	1
HigherEd	1,851	0.89	0.31	0	1
age	1,851	33.70	10.65	18	80
CatFirst	1,851	0.59	0.49	0	1
percentC	1,846	0.57	0.44	0.00	1.00
Clicks	1,851	1.77	1.55	1	16
ScrollDepth	1,684	10.61	29.27	1	828
ChoiceUnsorted	1,851	0.57	0.50	0	1
ChoiceSortedCat	1,851	0.09	0.29	0	1
ChoiceNonCat	1,851	0.34	0.47	0	1

Descriptive Statistics by Gender

Statistic	Men					Women				
	N	Mean	St. Dev.	Min	Max	N	Mean	St. Dev.	Min	Max
OpinionLeader	1,037	25.00	6.91	6	42	813	24.81	7.50	6	42
Social	1,038	3.08	1.18	1	5	813	3.11	1.15	1	5
FBfriends	1,038	223.42	380.36	0	5,000	813	243.08	354.51	0	3,500
PreviousTED	1,038	0.72	0.45	0	1	813	0.62	0.49	0	1
HigherEd	1,038	0.88	0.33	0	1	813	0.91	0.28	0	1
Age	1,038	31.94	9.57	18	80	813	35.95	11.51	18	76
CatFirst	1,038	0.59	0.49	0	1	813	0.59	0.49	0	1
percentC	1,034	0.56	0.44	0.00	1.00	812	0.57	0.44	0.00	1.00
Clicks	1,038	1.78	1.55	1	16	813	1.75	1.54	1	14
ScrollDepth	941	11.06	36.88	1	828	743	10.04	14.81	1	119
ChoiceUnsorted	1,038	0.55	0.50	0	1	813	0.58	0.49	0	1
ChoiceSortedCat	1,038	0.09	0.29	0	1	813	0.08	0.28	0	1
ChoiceNonCat	1,038	0.35	0.48	0	1	813	0.33	0.47	0	1

Correlations

	<i>Opinion Leader</i>	<i>Social</i>	<i>FB friends</i>	<i>Previous TED</i>	<i>Higher Ed</i>	<i>age</i>	<i>Cat First</i>	<i>Percent C</i>	<i>Clicks</i>	<i>Scroll Depth</i>	<i>Choice Non Cat</i>	<i>Choice Sorted Cat</i>
<i>OpinionLeader</i>	1	0.19	0.12	-0.01	0.02	0.02	0.01	0.02	0.05	0.01	0.03	-0.02
<i>Social</i>	0.19	1	0.25	-0.11	0.08	0.04	-0.06	-0.05	-0.04	-0.02	-0.03	0.002
<i>FBfriends</i>	0.12	0.25	1	0.002	0.08	-0.14	-0.07	-0.06	0.01	-0.02	-0.05	0.02
<i>PreviousTED</i>	-0.01	-0.11	0.002	1	0.06	-0.11	-0.03	-0.06	-0.02	0.05	-0.08	0.04
<i>HigherEd</i>	0.02	0.08	0.08	0.06	1	0.04	-0.004	0.004	-0.02	0.03	0.01	-0.001
<i>Age</i>	0.02	0.04	-0.14	-0.11	0.04	1	0.10	0.12	-0.04	0.04	0.10	-0.02
<i>CatFirst</i>	0.01	-0.06	-0.07	-0.03	-0.004	0.10	1	0.89	0.14	0.04	0.71	0.14
<i>percentC</i>	0.02	-0.05	-0.06	-0.06	0.004	0.12	0.89	1	0.10	0.04	0.89	-0.06
<i>Clicks</i>	0.05	-0.04	0.01	-0.02	-0.02	-0.04	0.14	0.10	1	-0.03	0.03	0.36
<i>ScrollDepth</i>	0.01	-0.02	-0.02	0.05	0.03	0.04	0.04	0.04	-0.03	1	0.03	-0.01
<i>ChoiceUnsorted</i>	0.03	-0.03	-0.05	-0.08	0.01	0.10	0.71	0.89	0.03	0.03	1	-0.36
<i>ChoiceSortedCat</i>	-0.02	0.002	0.02	0.04	-0.001	-0.02	0.14	-0.06	0.36	-0.01	-0.36	1

7.2 Opinion Leadership Questionnaire

Our study employs the Opinion Leadership scale adapted in Goldsmith et al. (2003) from Flynn et al. (1996). Users respond to the following opinion leadership questionnaire:

Please indicate your level of agreement with the following statements: (1-strongly disagree, 7-strongly agree) -

1. I often persuade other people to buy the products that I like.
2. Other people rarely come to me for advice about choosing what to buy.
3. People that I know pick their purchases based on what I have told them.
4. My opinion on what to buy seems not to count with other people.
5. I often influence people's opinions about buying things.
6. When they choose products to buy, other people do not turn to me for advice.

Note that 1, 3, and 5 are positive items (positively correlated with opinion leadership), while 2, 4 and 6 are negative items.

Let $OL(i) \in \{1..7\}$ denote the response to question i . The opinion leadership score for each respondent is given by: $OL \equiv OL(1) + (8 - OL(2)) + OL(3) + (8 - OL(4)) + OL(5) + (8 - OL(6))$.

The maximum value for OL is 42 and obtains when the user strongly agrees with OL(1), OL(3), and OL(5) and strongly disagrees with OL(2), OL(4), and OL(6). The minimum value for OL is 6 and obtains when the user strongly disagrees with OL(1), OL(3), and OL(5) and strongly agrees with OL(2), OL(4), and OL(6).