Analyzing Advertising Labels: Testing Consumers' Recognition of Paid Content Online

Jeff Johnson, Ph.D.
Computer Science Dept.
University of San Francisco
San Francisco, CA 94117, USA
jajohnson9@usfca.edu

Bernard J. Jansen, Ph.D
Qatar Computing Research Inst.
Hamad Bin Khalifa University
Qatar
jjansen@acm.org

Manoj Hastak, Ph.D.
Kogod School of Business
American University
Washington, D.C., 20016, USA
mhastak@american.edu

Devesh Raval, Ph.D.
U.S. Federal Trade Commission
600 Pennsylvania Ave, NW
Washington, D.C. 20580, USA
draval@ftc.gov

Abstract

In 2014-2015, the U.S. Federal Trade Commission (FTC) commissioned a study to assess consumers' ability to recognize ads and other paid content in online search results and news/article feeds. The co-authors designed the study, oversaw its execution, and analyzed the results, with support from FTC staff. The goals of the research were to assess the effectiveness of methods that online services use to label ads, and to see if alternative methods of labeling ads could improve consumers' ability to recognize them. In a controlled experiment, 48 consumers interacted with both desktop and mobile Web pages that were captured from search and online magazine websites. In half of the conditions, the Web pages were modified based on established Web design guidelines to improve the clarity of ad labeling. The participants' behavior, comments, and eye movements were recorded. Initial findings of this experiment are: (a) consumers cannot always distinguish ads, paid content, and paid search results from unpaid content, and (b) improving the salience and placement of labels based on established UI design guidelines can improve consumers' ability to recognize ads, paid content, and paid search results. We conclude with implications of the results and areas for future research.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.
CHI'18 Extended Abstracts, April 21–26, 2018, Montreal, QC, Canada
© 2018 Copyright is held by the owner/author(s).
ACM ISBN 978-1-4503-5621-3/18/04.
https://doi.org/10.1145/3170427.3188533

Author Keywords

User interface design; graphic design; search results; search engines; CHI policy; advertising; ads; native ads; deceptive ads; regulation; consumer; FTC

ACM Classification Keywords

D.2.2. Design tools and techniques, user interfaces; H.1.2. User/machine systems, human factors; H.5.2. User interfaces: theory and methods, Ergonomics, Screen design, User-centered design, Style-guides, Graphical user interfaces.

Introduction

Online search, commerce, news, commentary, and entertainment are now a part of daily life in much of the world, including the United States. Advertisements, paid content, and other pay-for-placement content are

an important part of that online experience: they provide revenue. However, as new forms of advertising have emerged, so have concerns about consumers' ability to recognize ads.

Historically, ads consisted of banner ads and ads embedded within search results. As consumer use of the Web grew, a new type of ad emerged, which the ad industry calls "native". Native ads are positioned with and designed to resemble non-ad content.

The U.S. Federal Trade Commission (FTC) is charged with, among other things, protecting U.S. consumers from deceptive advertising, including deceptively-formatted ads. According to the FTC's Enforcement Policy Statement on Deceptively Formatted Advertisements, "an ad is deceptive if it materially misleads consumers as to its commercial nature or source" [2]. The FTC's policy is that consumers should be able to distinguish ads from non-ads [4].

Some prior studies examined how ad labelling affects people's clicking or buying behavior or visual attention [5, 7]. A few used online surveys to ask consumers explicitly if they recognize a specific ad as an ad [10]. To our knowledge, no prior studies used controlled manipulation of ad labelling and collection of behavioral data to assess the effect of labeling on ad recognition.

To help inform the FTC's understanding of the issues, the FTC sponsored a study to answer these questions:

1. How well can people distinguish paid content from unpaid (natural or editorial) content, given prevailing methods of marking paid content?

2. Are there methods of marking ads and paid content that would be easier for people to recognize?

The FTC announced the study on Dec 15, 2017 in a press release [3] and a staff report [4]. This paper presents highlights of the study described in the report.

Method

Materials: Original Pages

The FTC captured eight sets of web pages in mid-2014 as representative samples of popular websites that display advertisements or paid content while providing search, online shopping, product review, news, and entertainment services. Each page set included a primary page and some site pages linked to the primary page. Four page sets were captured from popular search websites and four were captured from media websites providing news, product review, and entertainment article feeds (see Table 1).

