THE PINCHAS SAPIR CENTER
FOR DEVELOPMENT
Tel Aviv University

המרכז לפיתוח ע"ש פנחס ספיר
ליד אוניברסיטת תל אביב
"עמותה רשומה"
580010221

THE PINHAS SAPIR CENTER FOR DEVELOPMENT
TEL AVIV UNIVERSITY

Online Exploration when Search Topic and Popularity Ranking Are Decoupled: Insights on Echo Chambers

Sagit Bar-Gill and Neil Gandal

Discussion Paper No. 5-2019

February 2019

אוניברסיטת תל-אביב, תל אביב יפו 6997801 | טל': 03-6409202 | פקס: 03-6405815
Tel Aviv University, Tel Aviv Yafo 6997801 Israel | Tel. 972-3-6409202 | Fax 972-3-6405815 | email: sapircen@tauex.tau.ac.il

Online Exploration when Search Topic and Popularity Ranking Are Decoupled: Insights on Echo Chambers

Sagit Bar-Gill *(TAU)*

Neil Gandal *(Tel Aviv University and CEPR)*

November 2018

Abstract: Personalized search algorithms produce results that are both topically relevant and ranked by their general popularity and individual fit to users' previous searches and choices. New choices from such tailored lists feed back into the algorithms, over time creating content echo chambers, where content exposure is increasingly biased toward users' and their friends' interests and views. We create an online search environment for TED Talks, where topic and popularity are separately controlled, and study the relationship between users' characteristics and their reliance on own interests vs. crowd-based popularity sorting in content exploration. In topic-based searches, we randomly block/show popularity information to examine its impact on the tendency to explore. We find that high levels of sociability, previous experience with similar content, and a younger age are associated with an increased susceptibility to echo chambers, represented by little to no exploration and popularity sorting prior to content choice. Opinion leaders may alleviate echo chambers in their social circles as they conduct more topic-based exploration and exhibit lower popularity reliance. Showing popularity information increases opinion leaders' popularity sorting, but does not impact non-leaders' exploration. Our findings identify users' echo chamber risk factors, and suggest that reducing the salience of popularity information may contribute to more balanced content exposure facilitated by opinion leaders.

Keywords: Content exploration, online search, online content choice, decision making, opinion leadership, echo chamber, online behavior.

We are grateful to the Foerder Institute for Economic Research and The Sapir Center for Development 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 Israeli IO Day at Hebrew University, the IDEI Conference on Internet/Software (Toulouse), and the New York City Media Workshop for helpful comments and suggestions.

1 Introduction

Just as sound reverberates in an enclosed space, so do opinions and information in our social circles. Since “birds of a feather flock together,” these social circles consist mostly of like-minded individuals, making it harder for dissenting views to permeate our discourse, and creating an echo of our own information and views (1). This is the echo chamber phenomenon that has been drawing growing concern in post-election USA, post-Brexit UK, and pretty much everywhere in our increasingly connected world (2–4). A natural result of our homophilous social choices, echo chambers have always been around. But as our social networking continues its shift online, they seem to be getting worse (2, 4, 5).

Who is responsible for our online echo chambers – ourselves or filtering and ranking algorithms beyond our control? Filtering and ranking algorithms are implemented in many of our online environments, creating much-debated personalization of search results (6–9), and holding the potential to influence individual behaviors and aggregate outcomes (10). Yet the human factor carries substantial weight. Bakshy and colleagues at Facebook research labs (11) find that individuals’ own choices are more to blame for their echo chambers than Facebook’s newsfeed ranking algorithm. In line with this, recent Pew research (12) finds that 83% of social media users ignore political posts they disagree with, with 39% of users taking action such as changing settings, blocking or unfriending someone, because of posts related to politics. With our past choices and clicks feeding into personalization algorithms (2, 9), user actions and algorithms mutually reinforce, to create deeper echo chambers in the long run.

We study the role of user characteristics in creating biased content exposure. We ask what types of users are most likely to get caught in deeper content echo chambers, and what is the role of popularity information provision in facilitating echo chambers. Our online experiment examines the relationship between users’ personal and social characteristics, and their content exploration and choice patterns in a simplified search environment where users separately control two search dimensions – topic and popularity. In this setting, we find that susceptibility to echo chambers is well proxied by: (I) conducting relatively little exploration in the search process, and (II) relying on popularity in content choice. Intuitively, little exploration along with reliance on the crowd’s tastes will lead, over time, to reduced exposure to diverse content, thereby facilitating the creation of an echo chamber.

We examine the relationship between users' personal and social characteristics, exploration patterns, and subsequent content choices. Opinion leaders – individuals who influence others' opinions or choices (13–20)– are likely to affect their peers' content exposure. We thus specifically study their search patterns, comparing influencers to individuals who are not opinion leaders.

We further explore the effect of displaying popularity information, represented by Youtube view counts, on exploration and choice. Popularity information has been previously found to impact choices (21–25), and we extend its study by examining its impact on content search processes and its differential impact on opinion leaders and non-leaders.

2 The Experimental Setting

Our search environment, named *TED-it*,[1] allows users to explore the collection of TED[2] talks using two buttons – *Category* and *Popularity*.[3] The experimental task is to search *TED-it* for a talk using these buttons, and watch it for at least five minutes[4], after which a *Sign Out* button becomes active, and must be clicked to receive payment for participation. Figure 1 presents a screenshot of the first screen on TED-it (prior to any clicking), that includes instructions on the buttons' functionality. These instructions do not appear in following screens once the user begins exploration, but remained available at any point by clicking "Show Instructions" (see Figure 2). Explanations regarding the task and the search environment are therefore clearly provided (see SI for further details and screenshots), and summarized below.

---

[1] The URL for TED-it is http://ted-it.tau.ac.il/tedit/turk.php.

[2] 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/). The collection of TED talks comprised approximately 1600 short videos when the experiment took place.

[3] 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.

[4] This requirement has been shown to motivate non-trivial search activity in an early stage pilot.

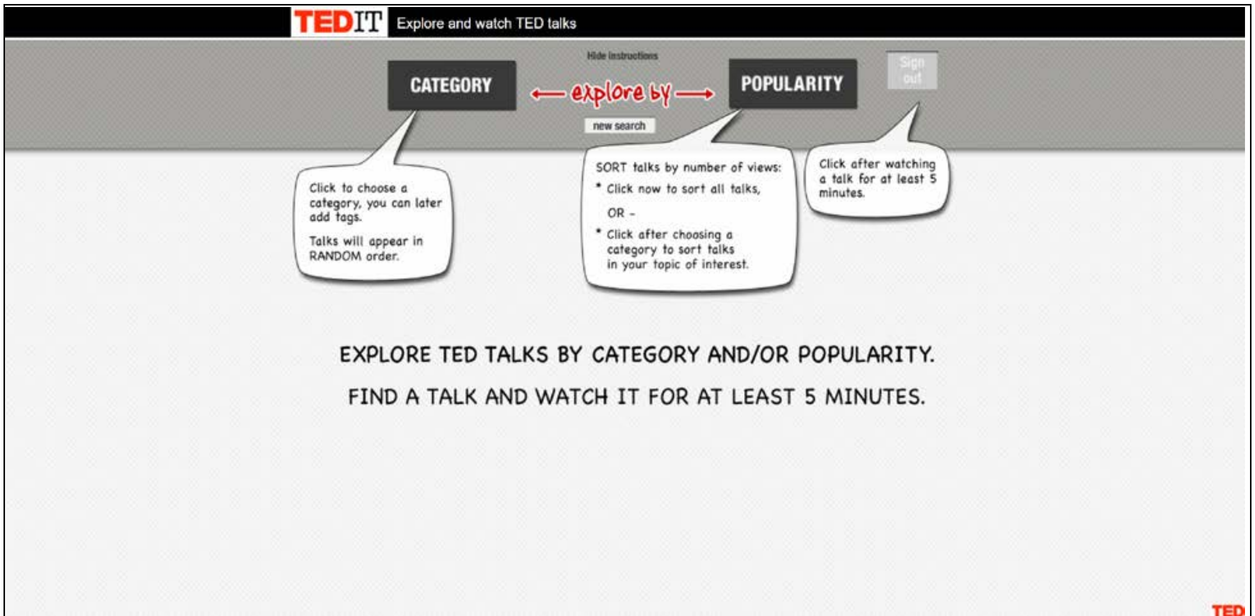

Figure 1. The TED-it Start Screen. Explanation on the experiment task and buttons' functionality are clearly provided.

The *Category* button enables topical search, and opens a dropdown list of fifteen categories. Choosing one of these produces a list of talks in the chosen category, in *random* order. Clicking the *Popularity* button sorts the displayed search results by their number of views on Youtube from most to least popular. When *Popularity* is clicked first (i.e., before a category has been chosen), the button produces a sorted list of all talks. Users may click each of the buttons as many times as they like, creating a *search sequence* with individual weights on topic and popularity. Figure 2 shows a screenshot with search results appearing after a user clicked on *Category* and chose the Entertainment category.

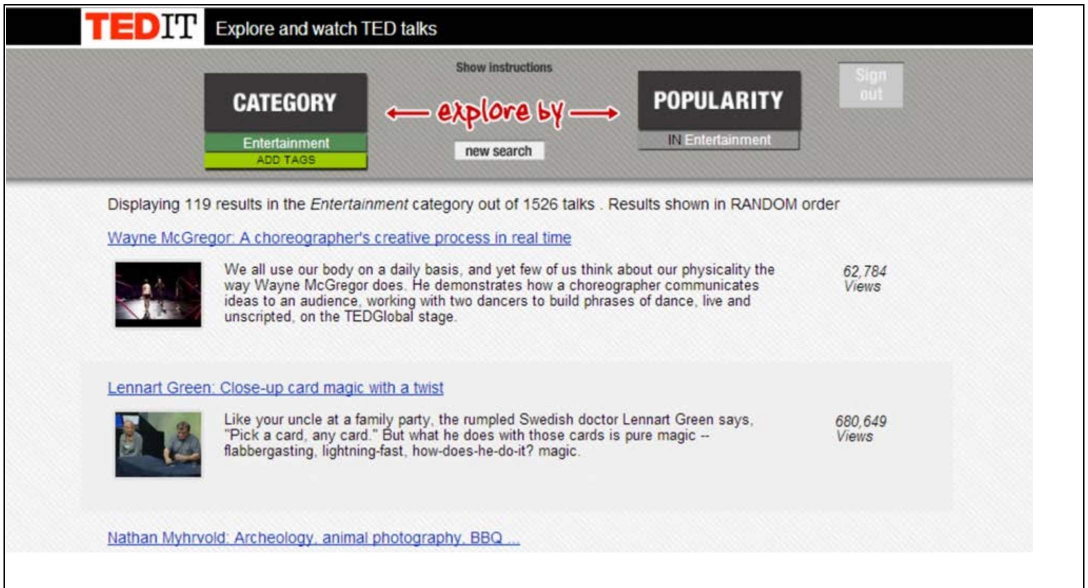

Figure 2. Search Results on TED-it, following a *Category* click and choice of the Entertainment category. Results are displayed in random order, as *Popularity* has not been clicked. This user has been randomly assigned to the information provision treatment, where view count information is shown alongside search results that follow a *Category* click.

