



Differentiation of online text-based advertising and the effect on users' click behavior



Jason T. Jacques^{a,*}, Mark Perry^b, Per Ola Kristensson^a

^a Department of Engineering, University of Cambridge, Cambridge, United Kingdom

^b Department of Computer Science, Brunel University, Uxbridge, United Kingdom

ARTICLE INFO

Article history:
Available online 16 May 2015

Keywords:
Advertising
Crowdsourcing
Behavior

ABSTRACT

Online syndicated text-based advertising is ubiquitous on news sites, blogs, personal websites, and on search result pages. Until recently, a common distinguishing feature of these text-based advertisements has been their background color. Following intervention by the Federal Trade Commission (FTC), the format of these advertisements has undergone a subtle change in their design and presentation. Using three empirical experiments, we investigate the effect of industry-standard advertising practices on click rates, and demonstrate changes in user behavior when this familiar differentiator is modified. Using three large-scale experiments ($N_1 = 101$, $N_2 = 84$, $N_3 = 176$) we find that displaying advertisement and content results with a differentiated background results in significantly lower click rates. Our results demonstrate the strong link between background color differentiation and advertising, and reveal how alternative differentiation techniques influence user behavior.

© 2015 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Advertising is the primary revenue stream for many of the “free” services provided to Internet users (Castro, 2012) and is essential to the continued economic sustainability of many services, such as email, news and search. The immense value for both the end-user and the commercial entity providing these services is demonstrated by organizations such as Google, which in 2013 derived over 90% of its \$55.5 billion revenue from advertising (Google., 2014).

While large organizations can directly negotiate with advertisers, smaller organizations typically outsource their advertising to advertising-networks, such as those offered by Google, Microsoft, and Yahoo (Evans, 2008). These networks provide syndicated, context-sensitive adverts to webpage publishers in the form of discrete ad units consisting of one or more text links, in exchange for a proportion of the revenue generated from users clicking on these adverts (Broder, Fontoura, Josifovski, & Riedel, 2007). Such strategies entrust much of the control over advertising content, presentation, and cost-per-click to the advertising network, which dynamically displays ad units directly to web users viewing the content, with little, if any, interference from the webpage

publisher. Advertising networks may offer advice to publishers on how to best present their adverts. For instance, in order to help publishers maximize their return from the AdSense program, Google offers suggestions to publishers to “customize the colors and fonts of your ads to match your site’s look and feel” (Google., 2010).

In addition to colors and fonts, the location of an advert is also considered an important factor in user recognition. Fox, Smith, Chaparro, and Shaikh (2009) suggest that adverts should be located high on the page and use a contrasting color scheme. By maximizing visibility, adverts are expected to build brand recognition and it is suggested that this increased recognition will result in increased click rates (Fox et al., 2009). However, for pay-per-click adverts, brand recognition alone is not sufficient to generate revenue for the website publisher. It is also necessary for the advert to be clicked. While presentation of text-based advertising is much practiced, little research can be found to suggest which styles might be appropriate for content publishers, or how presentation information might be integrated into advert selection methods used by advertising networks.

1.1. Presentation and user engagement

The literature differs on the importance of measuring actual user engagement with advertising. Hofacker and Murphy (1998) suggest that banner advertising can be effective for the advertiser

* Corresponding author.

E-mail addresses: jtj21@cam.ac.uk (J.T. Jacques), mark.perry@brunel.ac.uk (M. Perry), pok21@cam.ac.uk (P.O. Kristensson).

without actually being clicked, however, they also highlight the importance of measuring clicks as this can be a determining factor in choosing which advertisements to display when offered on a pay-per-click basis. Gathering click-stream data can improve the understanding of users' experience and their interactions with recommendations, personalization, and advertising. Using click data allows advertisers to form a more nuanced picture of how users move from content, to advertising, to purchase (Ting, Clark, Kimble, Kudenko, & Wright, 2007).

To generate revenue for publishers of pay-per-click advertising, users must act and click an advert. Improved recall of advertising does not necessarily change user behavior or lead to an increase in purchase intent (Goldfarb & Tucker, 2011). Simply making an advert more visible may not increase advertising-derived revenue. Goldfarb and Tucker (2011) suggest that users react negatively to more obtrusive targeted advertising including video, pop-ups, and overlays as users fear exposure of private information. While their work investigated the effects of high-profile multimedia adverts, they also hypothesize that the success of text-based, context-sensitive advertising might be due to its reduced intrusiveness.

Until the end of 2013, Google used two different differentiation techniques for text-based, contextual advertising. For advertisements to the right of the content area, a simple "Ads" header is used, as seen in Fig. 1. In the content-area, in addition to an "Ads" header, a background color is used to differentiate these advertisements from the content results that follow. While Google is legally obliged to use some method to indicate advertising content, under Federal Trade Commission guidelines (2000), the method of indicating this is entirely up to the publisher. However, more recent guidance has provided additional direction.

In an open letter to search engines, and other organizations dependent on Internet advertising, the FTC reiterated its stance on differentiating advertising from content (Federal Trade Commission., 2013). The FTC continued their encouragement of clarity and prominence of advertising disclosures, however new focus was directed in particular towards the contemporary use of visual cues and text labels. In the letter the FTC also raised concerns regarding the display of advertising on computer monitors

and mobile devices with their potentially differing display capabilities.

We have observed that, increasingly, search engines have introduced background shading that is significantly less visible or "luminous" and that consumers may not be able to detect on many computer monitors or mobile devices.

