



## Marketing Science

Publication details, including instructions for authors and subscription information:  
<http://pubsonline.informs.org>

### Sponsorship Disclosure and Consumer Deception: Experimental Evidence from Native Advertising in Mobile Search

Navdeep S. Sahni, Harikesh S. Nair

To cite this article:

Navdeep S. Sahni, Harikesh S. Nair (2020) Sponsorship Disclosure and Consumer Deception: Experimental Evidence from Native Advertising in Mobile Search. Marketing Science 39(1):5-32. <https://doi.org/10.1287/mksc.2018.1125>

Full terms and conditions of use: <https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact [permissions@informs.org](mailto:permissions@informs.org).

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2019, INFORMS

Please scroll down for article—it is on subsequent pages



With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

# Sponsorship Disclosure and Consumer Deception: Experimental Evidence from Native Advertising in Mobile Search

Navdeep S. Sahni,<sup>a</sup> Harikesh S. Nair<sup>a,b</sup>

<sup>a</sup>Stanford Graduate School of Business, Stanford University, Stanford, California 94305; <sup>b</sup>JD.com, Mountain View, California 94043

Contact: navdeep.sahni@stanford.edu,  <http://orcid.org/0000-0002-5637-5499> (NSS); harikesh.nair@stanford.edu,

 <http://orcid.org/0000-0003-1697-7767> (HSN)

Received: February 21, 2017

Revised: August 29, 2017; December 21, 2017;  
March 30, 2018

Accepted: April 11, 2018

Published Online in Articles in Advance:  
July 8, 2019

<https://doi.org/10.1287/mksc.2018.1125>

Copyright: © 2019 INFORMS

**Abstract.** Recent advances in advertising technology have lead to the development of “native advertising,” which is a format of advertising that mimics the other nonsponsored content on the medium. Whereas advertisers have rapidly embraced the format on a variety of digital media, regulators have expressed serious concerns about whether this format materially deceives consumers because the advertising disclosure is incomplete or inappropriate. This has reignited a longstanding debate about the distinction between advertising and content in media markets, and how it affects consumers. This paper contributes to this debate by providing empirical evidence from a randomized experiment conducted on native advertising at a mobile restaurant-search platform. We experimentally vary the format of paid-search advertising, the extent to which ads are disclosed to over 200,000 users, and track their anonymized browsing behavior including clicks and conversions. The research design we propose uses comparisons of revealed preferences under experimentally manipulated treatment and control conditions to assess the potential for consumer confusion and deception. A design based on revealed preference speaks to the “material” standard of regulators, helps assess confusion while avoiding directly questioning consumers, and may be useful in other settings. Implementing the design, we find that native advertising benefits advertisers and detect *no evidence* of deception under typically used formats of disclosure currently used in the paid-search marketplace. Further investigation shows that the incremental conversions due to advertising are not driven by users clicking on the native ads. Rather, the benefits from advertising are driven by users seeing the ads and later clicking on the advertiser’s “organic” listings. Thus, we find little support of native advertising tricking users into clicking and driving them to advertisers as typically feared; instead, users seem to view ads and deliberately evaluate the advertisers. Furthermore, mere exposure seems sufficient to produce most of the incremental effect of advertising.

**History:** Avi Goldfarb served as the senior editor and Anja Lambrecht served as associate editor for this article. This paper has been accepted for the *Marketing Science* Special Issue on Consumer Protection.

**Supplemental material:** Data are available at <https://doi.org/10.1287/mksc.2018.1125>.

**Keywords:** native advertising • disclosure • consumer deception • field experiments • restaurants • mobile • paid search • platforms • media

## 1. Introduction

Since its earliest days, the separation of content from advertising has been an important issue for media. Clear separation is seen as essential for the credibility of the media’s content, and for sustaining an ad-supported business (Carlson 2014, Coddington 2015). This separation is becoming blurred in modern digital settings, spurred by several forces besetting media markets. On the demand side of the market, worries over the effectiveness of traditional formats of online advertising, the increasing use of technology such as ad blockers by consumers to avoid ads, and the shifting of media consumption to small-screen mobile devices have pushed platforms to develop new schemes for delivering ads that capture consumer attention (Sonderman and Tran 2013, Mitchell 2015). On the supply side, the large supply of outlets on which to serve digital ads has increased

competition and reduced ad prices, forcing media companies to explore new avenues for monetization. Publishers have responded to these marketplace pressures by introducing “native ads”—advertising that matches the form, style, and layout of the media content into which it is integrated.<sup>1</sup> Reports in the trade press suggest that U.S. spending on such ads may grow to as high as \$21 billion in 2018, rising from under \$4.7 billion in 2013 (Hoetzel 2015). Leading digital platforms also report that native advertising is a large component of their mobile advertising efforts (e.g., Facebook reports that native ads comprise about 83% of all ads served on the Facebook Audience Network and predicts that at about \$53 billion in dollar volume, it will comprise over 60% of all global mobile display advertising by 2020; Cohen 2016).

