MIS Quarterly

RESEARCH ARTICLE

DO SEARCH ENGINES INFLUENCE MEDIA PIRACY? EVIDENCE FROM A RANDOMIZED FIELD STUDY[1]

Liron Sivan, Michael D. Smith, and Rahul Telang
Heinz College, Carnegie Mellon University, Pittsburgh, PA 15213 U.S.A.
{lsivan@andrew.cmu.edu} {mds@cmu.edu} {rtelang@andrew.cmu.edu}

Many papers in the literature have analyzed the role search rank plays in influencing user choices for the same product offered through different links. However, the literature has not analyzed whether search position can cause users to change their consumption between two ex ante distinct product categories.

We analyze this question in the context of the ongoing public policy debate surrounding the role search engines can play in anti-piracy efforts. Specifically, we analyze whether reducing the prominence of infringing links can impact users choices between legal and infringing content. To do this we design a customized search engine that allows us to manipulate the positions of infringing and legal links in users' search results. We then use this search engine to conduct experiments on a general population of users and on a subset of college-aged users.

Our data show that reducing the prominence of infringing links in search results causes users who otherwise would have consumed infringing content to switch their consumption to paid legal content, and that these results hold even among users whose initial search queries express an explicit preference for infringing content. These results suggest that even small changes in the cost of discovering pirated content can have a relatively large impact on user behavior. As such, our results inform an important public policy debate by showing that search engines have a vital role to play in the fight against online copyright theft.

Keywords: Search engines, piracy, copyright, anti-piracy, randomized experiment

Introduction

Many studies in the academic literature have established that search rank has a strong impact on users' click-through and purchase decisions. However, these studies assume that the search results all pertain to the same set of products. What is not known in the literature is whether search rank can cause users to switch consumption between two *ex ante* distinct sets of products.

We analyze this question in the context of the public policy debate regarding how to encourage users who consume pirated content to switch to legal channels. Specifically, we ask whether reducing the prominence of infringing links in search results can cause users who would have otherwise consumed free "pirate" content to switch their consumption to paid legal channels. Our study thus informs an important policy debate and extends the search rank literature by examining user behavior in the presence of two different types of content: legal and infringing.

The role search engines should play in fighting piracy has strong views on both sides. Search engines argue that they already do a great deal to stop piracy. For example, in August 2012, Google started taking valid copyright removal notices

[1]Bin Gu was the accepting senior editor for this paper. Siva Viswanathan served as the associate editor.

The appendices for this paper are located in the "Online Supplements" section of *MIS Quarterly*'s website (https://misq.org).

into account when ranking sites in search results,[2] and in September 2013, Google released a report titled "How Google Fights Piracy"[3] documenting the scale of the challenge Google faces: In January 2016 alone, Google responded to more than 68 million copyright removal notices affecting nearly 75,000 unique Internet domains.[4]

At the same time, content owners argue that search engines' efforts have been ineffective for most searches and that more intervention is needed. For example, a recent study funded by the Motion Picture Association of America (MPAA)[5] analyzed the role that search results play in the discovery of television and movie content, arguing that many queries resulting in infringing links do not contain keywords that indicate a specific intent to view pirated content.

However, given the differences between paid legal content and free pirated content, it is unclear whether making one type of content harder to find in search results will increase consumption of the other. In an editorial on the Recording Industry Association of America's (RIAA) desire for Google to make piracy harder to find, the technology blog *Techdirt* summarizes the view of many in the technology industries on this question, arguing that given the differences between these products it will be difficult to convince people interested in one type of product to switch to the other:

> The RIAA might not like it, but the simple fact is that when people are searching for [artist] [track] mp3 and [artist] [track] download, chances are they're not looking to buy, but to download for free . … Even if Google magically did show them Apple, Amazon and Emusic as the top results for every [artist] [track] mp3 and [artist] [track] download, the people doing those searches wouldn't go there, because they're not looking to buy.[6]

In this view, if an individual user is looking for infringing content, they have already made up their mind to consume infringing content, and minor changes in search results will not cause them to switch to legal channels. Thus, although it is well known in the literature that search rank matters when users are choosing among links for the same type of product, it is an open question whether this applies in our setting where pirate and legal links may belong to *ex ante* different sets of content. This view is summarized nicely in the words of the technology blog, *ExtremeTech*: "In general, if someone wants to pirate something, they're going to pirate it, even if they have to click through a couple of pages of search results."[7]

In short, although many copyright holders have called for Google and other major search engines to reduce the prominence of pirate links in search results, a fundamental question remains: Can search results cause users who otherwise would have consumed pirated content to switch their consumption to legal content? If anything, answering this question has become more important recently given Google's changes to its search algorithm to more aggressively demote pirated links,[8] the dramatic increase in takedown notices processed by Google,[9] and the entertainment industry's calls for continued action.[10]

In spite of the importance of these questions for academic researchers, managers, and policymakers, we are aware of no studies that empirically analyze the role search results play in the choice between legal or infringing content. One reason for the few studies in this area is the difficulty in assigning causation. By design, the top search results are likely to be the most "relevant" to the user. If a user searches for infringing content and infringing results are listed first, it is impossible to disentangle whether she clicks on that content because of her interest or because of its placement. In short, a user's desire to look for a particular type of content shapes their behavior, making it impossible to use observational or archival data alone to show whether search results influence user choices.

In this paper, we address this methodological challenge with randomized field experiments, using the experimental design

[2]See Google's press release on this program: http://insidesearch.blogspot.com/2012/08/an-update-to-our-search-algorithms.html.

[3]https://docs.google.com/file/d/0BwxyRPFduTN2dVFqYml5UENUeUE/edit.

[4]See http://www.google.com/transparencyreport/removals/copyright/.

[5]"Understanding the Role of Search in Online Piracy," prepared by Millward Brown Digital for the MPAA.

[6]"RIAA: Google Isn't Trying Hard Enough to Make Piracy Disappear from the Internet," Mike Masnick, February 21, 2013 (https://www.techdirt.com/articles/20130221/07560622055; accessed November 24, 2014).

[7]"Google Finally Decides to Demote 'Notorious' Piracy Sites in Search Results," Sebastian Anthony, October 20, 2014 (http://www.extremetech.com/extreme/192471-g; accessed November 24, 2014).

[8]See, for example, http://googlepublicpolicy.blogspot.com/2014/10/continued-progress-on-fighting-piracy.html (accessed December 9, 2014).

[9]From June 2012 to January 2016, the number of copyright notices Google processed per month increased nearly three-fold from 25 million to 68 million.

[10]See, for example, https://torrentfreak.com/google-counsel-sees-problems-with-take-down-stay-down-151121/.

to avoid contamination due to user intent. We then implement this design in two main experiments: one where participants are drawn from a representative panel of the U.S. population, and one with a panel of college-aged participants. Both groups are recruited through an independent company maintaining large survey panels. We expose these users to a potential task of finding a movie through online channels, and encourage them to use our custom-built search engine in place of the search engine they would normally use (e.g., Bing, Google).

In the first experiment (representative panel), our search engine displays results to users in one of three randomly assigned conditions. In the control condition users see the same results that would be displayed from a major search engine. In the first treatment condition for this experiment, infringing links are artificially promoted in the search results, and in the second treatment condition legal links are artificially promoted. The second experiment (college-aged users) adds two additional treatment conditions to test milder legal and milder infringing search manipulations.

We avoid the user intent fallacy by randomly assigning users to one of these treatment conditions. We also record and examine their search choices through our interface, providing additional detail on their behavior within the experiment. Finally, we ask them to complete a questionnaire at the end of the experiment to measure their attitudes regarding piracy.

Our results suggest that individual users consider legal and infringing links to belong to *ex ante* differentiated sets of products. Our clickstream data show that users who have consumed infringing content in the past and users whose search queries express a preferences for infringing content exhibit higher search intensity when placed in the legal treatment condition than they do in the control or infringing conditions, and likewise for users with less experience with infringing content or a stated preference for legal content when placed in the infringing treatment condition.

Having established this difference, we then find that the prominence of search results can cause users who would have otherwise purchased legally to switch to infringing content, and vice versa for users who otherwise would have consumed infringing content. In the control condition about 80% of users choose to buy the product legally versus 56% in the infringing treatment condition.[11] We see similar results in a second experiment with a college-aged population (18–24 year-old users): 62% of control users choose to purchase legally versus 39% in the infringing treatment condition. Our second experiment also shows that stronger treatments lead to stronger outcomes: In the "mild" infringing treatment the number of purchases rises from 39% to 48%.

Our data also allow us to test whether these results extend to users who have a stated preference for legal or infringing content. We do this by using a user's initial search terms to infer their intention to pirate or consume legally. Classifying user intentions in this way, we find that our main results hold among users with a stated preference for legal or infringing content: users who express an intention to consume legally are less likely to do so in the infringing treatment condition than in other conditions, and users who initially express an intention to consume pirated content are less likely to do so in the legal treatment condition than in the other conditions.

Together, our results suggest that search rank can influence users' choices between two *ex ante* sets of products: paid legal consumption and free illegal consumption. This finding in turn suggests that reducing the prominence of pirated links in search results can be a viable strategy for fighting intellectual property theft for both a general population of users and for younger (college-aged) users, and for both "undecided" users and users with a pre-existing preference for legal/pirated content.

Related Literature

Our research relates to several streams of the economics and information systems literatures. First, our research relates to studies analyzing the impact of piracy on sales. Within these literatures, the vast majority of papers find that piracy harms sales (for literature reviews, see Danaher, Smith, and Telang 2014; Liebowitz 2008; Oberholzer-Gee and Strumpf 2010). With this result well established in the literature, recent papers have analyzed the effectiveness of efforts to reduce the impact of piracy. These papers generally find that making content available in legal digital channels (e.g., Danaher et al. 2010; Danaher et al. 2015), targeting the demand-side of piracy (e.g., the HADOPI anti-piracy law in France; Danaher, Smith, et al. 2014), and targeting the supply-side of piracy (e.g., the shutdown of Megaupload; Danaher and Smith 2014) can all be effective in changing user consumption of pirated content. Beyond these legislative interventions, Reimers (2014) shows that industry-led notice and takedown strategies for eBooks can be effective at increasing legal sales, and Bhattacharjee et al. (2006) find that RIAA lawsuits lowered levels of file sharing, with a much greater impact on high-level sharers.

[11] As we note in more detail below, our experiment is designed to identify relative differences between the two treatment conditions and the control condition (as opposed to absolute levels of piracy).