***Search* page sets**
Google desktop search for "tablet": results page & click-through Shopping page
Google mobile search for "carpet cleaner": results page & shopping page
Bing desktop search for "air purifier": results page
Bing mobile search for "vitamin D supplements": results page
***Article feed* page sets with "native" ads**
Gear Patrol desktop home page & Subaru click-through article page
Yahoo desktop home page & Instaflex desktop click-through article
Time magazine mobile article-list page & ADT click-through article page
Chicago Tribune Embassy Studio mobile article page

Table 1: Pages captured for study

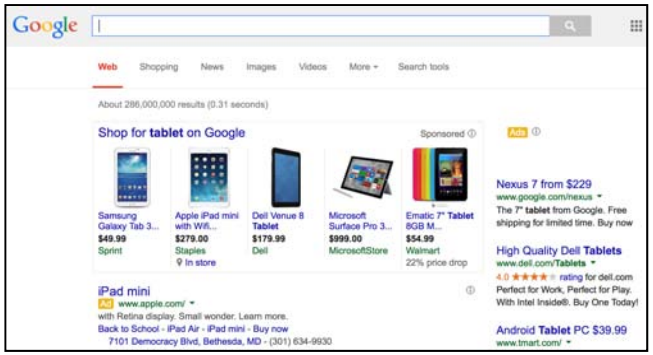

Fig. 1: Top of Google desktop results page, showing ad labels

The captured page sets marked ads in various ways (see Figs. 1 and 2). For example, the Google search-results page marked "Shop … on Google" links at the top of the page by enclosing them in a box with a gray border and displaying "Sponsored" in the top right corner of the box, and marked ads below and to the right of the box with a white-on-yellow "Ad" symbol.

Examples of Established Usability Design Guidelines Used for Modifications

- Make headings and labels more noticeable and legible by using large and bold fonts.
- Text color should contrast strongly with the background. Dark text on light backgrounds is preferable to light text on dark backgrounds.
- When color differences convey meaning, use colors that contrast well with each other, and provide redundant cues other than color.
- Visually distinguish different types of content with labeling, highly visible borders, or background colors.
- Position and space labels so it is clear what content-items they cover.
- Avoid having multiple terms or symbols for the same concept, or the same term or symbol that means multiple concepts.
- Use terms that are familiar and clear to users. Avoid technical or industry-internal jargon.
- Paid advertisements should look like advertisements, not site content or navigation.

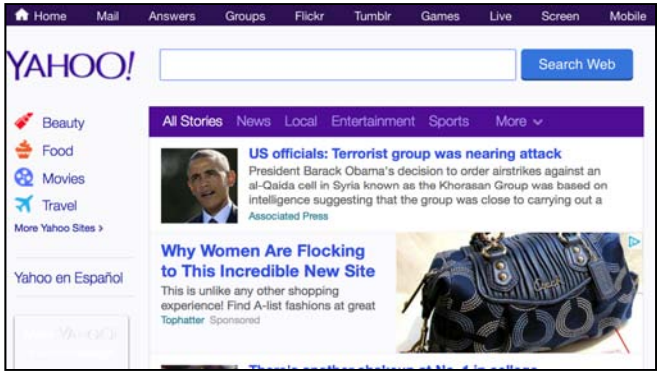

Fig. 2: Top of Yahoo home page, showing ad labels

Materials: Modified Pages

An alternate version of each captured page-set was created by modifying ad labels to improve their clarity, salience, and placement (see Figs 3 and 4). The modifications were made by copying the captured pages, then editing the HTML and CSS code. The intent was to bring the ad labels into better conformance with well-established user-interface and Web design guidelines [1, 6, 8, 9] (see sidebar, left).

An important principle guiding our modifications was that they should be minimal, improving the ad labelling while leaving the pages as unchanged as possible. This was to avoid changing the overall page design, which companies regard as integral to their brand and corporate image, and to facilitate attributing observed performance differences to our modifications.

The detailed modifications to the captured pages, along with the rationale and guidelines behind each change, are described in the full FTC study report [4].