[Federal Trade Commission, 2013]

Fig. 1 shows an example of the styling discussed. Since the publication of the FTC letter, Google has changed advertising presentation to use a simple line delimiter coupled with a yellow "Ad" marker, removing the background coloration identified as problematic by the FTC. With publishers adjusting practice to maintain compliance with the mandated guidance, it is important to validate their recommendations. If users have become accustomed to particular cues to identify advertising, the removal of background coloration could be detrimental to the FTC's intention of enhancing clarity.

1.2. Hypotheses and contributions

In this paper we investigate the changes in users' click behavior when users encounter advertising and content with and without different types of background differentiation, the most common method of distinguishing text adverts on websites. Investigations into user behavior are crucial to provide appropriate guidance and recent interest in presentation styles by legislative bodies and oversight organizations, such as the FTC, has made this more crucial than ever. For publishers displaying advertising, an increased understanding of the effect of visual features on click rates can help to achieve a more enticing, or less detracting, presentation while maintaining compliance with the relevant guidance and legislation.

We further investigate whether background differentiation of advertisements has led to a priming effect that has consciously or subconsciously trained users to avoid content results with background differentiation—even though these content results might be the most relevant. We investigate the following hypotheses:

H1. background differentiation of advertising reduces click rates; and

Ads related to **google advertising** ⓘ

Want To Advertise Here? - Get Your Website Listed Here, Fast
[google.advertisingwebservice.co.uk/](#)
 Risk Free Advertising - Start Now.

Pay Per Click Advertising 020 7234 0500
[www.periscopix.co.uk/](#) - ★★★★★ 57 seller reviews
 Search Engine Advertising Experts. Call Us Now For A Free Consultation
 Periscopix has 184 followers on Google+
 PPC Management - Display Advertising - Remarketing - Mobile PPC Advertising

AdWords Advertising - Have Experts Manage your Account
[www.latitude-express.com/AdWords](#)
 Contact Us Today for a Free Review

Google Ads
[www.google.com/intl/en/ads/](#) ▼
 Learn about advertising on Google and how to make money from your site.
 New Advertiser - Looking to grow your business? - AdSense - Video Ads

Google AdWords – Online advertising by Google
[adwords.google.com/](#) ▼
 Advertise with Google AdWords ads next to Google search results to boost website traffic and sales. With Google AdWords pay-per-click (PPC) keyword ...
 Get started - How it works - AdWords Express - Costs

Ads ⓘ

Use Google AdWords
[www.google.co.uk/AdWords](#)
 0800 169 0488
 Reach The Right Customers Online
 Create Your Google Ad Today

Easy Online Advertising
[www.prweb.com/](#)
 Cost-Effective Way to Attract New Customers. Free Sign Up!
 PRWeb has 489 followers on Google+

Pay Per Click Works
[www.bluelake.co.uk/Adwords-Specialist](#)
 8 Years Pay Per Click Experience.
 Make Your Web Business A Winner!

Wisdek Corp.
[www.wisdek.com/](#)
 The Online Advertising Experts.
 Bring More Leads to Your Business

AdWords Professional
[www.briancarteryeah.com/](#)
 7 Years Experience.
 Proven Results. Get help.

Fig. 1. Google Search result page for "google advertising". Note: additional 'content' results followed.

H2. this effect also impacts content results.

Finally, we investigate the effect of an unusual advertising differentiation strategy on users' click behavior. We expect unfamiliar methods of separating on-screen content to have a reduced effect on click rates and to reject:

H3. unfamiliar differentiation also reduces click behavior.

We show that:

- There is a statistically significant drop in advertising click rates when using background coloration from 17.3% to 9.1% ($F_{1,100} = 26.669$, $\eta_p^2 = 0.211$, $p < 0.0005$). See Experiment 1 and confirmation in Experiment 2.
- There is a statistically significant drop in content click rates when using background coloration from 21.7% to 12.3% ($F_{1,100} = 24.698$, $\eta_p^2 = 0.198$, $p < 0.0005$). See Experiment 1 and confirmation in Experiment 2.
- There is no statistically significant difference in click rates for advertising ($F_{1,175} = 1.550$, $\eta_p^2 = 0.009$, $p = 0.215$) or content ($F_{1,175} = 2.091$, $\eta_p^2 = 0.012$, $p = 0.150$) when using an unfamiliar differentiation strategy. See Experiment 3.

2. Experiment 1: background coloration

We designed a Human Intelligence Task (HIT) for Amazon Mechanical Turk, an online microtask market that can be used to crowdsource user studies (Kittur, Chi, & Suh, 2008; Heer & Bostock, 2010; Mason & Suri, 2011). The HIT consisted of 20 test-cases, two verification cases, and five demographic questions. The experiment was a within-subjects design, with a fully randomized presentation of test cases. Participants were shown a modeled search engine environment and asked to select the best result for a given search phrase from a list of results by clicking on the link for that result. The interface closely resembled that of existing search engines, as seen in Fig. 2. The objective of the experiment was to identify which type of link, those with or without background differentiation, participants would select in a familiar situation, at a natural junction point where clicking a link is required.

2.1. Method

We recruited 101 participants for the experiment using Amazon Mechanical Turk. Participants were 36.7% female, 63.3% male, with a mean reported age of 27.75 years (ranging from 18 to 50, with 83.2% between 25 and 35). A standard deviation of 6.45 highlights the narrow age range of the dataset. Participants indicated their educational level using the UNESCO International Standard Classification of Education (UNESCO, 1997), with 47.5% indicating degree-level education of some kind, and the majority of participants (86.1%) indicating completion of secondary education.