Our research also informs the academic literature on consumer behavior using search engines. Previous studies have examined how the prominence of search results influences user behavior, showing that position has a significant impact on click-through and conversion rates when consumers choose among links for the same products (In the context of sponsored search, see Agarwal et al. 2011; Yang and Ghose 2010; in the context of organic search, see Baye et al. 2012). For organic search results, Brooks (2004) shows that click-through rates are higher for links that are placed higher in the results listing. Similarly, an eye tracking experiment performed by Pan et al. (2007) revealed that college students trust Google's ability to rank results by their true relevance, such that users' decisions are strongly biased toward links higher in position even, if the abstracts themselves are less relevant. However, there have been relatively few studies analyzing the impact of position on conversion rates and revenue. Among these studies, Agarwal et al. (2011) find that although click-through rates decline with position, conversion rates (and therefore revenue) increases with position for more specific keywords and as a result "the topmost position is not necessarily the revenue- or profit-maximizing position" (p. 1057).

Our research extends these literatures by first analyzing how search engine ranking affects users' substitution between two *ex ante* different sets of products: legal content and infringing content. In this way, our study extends the existing literature which focuses on *ex ante* similar products, to a setting where the differences in the types of products could make the trade-offs between different links (legal and pirate) quite high, particularly for consumers who may have a stated preference for one type of content or the other. Our study also extends the literature on the effectiveness of anti-piracy measures by informing an active policy question: Can changes in search results can be an effective tool to switch users from piracy to legal consumption?

Experiment 1: General Population, Two Treatment Conditions

Experimental Design

All of our experiments utilize a custom-built search engine to test the impact of search results on media consumption choices among a general population of Internet users. In experiment 1, participants were recruited from the general population by a company maintaining a large survey panel. The study was performed online in three parts: (1) a screening phase in which the participants were asked to choose a movie they wish to acquire; (2) a search phase in which the participants were asked to use our custom search engine to search for a source to acquire the movie; and (3) a post-experiment questionnaire.[12]

During the screening phase, participants were asked whether they were interested in watching a movie, and were presented with a list of 50 alternatives (see Figure 1). Potential movies were selected in advance to make sure that our manipulation conditions could be implemented. Users were asked to select a movie (as opposed to assigning them a movie) to ensure that the users were motivated to find a movie they wished to view. Consistent with our goal of observing user behavior in search for movies, we excluded from the experiment any participants who stated that they were not interested in watching any of the movies in our list.

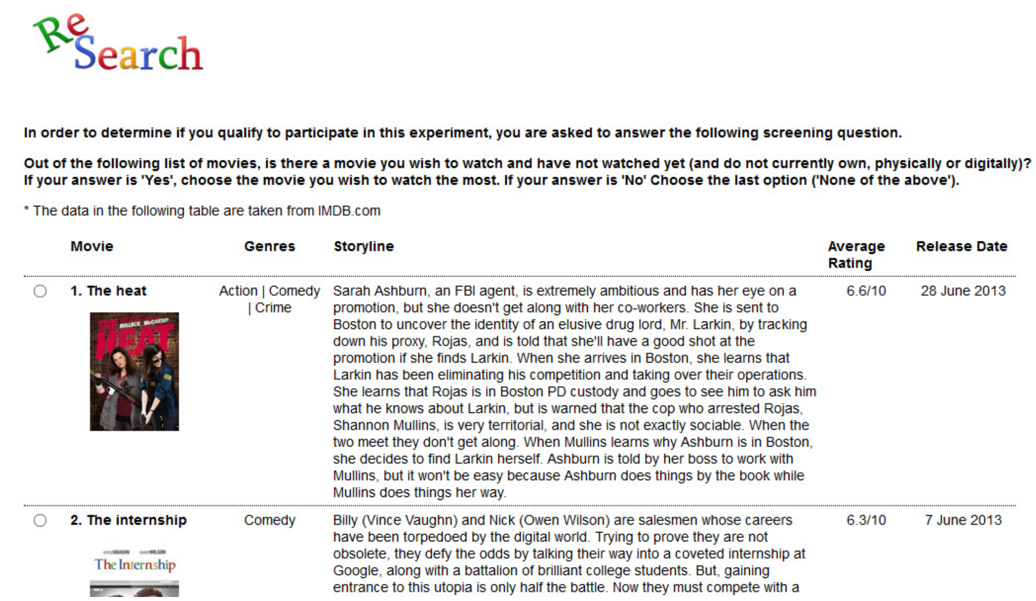
Participants who chose a movie were tasked to search for an online source (download, stream, purchase, or rent) to obtain the movie. In the instructions (see Figure 2), participants were told this was an experiment to test the effectiveness of our search engine and therefore they should use our search engine in place of whatever other search engine they would normally use. The instructions also stated that if the participant had a specific website in mind to obtain the movie, they could go directly to that site after initially trying to search for the movie using our search engine.

Consistent with standard practice in the experimental economics literature, participants were given a $20 prepaid virtual Visa card as compensation for their time in completing the experiment and were able to keep the movie they acquired and any remaining money from their $20 card.[13] Compensating experimental participants in this way is consistent with numerous studies in the economics, marketing, and information systems literature (see, for example, Dhar et al. 2007; Haisley et al. 2008; Haws and Winterich 2013; Isaac and Davis 2006; Löschela et al. 2013; Rucker et al. 2011; Suk et al. 2012; Tsai et al. 2011). As discussed in more detail in Appendix A, although an unrestricted endowment of this sort may change the overall likelihood of making a purchase, it will do so equally for both the treated and control group users. As such, any difference in response between these two groups can reliably be attributed to the experimental intervention.[14]

[12] More details on the experimental methodology are provided in Appendix A.

[13] We also stated that their identity is unknown to us and their behavior cannot be traced back to them.

[14] We also note that our results in this experiment, where users are given a $20 gift card, are consistent with results we present in experiment 3 below, where Amazon Mechanical Turk users are only compensated by $1.50 for their time, further suggesting that our main results are not driven by the endowment effect.



ReSearch

In order to determine if you qualify to participate in this experiment, you are asked to answer the following screening question.

Out of the following list of movies, is there a movie you wish to watch and have not watched yet (and do not currently own, physically or digitally)? If your answer is 'Yes', choose the movie you wish to watch the most. If your answer is 'No' Choose the last option ('None of the above').

* The data in the following table are taken from IMDB.com

Movie	Genres	Storyline	Average Rating	Release Date
1. The heat	Action | Comedy | Crime	Sarah Ashburn, an FBI agent, is extremely ambitious and has her eye on a promotion, but she doesn't get along with her co-workers. She is sent to Boston to uncover the identity of an elusive drug lord, Mr. Larkin, by tracking down his proxy, Rojas, and is told that she'll have a good shot at the promotion if she finds Larkin. When she arrives in Boston, she learns that Larkin has been eliminating his competition and taking over their operations. She learns that Rojas is in Boston PD custody and goes to see him to ask him what he knows about Larkin, but is warned that the cop who arrested Rojas, Shannon Mullins, is very territorial, and she is not exactly sociable. When the two meet they don't get along. When Mullins learns why Ashburn is in Boston, she decides to find Larkin herself. Ashburn is told by her boss to work with Mullins, but it won't be easy because Ashburn does things by the book while Mullins does things her way.	6.6/10	28 June 2013
2. The internship	Comedy	Billy (Vince Vaughn) and Nick (Owen Wilson) are salesmen whose careers have been torpedoed by the digital world. Trying to prove they are not obsolete, they defy the odds by talking their way into a coveted internship at Google, along with a battalion of brilliant college students. But, gaining entrance to this utopia is only half the battle. Now they must compete with a	6.3/10	7 June 2013

Note: The list of movies includes IMDB.com data on Genre, storyline, average reviewers rating, and theatrical release date. The movies included in the study were chosen from IMDB's weekly top DVDs lists during the months prior to the experiment.

Figure 1. Screening Question

We have designed a search engine which is especially efficient at searching content related to movies (URL: search.co.be). The goal of this experiment is to understand if (and how) people use search engines to search for movie content.

As part of the experiment you are asked to find a website (source) where you would like to get this movie. Please use our search engine (do not use Google, Bing, etc.) to search on the Internet. You can also use your own installed programs (if you have any) to reach the desired source or directly go to the specific source to acquire the movie (by entering appropriate URL on a browser for example), but please make sure to search with our search engine first. Once you reach the desired source, you should download/stream/purchase/rent the movie. If you need to purchase the movie you can use the provided virtual credit card to do so.

Figure 2. Experiment Task

Before the experiment started, participants were randomly assigned to one of the following three search engine conditions: no manipulation, non-infringing (legal) content manipulation, or infringing (piracy) content manipulation.

- **Condition 1: No manipulation (control)**: The first 100 search results were retrieved from a major search engine and were displayed to the searcher without manipulation. The search results consist of 10 pages with 10 search results on each page.

- **Condition 2: Legal content manipulation**: The first 100 search results were retrieved from a major search engine and were displayed such that the first 3 results on each of the 10 pages were replaced (if necessary) with results offering legal options to rent/purchase the movie. Additionally, all infringing links on the first page were replaced with legal options such as Amazon.com or iTunes.[15] See Figure 3 for sample search results in the legal treatment condition.

[15] Neutral results in positions 4–10 (such as IMDB.com, Wikipedia.com, etc.) were left unchanged.

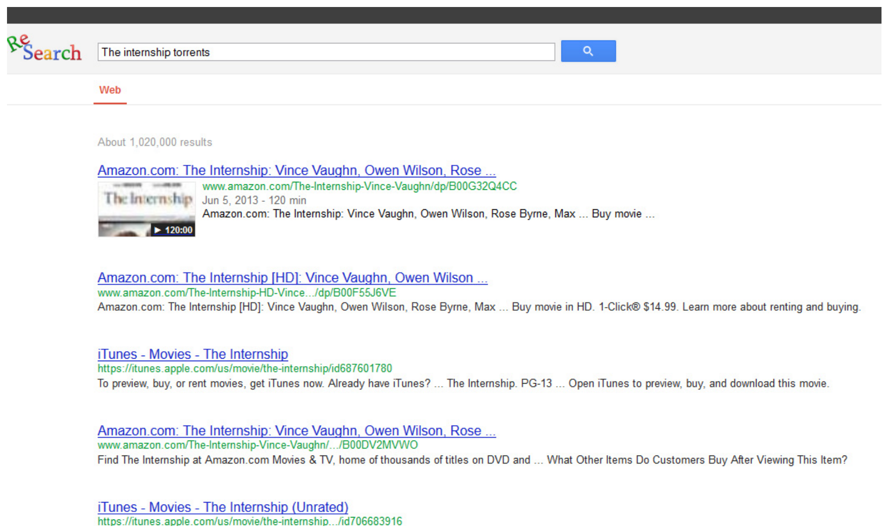


Figure 3. Example Search Results for the Legal Content Manipulation Condition

- **Condition 3: Piracy content manipulation**: The first 100 search results retrieved from the major search engine were displayed such that the first 3 results on each of the 10 pages are replaced (if necessary) with results offering piracy options to obtain the movie. Additionally, all legal options on the first page were replaced with piracy options.[16]

After completing the experiment, participants were given a post-experiment survey including questions about the source from which they acquired the movie, its price ($0 if acquired from an infringing source), several demographic questions, and questions about their media consumption preferences (see Figure 4 for sample post-experiment questions).