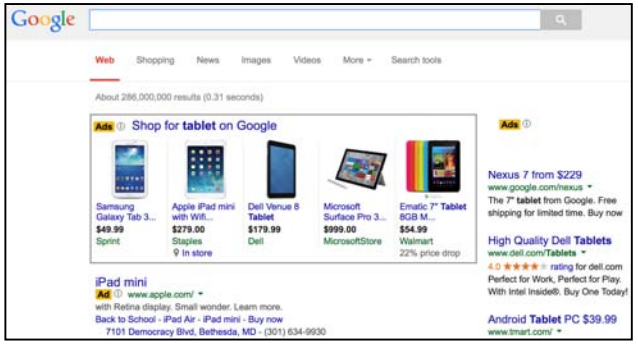

Fig. 3: Top of altered Google desktop results page, showing modified ad labels

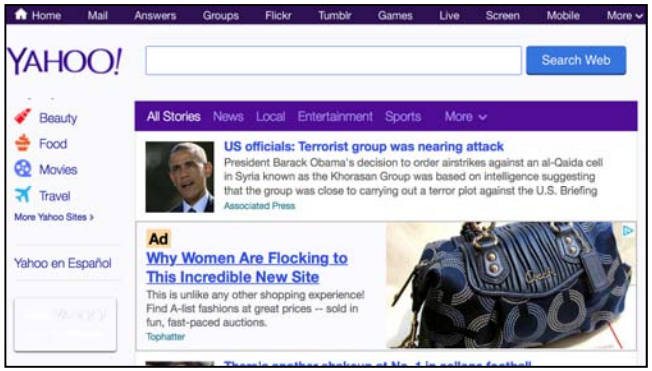

Fig. 4: Top of altered Yahoo home page, showing modified ad labels

Design

The experimental design was mixed: within and between subject. Each participant was shown all eight page sets, half original and half modified. No participant was shown both the original and modified version of the same page set. Page sets were assigned such that half of the participants saw four original page sets and four modified page sets, and the other half of the participants saw the other four original page sets

Sex
male: 22
female: 26
Age
18-24: 5
25-34: 9
35-44: 8
45-54: 10
55-64: 12
65+: 4
Education
high school: 4
some college: 13
college degree: 20
post graduate: 11
Ethnicity
Caucasian/White: 30
African American/Black: 13
Asian: 2
Hispanic/Latino: 2
Hawaiian/Pacific Islander: 1
Internet Usage
Frequent: 23
Occasional: 25

Table 2: Participant demographics

and the other four modified page sets. In addition, half of the page sets shown to each participant were desktop and half were mobile, and half were search and half were native ad. Every participant saw the eight page sets in a different order.

Participants

Participants were recruited from the metropolitan Washington, D.C. area and pre-screened by phone. To participate, they had to be regular users of computers, smartphones, search engines, and the Web. They could not: (a) have participated in a study or focus group in the past 6 months, (b) work in any of the industries represented in the captured websites, or (c) have a blog of their own that presented advertising. In addition, they could not have *nystagmus* (involuntary eye movements), *strabismus* (cross-eyes or wall-eyes), wear bifocals, trifocals, or progressive lenses, or wear contact lenses and reading glasses simultaneously.

Forty-eight (48) participants were chosen, representing a mix of sex, age, education level, ethnicity, and self-reported frequency of Internet use (see Table 2). The demographic categories were approximately balanced across conditions. Participants received $125.

Procedure

A moderator and an observer conducted each session, following a script developed by the co-authors in consultation with FTC staff. Test-sessions lasted about 90 minutes (1.5 hours), consisting of introductory instructions, calibration of the eye-tracking equipment, presentation of four page sets one at a time, a short break, recalibration of eye-tracking, presentation of the final four page sets, and a post-test debriefing in which participants were asked if they recognized any ads on the last page set, and in general how they respond to online ads and interpret common ad-label terms.

The presentation of each page set consisted of the participant (P) reading brief instructions, then pressing a start key, which displayed the first screen and activated eye-tracking. P was asked to "think aloud" while viewing each page. When P paused or was about to (or did) click to another page, the moderator (M) directed P back to the initial page and began asking probing questions. M then directed P's attention to a specific item of interest on the page and asked for comments. Some of these items were ads of interest in the study, while others were non-ad decoys to conceal the study's purpose. M asked P to click on the item – ad or decoy. As the click-through page appeared, P's eyes were tracked and their comments recorded. M asked P questions about the click-through page, then directed P back to the first page to offer any final comments. Finally, M transitioned to the next page set.

Data Collection

The following data was collected for each page set:

- Video recording of computer or smartphone screen with P's eye-movements and fixations superimposed.
- Time-sequenced eye-movements and fixations.
- Audio recording of what P and M said.
- Observer's notes about P's behavior and comments.