To study the effect of popularity information on content exploration, we randomly assign users to one of two conditions - the *no information provision* and *information provision* treatment groups. In the first, popularity information (i.e., talks' view count on Youtube) is not provided alongside search results following each *Category* click, and in the second, this information is provided. Note that results continue to be listed in random order for both groups following each click on *Category*. The screenshot in figure 2 shows results for a user randomly assigned to the *information provision* condition.

Users' personal and social characteristics, as well as their tendency to serve as opinion leaders were assessed using a series of short self-report questionnaires [5]. Specifically, experiment participants reported their demographic information, previous experience with TED content, and their subjective relative sociability level. They also completed an opinion leadership questionnaire (26, 27). See SI for further information on these questionnaires.

[5] The timing of presentation of the questionnaires was randomized, such that the questionnaires appeared before and after the search task with equal probability. This is to control for possible priming effects, whereby the questionnaires may impact task performance or vice versa.

Our study met ethical guidelines for experiments involving human subjects. Participation in the experiment presented no more than minimal risk to subjects and was approved by MIT's Committee on the Use of Humans as Experimental Subjects (COUHES) and Tel Aviv University's Institutional Review Board (IRB). Informed consent was obtained from all subjects (further details in the SI).

3 Content Exploration and the Distribution of Choices on TED-it

We recruited 1,846 experiment participants via Amazon Mechanical Turk (AMT)[6]. Full descriptive statistics on these subjects are reported in the SI.

The average TED-it user performed 0.60 *Popularity* clicks and 1.05 *Category* clicks. Fully 32% of users clicked only on *Popularity*, thereby sorting all talks, and choosing from this list.

Much like other distributions of media consumption, the distribution of views on TED-it appears long tailed, with a few talks receiving high view counts, and the "tail" of many less-popular talks - viewed by a handful of users. Our 1,846 users watched 532 different talks. The number of views for the top five talks is 123, 116, 69, 69, and 63, respectively. At the bottom end of the distribution, talks ranked between 107 and 532 had one or two views. The distribution of views on TED-it is graphed in panel (a) of figure 3.

While TED-it presents information on talks' number of Youtube views, and allows for sorting based on these view counts, the rank order of views on TED-it does not mirror that on Youtube. This is evident from panel (b) of figure 3, which shows Youtube view counts for talks ordered by their view count on TED-it. Indeed, the most popular talks on TED-it have very high Youtube view counts, yet talks chosen frequently by our users are not necessarily popular on Youtube. For instance, forty-two individuals in our experiment chose the TED talk: "Are games better than life?" which had only 97,808 Youtube views. This is especially true for users who explored based on topic.

[6] Mason and Suri (33) survey several studies, and conclude that the behavior of AMT workers is comparable to that of laboratory subjects in both offline and online contexts.

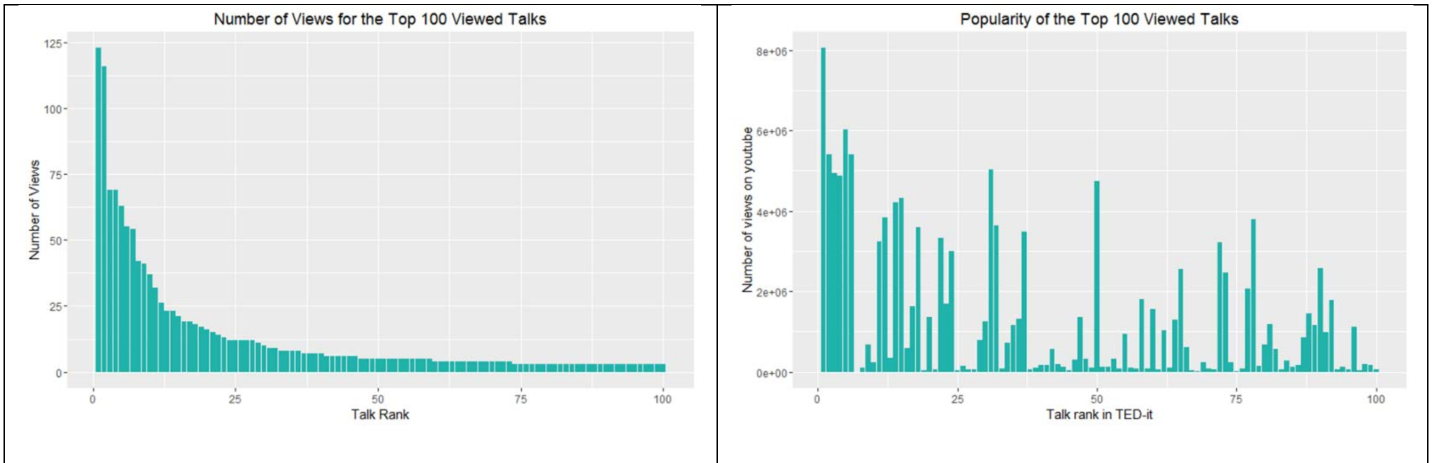

Figure 3. Distribution of view counts for the top 100 talks viewed on TED-it: (a) Distribution of number of views on TED-it, of experiment participants; (b) Distribution of number of views on Youtube, of Youtube users.

How does the view count distribution change with users' exploration and choice patterns? In figure 4, we plot three view count distributions, one for each of the three possible end-states of an exploration process in our environment. These are:

(a) Unsorted category: an unsorted list, within category. This is the end-state for 56% of participants.

(b) Sorted category: a popularity-sorted list, within category. This is the end-state for 9% of participants.

(c) All sorted: a sorted list of all talks (no category chosen). This is the end-state for 35% of participants.

Quite intuitively, popularity reliance in exploration increases monotonically from (a) to (c). Validating our notion of popularity reliance, figure 4 demonstrates that the popularity of chosen talks is largely determined by the end-state of exploration. Specifically, the mean Youtube view count is 521,007, 2,909,773 and 5,032,500 in end-states (a), (b) and (c) respectively, and the correlation between the TED-it and Youtube view counts, denoted $\rho_{i,Y}$, where $i \in \{a, b, c\}$, increases from $\rho_{a,Y} = 0.12$ for (a)-users, to $\rho_{b,Y} = 0.66$ for (b)-users, to $\rho_{c,Y} = 0.79$ for (c)-users.

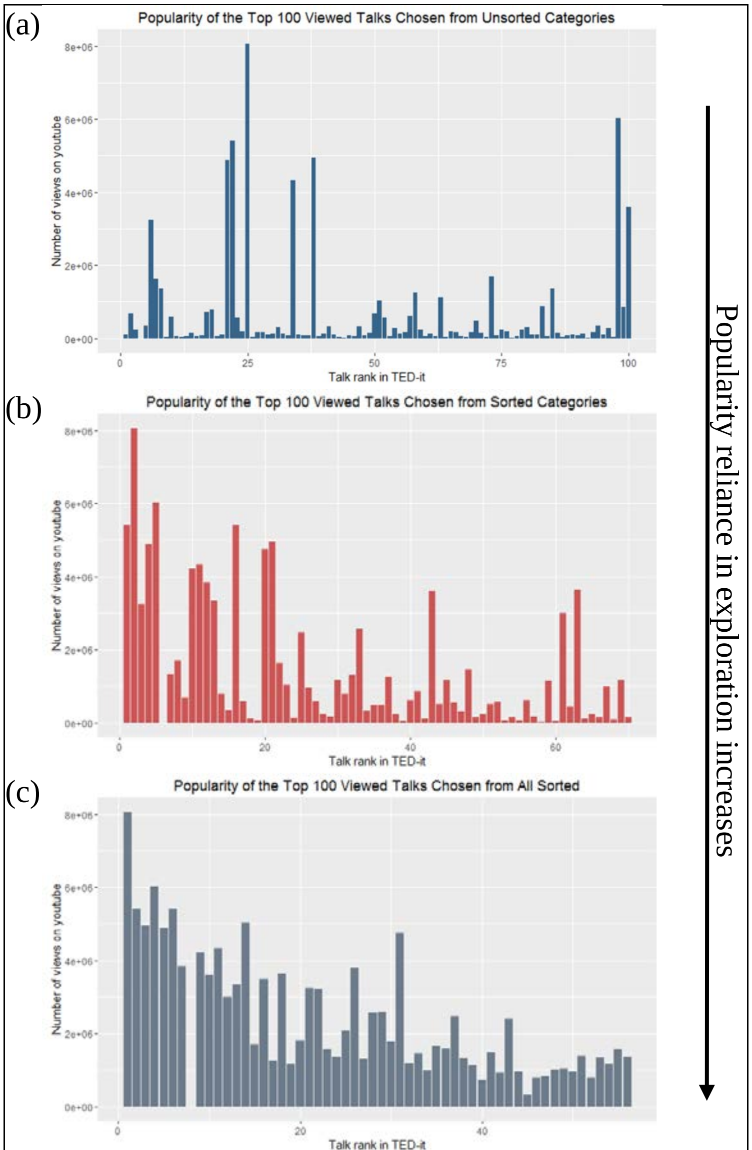

Figure 4. Distribution of Youtube view counts for top 100 talks viewed on TED-it, by end-state of exploration: (a) Unsorted category; (b) Sorted category; (c) All sorted.

Lower popularity reliance leads users to content that is less known by their peers, and is more suited to their personal interests. We argue that users who consistently search by subject matter and rely less on the crowd's interests are less likely to suffer from echo chambers.

Susceptibility to echo chambers in our setting is thus proxied by conducting relatively little exploration, and is represented by a low utilization of category search, along with a strong reliance on popularity sorting. We turn to examine which user characteristics are associated with a higher risk of echo chambers.

4 User Characteristics and Exploration Patterns

We now examine how user characteristics such as sociability, familiarity with content, opinion leadership and age affect content exploration. We operationalize "exploration", by studying users' search flow in our environment: from their first click, to follow-up clicks, to the type of results they choose from, and the chosen talk's location (or, scroll depth) in that list.

We are predominantly interested in users' *Category* vs. *Popularity* utilization, as representing exploration vs. reliance on the crowd's previous choices. Our main explanatory variables – sociability, content experience and opinion leadership – are expected to predict the relative utilization of these search dimensions, and are less likely to predict general search effort, such as number of clicks, or scroll depth. Indeed, the relationship between these user characteristics and overall clicking and scrolling is not statistically significant (S5), and we henceforth focus on the relationship between user characteristics and (dependent) variables describing the extent of topical exploration. Specifically, we examine how user characteristics predict their first click, share of *Category* clicks, and their end-state of exploration (i.e., type of results they choose from).

The first click represents exploration tendency. A first click on *Popularity* implies a very low tendency to explore, as the user simply sorts all talks by popularity. On the other hand, a first click on *Category* implies that the user examines the list of available topics and selects a subject of interest to him, thus taking a more exploratory approach. We thus define the binary variable *CategoryFirst* which equals 1 when the first click is *Category*, and 0 when it is *Popularity*.