To encourage accurate reporting, nationality was pre-selected using IP geo-location, although participants could change this if the identified country was incorrect or if they were physically in a location other than their home nation. Of 101 participants, 74 self-identified as being of Indian nationality, with the USA as the second largest nation represented, with three participants. A large variety of other nations were also represented with one or two participants, including Brazil, Canada, Greece, Kenya, Malaysia, Pakistan, Romania, Russia, South Africa, Thailand, and the United Kingdom. These demographics are consistent with previous studies using Mechanical Turk (Ross, Irani, Silberman, Zaldivar, & Tomlinson, 2010) and indicate a young, well-educated sample, familiar with the use of online services, suggesting previous exposure to Internet search tools and online advertising.

2.1.1. Task description and procedure

The purpose of the experiment was explained in an accompanying task description, and participants were assured that, although displayed, any advertising was included only to measure any effects and would not generate revenue for the HIT requester. To ensure participants understood the task, an example of the type of test case to be used was shown and the participant was required to provide a valid response (i.e. a click on any of the links) to continue.

While participants were at liberty to ask for additional information about the experiment, none did so. Participants were briefed with the following message (emphasis present in the original text):

The following academic survey requires you to review a selection of search results for a specified query, and indicate your selection of best result by clicking on the link you prefer.

Question 1

Please select, by clicking on the link, which result you believe is the most appropriate for the search **credit card**.

Your task is to select the best result using only the details provided below. Do **not** attempt to review the sites themselves. Clicking on any of the following links will **immediately** confirm that result as your selection and this action **cannot** be undone.

credit card

[Credit Cards | Personal | Visa USA](#)
usa.visa.com/personal/cards/credit/index.html
Review Visa's **credit card** options: Visa Standard **credit cards**, Visa Rewards **credit cards**, and Visa Premium Rewards **credit cards**, including the Visa Signature ...

[Credit Cards - Compare Credit Card Offers at CreditCards.com](#)
www.creditcards.com/
Compare **Credit Cards** & **Credit Card** Offers at CreditCards.com. Search **credit cards** and reviews about the best low interest, 0% balance transfer, reward, cash ...

[American Express® Cards | americanexpress.com](#)
www.americanexpress.com
View Special Offers on Our Most Popular **Cards**. Apply Now!

[Credit card - Wikipedia, the free encyclopedia](#)
en.wikipedia.org/wiki/Credit_card
A **credit card** is a small plastic card issued to users as a system of payment. It allows its holder to buy goods and services based on the holder's promise to pay ...

Fig. 2. Test-case as displayed to participants. Note: the window was scrolled to this position; additional scrolling would be required to see the fifth link.

The survey uses many commercial examples, however none of the results are live pay-per-click or pay-per-view advertisements and you will **not** be contributing clicks nor view counts to any advertising campaign.

After accepting the HIT, participants were asked to provide some general demographic information including their age, gender, nationality, level of education, and period of Internet usage in years. Participants then completed 20 test cases and two verification cases, with cases appearing in a random order, as seen in Fig. 2.

2.1.2. Test cases

Test cases were prepared in advance with content extracted from genuine Google search result pages for each search phrase to be shown to the participant. Google search result pages may contain up to three content area advertisements above the ten content results displayed per page, an advertising ratio of 3:10. To minimize the number of results that each participant must review before moving on and to maximize the number of samples gathered from each participant, each test case consisted of only five clickable items, one advert to four content results, a similar but slightly lower advertising ratio of 3:12.

Each test case consisted of the first displayed advertisement and the top four content results for the search phrase to accompany the content according to the Google search engine at the time of capture. The order of the items in each test case was randomized for each participant to control for any effects due to the ordering of the items on the page. A differentiating background color was applied to either zero or one item in each test case. The application of background coloration to the advert, one of the four content results, or neither, was determined randomly (uniform distribution) for each test case and for each participant. The expected nominal stimulus frequency was 5 differentiated adverts, 15 undifferentiated adverts, 5 differentiated content results, and 75 undifferentiated content results for each participant.

To most closely model the participants' probable real-world user experience, and to capture genuine reactions to background coloration, the color scheme and layout replicated the familiar presentation of Google search result pages. The subtle differentiation of the selected background color used on the highlighted items included a similarly subtle reduction in luminance (1.0, white, the control; 0.96, pale yellow, the differentiator), ensuring that color-blind participants would still perceive some form of differentiation. As color-blind individuals are not typically identified by website publishers, and are not presented different content from other Internet users, excluding users with this condition would not accurately represent user experience in a way that is typical of broader Internet usage and therefore no screening was performed.

Recognizing that participants might be from a variety of nations, test cases were designed to maximize their global applicability. Search phrases were selected for globally available branded products (e.g. iPad, Starbucks, Viagra) and commoditized items (e.g. flights, free music, online degrees). To ensure websites relevant to international participants would dominate the collected advertising and content results, both advertising and content results were gathered in advance using the selected search phrase on <http://www.google.com> through a publicly accessible US-based proxy server. Due to the randomized order in which test cases were displayed, it was essential to minimize any user targeting of the authors carried out by the search engine based on the order in which the searches were conducted. To prevent this type of targeting during the test-case data collection and preparation process a full browser reset and restart was conducted between each search carried out, removing any tracking cookies installed during the collection phase.

2.1.3. Verification cases and screening

Originally, data was gathered from 200 participants. However, to reduce noise introduced by invalid responses, such as systematically clicking the first link without looking at the text, participants were screened for experimental inclusion based on performance in two verification cases that were interleaved with the regular test cases. Any participant who failed either of the two verification cases was excluded from the data set. Due to the inclusion of the two test cases, this screening excluded 99 out of originally 200 participants, leaving 101 participants.