To ensure that participants performed the task correctly, we only included participants who claimed the $20 virtual Visa card, used our custom-built search engine, and reported the source from which they acquired the movie. In total, 235 completed the task, and 196 also completed the post-experiment questionnaire. As discussed in Appendix A, dropout rates are not statistically different across experimental conditions and the distribution of user characteristics (demographics, attitudes towards piracy) is similar across the control and treatment groups.

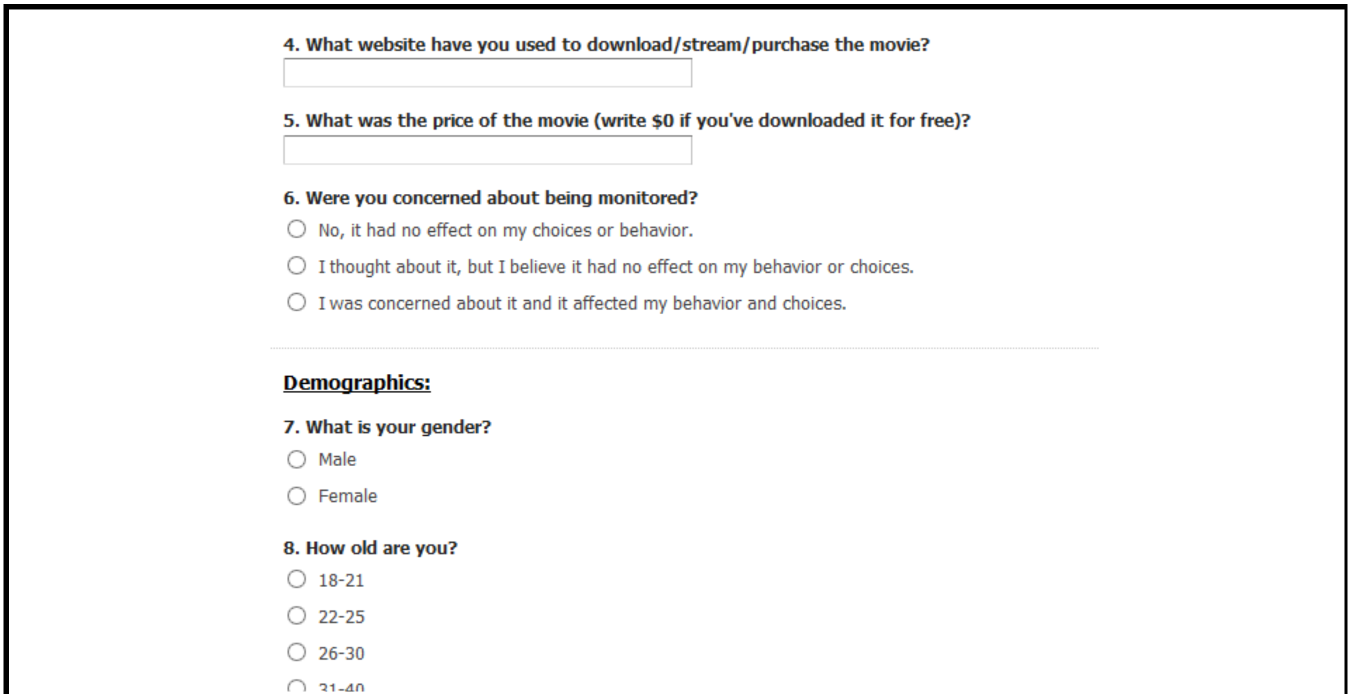
To verify similarity in preexisting attitudes toward piracy, we tested whether participants' initial intent for pirate or legal content is similar across the experimental conditions. We did this by classifying whether their initial search terms reflected neutral, legal, or infringing intent according to the degree to which pirate or legal links were present in the (unmodified) search results for similar queries commonly issued by our users.[17] Similar to the user characteristics described above, and consistent with the experimental assignment, the distributions are not statistically different across the control and treatment conditions.

Results

Table 1 compares the proportion and average price of legal purchases across the treatment conditions. In the control condition, 80% of participants acquired the movie through a legal channel (the remaining 20% chose a pirated channel). Relative to this baseline, participants in the legal treatment condition were significantly more likely to acquire the movie from a legal channel (P = 94.4%, SD = 23.2%) than were participants in the baseline condition (P = 80.0%, SD = 40.3%), a difference which is statistically significant ($t(121) = 2.85$, $p < .01$). Conversely, participants in the infringing content treatment condition were significantly less likely to acquire content from a legal channel (P = 56.9%, SD = 49.9%) relative to the baseline condition (P = 80.0%, SD = 40.3%). Again, a t-test confirms that the differences are statistically significant ($t(91) = 2.43$, $p < .01$).

[16]As in condition 2, neutral results in positions 4–10 (such as IMDB.com or Wikipedia.com) were left unchanged.

[17]The details of this classification can be found in Appendix A.



4. What website have you used to download/stream/purchase the movie?

5. What was the price of the movie (write $0 if you've downloaded it for free)?

6. Were you concerned about being monitored?

- No, it had no effect on my choices or behavior.
- I thought about it, but I believe it had no effect on my behavior or choices.
- I was concerned about it and it affected my behavior and choices.

Demographics:

7. What is your gender?

- Male
- Female

8. How old are you?

- 18-21
- 22-25
- 26-30
- 31-40

Figure 4. Post-Experiment Questionnaire

Table 1. Between-Conditions Comparison of the Proportion of Purchases Made from Non-Infringing Websites.

	n	# Legal Option	% Legal Otion	Average Price (conditional on acquiring a legal copy)
Condition 1: No manipulation	60	48	80.0%	$9.80
Condition 2: Legal content manipulation	71	67	94.4%	$9.89
Condition 3: Infringing content manipulation	65	37	56.9%	$9.93

These results, and the statistically significant differences across groups, strongly suggest that reducing the prominence of pirated content in search results can have a significant impact on a users' propensity to choose legal or pirated content, and thus is a viable anti-piracy strategy. This result holds in spite of the fact that pirate or legal links are still readily available to users beyond the first page of search results.[18] This suggests that, in contrast to the conventional wisdom about piracy behavior cited above, users in our sample are willing to substitute between legal and infringing consumption channels based on relatively small changes in search engine ranking.

To further explore the willingness to substitute between channels, we limit our analysis to only those users whose search terms indicate an initial preference for legal or pirated content. This is important because these users might be considered the most "committed" to consume through legal or pirated channels, and thus the least likely to change their behavior. We do this by following the methodology described in Appendix A, classifying each user's "intent" based on whether their initial search terms express a preference for legal or illegal content. In Table 2, we compare the purchase likelihood for users with pirated or legal intent in the different treatment groups.

[18]For example, in the control condition, pages 2–5 in the search results have, on average, 1.88 links per page to products sold by Amazon, iTunes, or Google Play compared to 1.72 links per page for search results in the infringing treatment condition.

Table 2. Legal Purchase Rates Across Treatment Conditions and Initial Intent			
First Search Term	**Control Group**	**Legal Content Manipulation**	**Infringing Content Manipulation**
Legal intent	31/31 (100%)	22/23 (96%)	**24/33 (73%)**
Infringing intent	4/12 (33%)	**15/17 (88%)**	5/11 (45%)

Each cell in the table reports the number of users observed in a particular experimental and "initial intent" condition, and the number and proportion of users who made a legal purchase. For example, Table 2 shows that 33% (4 of the 12) of control group users who initially expressed intent to consume through infringing channels, ultimately purchased through a legal channel.

This table shows that users who initially express an intent to consume legally are statistically significantly less likely to do so when placed in the infringing content manipulation (73%) than in either the legal (96%) or control (100%) conditions ($t(48) = 2.55$, $p < 0.01$; $t(32) = 3.46$, $p < 0.01$, respectively). Likewise, users who initially express an intent to consume pirated content are significantly more likely to purchase legally in the legal content manipulation (88%) than in the infringing (45%) or control (33%) conditions ($t(15) = 2.42$, $p < 0.05$; $t(18) = -3.36$, $p < 0.01$, respectively). These results demonstrate that the ranking of pirated and legal search results matters even among users with an initial preference for pirate or legal channels. In Appendix B, we show that these results hold in a logistic regression model controlling for observable participant characteristics. In appendix C, we show that these results hold if, instead of classifying intent based on search keywords, we classify legal and infringing intent based on whether a user had consumed pirated content in the prior 12 months.

One might be concerned that users don't distinguish between legal and infringing content, such that all users tend to click on the first link they see regardless of whether it is from a legal or infringing source. If that were true, our results in Table 2 could simply arise from the well-known fact that search rank matters and that users are more likely to click on the first links displayed in search results. To test whether users in our sample perceive a difference between legal and infringing content, we analyze our clickstream data with respect to the number of searches and the average and maximum position of search for users we classify as having a preference for legal or infringing content. If these users are indifferent between legal and infringing links, we would expect that their search intensity would be nearly the same regardless of their treatment condition.

Our clickstream data suggest that, in contrast to this hypothesis, users with a preference for a particular type of content search more when placed in their non-preferred treatment condition. Specifically, our clickstream data show that consumers with legal intent initiate on average 2.68 searches in the control condition ($n = 31$) and 4.22 searches in the infringing treatment condition ($n = 33$) (statistically significant difference at the 1% level). Likewise, consumers with infringing intent initiate, on average, 2.50 searches (12) in the control condition and 4.08 searches (13) in the legal treatment condition (statistically significant difference at the 10% level). We see similar results when we consider the average position of search instead of the number of searches. Consumers with legal intent click on results at an average position of 3.52 (31) in the control condition and 8.94 (27) in the infringing treatment condition (statistically significant difference at the 1% level), and consumers with infringing intent click on an average position of 2.97 (11) in the control condition and 3.79 (13) in the legal treatment condition (insignificant). Finally, we see consistent results if we consider the maximum position that each consumer clicks on during their searches. Averaging across consumers, consumers with legal intent search to a maximum position of 5.10 (31) in the control condition, and to a maximum position of 13.78 (27) in the infringing treatment condition (statistically significant difference at the 1% level). And consumers with infringing intent search to a maximum position of 4.64 (11) in the control condition and a maximum position of 7.09 (13) in the legal treatment condition (statistically significant difference at the 10% level). Together these results suggest that users treat legal and infringing content distinctly in their search behavior.[19] The fact that users choose to search more in their non-preferred treatment condition is evidence that they perceive a difference between legal and infringing links. The fact that some users eventually switch is evidence that lower search rank can cause users to switch from their preferred content (legal or pirate) to their non-preferred content.