The last one or two clicks represent the end-state of exploration, namely, the type of list users choose from. The possible end-states are (a), (b) and (c) defined above, in order of increasing popularity reliance, and our models include the binary variable *UnsortedCategory* that takes on the value 1 when the user's end-state is (a) and 0 otherwise. Using a dummy variable for end-state (a) as the key dependent variable for end-state of exploration is justified, as this end-state implies the highest degree of exploration in our setting, and yields a view count distribution that differs substantially from general viewing patterns on Youtube, and from those for the other two end-states.[7]

---

[7] Recall that the correlation between Youtube view counts and experiment view-counts for (a)-users ($\rho_{a,Y} = 0.12$) differs substantially from the correlations with view counts for (b)- and (c)-users ($\rho_{b,Y} = 0.66$ and $\rho_{c,Y} = 0.79$).

Furthermore, we define *ShareCategory* as the proportion of *Category* clicks out of all *Popularity* and *Category* clicks,[8] and use this variable as an additional proxy for exploration intensity. With 80% of participants conducting two search clicks or less, these three variables provide a comprehensive view of exploration in our environment.

Studying the relationship between user characteristics and exploration behavior, we run the following Logit regressions, in which variables are all at the experiment participant level, indexed by i (in subscript):

$$Logit(y_i) = \alpha_i + OL_i + Social_i + PreviousTED_i + HigherEd_i + Age_i + Info_i + OL_i * Info_i + \epsilon_i$$

Where y is one of $\{CategoryFirst, ShareCategory, UnsortedCategory\}$. OL_i is an indicator variable that equals 1 if participant i is an opinion leader and 0 otherwise, where an opinion leader is a participant whose opinion leadership score is in the top quartile (SI). $Social_i$ is i's subjective reported sociability level in $\{1,2,..,5\}$, and $PreviousTED_i$, representing content familiarity, equals 1 if i has previously seen a TED talk and 0 otherwise. $HigherEd_i$ takes the value 1 if participant i reports attaining at least some college education, and 0 if i reports no postsecondary education, and Age_i is i's age. $Info_i$ indicates whether or not i was randomly assigned to view popularity information alongside category-based results, and thus equals 1 if this information is shown and 0 if it is blocked. We further include the interaction term $OL * Info$ to study the impact of popularity information on opinion leaders' exploration. Since results are only displayed after the first click, $Info$ and $OL * Info$ are dropped when $y = CategoryFirst$. Detailed variable definitions and descriptive statistics are provided in the SI (tables S1,S2, S3, S4). Regression results, which are discussed below, appear in Table 1.

Our study was not designed to uncover gender differences in the relationship between user characteristics and content exploration patterns, however, these do exist. Namely, our regression models and relationships between variables of interest are largely statistically significant for males, and not significant for females, indicating the existence of consistent patterns for men, but not for women. We find no significant differences in descriptive statistics for men and women (S3) that

[8] We focus on these two main buttons since their usage directly relates to our main research questions. Moreover, the utilization of other buttons (such as, Add Tags), was extremely low (S2)

would explain the lack of significance of results for women, and leave this as a question for future research by gender scholars.[9] We, henceforth, present our results separately for men and women.

Sociability

We conjecture that highly social individuals will tend to “follow the herd” and suffer a risk of echo chambers, since they identify more strongly with their peers. Indeed, we find that highly social individuals show a weaker preference for exploration by topic, and rely more heavily on popularity sorting, compared to individuals who report lower sociability. We further find that highly social individuals are less likely to choose from unsorted category-specific results. These results are all statistically significant for men, and not significant for women (see Table 1). Namely, for male subjects, an increase of one point in the (reported five-point) sociability index decreases the odds of starting exploration with *Category* by 16% (p=0.002***), the ratio of *Category* to *Popularity* clicks by 14% (p=0.001***), and the odds of choosing a talk from unsorted category-specific results by 16% (p=0.004***), on average. Figure 5(a) depicts the stronger popularity reliance for those reporting higher sociability.

Familiarity with Content

Previous experience with a content space is expected to increase the use of heuristics and reduce the tendency to explore in that space. As a result, experience with content may increase the risk of echo chambers. Our results corroborate this hypothesis for male users. Figure 5(b) shows that users who have previously seen TED talks show stronger popularity reliance, as they are less likely to choose a talk from an unsorted list. Specifically, for men, having previously seen a TED talk decreases the odds of commencing search with a *Category* click by 22% (p=0.10*), the ratio of *Category* to *Popularity* clicks by 30% (p=0.003***), and the odds of choosing a talk from an unsorted topic-specific list by 37% (p=0.002***), on average. The direction of these effects are the same for women, though not statistically significant.

---

[9] Gender differences have been found in decision-making, information search, and web search (e.g., 29–32, to name but a few).

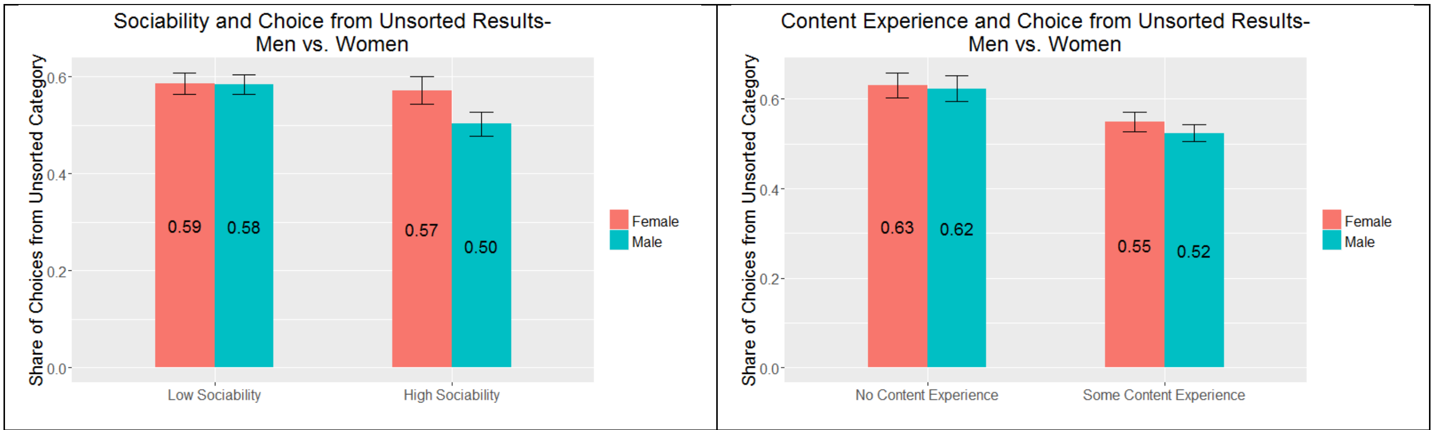

Figure 5. The relationship between choice from unsorted topical results and: (a) sociability; (b) content experience, for males vs. females. The results show that the tendency to explore decreases with sociability and content experience.

Age and higher education

Examining the relationship of various user demographics with the exhibited exploration patterns, we find that younger individuals, both male and female, are less likely to conduct topical search and rely mostly on popularity considerations. For men, being one year older increases the odds of starting exploration with a *Category* click by 3% (p=0.0001***), the ratio of *Category* to *Popularity* clicks by 2% (p=0.003***), and the odds of choosing a talk from an unsorted topical list by 3% (p=0.0001***). For women, point estimates are in the same direction, though the models are not statistically significant.[10] Younger users are thus more likely to get caught in content echo chambers than older users.

With respect to higher education, we find that the odds of ending the exploration process in an unsorted category-specific list is 51% higher (p=0.04**) for men with at least some college education, compared to their less educated counterparts. Note, however, that our sample includes predominantly users with at least some college education, with 88% of men and 91% of women reporting some higher education. In light of the very low variance in this variable, it is not surprising that we do not find statistically significant relationships of higher education with most of our dependent variables of interest.

Opinion Leaders May Alleviate Echo Chambers

[10] An exception is model (4) in appendix table S5, showing that, for women, another year of age increases scroll depth by 0.36 (p=0.0003***).

Opinion leaders are intrinsically motivated to influence others in their social circles. Since already-popular content decreases opinion leaders' capacity to act as thought leaders, we expect these individuals to exhibit a lower popularity reliance, and seek out content based on their topical interests, to create new avenues for influence. This implies that, on average, opinion leaders should invest more effort in content search and display a stronger *Category* preference compared to non-leaders in our environment.

In line with this, opinion leadership is found to affect exploration patterns, but only for men in our sample. For males, opinion leadership increases the odds of starting exploration with a *Category* click by 31% (p=0.10*), the ratio of *Category* to *Popularity* clicks by 72% (p=0.004***), and the odds of choosing a talk from unsorted topical results by 68% (p=0.03**). Figure 6(a) depicts male opinion leaders' higher tendency to explore by topic, and to choose from an unsorted category-specific list. These results suggest that opinion leaders can be expected to alleviate echo chambers in their social circles.

The Role of Popularity Information

We explore the role of popularity information provision in creating echo chambers. Popularity information such as Facebook like-counts and Youtube view counts, is baked into our online environments. How is popularity information affecting our choices and tendency to explore unfamiliar territories? On the one hand, popularity information can substitute for sorting, by providing a quality guarantee for content that is not "topping the charts" but still drawing some attention, i.e. generating thousands of views but not millions. On the other hand, popularity information can create increased awareness to popularity considerations, and complement them, leading to a stronger reliance on the crowd's tastes and more sorting. Whether providing popularity information is a substitute or a complement to sorting is thus an empirical question, which we answer by randomly assigning users to one of two conditions, where popularity information is shown/blocked in category-based searches.

Interestingly, only male opinion leaders are affected by the provision of popularity information - this information does not affect the exploration and choice patterns of non-leaders. This may be due to opinion leaders' higher tendency to conduct topical search, which increases the salience of the information-provision treatment to this group.

Male opinion leaders tend to complement popularity information with increased sorting, as their ration of *Category* to *Popularity* clicks is 40% lower (p=0.04**), and their odds of choosing from an unsorted topic-based list are 41% lower (p=0.10*) than their counterparts in the no-information provision condition. The strong topic preference observed for these opinion leaders is therefore weakened in the presence of salient information on the crowd's choices. This effect is shown in Figure 6(b).