The verification cases each consisted of search phrases coupled with five search results. One was the 'correct' result, which was the top Google result for that particular search. The remaining four were the top results from other search phrases. The two verification cases were presented to participants in the same format as the normal test cases, including random ordering of the clickable items, and were inserted into random positions in the series of test cases. No attempt to disguise the alternate search phrases used to collect incorrect results was made, and any contextually inappropriate emboldened words shown in the captured items were preserved. Verification cases were validated for filtering purposes using a small pilot study of five Mechanical Turk participants in which all but one participant selected the 'known good' answer for both verification cases. In this pilot study, the participant failing the test case had practiced systematic selection of the displayed links. In the experiment, the probability of passing both verification cases by random chance was 4% (0.2×0.2).

2.2. Results

101 participants each provided 20 data-points, for a total of 2020 click-samples. Participants saw a total of 2020 adverts, 495 with background differentiation, 1525 without; and 8080 content results, 533 with differentiation, 7547 without. To allow direct comparison of the presented stimulus, the *click rate* was calculated for each participant, defined as the ratio of clicks made by a participant:

$$\text{click rate} = \frac{\text{clicked links}}{\text{shown links}}$$

The click rate was calculated for each participant for each of the four categories of clicks:

- Differentiated adverts.
- Undifferentiated adverts.
- Differentiated content results.
- Undifferentiated content results.

These results are summarized in Fig. 3 and Table 1.

A review of the aggregated click counts indicated that out of the 2020 clicks recorded, 311 were on advertisements, 15.4% of all clicks, somewhat below the frequency of advertising appearing in the test cases, where it represented approximately 20% of all clickable items displayed. The mean click rate was 9.1% on pages where the advertising was presented with background differentiation and 17.3% on pages where the advertisement was undifferentiated. This suggests background differentiation is a strongly negative factor for user engagement with online advertisements. This effect was not limited only to advertising results; participants also demonstrated a clear preference for undifferentiated content results, which received a click rate of 21.7% compared with 12.3% for differentiated content results. Further, both the mode and median click rate was zero for both differentiated advertising and content results (Table 1), highlighting that this aversion was extremely common among participants, occurring in at least half of all cases. In general, background differentiation significantly reduced click rates.

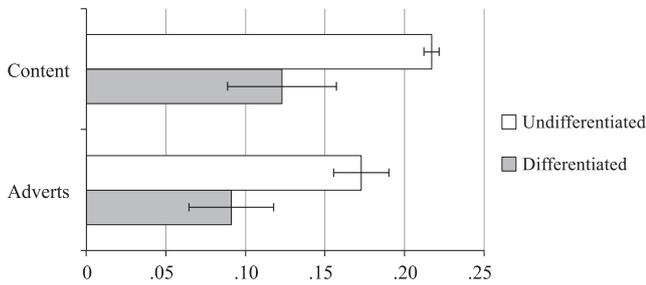


Fig. 3. Mean click rate for each variation in Experiment 1. Error bars indicate 95% confidence interval.

To understand how individual participants interpreted the differentiation a repeated-measures analysis of variance (ANOVA) at significance level $\alpha = 0.05$ was carried out on the per-participant click rates. The ANOVA showed that the difference in click rate between advertising with and without background differentiation was statistically significant ($F_{1,100} = 26.669, \eta_p^2 = 0.211, p < 0.0005$). A second repeated-measures ANOVA showed that same effect was statistically significant for content items ($F_{1,100} = 24.698, \eta_p^2 = 0.198, p < 0.0005$).

While our repeated measures design allowed us to compare individual participant behavior, for a typical publisher this may not be possible. To account for this we additionally considered the raw click data, ignoring the variance introduced by individual participant preferences. We carried out chi-squared tests to compare the distribution of clicks and stimulus seen. The aggregated click data showed a significant difference for both differentiated and undifferentiated advertising ($X^2(1, N = 311) = 15.863, p < 0.0005$) and content results ($X^2(1, N = 1709) = 16.542, p < 0.0005$).

3. Experiment 2: background coloration (US participants)

Due to the high percentage of Indian nationals in Experiment 1, a confirmatory run was essential to ensure potential cultural differences between our sample and the global Internet population was not influencing the results.

3.1. Method

To validate Experiment 1 a second HIT was launched requesting 100 assignments, with an eligibility requirement that workers be located in the United States. In all other details, the confirmatory experiment design was identical to Experiment 1.

3.2. Results

The HIT was requested with 100 assignments and 85 participants remained after a screening process that followed the same verification procedure as with Experiment 1. Of these, one participant indicated Georgian nationality, despite the filtering carried

Table 1
Summary of click data in Experiment 1.

Category	Adverts		Content results	
	Color	None	Color	None
Stimulus seen	495	1525	533	7547
Clicks	46	265	71	1638
Mean click rate	0.091	0.173	0.123	0.217
Median click rate	0	0.176	0	0.221
Mode click rate	0	0.188	0	0.227
Click rate SD	0.136	0.089	0.175	0.025

Table 2
Summary of click data in Experiment 2.

Category	Adverts		Content results	
	Color	None	Color	None
Stimulus seen	480	1200	384	6336
Clicks	51	209	55	1365
Mean click rate	0.110	0.174	0.140	0.215
Median click rate	0	0.176	0	0.215
Mode click rate	0	0	0	0.216
Click rate SD	0.158	0.094	0.195	0.025

out by Amazon Mechanical Turk, and was also removed, leaving 84 participants in the data set. As shown in Fig. 4 and Table 2, similar to Experiment 1, participants demonstrated a clear preference for undifferentiated advertising and content results.