Analysis & Results

Preprocessing

The session videos were transcribed, yielding written records of what was said in each session. The time-sequenced eye-tracking data files were processed to produce spreadsheets showing eye-movement statistics for various areas of interest (AOIs) in the page sets.

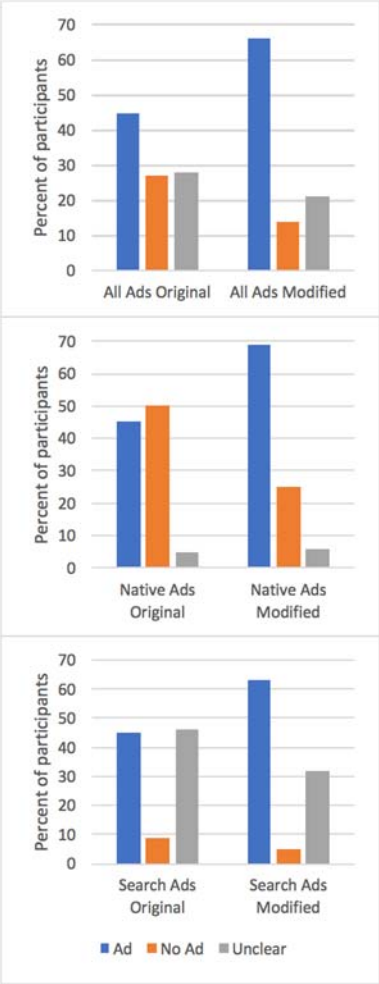

Fig. 5: Comparing distributions of coded ad recognition outcomes between original and modified page sets.

Ad Recognition Analysis

As described above, participants' behavior as they viewed the captured or modified pages and their comments about each page were recorded and examined. Three judges – one of the co-authors and two FTC staff members not involved in the study – independently reviewed the video and/or transcript of each scenario, focusing on segments in which participants (Ps) were exposed to ads. For each segment, the judges coded one of three outcomes: recognized ad ("Ad"), did not recognize ad ("No ad"), and unclear ("Unclear"). The agreement between coders was 79.2%. The remaining coding differences were resolved by deferring to the coding co-author. (Deferring to the FTC coder instead does not substantively change the results or conclusions.)

To see if the modified ad labels affected peoples' ability to recognize ads, we calculated the percentage distribution of the three coded ad-recognition outcomes for the original page sets and compared it with that for the modified page sets. (The percentages of the three outcomes for a group of page sets sums to 100%.) We did this separately for the native ad page sets and the search ad page sets, and for all page sets together.

The distributions for the modified page sets differed from those for the original page sets (see Fig. 5): the percentage of ads correctly recognized was higher (and of course that of ads *not* recognized was lower).

To determine if these differences in distributions were statistically significant, we used a logistic (categorical) regression analysis. This regression analysis controls for factors that could influence the likelihood of a participant correctly recognizing an ad as an ad, including the fact that some ads are more recognizable as ads than other ads, the tendency of an individual participant to correctly recognize an ad or to presume that all content is advertising, or the order in which a participant saw the ads. (See Appendix F in [4] for details of the regression analysis.)

Applying this analysis, the aggregated effects of the modified ad labelling across the different ads we tested are statistically significant ($p < .05$). Overall, the modified ad labelling increased the probability that a participant recognized an ad by 21 percentage points, with the 95% confidence interval ranging from 15 to 27 percentage points (see Fig. 6).

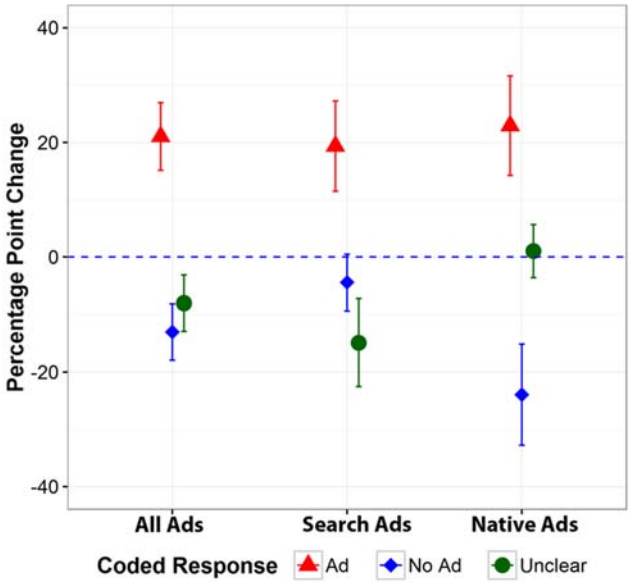

Fig. 6: 95% confidence intervals for percentage change in ad recognition due to modified ad labeling, for Search page sets and Native page sets separately, and for all page sets.