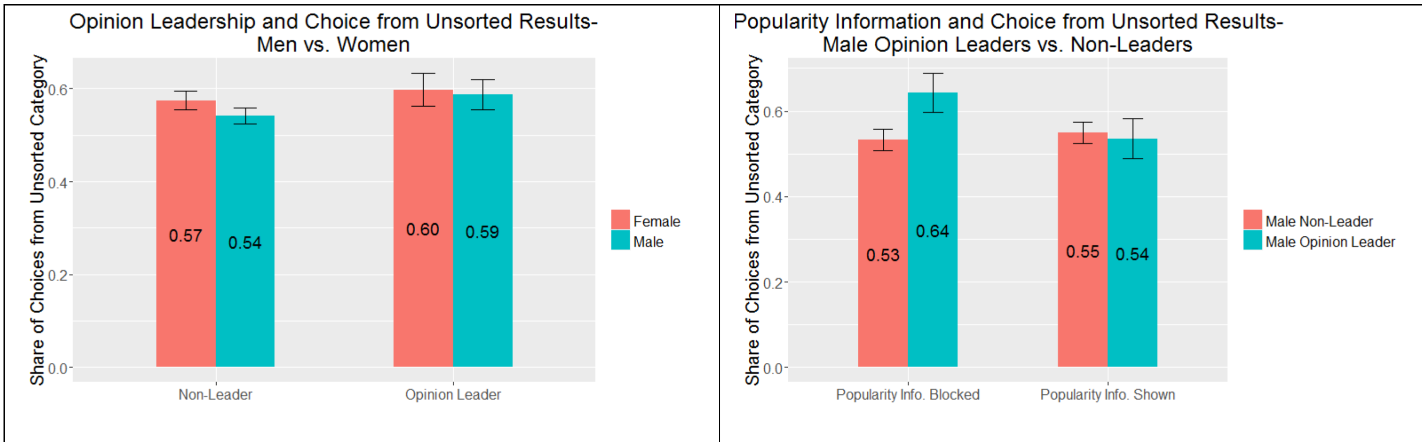

Figure 6. The relationship between choice from unsorted topical results and: (a) opinion leadership, for males vs. females; (b) availability of popularity information, for male opinion leaders vs. non-leaders. The results show that the tendency to explore increases with opinion leadership, and that, for male opinion leaders, display of popularity information leads to a decrease in exploration.

Table 1. Main regression results.

	Logit Regressions for dependent variable:					
	Men			**Women**		
	Category First	*Share Category*	*Unsorted Category*	*Category First*	*Share Category*	*Unsorted Category*
	(1)	(2)	(3)	(4)	(5)	(6)
OL	0.27*	0.54***	0.52**	-0.07	0.02	-0.02
	(0.16)	(0.19)	(0.23)	(0.17)	(0.19)	(0.25)
Social	-0.18***	-0.15***	-0.17***	-0.06	-0.09*	-0.01
	(0.06)	(0.04)	(0.06)	(0.06)	(0.05)	(0.06)
PreviousTED	-0.25*	-0.36***	-0.47***	-0.09	-0.19	-0.27*
	(0.15)	(0.12)	(0.15)	(0.15)	(0.12)	(0.15)
HigherEd.	0.20	0.21	0.41**	-0.40	-0.44**	-0.38
	(0.19)	(0.15)	(0.19)	(0.27)	(0.21)	(0.27)
Age	0.03***	0.02***	0.03***	0.01*	0.002	0.01*
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Info		0.15	0.08		0.05	-0.14
		(0.11)	(0.14)		(0.13)	(0.16)
*OL*Info*		-0.51**	-0.52*		0.25	0.18
		(0.24)	(0.31)		(0.27)	(0.34)
Constant	-0.06	0.39	-0.34	0.53	1.27***	0.53
	(0.35)	(0.28)	(0.35)	(0.40)	(0.33)	(0.40)
Observations	1,034	1,034	1,034	812	812	812
Log Likelihood	-682.12	-853.26	-689.35	-546.77	-669.75	-546.77
LR χ^2	32.575	35.601	43.833	7.878	13.277	11.226
$Prob > \chi^2$	4.57e-06 ***	8.62e-06 ***	2.30e-07 ***	0.163	0.066*	0.129

Notes:
1. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$
2. In models (2) and (5), the dependent variable is a proportion and we therefore use the total number of *Popularity* and *Category* clicks as weights.

Users' satisfaction

We use two outcome variables as proxies for users' satisfaction with their chosen talk: (1) *ExtraVideos*, and (2) *ExtraSeconds*. *ExtraVideos* is a binary variable which equals 1 for participants who watched another video after the mandated one, and 0 otherwise. Approximately 6% of participants proceeded to watch an extra video. *ExtraSeconds* is an indicator variable which takes on the variable 1 if the user was in the top 50% of excess viewing length, i.e., past the required minimum of viewing the chosen talk for five minutes. This approach assumes that if participants'

TED-it usage substantially exceeds the experiment requirements, then it is likely that they derived enjoyment from their content consumption.

Controlling for user characteristics, we do not find any indication that exploration and choice patterns, or specifically, popularity reliance therein, are associated with either of these user satisfaction metrics. This is evident from running our main specification with $ExtraVideos$ and $ExtraSeconds$ as dependent variables, additionally controlling for exploration characteristics – $CategoryFirst$ (S6a), $ShareCategory$ (S6b), and $UnsortedCategory$ (S6c). The results show that 95% confidence intervals for the effect of these exploration characteristics all include 0, implying no significant association with enjoyment.

5 Discussion and Concluding Notes

Tendencies to conduct limited exploration and rely on peers' past choices in own content choice are likely to lead users down echo chambers, with limited exposure to content that is not in line with their peers' and own views and opinions. Exposure to diverse content may be further limited over time, as user tendencies feed into personalization algorithms. We find that highly social users, those searching in a content space they are already familiar with, and young users, are all at a higher risk of such echo chambers. Opinion leaders may ameliorate echo chamber concerns within their social circles, due to their tendency to conduct more popularity-independent exploration compared to their followers. However, their inclination to explore is highly sensitive to the provision of popularity information, and curtailed by it.

Should we therefore suppress or, at least, reduce the visibility of popularity information in our online environments to increase the diversity of content consumed? To the extent that content diversity is a desired end, the answer is yes. Furthermore, in our analyses, there were no statistically significant correlations between users' exploration characteristics, and proxies for enjoyment, such as viewing talks past the mandated time, or watching an extra talk. This suggests that, at least in the realm of curated content, such as TED talks, popular content is not inherently superior to less popular content. Therefore, designing online environments that encourage exploration (e.g., with increased visibility of non-hit content, and reduced visibility of view counts) may alleviate content echo chambers, with little impact on user satisfaction.

While our experiment takes place in a non-standard search environment, it offers first insights as to users' personal preferences for exploration and to the extent to which they would actively seek out ranked results, if given a choice. Future research may extend these results to more organic settings and general content spaces. Still, our results may speak to general concerns raised regarding the growing role of algorithms in our lives (28, 29). Algorithmic ranking, based on popularity, users' previous searches, clicks, and other factors, is an integral part of search engines, social media, and other online environments. Ranking of results has been shown to have a major influence on user choices (30), has the potential of introducing biases (10), and these impacts easily go unnoticed, as users are largely unaware of the black-box mechanisms determining their search results and news feeds (10, 31, 32). The null effect we find for the relationship between ranking and user enjoyment, while limited to curated content, indicates that the user experience does not necessarily suffer when ranking is removed. Regulators may use this as support for the case for diversifying search results, at least in domains of national importance, such as news and elections.

References

1. McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a Feather: Homophily in Social Networks. *Annu Rev Sociol* 27(1):415–444.

2. El-Bermawy MM (2016) Your Echo Chamber is Destroying Democracy | WIRED. *Wired*. Available at: https://www.wired.com/2016/11/filter-bubble-destroying-democracy/ [Accessed August 1, 2017].

3. Chater J (2016) What the EU referendum result teaches us about the dangers of the echo chamber. *New Statesman*. Available at: http://www.newstatesman.com/2016/07/what-eu-referendum-result-teaches-us-about-dangers-echo-chamber [Accessed August 1, 2017].

4. Khalid A (2017) Tech Creates Our Political Echo Chambers. It Might Also Be A Solution. *NPR*. Available at: http://www.npr.org/sections/alltechconsidered/2017/04/12/522760479/tech-creates-our-political-echo-chambers-it-might-also-be-a-solution [Accessed August 1, 2017].

5. Aiello LM, et al. (2012) Friendship prediction and homophily in social media. *ACM Trans Web* 6(2):1–33.

6. Personalized Search for everyone (2009) *Google Off Blog*. Available at:

https://googleblog.blogspot.com/2009/12/personalized-search-for-everyone.html [Accessed August 1, 2017].

7. Sullivan D (2011) Bing Results Get Localized & Personalized. *Search Engine L*. Available at: http://searchengineland.com/bing-results-get-localized-personalized-64284 [Accessed August 1, 2017].

8. Pariser E (2011) *The Filter Bubble: What the Internet Is Hiding from You* (The Penguin Group) Available at: http://dl.acm.org/citation.cfm?id=2029079 [Accessed August 1, 2017].

9. Hannak A, et al. (2013) Measuring personalization of web search. *Proceedings of the 22nd International Conference on World Wide Web - WWW '13* (ACM Press, New York, New York, USA), pp 527–538.

10. Epstein R, Robertson RE (2015) The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. *Proc Natl Acad Sci U S A* 112(33):E4512-21.

11. Bakshy E, Messing S, Adamic LA (2015) Exposure to ideologically diverse news and opinion on Facebook. *Science (80-)* 348(6239):1130–1132.

12. Duggan M, Smith A, Page D (2016) *The Political Environment on Social Media. Pew Research Center* (Washington, D.C.) Available at: http://www.pewinternet.org/2016/10/25/the-political-environment-on-social-media/ [Accessed August 1, 2017].

13. Katz E, Lazarsfeld PF (1955) *Personal influence: The Part Played by People in the Flow of Mass Communication* (The Free Press, New York, NY) Available at: https://books.google.com/books?hl=en&lr=&id=rElW8D0D8gYC&oi=fnd&pg=PR1&dq=Personal+influence:+The+part+played+by+people+in+the+flow+of+mass+communications.&ots=ODyZ-Pg_dJ&sig=cmAFXEcFraJWINNm7XfssrELbIM [Accessed August 1, 2017].

14. Katz E (1957) The Two-Step Flow of Communication: An Up-To-Date Report on an Hypothesis. *Public Opin Q* 21(1):61–78.

15. Rogers EM (1995) *Diffusion of innovations* (Free Press, Glencoe) doi:citeulike-article-id:126680.

16. Weimann G (1994) *The influentials : people who influence people* (State University of New York Press) Available at: https://books.google.com/books/about/The_Influentials.html?id=ou5bTZGSOYAC [Accessed August 1, 2017].

17. Van den Bulte C, Joshi Y V. (2007) New Product Diffusion with Influentials and Imitators. *Mark Sci* 26(3):400–421.

18. Flodgren G, et al. (2007) Local opinion leaders: effects on professional practice and health care outcomes. *Cochrane Database of Systematic Reviews*, ed Eccles MP (John Wiley & Sons, Ltd, Chichester, UK). doi:10.1002/14651858.CD000125.pub3.

19. Godes D, Mayzlin D (2009) Firm-Created Word-of-Mouth Communication: Evidence from a Field Test. *Mark Sci* 28(4):721–739.

20. Iyengar R, Van den Bulte C, Valente TW (2011) Opinion Leadership and Social Contagion in New Product Diffusion. *Mark Sci* 30(2):195–212.

21. Salganik MJ, Dodds PS, Watts DJ (2006) Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market. *Science (80-)* 311(5762):854–856.