As native advertising has proliferated, a vociferous policy debate about its effect on consumers has erupted.

On the one side of the debate are regulators, who are concerned that consumers are harmed when the commercial nature of content is not properly disclosed. On the other side is the advertising industry, which is worried that overly prescriptive and onerous regulatory guidelines will hamper the efficiency and growth of well functioning ad markets.

Regulators worry about consumer deception. Under Section 5 (15 USC §45) of the U.S. Federal Trade Commission (FTC) Act of 1914 (Miller 1983), “an action or practice is deceptive if it’s likely to mislead consumers who are acting reasonably under the circumstances, and if it would be material to their decision to buy or use the product.” The FTC judges a misrepresentation as *material* if it is likely to affect consumers’ choices or conduct regarding an advertised product or the advertising for the product, irrespective of the medium.<sup>2</sup> Therefore, under these criteria, native ads raise the possibility of deception when the nature of sponsorship disclosure is (a) insufficient to help reasonable consumers recognize it as paid for by a third party *and* (b) this lack of recognition affects the consumers’ decisions.<sup>3</sup>

On account of its dominant influence on digital advertising spending, concerns about disclosure in paid search—on which this paper focuses—have been particularly dominant. In letters sent to search engines in 2013, the FTC observed that

[c]onsumers ordinarily expect that natural search results are included and ranked based on relevance to a search inquiry, not based on payment from a third party. Including or ranking a search result, in whole or in part, based on payment is a form of advertising. To avoid the potential for deception, consumers should be able to easily distinguish a natural search result from the advertising that a search engine delivers. (Engle 2013)

Concerns over disclosure practices related to native advertising have since accumulated, eventually culminating in the release of enforceable guidelines in December 2015 (Federal Trade Commission 2015a). In them, the FTC stipulates any disclosure used must be “sufficiently prominent and unambiguous to change the apparent meaning of the claims and to leave an accurate impression.”

Against this backdrop, advertisers and publishers have embraced native advertising while voicing concern over increased regulatory oversight (Interactive Advertising Bureau 2015). The digital advertising industry bemoans the inability to enforce and monitor disclosure protocols in all scenarios, worries about the potential for some guidelines (e.g., about the types of text that are understandable by consumers) to clash with platform look and style, is skeptical about government intervention in the “creative process,” and feels that current levels of self-regulation are sufficient. Reflecting

these considerations, Brad Weltman, vice president of public policy at the Interactive Advertising Bureau (2015), notes in its official response to the FTC’s guidelines that

[w]hile guidance serves great benefit to industry, it must also be technically feasible, creatively relevant, and not stifle innovation. To that end, we have reservations about some elements of the Commission’s Guidance. In particular, the section on “clarity of meaning” in native advertising disclosures is overly prescriptive, especially absent any compelling evidence to justify some terms over others.

Perhaps reflecting this discomfort, compliance with the guidelines in the digital marketing industry has been mixed (Swant 2016), even as the FTC increases enforcement (Coffee 2016).

What complicates the debate is the paucity of studies that explore the effect of native advertising on revealed consumer behavior in real field settings. Most studies in the area (reviewed later) have been survey based, querying consumers about whether they were able to distinguish between the content and paid ads they saw in the past, and assessing their stated attitudes toward native advertising. The survey evidence to date has been contradictory, with some suggesting consumer confusion about sponsorship status and others the opposite. A lacuna has been that the studies do not speak directly to the material impact of native advertising, because they lack data on actual consumer actions after been exposed to the native ads. Furthermore, some of the results are subject to typical concerns with inference based on stated preference data (like imperfect recall, framing effects, and small sample sizes). Nonetheless, regulation has proceeded on the basis of the existing evidence and on the basis of the broad powers endowed with the FTC to *presume* materiality in a variety of settings.<sup>4</sup> The commission’s stand has been that it “will always consider relevant and competent evidence offered to rebut presumptions of materiality” (Federal Trade Commission 1983), so more emerging evidence based on field-based data continues to be useful to the debate.