[19] As a further check of whether our results are driven by a lack of perceived differentiation between legal and pirate links, in Appendix E we conduct an additional experiment where we add two new treatment conditions that explicitly distinguish infringing links from other types of links through the use of a flag. Our results for links with the flag are statistically the same as those without the flag, suggesting that users treat legal and infringing links differently.

Experiment 2: Younger Audience, Four Treatment Conditions

Experimental Design

In a second experiment, we obtained participants from the same independent survey panel company used in experiment 1. However, in this experiment, we limited participation to users who were 18 to 25 years old and who were college students at the time of the experiment or who had at least 2 years of college education. We did this to analyze the degree to which our results extend to a college-aged population, a population that tends to be disproportionately more likely to pirate than other demographic segments (e.g., Rob and Waldfogel 2007, Vandiver et al 2012).

We also added two additional treatment conditions in this second experiment. We retain our original legal and infringing manipulations and add two additional manipulations that soften the impact of these manipulations by only changing the top three results on the first page and making no other changes to the displayed results:

- **Condition 2a: Mild legal content manipulation**: The first 100 search results were displayed to the searcher such that the first 3 results on the first page (and only the first page) were replaced (if necessary) with results that offer legal options to rent/purchase the movie.

- **Condition 3a: Mild piracy content manipulation**: The first 100 search results were displayed to the searcher such that the first 3 results on the first page (and only the first page) were replaced (if necessary) with results that offer infringing options to download/stream the movie.

Data and Results

Following the approach in experiment 1, a total of 234 participants qualified for our study.[20] Similar to experiment 1, the demographic characteristics, attitudes toward piracy, and initial search intent, have similar distributions across the control and treatment groups. More details on this experiment are included in Appendix D.

In Table 3, we compare the proportion of legal purchases (and the resulting average price of a legal purchase) made by participants in each of the experimental conditions. This table shows that in the control condition, where search results were not manipulated, 61.9% of the participants chose to acquire the movie through a legal channel (and the remaining 38.1% consumed through a pirated channel). Compared to the general population sample, where 80% of participants in the control condition acquired through a legal channel, these data suggest that younger users are less likely to acquire content through legal channels than are users in the general population.

Relative to the proportion of legal purchases in the control condition, Table 3 shows that participants who were assigned to the mild legal content treatment condition were more likely to acquire the movie from a legal channel (P = 75.5%, SD = 43.4%) than were participants in the baseline condition (P = 61.9%, SD = 43.4%), a difference that is statistically significant ($t(83) = -1.39$, $p < 0.1$). We also see that the effect of the more intense legal content manipulation (which is the same as the legal manipulation in experiment 1) resulted in a statistically higher proportion of legal purchases than milder legal manipulation condition (P = 91.7%, SD = 27.9%, t-test: $t(63) = -2.18$, $p < .05$). Conversely, the results show that participants in the mild infringing condition were significantly less likely to acquire content from a legal channel (P = 47.7%, SD = 50.5%) relative to the baseline condition (P = 61.9%, SD = 43.4%; t-test: $t(84) = 1.32$, $p < 0.1$), and that the more intense infringing content manipulation (which is the same as the infringing manipulation used in experiment 1) is stronger than that of the mild treatment, causing a directionally lower proportion of legal purchases versus the control (P = 38.2%, SD = 49%). However, this difference is not statistically significant (t-test: $t(90) = 0.83$, $p > 0.1$), possibly due to the relatively small sample size.

These results confirm the results in experiment 1 for a younger set of users, and also suggest that the effect varies with the intensity of the treatment, and is present even for relatively minor reductions in search engine ranking.

We next focus on users whose initial queries express intent to consume infringing or legal content using the same categorization method described in experiment 1. The results of this analysis, displayed in Table 4, suggest that users who initially express intent to consume legally are significantly less likely to do so when placed in the infringing content manipulation (63%) than in the other treatment conditions (p < 0.05%). Likewise, users who initially express intent to consume infringing content are significantly more likely to consume legally in the legal content manipulation (71%) than in the other treatment conditions (p < 0.10%). In Appendix D, we

[20] Approximately 650 participants were invited to participate in the study and expressed an interest in watching one of the movies in the study. A total of 550 participants logged into the system and were presented with the task details. Out of these participants, 270 completed the experimental task as instructed, and of these participants, 234 qualified for our study by also completing the post-experiment questionnaire. The dropout rates in each of the different stages described above are not statistically different across experimental conditions.

Table 3. Between-Conditions Comparison of the Proportion of Purchases Made from Non-Infringing Websites in Study 2

	n	# Legal option	% Legal option	Average Price (conditional on acquiring a legal copy)
Condition 1: No manipulation	42	26	61.9%	$12.26
Condition 2: Mild legal content manipulation	49	37	75.5%	$9.57
Condition 2a: Legal content manipulation	48	44	91.7%	$11.84
Condition 3: Mild infringing content manipulation	44	21	47.7%	$12.76
Condition 3a: Infringing content manipulation	51	20	39.2%	$14.75

Table 4. Legal Purchase Rates Across Treatment Conditions and Initial Intent

		Legal Content Manipulation		Infringing Content Manipulation	
First search term	Control	Mild	Intense	Mild	Intense
Legal intent	21/23 (91%)	18/19 (95%)	20/20 (100%)	16/18 (89%)	**17/27 (63%)**
Infringing intent	0/13 (0%)	4/14 (29%)	**10/14 (71%)**	7/15 (47%)	2/9 (22%)

show that these results are robust to classifying user intent based on whether a users had consumed significant amounts of pirated content in the prior 12 months instead of based on their initial queries. Appendix D also shows that our results are robust to a logistic regression that controls for differences between groups based on observed characteristics.

In summary, the results from experiment 2 are consistent with those in experiment 1 in that, by their observed search behavior, users consider legal and infringing content to be distinct sets of products, that the ordering of results can cause users who would have otherwise pirated to switch their consumption to legal content and vice versa, and that these results hold even among users who initially state a preference for legal or infringing content. In addition, experiment 2 shows that these results hold for younger users and that stronger treatments yield stronger responses.

Discussion

Our goal in this research is to analyze whether changes in search rank can cause users to switch their consumption between two *ex ante* distinct categories of products. This question extends the prior literature that has analyzed the role search rank plays in influencing user choice among *ex ante* similar products.

We conduct our research in the context of consumption through paid legal channels and free pirate channels. Our empirical approach uses an experimental design to simulate consumers' online movie search and consumption processes. Our custom search engine allows us to experimentally manipulate the rank and positioning of pirate and legal links in search results.

In the first experiment, we studied a representative sample of the population recruited through an independent survey company. We then exposed these users to three randomly assigned search conditions: a control condition, which displayed search results from a major search engine; an infringing content treatment condition, which artificially promoted infringing sites in the search results; and a legal content treatment condition, which artificially promoted legal sites in the search results. In the second experiment, we recruited an additional sample of college-aged students (18–25 year olds) as participants, and added two additional treatment conditions: a mild legal treatment and a mild infringing treatment.

Our results in these two experiments suggest that the ordering of search results for pirated content strongly impacts users' decisions of whether to purchase legal content. Our first experiment shows that 80% of users in the control condition choose to purchase content through legal channels. Relative to this baseline, we find that 94% of users in the legal treatment and 57% of users in the infringing treatment purchase content through legal channels. Similarly, in the second experiment with a college-aged audience, 62% of users in the control condition purchase content versus 92% in the legal treatment and 39% in the infringing treatment. The second experiment also shows that stronger treatments lead to stronger outcomes: 76% of users purchase in the mild legal treatment (versus 92% in the legal treatment) and 48% of users purchase in the mild infringing treatment (versus 39% in the infringing treatment).

We also show that these results hold even among users whose initial search terms reflect a preference for legal or infringing content. Users in our study who initially express the intent to consume legally are less likely to purchase legally in the infringing treatment condition than in other conditions, and users who initially express the intent to consume through pirate channels are more likely to consume legally when they are placed in the legal treatment condition.

Together our results show that reducing the prominence of piracy links in search results can cause users who would have otherwise consumed through free infringing channels to switch their consumption to paid legal channels. This change in behavior among users could be driven by one of three factors: an increase in the search cost necessary to find a user's preferred type of content (legal or infringing), a change in the user's perception of the quality of pirate versus infringing content, or a change in the user's perception of the social norms associated with consuming a particular type of content.

Regardless of the specific driver of the change in user behavior, our results have important implications for policy-makers. Our results show that relatively small changes in the ease of finding pirated content can have relatively large impacts on consumer behavior, even among consumers with a stated preference for pirated content. Specifically, by reducing the prominence of pirated links in search results, search engines can cause users who would have otherwise pirated content to switch to legal consumption. As such, our results show that changing the visibility of pirated content is a viable policy alternative in the fight against copyright theft and one that should be considered alongside other policy options such as ISP-level blocking, site shutdowns, and graduated response notice sending.

Of course, our study is not without limitations. The first limitation is that we are not able to separately identify the underlying mechanisms *vis-à-vis* search cost, perceived reputation, and perceived social norms in our experimental design. Second, although our custom search engine closely replicates the functionality of standard search engines, the experimental results could be perceived differently than results displayed in a non-experimental setting. Finally, our experiment does not allow us to test potential longer-term changes in user behavior from manipulating search results.

Acknowledgments

The authors thank seminar participants at the National Bureau of Economic Research Summer Institute, the Conference on Digital Experimentation @ MIT, Google, the University of California at San Diego, the New York University, and the University of Michigan for helpful comments on this research. This research was conducted as part of Carnegie Mellon University's Initiative for Digital Entertainment Analytics (IDEA) which receives unrestricted (gift) funding from the Motion Picture Association of America. This research was conducted independently without any oversight or editorial control.

References

Agarwal, A., Hosanagar, K., and Smith, M. D. 2011. "Location, Location, Location: An Analysis of Profitability of Position in Online Advertising Markets," *Journal of Marketing Research*. 48(6), pp. 1057-1073.