22. Salganik MJ, Watts DJ (2008) Leading the Herd Astray: An Experimental Study of Self-fulfilling Prophecies in an Artificial Cultural Market. *Soc Psychol Q* 71(4):338–355.

23. Cai H, Chen Y, Fang H (2009) Observational Learning: Evidence from a Randomized Natural Field Experiment. *Am Econ Rev* 99(3):864–882.

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

25. Muchnik L, Aral S, Taylor SJ (2013) Social Influence Bias: A Randomized Experiment. *Science (80-)* 341(6146):647–651.

26. Goldsmith RE, Flynn LR, Goldsmith EB (2003) Innovative Consumers and Market Mavens. *J Mark Theory Pract* 11(4):54–65.

27. Flynn LR, Goldsmith RE, Eastman JK (1996) Opinion Leaders and Opinion Seekers: Two New Measurement Scales. *J Acad Mark Sci* 24(2):137–147.

28. Dickey MR (2017) Algorithmic accountability. *Tech Crunch*. Available at:

https://techcrunch.com/2017/04/30/algorithmic-accountability/ [Accessed August 3, 2017].

29. Pasquale F *Black box society : the secret algorithms that control money and information*.

30. Pan B, et al. (2007) In Google We Trust: Users' Decisions on Rank, Position, and Relevance. *J Comput Commun* 12(3):801–823.

31. Schwab P-N (2016) Facebook : how do people perceive NewsFeed curation algorithm. *Into Minds*. Available at: http://www.intotheminds.com/blog/en/facebook-how-do-people-perceive-newsfeed-curation-algorithm/ [Accessed August 3, 2017].

32. Powers E (2017) New Research Shows Students Are Largely Unaware of News Personalization. *Media Shift*. Available at: http://mediashift.org/2017/03/new-research-shows-students-largely-unaware-news-personalization/ [Accessed August 3, 2017].

33. Mason W, Suri S (2012) Conducting behavioral research on Amazon's Mechanical Turk. *Behav Res Methods* 44(1):1–23.

34. Benko C, Pelster B (2013) How Women Decide. *Harv Bus Rev* (September 2013). Available at: https://hbr.org/2013/09/how-women-decide [Accessed August 3, 2017].

35. Laroche M, Saad G, Cleveland M, Browne E (2000) Gender differences in information search strategies for a Christmas gift. *J Consum Mark* 17(6):500–522.

36. Barber N, Dodd T, Kolyesnikova N (2009) Gender differences in information search: implications for retailing. *J Consum Mark* 26(6):415–426.

37. Zhou M (2014) Gender difference in web search perceptions and behavior: Does it vary by task performance? *Comput Educ* 78:174–184.

Supporting Information

Participant Recruitment Text and Mechanism for Obtaining Informed Consent

Experiment instructions provided on Mechanical Turk:

The TED-it Search Experiment

We are conducting an academic experiment about online search.

You are invited to search the TED talks* collection, and find a great talk to watch.

Select the link below to enter our experiment. After watching a talk for at least 5 minutes, you will receive a worker code to paste into the box below to receive credit for participation.

Note: Your anonymity is guaranteed - no identifying information will be collected. The results of this study may be published in a book or journal, or used in teaching materials. For questions about this study please e-mail us at tedit.experiment@gmail.com.

Consent: By clicking the link below, you express your consent to participate in this study, and state that its purpose and nature have been sufficiently explained. You are free to withdraw at any time during the experiment, by simply navigating to another website or closing your browser.

This HIT is part of a MIT scientific research project. Your decision to complete this HIT is voluntary. There is no way for us to identify you. The only information we will have, in addition to your responses, is the time at which you completed the survey. The results of the research may be presented at scientific meetings or published in scientific journals. Clicking on the 'SUBMIT' button on the bottom of this page indicates that you are at least 18 years of age and agree to complete this HIT voluntarily.

* TED is a nonprofit devoted to Ideas Worth Spreading. TED talks' videos are released under a Creative Commons license, so they can be freely shared and reposted. We use TED talks in full compliance with this license.

Experiment link: http://ted-it.tau.ac.il/tedit/turk.php

Provide the worker code here: e.g. 123456

Submit

Recruitment text on Mechanical Turk:

Describe your HIT to Workers

Title	Search TED talks in an experimental environment Describe the task to Workers. Be as specific as possible, e.g. "answer a survey about movies", instead of "short survey", so Workers know what to expect.
Description	Fill out a questionnaire, then explore TED talks, and watch a talk for at least 5 minutes. Give more detail about this task. This gives Workers a bit more information before they decide to view your HIT.

Experiment screenshots

Experiment instructions screen. The timing of presentation of the questionnaires was randomized, such that the questionnaires appeared before and after the search task with equal probability. This controlled for possible priming effects, whereby the questionnaires may impact task performance or vice versa. In line with the variation on timing of questionnaires, there are two versions of the instructions screen, version (a) for users who answered the questionnaires after completing the exploration task, and version (b) for users who answered the questionnaire before conducting exploration. These are shown in Figures S1 and S2.

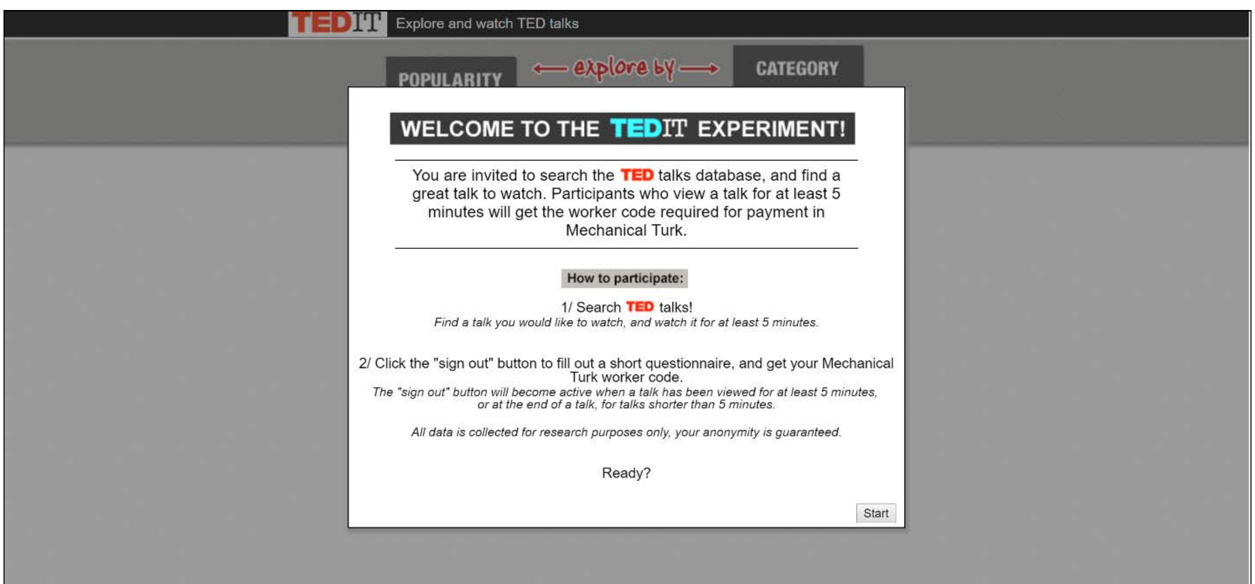

Figure S1. TED-it instructions screen version (a), for users randomly assigned to answer the questionnaires after completing the exploration task.

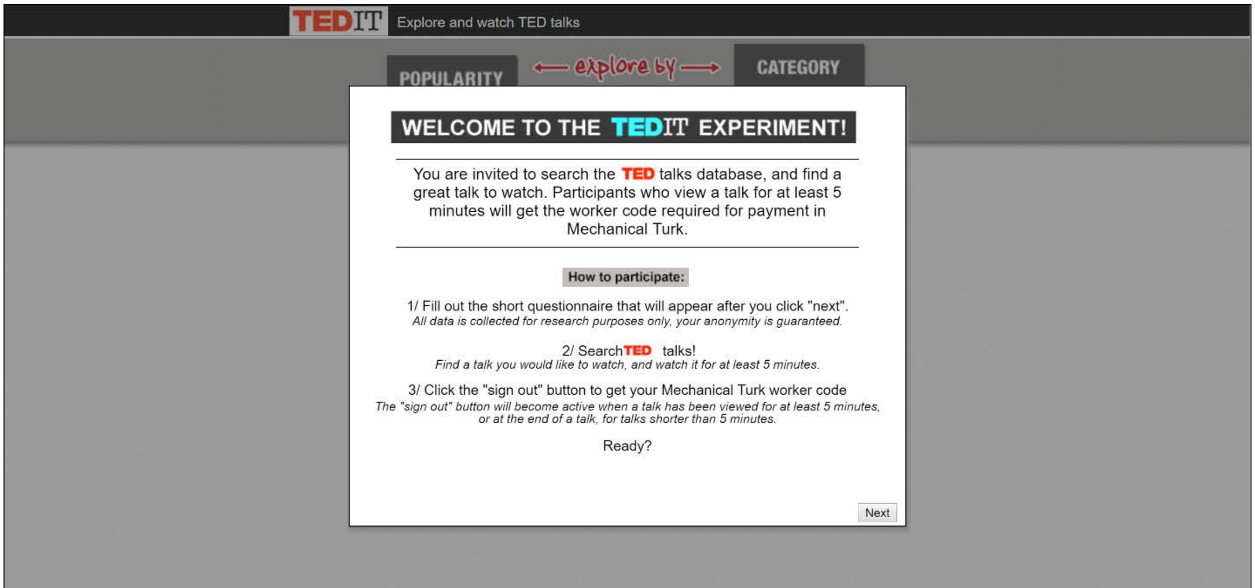

Figure S2. TED-it instructions screen version (b), for users randomly assigned to answer the questionnaires before conducting the exploration task.

Search-flow screenshots. Search flow on TED-it is presented below in figures S3-S5. Note the randomization of the two main buttons' location: figures S1-S2 and S8 show screens for a user randomly assigned to the version with *Popularity* on the left, whereas figures S3-S7 show screens for a user randomly assigned to the version with *Popularity* on the right.

Figure S3 presents the results screen after a first click on *Popularity*, or equivalently, after any *Popularity* click that follows a click on the "new search" button. This sorts all talks by their Youtube view count (shown on the right). For ease of implementation, and since users' scrolling is naturally limited, the top 100 most viewed talks are displayed (rather than all talks). For clarity, under the *Category* button the text "All Talks" appears, indicating that no category has been chosen, and similarly the text "In All Talks" appears under the *Popularity* button. Furthermore, a sentence describing the list of search results appears above the list, with the word SORTED in all capital letters.