The contribution of this paper to this debate is the development of a field-experiment implemented on a restaurant-search platform to assess consumer response to native search advertising, and the development of a new experimental design to assess deception based on revealed preference. The field experiment is implemented on the mobile search app of Zomato, a large global restaurant-search platform. Users of the platform search for restaurants to visit and are shown listings of restaurants that match their search criteria. The app environment enables randomizing users into conditions in which they are or are not exposed to advertised listings of differing formats. It also allows tracking their subsequent search and conversion activity

using anonymized click-stream data. This has a few advantages. The randomization avoids the usual confounds of user self-selection into exposure plaguing inference in search settings, wherein advertising is served in response to user's search interest (Goldfarb and Tucker 2011, Blake et al. 2015, Narayanan and Kalyanam 2015, Sahni 2015). The opportunity to run an experiment at a large scale helps obtain the sample sizes required to measure ad effects precisely, and to assess the mechanism behind the detected effects by exploring heterogeneity across advertisers and consumers of differing profiles. Finally, the availability of data on consumer behavior after ad exposure helps assess whether the exposure produces effects that materially affect the exposed consumers. The experiment involves 622 advertising restaurants and 265,975 individuals and is implemented over a period of about six weeks in 2014.

To understand the research design, note that deception, as defined by the regulators, occurs when the consumer's behavior differs from what it would have been if the consumer were best informed. Thus, assessment of deception requires comparing behavior under the typical ad format relative to a benchmark in which the ad is presented in a clear and conspicuous manner. The idea in this study is to present a research design that satisfies this criterion, by basing assessment on (1) observed behavior (2) that is linked to clearly posed "counterfactual" situations.

Based on this rationale, the research design randomizes consumers into two extreme conditions: one where advertisements are shown without any indication that they are sponsored by a paying advertiser (the No Disclosure condition), and the other extreme, where ads are shown with a clear and prominent disclosure that they are sponsored by a paying advertiser (the Prominent Disclosure condition). If observed behavior under typical disclosure looks closer to that in the No Disclosure condition, we infer that consumers are indeed deceived by typical disclosure protocols; if, on the other hand, observed behavior under typical disclosure looks closer to that in the Prominent Disclosure condition, we infer there is no evidence in the data for such deception. In this setup, we assume the Prominent Disclosure condition represents behavior in a counterfactual world with no deception. Therefore, apart from making the ad more conspicuous, it is important the prominent disclosure does not change other aspects of the ad that may affect behavior. For example, it should not imply something about the appeal of the advertiser, cause unusual deliberation, or make the ad more or less annoying to the consumer.<sup>5</sup>

Our experiment executes this research design in our empirical context as follows. It randomizes users into one of many conditions. The Typical Disclosure condition discloses ads in the typical manner applied in

the industry, using an ad label that discloses the paid status of the listing. The No Disclosure condition removes the ad label and displays ads as "organic" content. The Prominent Disclosure condition adds a bold yellow highlight to the typical disclosure, making the ads stand out in the users' search results feed. This manner of making the ad prominent is consistent with FTC recommendations to businesses (Federal Trade Commission 2015b). Additionally, to assess the effect of the ad label on consumer behavior, we test two typical ad labels used widely in the industry. The first ad label is a yellow label with the word "Ad" disclosing the paid nature of the listing. This ad label is similar to the ones used by search engines like Google and Yelp. The second ad label is the text "Sponsored," which is similar to ad labels commonly used by platforms such as Facebook. Screenshots of implementations of these conditions are shown in Figures 4 and 5 along with a more detailed description.

Analyzing the data from the experiment, we first find that natively formatted advertising is beneficial to advertising restaurants on the platform: relative to a no-advertising control condition, visits to an advertising restaurant's information page hosted on the Zomato platform go up by 41%, on average, in conditions in which native advertising is served. Furthermore, *calls* to the restaurant—the measure of conversion activity—go up by 67%, on average. These results show that the widely held belief among advertisers about the efficacy of the native advertising format holds in our setting.

Our analysis finds consumer behavior under the Typical Disclosure condition to be closer to that under the Prominent Disclosure condition than that under the No Disclosure condition. Comparing observed consumer behavior in the typical versus prominent disclosure format, we find that the outcome measures—*calls* and *page visits* for advertised restaurants—are statistically *indistinguishable* between these conditions. Therefore, in terms of the outcomes we observe, we detect no evidence of material deception under typical disclosure, based on the criteria above. Furthermore, we find that outcomes are significantly *different* in the Typical Disclosure condition and the No Disclosure condition, indicating a significant scope of deception in our setting—that is, not disclosing ads can affect consumer behavior. Finally, we find that outcomes are also statistically *indistinguishable* between the "Ad" and the "Sponsored" text formats, providing no evidence of consumers interpreting these wordings of ad disclosure differently. We present confidence intervals for these comparisons and discuss the bounds in estimated changes.<sup>6</sup>

Delving deeper into the consumer behavior, we then assess whether there is any evidence for a commonly articulated viewpoint that unsuspecting consumers are "tricked" by native formatting into clicking on listings



they would otherwise not consider, which then materially affects their choices. Whether this concern has support in the data depends on how consumers respond to ads. If consumers are naive, unable to recognize the ads, and/or have low propensity to explore their options either because they are gullible or have a high search cost, ads disguised as nonadvertised content can induce them to click on listings and make them buy the advertised product. On the other hand, if consumers are able to recognize ads and make their decisions mindfully, ads can serve as additional options that improve consumers' decisions. Consumers would then know that they are clicking on ads, and accordingly interpret the information they get from the exposure.