To confirm our initial results, and again understand how individual participants interpreted the differentiation, a repeated-measures analysis of variance (ANOVA) at significance level $\alpha = 0.05$ revealed that the difference in click rate between advertising augmented with or without background differentiation was statistically significant ($F_{1,83} = 9.446, \eta_p^2 = 0.102, p < 0.005$). A second repeated-measures ANOVA showed that same effect was statistically significant for content items ($F_{1,83} = 10.922, \eta_p^2 = 0.116, p < 0.005$).

As before, we additionally carried out chi-squared tests to compare the distribution of clicks and stimulus seen. The aggregated click data showed a significant difference for both differentiated and undifferentiated advertising ($X^2(1, N = 260) = 10.219, p < 0.005$) and content results ($X^2(1, N = 1420) = 8.933, p < 0.005$).

In summary, the confirmatory experiment identified the same significant effects as our initial study.

4. Experiment 3: alternative differentiation strategies

To better understand any priming effect due to the users' frequent exposure to background coloration as an indicator of advertising, a final experiment was conducted to identify how users might react to an unfamiliar identification strategy. If users are avoiding—or not avoiding—a particular style of differentiation, then alternative strategies may be attractive to advertisers, publishers, and regulators, depending on the intended effect on user behavior.

4.1. Method

For this experiment, instead of using background coloration, we used a dotted-line border around the advertising or content to be differentiated. The border was defined using CSS and specified as “dotted”, 1 pixel wide, and black in color. We retained the no-border, white background control stimulus. For this final HIT

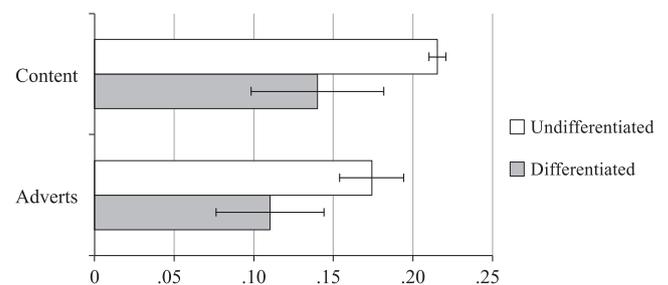


Fig. 4. Mean click rate for each variation in Experiment 2. Error bars show 95% confidence interval.

we requested 200 assignments. In all other details the experiment design was identical to Experiment 2 and retained the eligibility requirement requiring workers to be located in the United States. An example of the alternate differentiation strategy used in this experiment is seen in Fig. 5.

4.2. Results

The HIT was requested with 200 assignments and after a screening process, following the same verification procedure used with both Experiment 1 and Experiment 2, 180 participants remained. Of these, four participants indicated non-US nationality, despite the filtering carried out by Amazon Mechanical Turk, and were also removed leaving 176 participants in the data set.

Fig. 6 shows the reduced impact of unfamiliar differentiation compared to the more familiar background differentiation used in Experiment 1 and 2 (Figs. 3 and 4). Participants demonstrated a lack of clear preference for both undifferentiated advertising and content results. A review of the aggregated click counts indicated that out of these 3520 clicks, 614 (17.4%) were on advertisements. While still below the frequency of advertising appearing in the test cases (20%), this is higher than seen for background coloration (15.4% in Experiment 1).

On pages where the advertising was presented with border differentiation, participants selected this link as the best result with a mean click rate of just 15.8% of the time, compared with 18.0% on pages where the advertisement was not differentiated in this way. The effect of border differentiation (2.2% drop in click rate) is much less pronounced than seen for background differentiation (8.2% drop in Experiment 1). While the mode click rate for both differentiated advertising and content results was zero (Table 3), this common aversion was not nearly as overwhelming as a median of 16.7% for differentiated content results shows. When comparing with Experiments 1 and 2, this suggests that familiarity is likely a strong differentiator may be a factor in user behavior with advertising.

To understand how individual participants interpreted the differentiation a repeated-measures analysis of variance (ANOVA) at significance level $\alpha = .05$ was carried out on the per-participant

Table 3

Summary of click data from Experiment 3, where differentiation was by a dotted-line border.

Category	Adverts		Content results	
	Border	None	Border	None
Stimulus seen	910	2610	865	13215
Clicks	138	469	165	2748
Mean click rate	0.158	0.180	0.185	0.208
Median click rate	0	0.177	0.167	0.212
Mode click rate	0	0	0	0.224
Click rate SD	0.212	0.098	0.200	0.027

click rates. The ANOVA showed that the difference in click rate between advertising with and without border differentiation was not statistically significant ($F_{1,175} = 1.550$, $\eta_p^2 = 0.009$, $p = 0.215$). A second repeated-measures ANOVA showed that same effect was not significant for content items either ($F_{1,175} = 2.091$, $\eta_p^2 = 0.012$, $p = 0.150$). We additionally carried out chi-squared tests to compare the distribution of clicks and stimulus seen. The aggregated click data was not statistically significant for either differentiated and undifferentiated advertising ($X^2(1, N = 607) = 3.078$, $p = 0.079$) or content results ($X^2(1, N = 2913) = 1.160$, $p = 0.281$). This indicates that while in aggregate users *might* avoid content and advertising differentiated by a border individual behavior was much less consistent and the difference far less pronounced. Participants, given the opportunity, were no more likely to avoid the links differentiated by a border. This insight is particularly timely given the FTC's specific suggestion to provide "a clear outline" or "prominent border" around advertising (Federal Trade Commission, 2013).