Baye, M. R., De los Santos, B., and Wildenbeest, M. R. 2016. "What's in a Name? Measuring Prominence and its Impact on Organic Traffic from Search Engines," *Information Economics and Policy* (34), pp. 44-57.

Bhattacharjee, S., Gopal, R. D., Lertwachara, K., and Marsden, J. R. 2006. "Impact of Legal Threats on Online Music Sharing Activity: An Analysis of Music Industry Legal Actions," *Journal of Law and Economics* (49), pp. 91-114.

Brooks, N. 2004. "The Atlas Rank Report II: How Search Engine Rank Impacts Conversions," Atlas Institute, New York.

Danaher, B., Dhanasobhon, S., Smith, M. D, and Telang, R. 2010. "Converting Pirates without Cannibalizing Purchasers: The Impact of Digital Distribution on Physical Sales and Internet Piracy," *Marketing Science* (29:6), pp. 1138-1151.

Danaher, B., Dhanasobhon, S., Smith, M. D, and Telang, R. 2015. "Understanding Media Markets in the Digital Age: Economics and Methodology," Chapter 13 in *Economic Analysis of the Digital Economy*, A. Goldfarb, S. M. Greenstein, and C. E. Tucker (eds), Chicago: University of Chicago Press.

Danaher, B., and Smith, M. D. 2014. "Gone in 60 Seconds: The Impact of the Megaupload Shutdown on Movie Sales," *International Journal of Industrial Organization* (33), pp. 1-8.

Danaher, B., Smith, M. D, and Telang, R. 2014. "Piracy and Copyright Enforcement Mechanisms," in *Innovation Policy and the Economy, Volume 14*, J. Lerner and S. Stern (eds.), ChicagoL University of Chicago Press.

Danaher, B., Smith, M. D, Telang, R., and Chen, S. 2014. "The Effect of Graduated Response Anti-Piracy Laws on Music Sales: Evidence from an Event Study in France," *Journal of Industrial Economics* (62:3), pp. 541-553.

Dhar, R., Huber, J., and Khan, U. 2007. "The Shopping Momentum Effect," Journal of *Marketing Research* (44:3), pp. 370-378.

Haisley, E., Mostafa, R., and Loewenstein, G. 2008. "Myopic Risk-Seeking: The Impact of Narrow Decision Bracketing on Lottery Play," *Journal of Risk and Uncertainty* (37:1), pp 57-75.

Haws, K. L., and Winterich K. P. 2013. "When Value Trumps Health in a Supersized World," *Journal of Marketing* (77:3), pp. 48-64.

Isaac, M. R., and Davis, D. D. (eds.). 2006. *Research in Experimental Economics, Volume 11: Experiments Investigating Fundraising and Charitable Contributions*, Bingley, UK: Emerald Group Publishing Limited, pp. 157-175.

Liebowitz, S. J. 2008. "File-Sharing: Creative Destruction or just Plain Destruction?," *Journal of Law and Economics* (49), pp. 1-28.

Löschela, A., Sturma, B., and Vogtc, C. 2013. "The Demand for Climate Protection—Empirical Evidence from Germany," *Economics Letters* (118:3), pp. 415-418.

Oberholzer-Gee, F., Strumpf, K. 2010. "File Sharing and Copyright," in *Innovation Policy and the Economy, Volume 10*, J. Lerner and S. Stern (eds.), Chicago: University of Chicago Press, pp. 19-55.

Pan, B., Hembrooke, H., Joachims, T., Lorigo, L., Gay, G., and Granka, L. 2007. "In Google We Trust: Users' Decisions on Rank, Position, and Relevance," *Journal of Computer-Mediated Communication* (12:3), Article 3.

Reimers, I. 2014. "The Effect of Piracy Protection in Book Publishing," Working Paper, Northeastern University, Boston, MA.

Rob, R., and Waldfogel, J. 2007. "Piracy on the Silver Screen," *The Journal of Industrial Economics* (55:3), pp. 379-395.

Rucker, D. D., Dubois, D., and Galinsky, A. D. 2011. "Generous Paupers and Stingy Princes: Power Drives Consumer Spending on Self versus Others," *Journal of Consumer Research* (37:6), pp. 1015-1029.

Rutz, O. J., and Trusov, M. 2011. "Zooming in on Paid Search Ads—A Consumer-Level Model Calibrated on Aggregated Data," *Marketing Science* (30:5), pp. 789-800.

Suk, K., Lee, J., and Lichtenstein, D. R. 2012. "The Influence of Price Presentation Order on Consumer Choice," *Journal of Marketing Research* (49:5), pp. 708-717.

Tsai J. Y., Egelman, S., Cranor, L., and Acquisti, A. A. 2011. "The Effect of Online Privacy Information on Purchasing Behavior: An Experimental Study," *Information Systems Research* (22:2), pp. 254-268.

Vandiver, D. M., Bowman, S., and Vega, A. 2012. "Music Piracy Among College Students: An Examination of Low Self-Control, Techniques of Neutralization, and Rational Choice," *The Southwest Journal of Criminal Justice* (8:2), pp. 92-111.

Yang, S., and Ghose, A. 2010. "Analyzing the Relationship Between Organic and Sponsored Search Advertising: Positive, Negative, or Zero Interdependence?," *Marketing Science* (29:4), pp. 602-623.

About the Authors

Liron Sivan is a researcher at Carnegie Mellon University's Heinz College. She received a Bachelor's of Science in Computer Science, a Master's of Science in Finance, and a Ph.D. in Management from Tel-Aviv University. Her research focuses on engagement in social media, electronic commerce networks, and the digital media industry. Liron has published her work in journals including *MIS Quarterly, Journal of Marketing, Behavior Research Methods,* and *Marketing Letters*. Her work won best paper awards from *Journal of Marketing*, and the International Conference on Information Systems.

Michael D. Smith is the J. Erik Jonsson Professor of Information Technology and Marketing and co-director of the Initiative for Digital Entertainment Analytics at Carnegie Mellon University's Heinz College. He received a Bachelor's of Science in Electrical Engineering (summa cum laude) and a Master's of Science in Telecommunications Science from the University of Maryland, and received a Ph.D. in Management Science from the Sloan School of Management at MIT. His research specializes in entertainment analytics, marketing, and management and he is a co-author of the book *Streaming, Sharing, Stealing: Big Data and the Future of Entertainment* (MIT Press, September 2016).

Rahul Telang is a professor of Information Systems at the Heinz College, Carnegie Mellon University. His major research focus is on economics of digitization, studying how content digitization is affecting firms, users, and our policy. His work has appeared in top scientific journals and won numerous awards. He is co-author of *Streaming, Sharing, Stealing: Big Data and the Future of Entertainment* (MIT Press, September 2016).

MIS Quarterly

RESEARCH ARTICLE

Do Search Engines Influence Media Piracy? Evidence from a Randomized Field Study

Liron Sivan, Michael D. Smith, and Rahul Telang
Heinz College, Carnegie Mellon University, Pittsburgh, PA 15213 U.S.A.
{lsivan@andrew.cmu.edu} {mds@cmu.edu} {rtelang@andrew.cmu.edu}

Appendix A

Experiment 1: Empirical Approach

Endowment Effect

The experimental design we present in the paper is similar to that used previously by Tsai et al. (2011), which conducted a randomized experiment in which participants were solicited to "test a new search engine interface" (p. 259) and were paid $45 for their participation. As in our study, Tsai et al.'s study asked its participants to use a search engine to find several products and to purchase them, allowing the participants to keep both the products and any money left over after the purchases were made. This design, in the words of the authors, "created a price incentive, encouraging participants to purchase from merchants with lower prices" (p. 262). Providing compensation to experimental participants is a common and accepted practice in the experimental economics, marketing, and information systems literatures (see, for example, Dhar et al. 2007; Haisley et al. 2008, Isaac and Davis 2006; Haws and Winterich 2013; Löschela et al. 2013; Rucker et al. 2011; Suk et al. 2012). As it pertains to our experimental design, it is important to note that although providing subjects with an unrestricted endowment of money may change the overall likelihood of any user making a purchase, it will do so equally for both the treated and control group users such that any difference in response between these two groups can reliably be attributed to the experimental intervention. Further, the fact that participants can retain the money after the experiment has ended, and use the money for any purpose they desire, will strengthen their perceived ownership of it and will do so equally for both the treated and control group users. Again, this means that any difference in response between these two groups can reliably be attributed to the experimental intervention.

We also note that our results in experiment 1, where our users are given a $20 gift card, are consistent with results in experiment 3, where Amazon Mechanical Turk users are only compensated by $1.50 for their time, further suggesting that the differences in responses between the control and treatment groups are not driven by an endowment effect.

Recall Bias

It is important to note that we can only directly observe user behavior while they are on our search engine: we do not observe their behavior outside of the search engine. Thus, our observation of where the user obtained their content is based on their survey answers. While this may introduce some recall bias, the degree of recall bias should not vary across the control and treatment conditions and the survey occurs immediately after the participant obtains the content, significantly reducing the overall possibility of recall bias. It is also possible, based on prior studies that found survey respondents are likely to underreport socially undesirable activities, that participants who choose a pirated option could be less likely to reveal that that in their survey answers (Cannell et al. 1965; Means et al. 1992; Warner 1978; Wyner 1980). However, again, the propensity to misreport should not vary across the control and treatment conditions, and any underreporting bias would lead to an underestimate of the degree that infringing links induce more piracy. We also note that we can use observed search and clicking behavior to validate a user's survey answers (including in some cases verifying price). In the results section we show that the users' survey responses were consistent with their observed behavior.

Dropout Rates

In order to ensure that participants understood the task correctly, we included in our analysis only those users who claimed the $20 virtual Visa card, used our custom-built search engine, and who reported the source from which they acquired the movie and its price. Approximately 1,000 participants were invited to participate in the study. Of these participants, 770 met the initial qualifications for participating in the study by expressing an interest in watching one of the movies on our list. These participants were invited to participate in our study and 632 of them logged into the system and were presented with the task details. Out of these participants, 235 completed the task as instructed, and 196 qualified for our study by also completing the post-experiment questionnaire. The dropout rates across each of the different stages described above are not statistically different across experimental conditions.

Comparison of User Characteristics and Attitudes toward Piracy across Treatment Conditions

Prior to analyzing the results of the experiment, we confirmed that the distribution of user characteristics (demographic characteristics, attitudes toward piracy) is similar across the control and treatment groups. The average values for each experimental condition are presented in Table A1. Chi squared tests show that there are no statistically significant differences in demographic characteristics or attitudes toward piracy between the experimental conditions (p-value > 0.05). This confirms that our randomization worked as intended.