We do not find support for the naive-consumer viewpoint in our data. Analyzing consumer search patterns, we find that the effect of advertising is not driven by users who click on advertised listings. To establish this, we develop a new test for the role of ad clicks in mediating responsiveness to advertising. The test compares the unconditional probability of calling without clicking on ads between treatment and control groups to assess the possibility of a direct effect of ad exposure on calling. Based on the test, we find ads appear to work directly through exposure in our setting—individuals view the ads, update their impression of the advertiser, and continue to search. Eventually, if they decide to pick the advertised option, consumers reach it through search or organic clicks. The effect of ads is almost entirely driven by consumers who call the restaurant after clicking on its organic listing on being exposed to the restaurant's advertisement. Consequently, if a product is advertised, conveying clearly to the consumers that it is advertised benefits the firm. A companion paper (Sahni and Nair 2016) documents that such effects can be rationalized by canonical signaling theories of advertising.

These findings have two main implications. First, users in our data are sophisticated consumers of advertising. Advertising does change their behavior. But this change is not likely because they are being tricked into clicking on native ads. Second, the incentives of the platform, advertisers, and regulators are aligned: consumers value the clear disclosure regulators demand, and it benefits advertisers and improves monetization for the platform. Furthermore, currently used native formats in paid search do not seem to deceive consumers.

To close the introduction, although these findings are consistent with beliefs in the industry (e.g., Sebastian 2014), we caution the reader that we do not mean to imply that it generalizes across all scenarios found in the digital economy, especially in contexts where ads are annoying, out of context, or inappropriately disclosed (unlike the formats we have considered in this paper). A distinguishing feature of our setting is that ads are relevant to the search, and hence useful for consumers to make better decisions about restaurant choice. They are

also not annoying because they are fluidly blended into the in-feed of the search listings. To thoughtful consumers in search mode, search ads serve as additional options that improve consumer decisions and reduce search costs. These are likely to be important considerations for the success of native advertising. In other settings, such as particular types of paid advertorials on news websites, consumers may be more annoyed by native ads because they interfere with their goals from visiting and consuming the medium. In such scenarios, it is possible they respond differently. Further research and inquiry on these aspects is warranted to generalize beyond our context. Having said that, we believe our results do warrant a rethinking of a default presumption that consumers are easily tricked and fooled by ads, and hence are highly susceptible to deception. At least in our setting, that does not appear to be the case, and this may be relevant in other settings as well.

The rest of this paper discusses the relevant literature, the empirical setting, the experimental design, the main results, and statistical power considerations. The last section concludes.

## 2. Literature Review

As noted in the introduction, most studies on native advertising have been survey based, either querying consumers *ex post* about their attitudes and/or confusion regarding the distinction between paid and unpaid content, or asking them to imagine their actions in new situations with counterfactually posed native formatting. Summarizing the research in this area, Bakshi (2015, p. 9) notes, "there are no published, empirical studies on the association between native advertising and consumer deception, but some in-progress research supports the intuitive notion that consumers often mistake native advertisements for independently created editorial content." In perhaps the most comprehensive survey research to date, Franklyn and Hyman (2013) ran online surveys in three waves in 2010 and 2012 and reported that only 42% of participants in their first survey understood the difference between sponsored and unsponsored search results (36% in second), and only 35% reported they found it easy to distinguish between paid and unpaid search results. Both these suggest the possibility of confusion. They also reported that a near majority stated they simply click on the first link for which they see the brand they are interested in, irrespective of whether the link is paid or unpaid, suggesting that consumers are not much concerned about whether a link is an ad, and that clicks may drive consumers into subsequent conversion. In a similar vein, Hoofnagle and Meleshinsky (2015) showed internet users a labeled advertorial embedded in a blog and queried them open-ended and closed-ended questions about the articles' content. Twenty-seven percent of the users thought that the advertorial was written by a reporter or an editor

even after displaying a label stating “sponsored content,” again suggesting the possibility of consumer confusion.