5. Discussion

Our experiments confirm all three hypotheses we set out in the beginning of this paper. First, differentiating text-based advertisements via background coloration significantly reduces click rates. Second, differentiating text-based relevant content results via background coloration also significantly reduces click rates. To

Question 1

Please select, by clicking on the link, which result you believe is the most appropriate for the search **free music**.

Your task is to select the best result using only the details provided below. Do **not** attempt to review the sites themselves. Clicking on any of the following links will **immediately** confirm that result as your selection and this action **cannot** be undone.

[Grooveshark - Free Music Streaming, Online Music](#)
grooveshark.com/
 Grooveshark provides **free music** streaming, online radio stations, and lets you connect with artists and friends.

[Pandora Radio - Listen to Free Internet Radio, Find New Music](#)
www.pandora.com/
 Pandora radio is the personalized internet radio service that helps you find new **music** based on your old and current favorites. Create custom web radio stations ...

[Freeplay Music, Broadcast Production Music](#)
freeplaymusic.com/
 Freeplay Music, Highest Quality Broadcast Production Music
 Download, See Terms Of Use, Country, Hip Hop, Rap, Tech

[Jango - Free Music -- Listen to Music Online](#)
www.jango.com/
 Jango internet radio is the easy, legal way to play music from
 to listen to **free music** and make online radio stations.



Fig. 5. Alternative differentiation strategy using a dotted-line border, as used in Experiment 3. Detail of the alternative differentiation style is shown inset. Note: the window was scrolled to this position; additional scrolling would be required to see the fifth link.

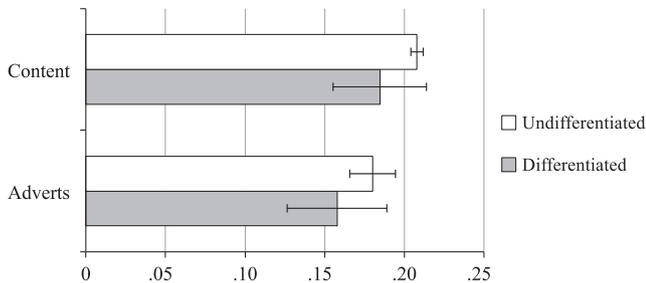


Fig. 6. Mean click rate for each variation in Experiment 3. Error bars show 95% confidence interval.

the best of our knowledge this is the first empirical evidence to demonstrate these effects on users' click behavior. Finally, by using an unfamiliar differentiation strategy, we show no evidence of a statistically significant impact of unfamiliar border differentiation on click rates.

5.1. User engagement

Our experiments show that users are less willing to engage with textual context-sensitive advertising when advertisements are displayed with a differentiating background. These results corroborate the predictions of Goldfarb and Tucker (2011), in their related but distinct study of obtrusive multimedia advertisements. However, obtrusiveness without a pre-existing connotation—as seen in Experiment 3—does not impact user behavior to the same degree. Disinclination to engage with differentiated content was not limited to advertising: we also identified a strong negative reaction to *content results* displayed with background differentiation. The strength of this effect with such a subtle and minor differentiation exceeded our expectations and focuses attention on the importance of presentation both for text-based advertising, but also for content, and how presentation may affect user behavior.

Studies of obtrusiveness of online advertising have suggested negative engagement might be caused by users' privacy concerns due to the highlighting of sensitive personal information (Goldfarb & Tucker, 2011). However, this is probably not a contributing factor in this study due to the pre-prepared nature of the test cases. Each participant was presented the same advertising and content results based on a fixed set of search phrases, unrelated to the participant or their preferences. Due to the pre-gathered, fixed content, no personal information was or could be featured in the results. Further, for advertising of a sensitive nature, such as in our "viagra" test case, participants demonstrated a click rate of above 20%—as might be expected from a random selection of links—when the ad was displayed both with and without background differentiation. This suggests that the lower click rate of differentiated advertisements is probably due to other factors.

5.2. Possible priming effect

Schmitt (1994) discusses the effect of context on advertising and how visual information prefixing, or surrounding content, may impact user interpretations of the content. Repeated exposure to differentiation techniques may lead to a priming effect. Additionally, consumers tend to make associations between color and function, often selecting safe and conventional choices when choosing items, and the consumer interpretation of color may be different from those intended (Grossman & Wisenblit, 1999). These difficulties in communicating meaning can lead to confusion for Internet users, something regulatory bodies would like to avoid.

As Internet users experience repeated exposure to advertising, click rates decrease (Hofacker & Murphy, 1998). Internet users may associate advertising with deceptive practices (Goldfarb & Tucker, 2011), taking them away from 'genuine' content, and they may feel manipulated when identifying the mechanism used to encourage their engagement with the advert. The perception of manipulation by users may be contradictory to the actual intent of advertisers who are compelled to differentiate their advertising from content (Federal Trade Commission., 2000). This inverse relationship between recognition of a technique and engagement leads us to hypothesize that users develop a learned behavior to avoid clicking on links differentiated by their background in order to protect themselves from their negative perceptions of advertising, and this perception may impact future behavior when viewing differentiated content. Our final experiment investigated an alternative technique and presentation, not currently used by any of the major search engine providers. Here we demonstrate that using an alternative differentiation style does not significantly impact user behavior, strengthening the case for a learned behavior with more familiar differentiation techniques.