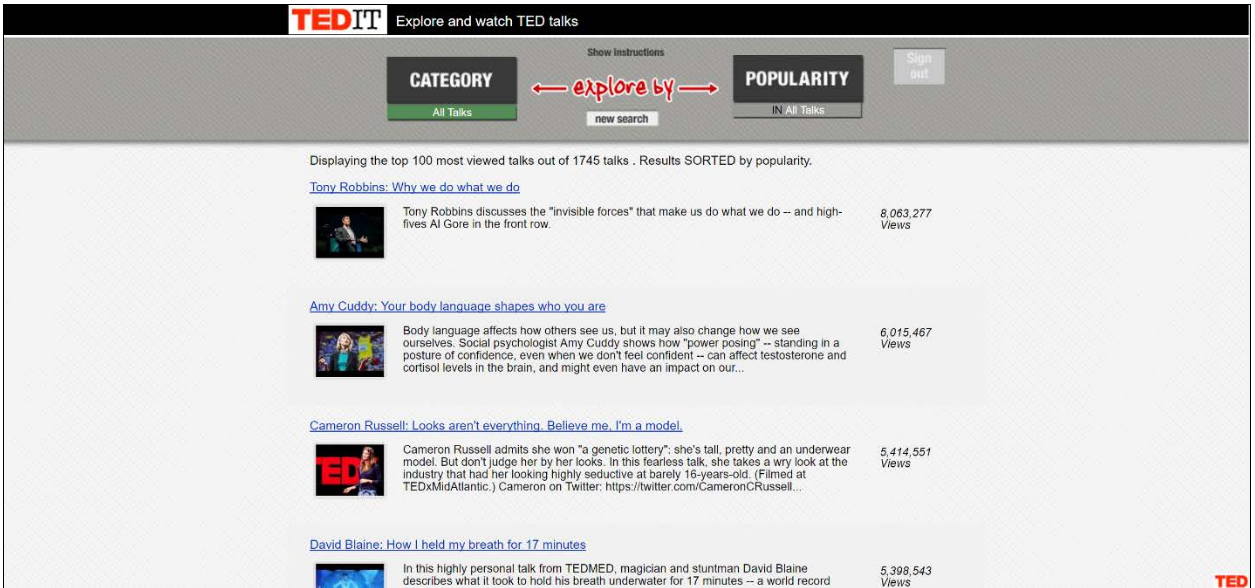

Figure S3. The search results screen following a click on *Popularity* when no category has been chosen. All talks are sorted by their Youtube view count.

Figure S4 presents the results screen after a click on *Category*, and a choice of the "News & Politics" category. This produces a list of all talks in the chosen category, shown in random order. Note that figure S4 shows a screen for a user randomly assigned to the condition where view count information is blocked following *Category* clicks. For clarity, under the *Category* button the text "News & Politics" appears, indicating the chosen category. The text "In News & Politics" appears under the *Popularity* button, indicating that a click on this button will sort the talks in the News & Politics category. Furthermore, a sentence describing the list of search results appears above the list, with the word RANDOM in all capital letters.

The following figure S5, shows the list of results in figure S4 after a click on *Popularity* which sorted the talks in the chosen category by their Youtube view count. The sentence above the results list is updated accordingly, with the word SORTED in all capital letters, and view count information is now presented on the right (this information is always presented after *Popularity* clicks, the randomization applies only to showing or blocking this information after *Category* clicks).

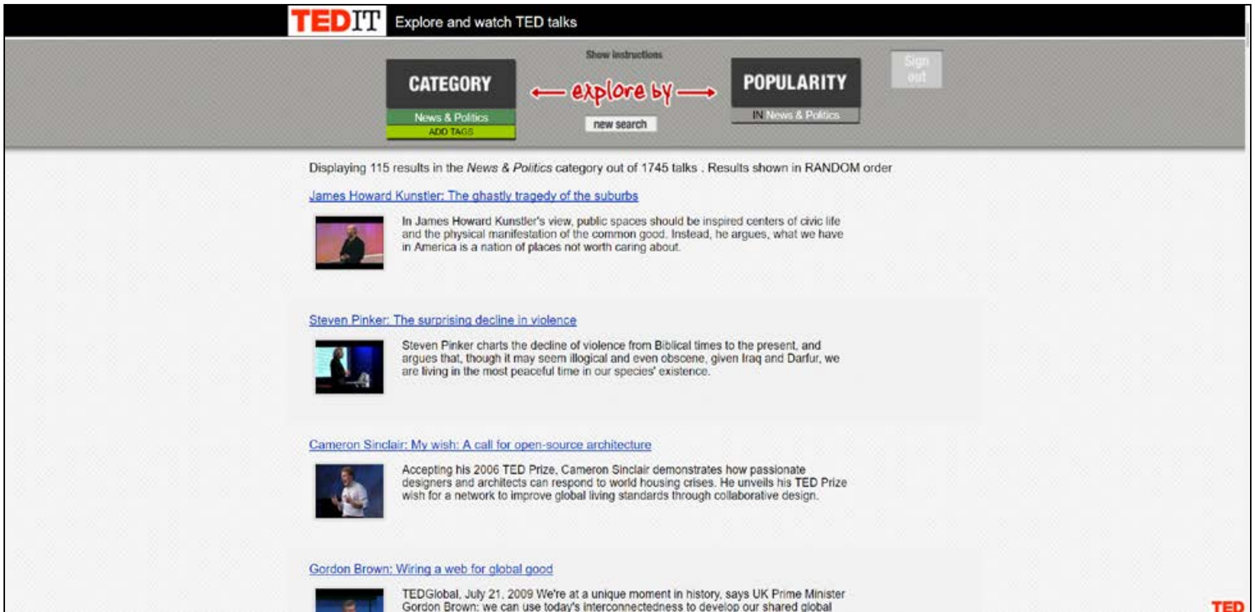

Figure S4. The search results screen following a click on *Category* when the News & Politics category has been chosen. Results are shown in random order, and view count information has been blocked (based on random assignment of this user).

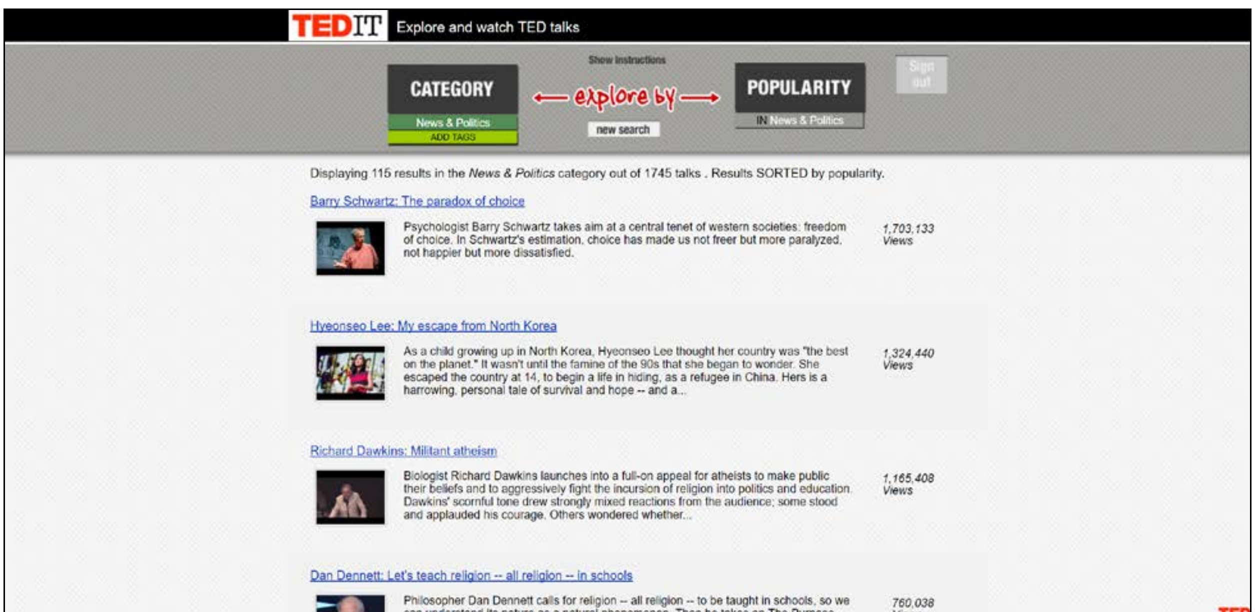

Figure S5. The search results screen following a click on *Category* with the News & Politics category chosen, and another click on *Popularity*. Results are sorted by their Youtube view count, which is shown on the right.

Watching a video. Figures S6 and S7 show a video being watched on TED-it, following a click on one of the videos in a list of search results. The video opens in a pop-up screen that streams the

TED video from the TED Youtube channel. Videos may be viewed for any length of time, but the *Sign Out* button only becomes active when a user has viewed a talk for the mandated time (5 minutes) or when a user views the full length of a talk (for several talks that are shorter than 5 minutes). Figure S6 shows the viewing pop-up window when a talk is being watching, and the user is not yet eligible to sign out, such that the *Sign Out* button is turned off (it is gray). Figure S7 shows the same video playing, after 5 minutes have elapsed, when the *Sign Out* button is active (it is yellow). Note that when the *Sign Out* button becomes active it flashes on and off for 5 seconds, to ensure that the user is aware of his completion of the experimental task and eligibility to sign out.

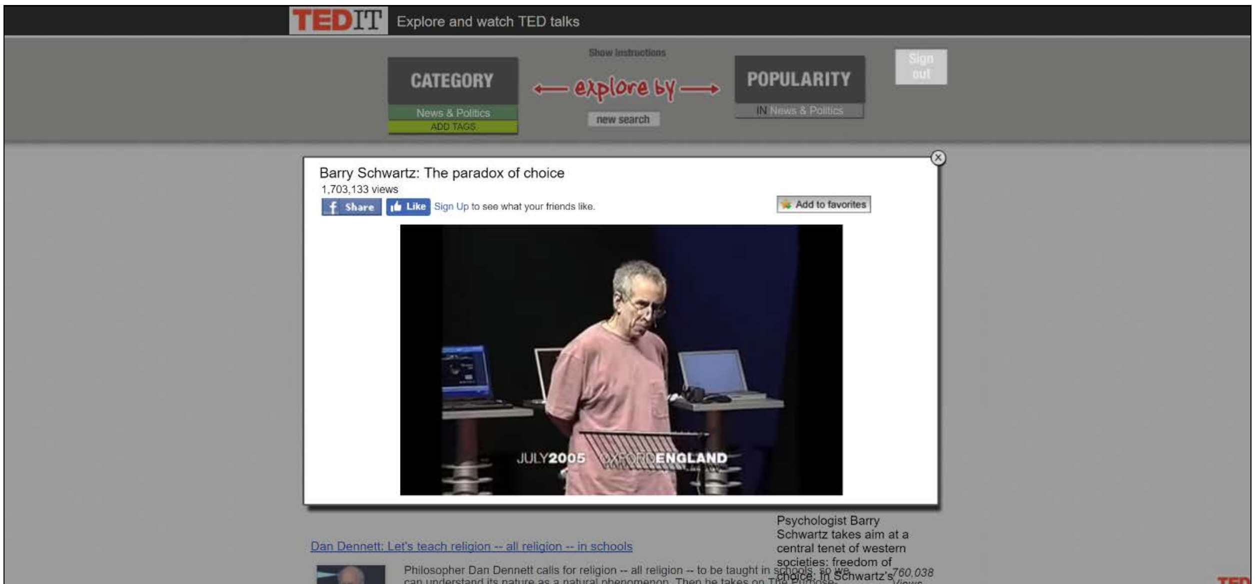

Figure S6. Watching a video on TED-it, in a pop-up window that streams the chosen video from the TED Youtube channel. This screen shows viewing before 5 minutes have elapsed, such that the experimental task is not yet complete and the *Sign Out* button is not yet active.