Other studies exposed users to native advertising in lab settings. Using responses from university participants to a lab moderator who showed them the listings from six web searches, Jansen et al. (2007) report that online searchers prefer to click links that they believe are algorithmic results rather than advertisements. In another lab experiment, Tutaj and van Reijmersdal (2012) report that participants find sponsored content more informative, more amusing, and less irritating than traditional banner ads, and that ad skepticism seems to be strongly related to perceived advertising value. Wojdyski and Evans (2015) recruited adult U.S. residents via Amazon Mechanical Turk and exposed them to 12 versions of a news story web page with native formatting. Querying participants after they finished reading both stories, they report that disclosures using the words “advertising” or “sponsored” increased advertising recognition compared with other conditions, and ad recognition generally led to more negative evaluations.

Although suggestive, the lab studies may not capture adequately the actual nature of deliberate consumer response to information in real-world settings, where the use of the information leads to consequential choices. Similarly, reflecting on the survey methodologies employed, Franklyn and Hyman (2013, p. 528) write, a

limitation is that the surveys asked participants what they had done previously, or would do in response to a specified situation. Asking people to remember or predict their own behavior is quite different than observing their actual behavior. Finally, responses to particular questions may be affected by survey respondents’ interpretation of the goals of the survey.... Additional work will be required to address these limitations, to the extent they are remediable.

Furthermore, past empirical studies have shown that ads affect a very small fraction of the people exposed to them (e.g., Lewis and Reiley 2014, Sahni 2015). Therefore, with sample sizes of the order of hundreds of individuals, there is little chance of recording responses from individuals who may actually be affected by the ads and their formats.

We are aware of only three studies that report revealed consumer behavior in response to changes in digital advertising labeling or content of similar nature. Edelman and Gilchrist (2012) recruited online participants to do their internet search through a web browser window they created, in which the label “Sponsored links” is randomly replaced by the label “Paid Advertisement” or “Ads.” Users assigned to the “Paid Advertisement” label showed about a 25% reduction in ad clicks relative to those who saw the other labels, and more correctly reported *ex post* that they clicked on fewer advertisements. Their results suggest that the

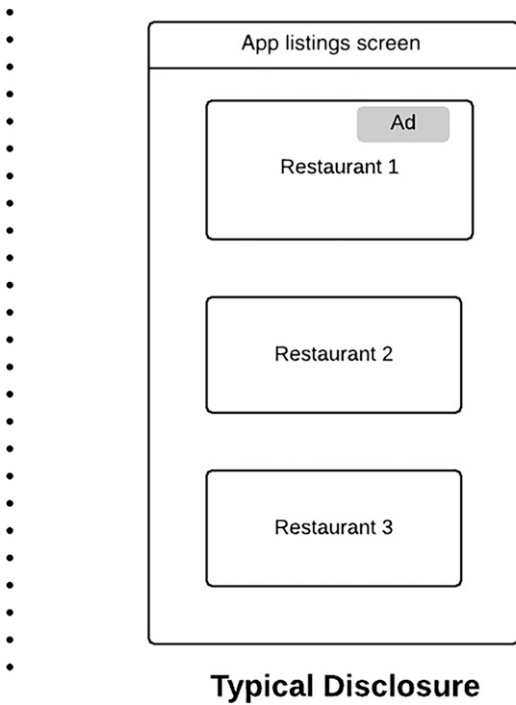
“Paid Advertisement” label may be more salient to searching users. In an article that examines the efficacy of “social advertising” by a nonprofit firm, Tucker (2012) mentions that click-through rates on the firm’s Facebook posts increased when the words “Please read this ad” were added to the campaign post. This is consistent with the users responding positively when explicitly told the sponsored post is an ad, though Tucker (2012, p. 22) cautions that “the sample size here is very small, making more definitive pronouncements unwise.” In a recent paper, Aribarg and Schwartz (2017) documented that increasing the prominence of an ad inserted in an email newsletter decreased clicks on it. Additionally, they analyzed survey and eye-tracking outcomes to compare consumer responses to banner versus native ads in controlled lab experiments, and found that consumers pay more attention to native ads. These studies explore only the effects on ad clicks and do not assess the effects of disclosure on visitation or conversion. Finally, both this paper and Sahni and Nair (2016) use subconditions of the broader experiment to assess two distinct issues related to advertising. Sahni and Nair (2016) focused on using the experimental data (as well as other supply-side information) to test signaling theories of advertising, whereas this paper’s focus is on developing a design to test whether sponsorship disclosure materially deceives consumers. The key effects in this paper are pinned down by the contrast between typical and prominent disclosure (i.e., based explicitly on differences in the format of disclosure), which is not used in Sahni and Nair (2016).