5.3. Implications for design

As a differentiation strategy with unclear intent, our paper shows that background differentiation creates a negative perception of presentation, including differentiated content. Additionally, we show that an alternative differentiation strategy, dotted borders, has a lessened impact on user behavior. These results suggest that highlighting of important content in results, be this for financial or regulatory reasons in our advertising test cases, or to indicate the importance of particular content, may be detrimental if the purpose is not clear to users and differentiation carefully applied. Where user experience is a high priority for publishers, differentiation strategies might focus on a standardized indication of advertising and an explanation of the mutual benefits of advertising, such as access to relevant products and services for users, and as a source of revenue for the website. For regulators, these findings question the benefit of encouraging providers to vary their advertising styling from the established industry-wide practice and point to the importance of consistency with differentiation techniques to encourage recognition and to match user expectations.

For advertising networks, these findings highlight potential opportunities for improving user engagement. Understanding differentiation, as a factor of user engagement, can also be useful for advertising networks syndicating these adverts, and who receive a portion of the revenue paid by advertisers for clicks. By dynamically reviewing the website where adverts will be published, either by automatically reviewing the page, or as part of the publisher's request for the advert, the network can factor the expected presentation of the adverts into its dynamic selection process. Improving the intelligence of these algorithms increases the chance of selecting an enticing advert that both benefits the user and encourages engagement. Appropriate application of these techniques could result in increased click rates and click-derived revenue to be shared between the advertising network and the website publisher.

5.4. Using crowdsourcing for experiments

Prior studies, such as Kittur et al. (2008) and Mason and Suri (2011), highlight some of the difficulties in carrying out user studies via crowdsourcing, such as 'gaming' of the system, where participants provide systematic or minimal responses and do not follow instructions. Amazon. (2011) goes further still and suggests that failure of 'workers' to follow instructions, and any attempt to

'game' the system, might be handled by non-payment to discourage intentionally poor quality work.

In previous comparative studies, Mechanical Turk has been shown to be a reliable way of gathering data for visualization studies (Heer & Bostock, 2010), behavioral studies (Mason & Suri, 2011), and ranking of perceptual data (Rosenthal & Dey, 2010). Further, it significantly increases the number of participants used for data collection at very low cost (Kittur et al., 2008). The use of Mechanical Turk for crowdsourced studies has proven viable for presenting new insights (Heer & Bostock, 2010), with results being comparable to those gathered from laboratory subjects (Heer & Bostock, 2010; Mason & Suri, 2011). However, appropriate care in the form of validation and verification must be taken to ensure valid responses (Kittur et al., 2008). To minimize this unwanted interference to our data, our experiments included verification cases to eliminate questionable responses, and preventing our sample from becoming contaminated by noise.

Despite an overwhelming proportion of Indian participants in our first study, these demographics are generally consistent with those reported in published Mechanical Turk studies (Mason & Suri, 2011; Ross et al., 2010). However, to control for potentially atypical demographics we carried out a confirmatory experiment with participants from the United States. This second study confirmed the significant effects we observed in our first study.

6. Conclusions and future work

In this paper we studied the effect of differentiation strategies with text-based advertisements on users' click behavior. Specifically, we have investigated three hypotheses. Our first hypothesis: that click rates are lower for background differentiated advertisement results. Our second hypothesis: click rates are also lower for background differentiated content results. Our empirical results confirm these hypotheses and find that displaying advertisement and content results with a differentiated background results in significantly lower click rates. Finally, our third hypothesis: unfamiliar differentiation reduces click rates. Here we reject the hypothesis since we identified no statistically significant difference. This suggests that an unfamiliar differentiation strategy has a more mixed user response. We suggest our empirical results are due to a priming effect that has trained users to avoid background differentiated content results, *even though these results might be the most relevant*. By using unfamiliar differentiation strategies, untainted by this priming, impact on user behavior is not statistically significant.

This paper has established background color differentiation as a significant contributing factor for user engagement in text-based advertising and content. Due to the myriad of possible variations, and without any pre-existing perception as to the effect of this change, we first focused on a single background differentiation strategy, altering the background color, in a familiar environment. We further investigated an unusual background differentiation strategy, a dotted-line border, to confirm the suspected priming effect. To further generalize these findings, a fruitful avenue for exploration might contrast a number of differentiating background styles and colors, thus allowing the identification of any user associations between various colorations and the type of content being displayed. This is useful because our results indicate border differentiation also provides many opportunities to encourage user interaction while maintaining compliance with regulator guidance. It is however important to note that we have not established a causal relationship between an unfamiliar background differentiation strategy and click rate. Our failure to reject the null hypothesis for our third hypothesis does not provide evidence that the null hypothesis is true. A different experimental design might provide evidence of a causal relationship.

In Experiment 1 and Experiment 2, we displayed the same background differentiation to participants of all nationalities—as can be observed in the Google search interface. In Experiment 3 we used an unfamiliar presentation, unused by all of the major contemporary advertisement-based search providers. While this paper highlights some important specific observations, the generality of this work might be further confirmed by attempting to stratify data collection using a series of nationally targeted studies, identifying any significant differences between national subsets of Internet users. By closely simulating Google's search interface, these experiments would allow investigation of the impact of any pre-existing associations made by users of this environment. Users may be aware of the increased prevalence given to advertising highlighted in this way and might actively avoid this type of link to prevent the manipulation they may perceive (Goldfarb & Tucker, 2011). With the combinatorial nature of visual presentations, our paper has focused on establishing the impact of existing industry practice, and how recently published guidance may affect user interaction with online advertising. Future work might investigate the display of more differentiation strategies such as boxes, hyperlinks, and animation (Federal Trade Commission, 2000), and ensuring that these changes would not cause dissatisfaction among internet users (Fallows, 2005). Such studies could verify any negative user associations with differentiation while breaking away from any existing preconceptions, practices, and guidelines, thereby identifying user preferences for content differentiation outside of these frameworks.