Table A1. Between-Conditions Comparison of Demographic Characteristics and Attitude toward Piracy

	% Women	Average Age Group	Average Household Size	Average Household Income Group	Average Attitude Against Piracy (Likert Scale)	% Downloads Infringing
Condition 1: No manipulation	70.0%	4.20	2.85	2.82	4.60	33.3%
Condition 2: Legal content manipulation	59.2%	4.30	2.85	2.77	4.75	39.4%
Condition 3: Infringing content manipulation	64.6%	3.97	2.68	3.02	4.58	44.6%

Note: There are 7 possible age group values in the questionnaire: 1 (18–21), 2 (22–25), …, 7 (61 and over); 10 possible household size values: 1, 2, …, 9, 10 or more; and 6 possible household income group values: 1 (less than $30,000), 2 ($30,000–$50,000), …, 6 (over $150,000).

To further verify similarity in preexisting attitudes toward piracy, we tested whether the participants' initial intent for pirate or legal content is similar across the three experimental conditions. We did this by classifying whether their initial search terms reflected neutral, legal, or infringing intent according to the degree to which pirate or legal links were present in the (unmodified) search results for similar queries commonly issued by our users.

We discovered that search terms using only the movie's name contained almost exclusively "neutral" results (i.e., results that neither promote legal or pirate sources), and thus we classify these searches as neutral. However, when search terms included the words "buy," "rent," or "purchase," the search results contained 38% more legal links than pirate links, and when the search term contained a legal domain name (e.g., Amazon), the search results contained 78% more legal links than pirate links. Thus, we classify these search terms as representing "legal" intent. Conversely, when search terms included the words "download," "stream," or "full movie," there were 33% more pirate links in the search results than legal links, and including the domain name of an infringing site (e.g., piratebay) resulted in search results that included 65% more pirate links than legal links. Because of this, we classify these search terms as representing "infringing" intent.

We then classify intent based on the initial intent reflected in each user's search terms (or neutral if the user did not express intent in their searches). As above, reassuringly the distributions are not statistically different across the control and treatment conditions (see Table A2 for frequencies across conditions).

Table A2. Between-Conditions Comparison of the Initial Intent (based on the first keyword each user entered)				
	N	Neutral Searches	Legal Intent	Infringing Intent
Condition 1: No manipulation	60	17	31	12
Condition 2: Legal content manipulation	68	28	23	17
Condition 3: Infringing content manipulation	65	21	33	11

Finally, we note that although the characteristics of users who participated in the experiment were similar across the control and treatment conditions, it is possible that our participant pool skews toward being more media or tech savvy than the general population. As such, in interpreting our results, one should focus on the difference between sales/piracy in the control and treatment conditions as opposed to the absolute levels of sales/piracy within any particular condition.

Appendix B

Logistic Regression

Although the tests reported in the body of the paper are sufficient to determine if there are differences between the control and treatment groups based on our experimental manipulations, we can also use a logistic regression model to control for and analyze differences between groups based on observed characteristics.

Specifically, we use the following logistic regression model to control for observable participant characteristics:

$$\log\frac{PR\left(Legal_i\right)}{1-PR\left(Legal_i\right)} = \alpha+\beta_1\bullet NI_i+\beta_2\bullet I_i+\sum_{j=3}^{5}\beta_j\bullet DC_i^j+\sum_{j=6}^{8}\beta_j\bullet MCP_i^j+\sum_{j=9}^{10}\beta_j\bullet ATP_i^j+\beta_{11}\bullet intent+\varepsilon_i \tag{3}$$

where $Legal_i$ denotes whether the movie was acquired from a non-infringing source; NI_i is an indicator variable denoting whether participant i was treated with the non-infringing search condition; I_i is an indicator variable denoting whether participant i was treated with the infringing search condition; $\sum_{j=3}^{5}\beta_j\bullet DC_i^j$ includes the following demographic characteristics: Gender (an indicator variable for whether the participant was a woman), Age ≤ 40, Household size, and Income; $\sum_{j=6}^{8}\beta_j\bullet MCP_i^j$ includes the following media consumption preferences: Time online (the average hours spent online per day), Acquires movies online (whether the participant ever downloaded or streamed a movie, including pirated movies), and Movies per year (the number of movies the participant watched in the last 12 months); $\sum_{j=9}^{10}\beta_j\bullet ATP_i^j$ includes the following attitude toward piracy variables: Against piracy (on a six-point Likert Scale ranging from 1, "There is nothing wrong with it," to 6, "It is the same as stealing"), Downloads infringing (whether the participant indicated that s/he uses torrents or other free online downloads/ streaming/file-sharing); and *Non-infringing intent*$_i$ (whether the first search term that the user entered indicates that his/her intent is to acquire a legal copy). We present the results of this model in Table B1.

These results are consistent with our means comparison results in that the treatment variables are statistically different from the control condition and that they have the expected sign (the legal treatment condition increases the likelihood of purchasing legally and the infringing treatment condition decreases the likelihood of purchasing legally). These results also confirm, as one would expect, that participants who consumed infringing content in the past are less likely to purchase the movie legally and that those who use a search term that implies legal intent are more likely to purchase legally.

Table B1. Logistic Regression Results for Equation (3)

Dependent Variable: Acquired Legally	(1) Basic Model	(2) Including Demographic Characteristics	(3) Including Media Consumption Preferences	(4) Including Attitude Towards Piracy	(5) Including Intent to Acquire Legally
Constant	1.386***	1.637**	-0.333	0.148	0.0575
	(0.323)	(0.663)	(1.288)	(1.511)	(1.533)
Non-infringing mode	1.432**	1.458**	1.606**	1.855***	1.905***
	(0.608)	(0.613)	(0.634)	(0.673)	(0.685)
Infringing mode	-1.108***	-1.152***	-1.232***	-1.235**	-1.261**
	(0.409)	(0.422)	(0.438)	(0.507)	(0.516)
Woman		-0.00896	0.0163	0.0167	0.0421
		(0.406)	(0.416)	(0.482)	(0.491)
Age $\leq$ 40		-0.603	-0.744*	-0.669	-0.492
		(0.398)	(0.436)	(0.491)	(0.500)
Household size		-0.134	-0.216	-0.0772	-0.106
		(0.146)	(0.151)	(0.186)	(0.193)
Income		0.126	0.132	-0.0719	-0.0333
		(0.142)	(0.147)	(0.166)	(0.173)
Time online			0.0287	0.235	0.169
			(0.213)	(0.249)	(0.250)
Acquires movies online			0.191	0.453*	0.412
			(0.231)	(0.275)	(0.282)
Movies per year			0.409**	0.229	0.255
			(0.178)	(0.210)	(0.213)
Against piracy				0.554	0.404
				(0.486)	(0.504)
Downloads infringing				-2.424***	-2.410***
				(0.516)	(0.523)
Non-infringing intent					1.266**
					(0.621)
Number of obs.	196	196	196	196	196
Pseudo R^2	0.1392	0.1611	0.1933	0. 3529	0. 3758

Notes: Standard errors in parentheses; ***p < 0.01, **p < 0.05, *p < 0.1

To see the effect of intent on user behavior, we use the method described in Appendix C to split the data based on user intent (following Question 15f in our survey) and rerun the regression. While the number of observations in the infringing intent are small, the results shown in Table B2 using the survey question provide the same insight as those presented in the body of the paper: Even users with infringing intent are affected by legal mode and vice versa.

Table B2. Logistic Regression Results for Equation (3) by Intent		
	(1)	**(2)**
	15F Legal Intent - % Legal b/se	**15F Infringing Intent - % Legal b/se**
Constant	-1.591	11.120
	(1.809)	(8.246)
Non-infringing mode	2.228*	1.795
	(0.922)	(1.293)
Infringing mode	-1.334*	-1.172
	(0.599)	(1.711)
Woman	-0.324	1.329
	(0.584)	(1.357)
Age $\leq$ 40	-0.999	-0.119
	(0.619)	(1.213)
Household size	0.069	-0.485
	(0.224)	(0.438)
Income	-0.051	0.067
	(0.198)	(0.407)
Time online	0.503+	-0.507
	(0.305)	(0.671)
Acquires movies online	0.531+	0.000
	(0.306)	(.)
Movies per year	0.398+	-1.841
	(0.236)	(1.577)
Against piracy	0.370	1.545
	(0.572)	(1.700)
Downloads infringing	-2.357**	0.000
	(0.583)	(.)
Number of obs.	165	28
Pseudo R^2	0.3957	0.3152

Note: Standard errors in parentheses; ***p < 0.01, **p < 0.05, *p < 0.1

Appendix C

Classifying Legal and Infringing Intent Using the Post-Experiment Questionnaire

In the body of the paper, we analyzed the responses of users who, based on their keyword choices, expressed the intent to consume legal or infringing content. Another way to classify intent is based on a user's prior experience with consuming pirated or legal content. To do this, we use question 15f in our post-experiment survey of users. This question asks users, "Of all the movies you watched, can you provide some details on what channel did you use to watch them?" One of the options was "Torrents (or other free online downloads/stream, file-sharing [channels])." Users were able to choose "0–1 movies," "2–5 movies," "5–10 movies," "10–15 movies," or "15+ movies" consumed through this method.

In Table C1 we replicate the results from Table 2 in the body of the paper, but in this case instead of using search queries to classify intent, we classify users who stated that they had consumed 0–1 movies through torrents or other free download/streaming channels as those that have legal intent, and users who had consumed two or more movies through torrents or other free download/streaming channels as those having infringing intent.

Table C1. Legal Purchase Rates across Treatment Conditions and Initial Intent

Post-Experiment Questionnaire	Control Group	Legal Content Manipulation	Infringing Content Manipulation
Legal Intent	43/51 (84%)	57/57 (100%)	**34/57 (60%)**
Infringing Intent	5/8 (62%)	**10/12 (83%)**	3/8 (38%)

This table shows that, consistent with the results presented in the body of the paper, users with legal intent are significantly less likely to consume legally when placed in the infringing content manipulation (60%) than in either the legal (100%) or control (84%) conditions (t-tests: $t(56) = -5.34$, $p < 0.01$; $t(50) = -2.75$, $p < 0.001$, respectively). Likewise, users with infringing intent are significantly more likely to purchase legally in the legal content manipulation (83%) than in the infringing (38%) or control (62%) conditions (t-tests: $t(7) = 2.07$, $p < 0.05$; $t(7) = 1.057$, $p = 0.16$, respectively). These results confirm the results from the body of the paper that the ranking of pirated and legal search results matters even among users with an initial preference for pirate or legal channels.