This paper also relates to a broader literature on deceptive claims in advertising content (not specifically on the nature of disclosure of the sponsorship status of ads). There is a large effort at the intersection of marketing and policy research that has focused on developing frameworks to evaluate deceptive claims in ads (e.g., Russo et al. 1981, Burke et al. 1988). Several researchers have pointed out the issues involved in the methodologies used to judge deception; for a general discussion, please see, for instance, Beales et al. (1981) and Pappalardo (1997). The regulator’s concerns about the reliability of survey-based research (e.g., Owen and Plyler 1991, Craswell 1997) and specific difficulties in establishing materiality (e.g., Richards and Preston 1992) have also received significant attention in this area. Apart from deception in advertising, researchers have also studied the existence of deceptive claims made by sellers and their impact on buyer decisions (see, e.g., Jin and Kato 2006, Rao and Wang 2017).

### 3. Logic of the Experimental Design

Though the ideas hold more generally, the discussion below is tailored to the setting of in-app search reflecting the empirical application of this paper. By an “in-app search ad,” we mean a paid-search advertisement

**Figure 1.** Stylized Example of In-Feed Search Ad



*Notes.* The figure depicts a search ad delivered within the feed of a mobile app. The format by which the sponsorship status of the ad is disclosed is referred to as typical disclosure. The researcher’s task is to assess whether this disclosure materially deceives consumers.

delivered into the content feed of a third-party app installed on a consumer’s mobile device (prominent examples of which include Yelp, Facebook, NYT Now from the *New York Times*, and apps of many media companies). To fix ideas, consider the stylized in-app search ad

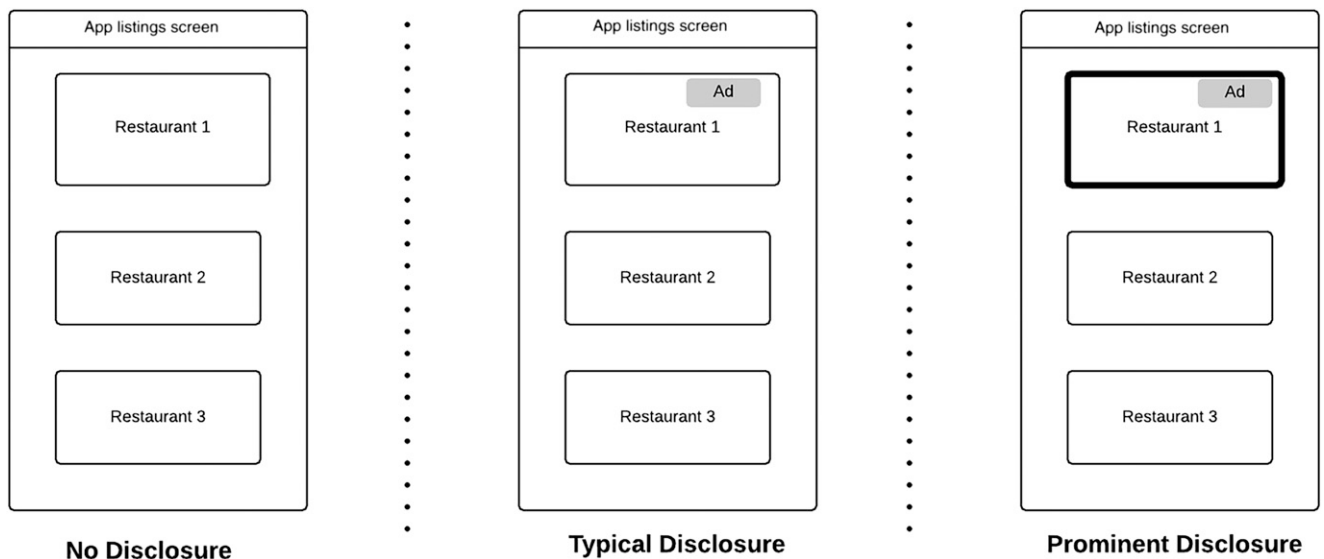
shown in Figure 1. The format by which the sponsorship status of the ad is disclosed is referred to as typical disclosure. The researcher’s task is to assess whether this disclosure materially deceives consumers. The idea of this paper is to randomize consumers into this Typical Disclosure condition and into two additional conditions, shown in Figure 2, No Disclosure and Prominent Disclosure. The assessment of material deception is then based on comparing the behavior of consumers in the three experimental conditions.

The experimental condition on the left, No Disclosure, shows the same advertised listing as in the Typical Disclosure regime in the feed of the app, but with no indication that it is paid for. The experimental condition on the right, Prominent Disclosure, shows the same advertised listing as in the Typical Disclosure regime in the feed of the app, but with a prominent indication that it is paid for. (The precise definition of what is prominent is context specific—more on this below.) The two experimental conditions are meant to mimic two counterfactually extreme worlds—one in which consumers have no awareness that the listing is paid for and another in which the advertised nature of the listing is clearly disclosed to the consumers. Everything else is held constant between the conditions.