Acknowledgement

This work was supported by a studentship from the Engineering and Physical Sciences Research Council. Additional data related to this publication is available at the University of Cambridge data repository: <http://www.repository.cam.ac.uk/handle/1810/247391>.

References

- Amazon. (2011). *Requester Best Practices Guide*. Retrieved from <http://mturkpublic.s3.amazonaws.com/docs/MTURK_BP.pdf>.
- Broder, A., Fontoura, M., Josifovski, V., & Riedel, L. (2007). A semantic approach to contextual advertising. In *Proceedings of the 30th annual international ACM SIGIR conference on research and development in information retrieval* (pp. 559–566). New York, NY, USA: ACM. <<http://doi.org/10.1145/1277741.1277837>>.
- Castro, D. (2012, September 27). Why ITIF Rejects Your "Do Not Track" Request. Retrieved from <<http://www.itif.org/publications/why-itif-rejects-your-do-not-track-request>>.
- Evans, D. S. (2008). The economics of the online advertising industry. *Review of Network Economics*, 7(3), 23–55.
- Fallows, D. (2005). *Search Engine Users*. Pew Internet & American Life Project. Retrieved from <<http://www.pewinternet.org/Reports/2005/Search-Engine-Users.aspx>>.
- Federal Trade Commission. (2000, May). Dot Com Disclosures: Information About Online Advertising. Federal Trade Commission. Retrieved from <<http://business.ftc.gov/sites/default/files/pdf/bus41-dot-com-disclosures-information-about-online-advertising.pdf>>.
- Federal Trade Commission. (2013). Sample Letter to General Purpose Search Engines. Retrieved from <<http://www.ftc.gov/os/2013/06/130625searchenginegeneralletter.pdf>>.
- Fox, D., Smith, A., Chaparro, B. S., & Shaikh, A. D. (2009). Optimizing Presentation of AdSense Ads within Blogs. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 53, pp. 1267–1271). Sage. <<http://doi.org/10.1177/154193120905301822>>.
- Goldfarb, A., & Tucker, C. (2011). Online display advertising: Targeting and obtrusiveness. *Marketing Science*, 30(3), 389–404. <<http://doi.org/10.1287/mksc.1100.0583>>.
- Google. (2010). 10 Things Every AdSense Publisher Should Know. Retrieved from <http://www.google.com/adsense/start/pdf/AdSense_10Things.pdf>.
- Google. (2014). 2014 Financial Tables – Investor Relations – Google. Retrieved May 20, 2014, from <<http://investor.google.com/financial/tables.html>>.
- Grossman, R. P., & Wisenblit, J. Z. (1999). What we know about consumers' color choices. *Journal of Marketing Practice: Applied Marketing Science*, 5(3), 78–88. <<http://doi.org/10.1108/EUM00000000004565>>.
- Heer, J., & Bostock, M. (2010). Crowdsourcing graphical perception: Using mechanical turk to assess visualization design. In *Proceedings of the SIGCHI*

- Conference on Human Factors in Computing Systems (pp. 203–212). New York, NY, USA: ACM. <<http://doi.org/10.1145/1753326.1753357>>.
- Hofacker, C. F., & Murphy, J. (1998). World Wide Web banner advertisement copy testing. *European Journal of Marketing*, 32(7/8), 703–712. <<http://doi.org/10.1108/03090569810224092>>.
- Kittur, A., Chi, E. H., & Suh, B. (2008). Crowdsourcing user studies with Mechanical Turk. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 453–456). New York, NY, USA: ACM. <<http://doi.org/10.1145/1357054.1357127>>.
- Mason, W., & Suri, S. (2011). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*, 44(1), 1–23. <<http://doi.org/10.3758/s13428-011-0124-6>>.
- Rosenthal, S. L., & Dey, A. K. (2010). Towards maximizing the accuracy of human-labeled sensor data. In *Proceedings of the 15th International Conference on Intelligent User Interfaces* (pp. 259–268). New York, NY, USA: ACM. <<http://doi.org/10.1145/1719970.1720006>>.
- Ross, J., Irani, L., Silberman, M. S., Zaldivar, A., & Tomlinson, B. (2010). Who are the crowdworkers?: shifting demographics in mechanical turk. In *Extended Abstracts of the 28th ACM Conference on Human Factors in Computing Systems* (pp. 2863–2872). New York, NY, USA: ACM. <<http://doi.org/10.1145/1753846.1753873>>.
- Schmitt, B. H. (1994). Contextual priming of visual information in advertisements. *Psychology and Marketing*, 11(1), 1–14. <<http://doi.org/10.1002/mar.4220110103>>.
- Ting, I., Clark, L., Kimble, C., Kudenko, D., & Wright, P. (2007). APD-A Tool for Identifying Behavioural Patterns Automatically from Clickstream Data. In B. Apolloni, R. Howlett, & L. Jain (Eds.), *Knowledge-Based Intelligent Information and Engineering Systems* (Vol. 4693, pp. 66–73). Springer. Retrieved from <<http://www.springerlink.com/content/3546913450q8vpx3/abstract/>>.
- UNESCO. (1997). International Standard Classification of Education. Retrieved from <<http://unesdoc.unesco.org/images/0011/001113/111387eo.pdf>>.