As with the results in the body of the paper, we can use our clickstream data to test whether, when using the post-experiment questionnaire to classify user intent, users with low infringing intent search more when they are placed in the infringing treatment condition than they do in the control condition and likewise for users with high infringing intent who are placed in the legal treatment condition.

Consistent with the results presented in the body of the paper, our clickstream data suggest that users with high piracy intent search more when placed in the legal treatment condition than in the control condition and users with low piracy intent search more when placed in the piracy treatment condition than in the control condition. Specifically, the post-experiment survey data show that consumers with low piracy intent initiate on average 2.39 searches in the control condition ($n = 51$) and 3.42 searches in the infringing treatment condition ($n = 57$) (statistically significant at the 1% level). Likewise consumers with high piracy intent initiate, on average, 2.00 searches (9) in the control condition and 3.33 searches (12) in the legal treatment condition (insignificant). When we consider the average position of search instead of the number of searches, consumers with low piracy intent click on results at an average position of 4.21 (50) in the control condition and 8.45 (47) in the infringing treatment condition (1%), and consumers with high piracy intent click on results at an average position of 2.74 (9) in the control condition and 3.54 (12) in the legal treatment condition (insignificant). These results also hold if we consider the maximum position that each consumer clicks on during their searches. Averaging across consumers, consumers with low piracy intent search to a maximum position of 6.02 (50) in the control condition, and to a maximum position of 13.49 (47) in the infringing treatment condition (1%). Consumers with high piracy intent search to a maximum position of 4.22 (9) in the control condition and a maximum position of 6.58 (12) in the legal treatment condition (statistically insignificant).

Appendix D

Experiment 2: Empirical Approach and Selected Results

Data and Results

Following our approach in experiment 1, in our second experiment we included in our analysis only those users who claimed the $20 virtual Visa card, used our custom-built search engine for at least one search, and reported the source from which they acquired the movie and how much they paid for it. A total of 234 participants qualified for our study.[1] In Table D1 we report average statistics for these participants across the treatment conditions. These statistics show that the distribution of demographic characteristics and attitudes toward piracy are similar across the control and treatment groups, as one would expect given the experimental assignment.

Table D1. Business-Conditions Comparison of Demographic Characteritics and Attitude Toward Piracy

	% Women	Average Age Group	Average Household Size	Average Household Income Group	Average Attitude Against Piracy (Likert Scale)	% Downloads Infringing
Condition 1: No manipulation	57.14%	1.69	3.02	2.60	3.40	80.95%
Condition 2: Mild legal content manipulation	73.47%	1.71	3.18	2.35	3.73	69.39%
Condition 2a: Legal content manipulation	62.50%	1.75	2.79	2.13	3.85	62.50%
Condition 3: Mild infringing content manipulation	54.55%	1.68	3.41	2.18	3.66	75.00%
Condition 3a: Infringing content manipulation	66.67%	1.73	3.29	2.31	3.76	64.71%

Note: There are 7 possible age group values in the questionnaire: 1 (1–21), 2 (22–25), …, 7 (61 and over); 10 possible household size values: 1, 2, …, 9, 10 or more; and 6 possible household income group values: 1 (less than $30,000), 2 ($30,000–$50,000), …, 6 (over $150,000).

As in experiment 1 (and using the method described in Appendix A), we compare the initial search intent expressed by users across the different treatment conditions (Table D2), finding no significant differences in expressed intent across conditions.

[1]Approximately 650 participants were invited to participate in the study and expressed an interest in watching one of the movies in the study. A total of 550 participants logged into the system and were presented with the task details. Out of these participants, 270 completed the experimental task as instructed, and of these participants 234 (86 men, 148 women) qualified for our study by also completing the post-experiment questionnaire. The dropout rates in each of the different stages described above are not statistically different across experimental conditions.

Table D2. Between-Conditions Comparison of the Initial Intent (based on the first keyword each user entered)				
	N	Neutral Searches	Legal Intent	Infringing Intent
Condition 1: No manipulation	42	6	23	13
Condition 2: Mild legal content manipulation	49	16	19	14
Condition 2a: Legal content manipulation	48	14	20	14
Condition 3: Mild infringing content manipulation	44	11	18	15
Condition 3a: Infringing content manipulation	51	15	27	9

Logistic Regression

We use a logistic regression model to control for and analyze differences between groups based on observed characteristics. This model is similar to the model from experiment 1, except that it includes dummy variables for the additional treatment conditions.

Specifically, we use the following logistic regression model to control for observable participant characteristics:

$$\log \frac{PR\left(Legal_i\right)}{1-PR\left(Legal_i\right)}$$

$$= \alpha + \beta_1 \bullet MNI_i + \beta_2 \bullet NI_i + \beta_3 \bullet MI_i + \beta_4 \bullet I_i + \sum_{j=5}^{7} \beta_j \bullet DC_i^j + \sum_{j=8}^{10} \beta_j \bullet MCP_i^j \tag{5}$$

$$+ \sum_{j=11}^{12} \beta_j \bullet ATP_i^j + \beta_{13} \bullet intent + \varepsilon_i$$

where MNI_i is an indicator variable denoting whether participant i was treated with the mild non-infringing search condition; MI_i is an indicator variable denoting whether participant i was treated with the mild infringing search condition; and the other variables are the same as before. We present the results of this model in Table D3.

The results in Table D3 are consistent with our means comparison results in that the intense treatment variables are statistically different from the control condition and that all treatment variables have the expected signs (the non-infringing treatment condition increases the likelihood of purchasing legally and the infringing treatment condition decreases the likelihood of purchasing legally). As with experiment 1, the results show that participants who consumed infringing content in the past are less likely to purchase the movie legally, and that those who use a search term that reveals non-infringing intent are more likely to purchase legally.

Table D3. Logistic Regression Results					
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
Dependent Variable: Acquired Legally	**Basic Model**	**Including Demographic Characteristics**	**Including Media Consumption Preferences**	**Including Attitude Toward Piracy**	**Including Intent to Acquire Legally**
Constant	0.486 (0.318)	0.772 (0.529)	3.077** (1.215)	3.265** (1.295)	2.932** (1.430)
Mild non-infringing mode	0.641 (0.460)	0.741 (0.470)	0.733 (0.479)	0.694 (0.484)	1.046** (0.529)
Intense non-infringing mode	1.912***	1.975***	1.977***	1.942***	2.321***
	(0.611)	(0.620)	(0.628)	(0.641)	(0.671)
Mild infringing mode	-0.576 (0.438)	-0.644 (0.452)	-0.700 (0.462)	-0.732 (0.469)	-0.727 (0.528)
Intense infringing mode	-0.924** (0.428)	-0.932** (0.438)	-1.152** (0.469)	-1.228** (0.477)	-1.280** (0.530)
Woman		-0.626* (0.321)	-0.824** (0.341)	-0.940*** (0.350)	-1.178*** (0.390)
Younger (Age < 22)		-0.462 (0.338)	-0.593* (0.349)	-0.518 (0.356)	-0.383 (0.394)
Household size		0.0843 (0.0977)	0.0831 (0.1000)	0.0409 (0.102)	-0.00624 (0.109)
Income		-0.0100 (0.118)	-0.0267 (0.122)	-0.0174 (0.123)	-0.119 (0.136)
Time online			-0.317* (0.187)	-0.319* (0.190)	-0.229 (0.208)
Acquires movies online			-0.442* (0.264)	-0.209 (0.284)	-0.324 (0.302)
Movies per year			0.0780 (0.146)	0.0928 (0.150)	0.132 (0.165)
Against piracy				0.148 (0.410)	0.425 (0.448)
Downloads infringing				-0.968** (0.416)	-0.889* (0.458)
Non-infringing intent					3.127*** (0.693)
Number of obs.	234	234	234	234	234
Pseudo R^2	0.1321	0.1519	0.1755	0. 1987	0. 3107

Notes: Standard errors in parentheses; ***p < 0.01, **p < 0.05, *p < 0.1

Last, we use our post-experiment questionnaire to classify infringing intent as a robustness test for the classification based on initial query words used in the body of the paper. As with experiment 1, to conduct this classification we use question 15f in our post-experiment survey of users and classify users who stated that they had consumed 0–1 movies through torrents or other free download/streaming channels as those that have legal intent and users who had consumed two or more movies through torrents or other free download/streaming channels as those having infringing intent.

Table D4. Legal Purchase Rates across Treatment Conditions and Initial Intent			
Post-Experiment Questionnaire	**Control Group**	**Legal Content Manipulation**	**Infringing Content Manipulation**
Legal Intent	16/19 (84%)	32/32 (100%)	**15/35 (43%)**
Infringing Intent	10/23 (43%)	**12/16 (75%)**	5/16 (31%)

Table D4 shows that, consistent with the results presented in the body of the paper, users with legal intent are significantly less likely to consume legally when placed in the infringing content manipulation (43%) than in either the legal (100%) or control (84%) conditions (t-tests: $t(31) = -5.01$, $p < 0.001$; $t(18) = -2.91$, $p < 0.005$, respectively). Likewise, users with infringing intent are significantly more likely to purchase legally in the legal content manipulation (75%) than in the infringing (31%) or control (43%) conditions (t-tests: $t(15) = 2.49$, $p < 0.05$; $t(15) = 1.98$, $p < 0.05$, respectively). These results confirm the results from the body of the paper that the ranking of pirated and legal search results matters even among users with an initial preference for pirate or legal channels.

Appendix E

Experiment 3: Flagging Infringing Links, Empirical Approach and Results

In the first two experiments, there was no explicit differentiation between the legal and infringing links presented to users. In this experiment we make an explicit distinction between legal and infringing links to draw a stronger differentiation between these two types of content. We, do this in part to validate our results in experiments 1 and 2, and as a partial test of whether our results are driven by factors other than user indifference between legal and infringing content.

For this experiment, we recruited 666 participants from Amazon Mechanical Turk (AMT). We included the three treatment conditions that were used in Experiment 1 and added two additional treatment conditions in order to understand what drives our previous results. Similar to our previous experiments before the experiment started, each participant was randomly assigned to one of the following five search treatment conditions: no manipulation (control), non-infringing (legal) content manipulation, infringing (piracy) content manipulation, flagging infringing results, and both flagging and promoting infringing results. We flagged infringing results (Figure E1) in a similar manner to the way Google flags results that may harm one's computer (Figure E2), but instead of saying "This site may be hacked" as Google does, our text indicated "This webpage may contain infringing materials."