### 3.1. Assessing Deception from the Design

Figure 3 shows how comparison of consumer behavior with the two extreme conditions can help assess material deception in the Typical Disclosure regime. The assessment flows from the way in which the materiality of a deceptive representation is defined under the law. Recall the law states that an advertisement is materially

**Figure 2.** Stylized Depiction of In-Feed Search Ad Under Typical and Two Extreme Experimental Disclosure Conditions



*Notes.* The figure depicts search ads delivered within the feed of a mobile app. The format by which the sponsorship status of the ad is disclosed is referred to as typical disclosure. To assess whether this disclosure materially deceives consumers, the research design proposes randomizing consumers into the Typical and No/Prominent Disclosure conditions.

deceptive if it misleads consumers *and* changes their actions with respect to the product or the ad.<sup>7</sup> To make the logic clear, call the outcome that the disclosure is materially deceptive event  $\mathcal{A}$ ; that it misleads consumers, event  $\mathcal{B}$ ; and that it changes their action, event  $\mathcal{C}$ , so that  $\mathcal{B} \cap \mathcal{C} \Rightarrow \mathcal{A}$ . A criticism of the survey based approaches can be articulated as asserting  $\mathcal{B} \Rightarrow \mathcal{A}$  without verifying whether event  $\mathcal{C}$  occurs.

What would lead a researcher to infer *no* material deception? If actions are found *not* to change when consumers are made fully aware that the listing is paid for, the typical disclosure cannot be materially deceptive, that is,  $\neg \mathcal{C} \Rightarrow \neg \mathcal{A}$ .<sup>8</sup> This is shown in the left column of Figure 3, corresponding to the state of the world where outcomes under the Typical Disclosure and Prominent Disclosure regimes are similar. Within this column, comparing the Typical Disclosure to the No Disclosure regime then helps assess the *scope* of sponsorship disclosure to deceive in that market setting. If actions are similar in all three regimes (top panel of the left column), then the sponsorship disclosure has little scope to meaningfully deceive consumers. Intuitively, consumers are not changing their actions because of disclosure, which suggests that disclosure does not affect their behavior regarding the advertised

product. Hence, issues related to disclosure have little scope to induce sizable deception. On the other hand, if actions under the No Disclosure and Typical Disclosure regimes look significantly different (bottom panel of the left column), then the sponsorship disclosure has a high scope to deceive consumers. Intuitively, consumers are changing their actions significantly in worlds with and without disclosure, which suggests that disclosure is something that affects how they react to the advertisement or buy the advertised product. Hence, in that market setting, issues related to disclosure have high scope to induce sizable deception. This can inform the need for regulation—the need for active regulation is higher in the latter case.<sup>9</sup>

What would lead a researcher to infer material deception? If actions are found to change significantly when consumers are made fully aware that the listing is paid for (i.e.,  $\mathcal{C}$  is true) relative to the typical disclosure scenario, the typical disclosure materially deceives consumers. We are able to infer deception based on  $\mathcal{C}$  alone because if prominent disclosure—making consumers fully aware of the sponsorship status of the listing—changes actions, then consumers were misled without prominent disclosure (i.e.,  $\mathcal{C}$  is true implies  $\mathcal{B}$  is true). This situation is shown in the right column of

**Figure 3.** Inference of Occurrence of Material Deception from Comparison of Typical Disclosure Regime to Two Extreme Disclosure Regimes

		Observed consumer behavior in Typical Disclosure vs. Prominent Disclosure regimes	
		Similar	Different
Observed consumer behavior in Typical Disclosure vs. No Disclosure regimes	Similar	No material deception (+ low scope for nondisclosure to deceive)	Material deception
	Different	No material deception (+ high scope for nondisclosure to deceive)	Material deception

*Notes.* The figure depicts how we assess whether a disclosure regime materially deceives consumers using the experimental design. The format by which the sponsorship status of the ad is currently disclosed is referred to as typical disclosure. To assess whether this disclosure materially deceives consumers, the research design proposes randomizing consumers into the Typical and No/Prominent Disclosure conditions and comparing their behavior.



Figure 3, corresponding to the state of the world where consumer behavior under the Typical Disclosure and Prominent Disclosure regimes is different. The comparison of the Typical Disclosure regime to the No Disclosure regime has less informational content in this case. High scope of deception is already established when actions under the Typical Disclosure and Prominent Disclosure regimes are different.