Avenues for Further Research

- Try to isolate the effects of ad-labelling factors such as label location, language, color, size on ad recognition.
- Effect of demographic variables (age, gender, Internet experience) on ad recognition and eye-movements.
- Analyze answers to post-test questions about level of trust in content.
- Conduct a survey with many respondents, asking some of the present study's post-test questions.
- Repeat study or conduct survey with a different protocol in which participants are asked to make explicit forced-choice assessments of whether item are ads or not, eliminating or sharply decreasing the possibility of "unclear" assessments.
- Repeat study with search tasks that are not shopping, to see if participants' ad recognition becomes clearer when the distinction between ads and non-ads is more important.
- Repeat study with multiple separated modifications for each original page to allow effect of specific modifications to be separated. For example, one modified page could relocate the ad disclosure, while another modified page could re-word it.
- Update study using more current ads.

We also found substantial improvements in ad recognition when we separately examined the search and native ad condition groups (see Fig 6). In the search ad conditions, we found that the probability that participants recognized the ads as ads increased by 19 percentage points (95% confidence interval of 11 to 27 points). The effects of the ad labeling modifications in the native ad conditions was a 23-percentage-point increase (95% confidence interval of 14 to 32 points). These results were significantly different from zero. The effects are similar when the page sets were analyzed in different groupings, e.g., ads viewed on a desktop computer versus a smartphone.

Conclusions

In this study, we found, both in search results with paid result items and in article-feeds with embedded "native" ads, that consumers cannot always recognize whether content they were viewing was a paid ad or not. Additionally, we found that small improvements based on widely-accepted Web design guidelines can significantly improve consumers' ability to recognize ads and paid search results.

The full FTC report [4] describes and analyzes the ad recognition results in greater detail, as well as analyzing the eye-tracking data. We plan to conduct further analysis of the data from this study. We hope that subsequent research (see sidebar, left) will refine and augment these findings. For example, native ads are more common now than in 2014 (when the ads used in this study were captured), so it would be useful to conduct studies using more up-to-date ads.

Acknowledgements

The authors wish to thank their colleagues at the FTC for their collaboration and guidance on this research. In particular, Laura Sullivan at the FTC was a key resource in guiding this study from inception to completion.

Disclaimer

The views expressed in this article are those of the authors. They do not necessarily represent those of the Federal Trade Commission or any of its Commissioners.

References

1. Arch, A. (2008) "Web Accessibility for Older Users: A Literature Review", WorldWideWeb Consortium (W3C)
2. FTC (2016), *FTC Enforcement Policy Statement on Deceptively Formatted Advertisements*, 81 Fed. Reg. 22596 (Apr 18).
3. FTC (2017), Press Release: FTC Publishes Staff Report on Exploratory Research Examining Consumer Recognition of Paid Search and Native Advertising Online (Dec 15).
4. FTC (2017), *Blurred Lines: An Exploration of Consumers' Advertising Recognition in the Contexts of Search Engines and Native Advertising*, (Dec 15).
5. Jansen, B.J. and Spink, A. (2009) Investigating Customer Click-through Behavior with Integrated Sponsored and Non-Sponsored Results, Int'l J. of Internet Marketing and Advertisement, 5(1/2), 74-94.
6. Johnson J. (2014) *Designing with the Mind in Mind, 2nd Ed.*, Morgan Kaufman Publishers.
7. Navdeep S. Sahni, and Harikesh S. Nair (2017), Native Advertising, Sponsorship Disclosure and Consumer Deception: Evidence from Mobile Search-Ad Experiments, (Feb 23), SSRN.
8. Pernice, K., Estes, J. and Nielsen, J. (2013) Senior Citizens on the Web, Nielsen Norman Group Report.
9. Usability.gov, (2015) Research-Based Web Design and Usability Guidelines.
10. Wojdynski, B. (2017), The Deceptiveness of Sponsored News Articles: How Readers Recognize and Perceive Native Advertising, *American Behavioral Scientist*, 60(12), 1475-1491.