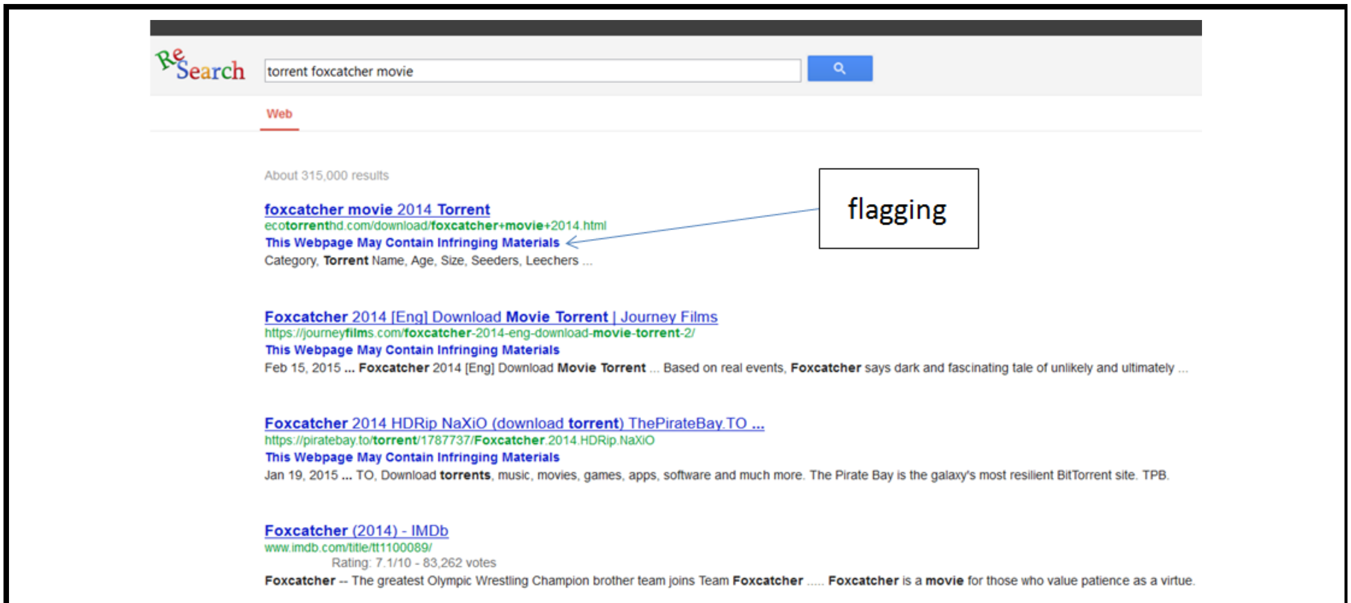


Figure E1. Flagging Infringing Results

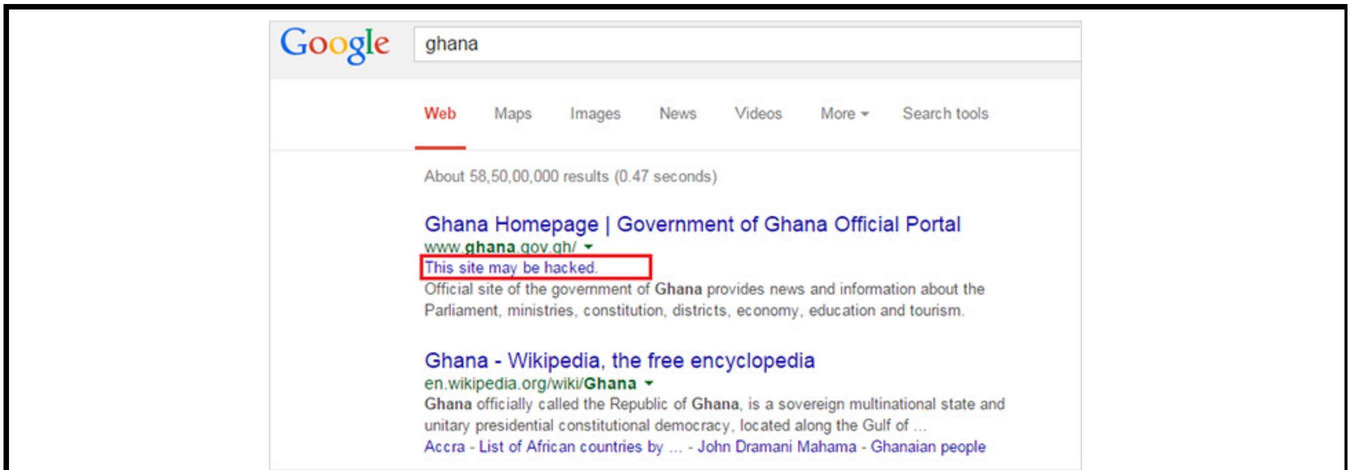


Figure E2. Google's "This Site May Be Hacked" Warning

Experimental Design

There are three additional differences between this experiment and the two experiments in the body of the paper. First, the participants in this experiment did not receive the $20 Visa virtual card used in the first two experiments. Instead they received $1.50 through AMT for their participation. Second, the participants were not asked to actually acquire the movie of their choice, but to search as they normally would if they wished to watch the movie, and to find a desired source from which they would have downloaded/streamed/purchased/rented it. Third, in the post experiment questionnaire the participants were only asked to name the source they would have selected and the amount they would have spent.

Results

In Table E1,we compare the proportion of legal purchases across the different experimental conditions and search types. Consistent with experiments 1 and 2, Table E1 shows that users whose search terms express legal intent[2] are significantly *less* likely to purchase legally when placed in the infringing content manipulation (19%) than in either the legal (100%) or control (98.73%) conditions ($p < 0.001\%$ in both cases). Likewise, users whose search terms express infringing intent are significantly *more* likely to purchase legally in the legal content manipulation (81%) than in the infringing (15%) or control (53%) conditions ($p < 0.01\%$ in both cases).

However, we can also use these results to better understand whether user behavior changes when infringing links are highlighted in search results through the use of the flag. In this regard, Table E1 shows that the difference between the rate of legal purchases of users who have infringing intent in the control condition (53%) and in the treatment condition in which infringing results are being flagged (48%) is not statistically significant ($p > 0.3$). Meaning, although promoting legal results increased the purchase rate from 53% (control condition) to 81% (legal content manipulation), labeling infringing results without promoting them had no effect on behavior. This is also true when comparing the legal purchase rate of users who have infringing intent in the two infringing content manipulation conditions: the difference between the condition that includes flagging (17%) and the condition that does not include flagging (15%) is not statistically significant at any reasonable level of significance. Also, when comparing the legal purchase rate of users who have legal intent in the two infringing content manipulation conditions, the difference between the condition that includes flagging (27%) and the condition that does not flagging (19%), is not statistically significant ($p > 0.1$). Taken together, the results of this experiment suggest that making the infringing nature of some links more noticeable to users through the use of a flag has no impact on their propensity to consume infringing versus legal content. This, in turn, sheds light on the driver of our results. If our results were driven by a lack of perceived differentiation between legal and pirate links, then one would expect that user behavior would change if a flag were used to explicitly distinguish infringing links from other links. The fact that this does not occur, while not conclusive, is suggestive that users do perceive a difference between legal and infringing links.[3] This inference is strengthened when combined with the clickstream data above suggesting that users with a stated preference for legal or infringing content search more intensely when placed in their non-preferred treatment condition. We discuss the implications of this result in more detail in the body of the paper.

Table E1. Purchase Rates across Treatment Conditions and Search Types

	N	% Legal	Legal IIntent	Infringing Intent	Legal Intent - % Legal	Infringing Intent - % Legal
Condition 1: No manipulation	138	78.99%	79	59	98.73%	52.54%
Condition 1a: Flagging infringing	143	71.33%	66	77	98.48%	48.05%
Condition 2: legal content manipulation	110	94.55%	78	32	100.00%	81.25%
Condition 3: infringing content manipulation	131	17.56%	78	53	19.23%	15.09%
Condition 3a: Infringing content manipulation & Flagging infringing	144	24.31%	102	42	27.45%	16.67%

[2]Intent here is defined following the definition used in the body of the paper: When a user's search results produce no infringing links, we classified it as legal intent and likewise for infringing intent.

[3]We thank the Associate Editor for making this observation.

References

Cannell, C. F., Fisher, G., and Bakker, T. 1965. "Reporting of Hospitalization in the Health Interview Survey," *Vital and Health Statistics*, PHS Pub. No. 1000, Ser. 2, No. 6, Washington, DC: U.S. Government Printing Office.

Dhar, R., Huber, J., and Khan, U. 2007. "The Shopping Momentum Effect," Journal of *Marketing Research* (44:3), pp. 370-378.

Haisley, E., Mostafa, R., and Loewenstein, G. 2008. "Myopic Risk-Seeking: The Impact of Narrow Decision Bracketing on Lottery Play," *Journal of Risk and Uncertainty* (37:1), pp. 57-75.

Haws, K. L., and Winterich K. P. 2013. "When Value Trumps Health in a Supersized World," *Journal of Marketing* (77:3), pp. 48-64.

Isaac, M. R., and Davis, D. D. (eds.). 2006. *Research in Experimental Economics, Volume 11: Experiments Investigating Fundraising and Charitable Contributions*, Bingley, UK: Emerald Group Publishing Limited.

Löschela, A., Sturma, B., and Vogtc, C. 2013. "The Demand for Climate Protection—Empirical Evidence from Germany," *Economics Letters* (118:3), pp. 415-418.

Means, B., Habina, K., Swan, G., and Jack, L. 1992. "Cognitive Research on Response Error in Survey Questions on Smoking," *Vital Health & Statistics*, Hyattsville, MD: National Center for Health Statistics, Series 6, Number 5.

Rucker, D. D., Dubois, D., and Galinsky, A. D. 2011. "Generous Paupers and Stingy Princes: Power Drives Consumer Spending on Self Versus Others," *Journal of Consumer Research* (37:6), pp. 1015-1029.

Suk, K., Lee, J., and Lichtenstein, D. R. 2012. "The Influence of Price Presentation Order on Consumer Choice," *Journal of Marketing Research* (49:5), pp. 708-717.

Tsai J. Y., Egelman, S., Cranor, L., and Acquisti, A. A. 2011. "The Effect of Online Privacy Information on Purchasing Behavior: An Experimental Study," *Information Systems Research* (22:2), pp. 254-268.

Warner, K. E. 1978. "Possible Increases in the Under-Reporting of Cigarette Consumption," *Journal of the American Statistical Association* (73), pp. 314-318.

Wyner, G. 1980. "Response Errors in Self-Reported Number of Arrests," *Sociological Methods and Research* (9), pp. 161-177.