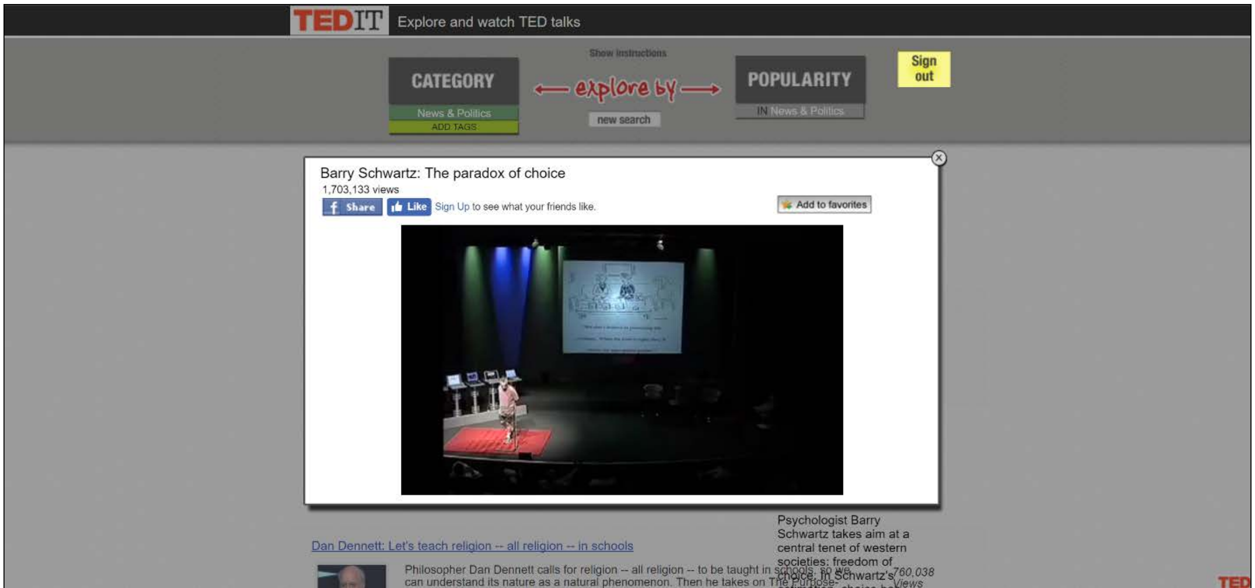

Figure S7. Watching the same video as in figure S6, after 5 minutes have elapsed. The experimental task is now complete and the *Sign Out* button is active (changed color to yellow).

Signing out. Figure S8 shows the TED-it sign out screen, that appears following a click on *Sign Out*. The worker code necessary for receiving payment on AMT is provided on this screen, along with links to continue using TED-it, if the user so chooses (utilization of these links was very low).

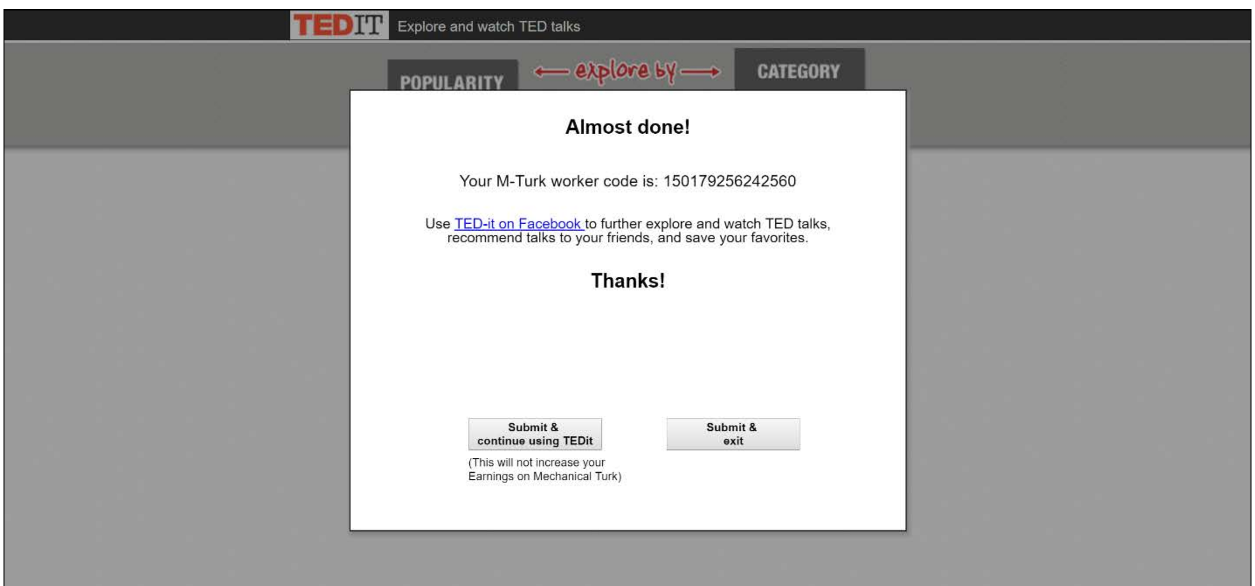

Figure S8. The TED-it sign out screen, appearing after the user clicks *Sign Out*. The worker code necessary for receiving payment on AMT is hereby provided, along with links to continue using TED-it, if the user so chooses.

Questionnaires

Opinion Leadership Questionnaire. Our study employs the following Opinion Leadership scale adapted in (26) from (27).

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: $OLscore \equiv OL(1) + \big(8 - OL(2)\big) + OL(3) + \big(8 - OL(4)\big) + OL(5) + \big(8 - OL(6)\big)$.

The maximum value for *OLscore* 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).

Demographic, content experience and social questions. Participants were further required to provide the following information.

- Gender – Male / Female
- Age (*you must be at least 18 to participate) – (dropdown menu 18-100)
- Country of residence – (drop down menu)
- Education – Some high school (or less) / High school diploma / Some college / Bachelor’s degree / Master’s of Ph.D
- Have you ever watched a TED talk? Yes / No
- Approximately how many friends do you have on Facebook?
- Do you consider yourself a social person? (Answer on a scale of 1-5, where 1=not social, 5=very social)

Tables

Table S1. Variable Definitions

Statistic	Definition
OLscore	6-42 score in the opinion leadership questionnaire (see above for calculation).
OL	Indicator variable that equals 1 if a participant's *OLscore* is in the top quartile and 0 otherwise.
Social	1-5 Likert scale response to the subjective question "Do you consider yourself a social person? (1-not social, 5-very social)."
PreviousTED	Indicator variable that equals 1 if a participant reports previously seening a TED talk and 0 otherwise.
HigherEd	Indicator variable that equals 1 if a participant reports attaining at least some college education, and 0 otherwise.
Age	Reported age, in years (must be at least 18 to participate).
Info	Indicator variable that equals 1 if a participant was randomly assigned to see view count information alongside unsorted category based results, and 0 otherwise.
CategoryFirst	Indicator variable that equals 1 when a participant's first click is *Category*, and 0 when it is *Popularity*.
Clicks	Total number of exploration clicks.
ShareCategory	Proportion of *Category* clicks out of all *Popularity* and *Category* clicks
ShareT	Proportion of *Add Tags* clicks out of all exploration clicks.
Scroll Depth	Integer indicating the chosen talk's location in the list of search results, where 1 is the first entry, and higher numbers indicate more scrolling.
UnsortedCategory	Indicator variable that equals 1 when a participant's end-state of exploration is unsorted category-specific results, and 0 otherwise (end-state (a)).
SortedCategory	Indicator variable that equals 1 when a participant's end-state of exploration is sorted category-specific results, and 0 otherwise (end-state (b)).
AllSorted	Indicator variable that equals 1 when a participant's end-state of exploration is a sorted list of all talks, and 0 otherwise (end-state (c)).
Extra Videos	Indicator variable that equals 1 when a participant watched another video after the mandated one, and 0 otherwise.
Extra Seconds	Indicator variable that equals 1 when a participant is in the top 50% of excess viewing length, i.e., past the required minimum of viewing the chosen talk for five minutes.

Table S2. Descriptive Statistics - All Users

Statistic	N	Mean	St. Dev.	Min	Max
OLscore	1,846	24.91	7.18	6	42
OL	1,846	0.23	0.42	0	1

Social	1,846	3.09	1.17	1	5
PreviousTED	1,846	0.68	0.47	0	1
HigherEd	1,846	0.89	0.31	0	1
Age	1,846	33.69	10.65	18	80
Info	1,846	0.51	0.50	0	1
CategoryFirst	1,846	0.59	0.49	0	1
Clicks	1,846	1.85	1.65	1	16
ShareCategory	1,846	0.59	0.45	0.00	1.00
ShareT	1,846	0.01	0.04	0.00	0.50
ScrollDepth	1,792	11.22	34.78	1	828
UnsortedCategory	1,846	0.56	0.50	0	1
SortedCategory	1,846	0.09	0.28	0	1
AllSorted	1,846	0.35	0.48	0	1
ExtraVideos	1,846	0.06	0.23	0	1
ExtraSeconds	1,787	0.50	0.50	0	1

Table S3. Descriptive Statistics by Gender

	Men					**Women**				
Statistic	N	Mean	St. Dev.	Min	Max	N	Mean	St. Dev.	Min	Max
OL	1,034	25.00	6.92	6	42	812	24.80	7.49	6	42
OpinionLeader	1,034	0.22	0.41	0	1	812	0.24	0.43	0	1
Social	1,034	3.08	1.18	1	5	812	3.11	1.15	1	5
PreviousTED	1,034	0.73	0.45	0	1	812	0.62	0.48	0	1
HigherEd	1,034	0.88	0.33	0	1	812	0.91	0.28	0	1
age	1,034	31.91	9.54	18	80	812	35.95	11.52	18	76
Information	1,034	0.51	0.50	0	1	812	0.50	0.50	0	1
CatFirst	1,034	0.59	0.49	0	1	812	0.59	0.49	0	1
Clicks	1,034	1.87	1.66	1	16	812	1.83	1.65	1	14
ShareCategory	1,034	0.58	0.45	0.00	1.00	812	0.59	0.45	0.00	1.00
ShareT	1,034	0.01	0.04	0.00	0.50	812	0.01	0.05	0.00	0.50
ScrollDepth	995	11.24	36.83	1	828	797	11.19	32.05	1	813
UnsortedCategory	1,034	0.55	0.50	0	1	812	0.58	0.49	0	1
SortedCategory	1,034	0.09	0.28	0	1	812	0.08	0.27	0	1
AllSorted	1,034	0.35	0.48	0	1	812	0.34	0.47	0	1
ExtraVideos	1,034	0.05	0.22	0	1	812	0.07	0.25	0	1
ExtraSeconds	992	0.47	0.50	0	1	795	0.53	0.50	0	1

Table S4. Correlations – All Users

	OL score	*OL*	*Social*	*Previous TED*	*Higher Ed*	*Age*	*Info*	*Cat. First*	*Clicks*	*Share Cat.*	*ShareT*	*Scroll Depth*	*Unsorted Cat.*	*Sorted Cat.*	*All Sorted*	*Extra Videos*	*Extra Seconds*
OLscore	1	0.70	0.32	0.01	0.02	-0.06	-0.01	-0.05	0.02	-0.04	-0.01	0.03	-0.02	-0.04	0.04	0.01	0.03
OL	0.70	1	0.19	-0.01	0.02	0.02	0.03	0.01	0.04	0.02	0.02	0.001	0.03	-0.03	-0.02	0.01	0.02
Social	0.32	0.19	1	-0.11	0.08	0.04	0.01	-0.06	-0.03	-0.05	-0.05	0.01	-0.03	-0.0001	0.03	-0.02	0.03
PreviousTED	0.01	-0.01	-0.11	1	0.06	-0.11	0.04	-0.04	-0.03	-0.07	-0.02	0.05	-0.09	0.04	0.06	0.01	0.02
HigherEd	0.02	0.02	0.08	0.06	1	0.04	0.03	-0.003	-0.02	0.002	-0.04	0.03	0.02	0.01	-0.02	0.01	0.02
Age	-0.06	0.02	0.04	-0.11	0.04	1	-0.004	0.10	-0.05	0.11	0.0003	0.07	0.11	-0.03	-0.09	0.03	0.11
Info	-0.01	0.03	0.01	0.04	0.03	-0.004	1	-0.004	0.02	-0.01	0.0003	0.005	-0.02	-0.01	0.02	-0.01	0.06
CategoryFirst	-0.05	0.01	-0.06	-0.04	-0.003	0.10	-0.004	1	0.13	0.90	0.09	0.04	0.70	0.16	-0.83	0.05	-0.04
Clicks	0.02	0.04	-0.03	-0.03	-0.02	-0.05	0.02	0.13	1	0.13	0.29	-0.03	0.01	0.36	-0.23	-0.01	0.04
ShareCategory	-0.04	0.02	-0.05	-0.07	0.002	0.11	-0.01	0.90	0.13	1	0.09	0.04	0.90	-0.03	-0.92	0.04	-0.03
ShareT	-0.01	0.02	-0.05	-0.02	-0.04	0.0003	0.0003	0.09	0.29	0.09	1	-0.01	0.09	0.01	-0.10	0.02	0.01
ScrollDepth	0.03	0.001	0.01	0.05	0.03	0.07	0.005	0.04	-0.03	0.04	-0.01	1	0.04	-0.02	-0.03	0.07	0.03
UnsortedCategory	-0.02	0.03	-0.03	-0.09	0.02	0.11	-0.02	0.70	0.01	0.90	0.09	0.04	1	-0.35	-0.83	0.01	-0.02
SortedCategory	-0.04	-0.03	-0.0001	0.04	0.01	-0.03	-0.01	0.16	0.36	-0.03	0.01	-0.02	-0.35	1	-0.22	0.01	-0.01
AllSorted	0.04	-0.02	0.03	0.06	-0.02	-0.09	0.02	-0.83	-0.23	-0.92	-0.10	-0.03	-0.83	-0.22	1	-0.03	0.03
ExtraVideos	0.01	0.01	-0.02	0.01	0.01	0.03	-0.01	0.05	-0.01	0.04	0.02	0.07	0.01	0.01	-0.03	1	0.02
ExtraSeconds	0.03	0.02	0.03	0.02	0.02	0.11	0.06	-0.04	0.04	-0.03	0.01	0.03	-0.02	-0.01	0.03	0.02	1

Table S5. Clicks and scroll depth - OLS

	Dependent variable:			
	Men		**Women**	
	Clicks	*Scroll Depth*	*Clicks*	*Scroll Depth*
	(1)	(2)	(3)	(4)
OL	0.14	4.70	0.31	-0.78
	(0.18)	(4.09)	(0.20)	(3.95)
Social	-0.08*	-0.64	-0.02	1.57
	(0.05)	(1.04)	(0.05)	(0.99)
Previous TED	-0.19	5.71**	-0.10	3.15
	(0.12)	(2.67)	(0.12)	(2.37)
Higher Ed.	0.06	3.41	-0.33	2.37
	(0.16)	(3.54)	(0.21)	(4.08)
Age	-0.005	0.10	-0.01**	0.36***
	(0.01)	(0.12)	(0.01)	(0.10)
Info	0.07	-1.05	0.11	3.51
	(0.12)	(2.64)	(0.13)	(2.59)
*OL*Info*	0.13	-7.67	-0.30	-0.79
	(0.25)	(5.67)	(0.27)	(5.33)
Constant	2.27***	3.15	2.54***	-12.36**
	(0.28)	(6.27)	(0.32)	(6.23)
Observations	1,034	995	812	797
R^2	0.01	0.01	0.01	0.03
Adjusted R^2	0.001	0.003	0.004	0.02
Residual Std. Error	1.66 (df=1026)	36.77 (df=987)	1.65 (df=804)	31.78 (df=789)
F Statistic	1.19 (df=7; 1026)	1.45 (df=7; 987)	1.50 (df=7; 804)	2.94*** (df=7; 789)
Note:				*$^{*}p<0.1$, $^{**}p<0.05$, $^{***}p<0.01$*

Table S6a. Extra videos and extra seconds – Logit. Controlling for *CategoryFirst*

	Dependent variable:			
	Men		**Women**	
	Extra Videos	*Extra Seconds*	*Extra Videos*	*Extra Seconds*
	(1)	(2)	(3)	(4)
Category First	0.46	-0.21	0.39	-0.17
	(0.31)	(0.13)	(0.30)	(0.15)
OL	-0.48	0.08	0.38	0.06
	(0.56)	(0.23)	(0.47)	(0.25)
Social	-0.27**	-0.04	0.13	0.16**
	(0.13)	(0.06)	(0.13)	(0.06)
Previous TED	-0.10	0.21	0.34	0.07
	(0.32)	(0.15)	(0.31)	(0.15)
Higher Ed.	-0.001	0.04	0.17	-0.04
	(0.43)	(0.19)	(0.55)	(0.26)
Age	0.01	0.02***	0.01	0.02***
	(0.01)	(0.01)	(0.01)	(0.01)
Info	-0.48	0.12	0.18	0.44***
	(0.32)	(0.15)	(0.33)	(0.17)
*OL*Info*	1.24*	0.17	-0.54	-0.35
	(0.71)	(0.31)	(0.65)	(0.34)
Constant	-2.35***	-0.82**	-4.14***	-1.21***
	(0.75)	(0.35)	(0.85)	(0.42)
Observations	1,034	992	812	795
Log Likelihood	-206.29	-677.85	-195.83	-536.42
LR χ^2	11.393	15.87	5.396	26.233
$Prob > \chi^2$	0.1804	0.0443**	0.7145	0.00096***
Akaike Inf. Crit.	430.57	1,373.70	409.67	1,090.85
Note:		*p<0.1, **p<0.05, ***p<0.01		

Table S6b. Extra videos and extra seconds – Logit. Controlling for *ShareCategory*.

	Dependent variable:			
	Men		**Women**	
	Extra Videos	*Extra Seconds*	*Extra Videos*	*Extra Seconds*
	(1)	(2)	(3)	(4)
Share Category	0.51	-0.16	0.18	-0.20
	(0.34)	(0.15)	(0.32)	(0.16)
OL	-0.47	0.07	0.35	0.07
	(0.56)	(0.23)	(0.46)	(0.25)
Social	-0.27^{**}	-0.04	0.13	0.16^{**}
	(0.13)	(0.06)	(0.13)	(0.06)
Previous TED	-0.09	0.21	0.34	0.06
	(0.32)	(0.15)	(0.31)	(0.15)
Higher Ed.	-0.01	0.04	0.15	-0.04
	(0.43)	(0.19)	(0.54)	(0.26)
Age	0.01	0.02^{***}	0.01	0.02^{***}
	(0.01)	(0.01)	(0.01)	(0.01)
Info	-0.47	0.11	0.17	0.45^{***}
	(0.32)	(0.14)	(0.33)	(0.17)
*OL*Info*	1.23^{*}	0.17	-0.52	-0.35
	(0.71)	(0.31)	(0.65)	(0.34)
Constant	-2.38^{***}	-0.85^{**}	-3.99^{***}	-1.19^{***}
	(0.75)	(0.35)	(0.85)	(0.42)
Observations	1,034	992	812	795
Log Likelihood	-206.29	-678.55	-196.55	-536.31
LR χ^2	11.38	14.48	3.97	26.45
$Prob > \chi^2$	0.18	0.07*	0.86	0.0009***
Akaike Inf. Crit.	430.59	1,375.10	411.09	1,090.63

Note: $^{*}p<0.1$, $^{**}p<0.05$, $^{***}p<0.01$

Table S6c. Extra videos and extra seconds – Logit. Controlling for *UnsortedCategory*.

	Dependent variable:			
	Men		**Women**	
	Extra Videos	*Extra Seconds*	*Extra Videos*	*Extra Seconds*
	(1)	(2)	(3)	(4)
Unsorted Category	0.38	-0.09	-0.18	-0.17
	(0.30)	(0.13)	(0.29)	(0.15)
OL	-0.47	0.07	0.35	0.08
	(0.56)	(0.23)	(0.46)	(0.25)
Social	-0.28**	-0.03	0.12	0.16**
	(0.13)	(0.06)	(0.13)	(0.06)
Previous TED	-0.09	0.21	0.32	0.06
	(0.32)	(0.15)	(0.31)	(0.15)
Higher Ed.	-0.02	0.04	0.12	-0.04
	(0.43)	(0.19)	(0.54)	(0.26)
Age	0.01	0.02***	0.01	0.02***
	(0.01)	(0.01)	(0.01)	(0.01)
Info.	-0.47	0.11	0.15	0.45***
	(0.32)	(0.14)	(0.33)	(0.17)
*OL*Info.*	1.22*	0.17	-0.49	-0.36
	(0.71)	(0.31)	(0.65)	(0.34)
Constant	-2.28***	-0.88**	-3.77***	-1.21***
	(0.74)	(0.35)	(0.84)	(0.42)
Observations	1,034	992	812	795
Log Likelihood	-206.61	-678.89	-196.51	-536.42
LR χ^2	10.757	13.795	4.0441	26.241
$Prob > \chi^2$	0.216	0.087*	0.853	0.00096 ***
Akaike Inf. Crit.	431.21	1,375.78	411.02	1,090.84
Note:				**p<0.1, **p<0.05, ***p<0.01*