### 3.2. Characteristics of Prominent Disclosure

The above discussion shows that the Prominent Disclosure regime serves as an important benchmark to detect deception. We do not specify what constitutes prominent disclosure across contexts; its exact form depends on the setting. However, it should satisfy two conditions for the design's rationale to hold:

- *Conspicuous visibility.* Ideally, prominent disclosure should be clear enough that any reasonable consumer is able to spot the ad in this format. If less than total visibility is achieved in prominent disclosure, material deception may exist even when consumer behavior does not change between the Typical Disclosure and Prominent Disclosure regimes—it is possible that some individuals who did not spot the ad would have changed their behavior had they spotted it. A regulator worried about deception under typical disclosure would want to make sure that conspicuous visibility is achieved by the prominent disclosure used in the test.

- *Other aspects held constant.* It is also important that increased prominence is achieved without causing annoyance, changing the meaning of the ad, or anything else that its format may imply. For example, adding shiny and unusual symbols (such as bright stars) next to the ad may make it prominent, but may also puzzle a reasonable consumer, or cause unusual deliberation, causing unintended Hawthorne effects. In the presence of such effects a change in behavior may not signify deception. It is also possible that unintended Hawthorne effects neutralize the change in behavior caused by prominence; therefore, material deception may exist even when no change in behavior is detected between the Typical and Prominent Disclosure regimes.

All parties to the debate would need to agree on the Prominent Disclosure regime a priori, to move forward with the design.

### 3.3. Discussion and Other Considerations

To summarize, the suggested experimental design uses “revealed preference” arguments to assess material deception as defined under the law. Making the case that there is no detectable evidence of deception requires that actions under prominent disclosure look similar to those under the typical disclosure format; establishing that deception has occurred requires that actions under prominent disclosure look different from

those under the typical disclosure format. Establishing both requires arguing that the prominent disclosure manipulation manipulates the awareness of the consumer only about the content's sponsorship status. Finally, although only the Typical Disclosure and Prominent Disclosure conditions are required to assess deception, the ability to randomize consumers into a No Disclosure condition, when possible, is useful to assess the need for active regulation of sponsorship disclosure in that market setting.

Two other aspects are worth noting. A challenge in implementing the design is that only few individuals generate outcomes such as a website visit or sale in the population targeted by typical ad campaigns. Those who *change* their behavior because of advertising disclosure format may be even smaller. Therefore, the sample size of targeted individuals required for the study would be large, and hence, more difficult to obtain relative to a typical survey. Additionally, the regulator would have to provide guidance on the metrics and threshold effect sizes (the proportion of population exposed to ads that changes behavior) that would signify substantial deception.<sup>10</sup> The current legal precedence provides benchmark effects on survey outcomes but not behavioral outcomes.<sup>11</sup>

## 4. Empirical Setting

Our study is conducted in collaboration with Zomato, which is one of the largest restaurant-search portals worldwide. Zomato operates in 22 countries and provides a platform for consumers to search and browse through information about thousands of restaurants in many local markets. In 2014, when the data for this study were collected, 30 million unique users used Zomato monthly to search for restaurants. Compared with the overall internet population, Zomato users are more likely to be female, between the ages of 25 and 34, and educated beyond college, and less likely to have children. The users in our data are located in large cities in South Asia and the Middle East.<sup>12</sup>

The Zomato platform provides its users the ability to search for restaurants on their internet website or their mobile app available on Android or Apple iOS smartphones and tablets. On the platform, a user can filter down to their specific geographic location and/or conduct text-based search while applying various criteria they prefer to the search, including the cuisines they might be looking for. In addition to the cuisine, the consumer can choose to specify a search category (home delivery/dine out/nightlife), which tells Zomato more about the consumer's intention behind the search.

The data come from users browsing Zomato's Android app, so we describe a user's search experience on the app in more detail. Searching for a restaurant on the app takes the user to a page that displays search results that satisfy the user's criteria. The left panel in Figure 4

shows an example. The search results are sorted by the search engine's proprietary measure of popularity of a restaurant, unless the user specifies alternative sorting criteria. Clicking on one of the listings in the search result takes the user to a restaurant's page, which provides more information about the restaurant. This page allows the user to view the detailed menu of the restaurant, its photos, and its location on a map. It also shows the restaurant's average rating and allows the user to browse through reviews given by other users on the platform.

## 5. Implementation of Experiment

At the time of this study, thousands of restaurants engaged in advertising on Zomato's website, displaying banners for their restaurants. There was, however, no advertising on the mobile app prior to our experiment. We collaborated with the platform to start experimental mobile advertising, testing various formats of advertising as part of the firm's prelaunch "A/B" testing. The experiment was launched in the form of a new update of Zomato's Android app hosted on the Google Play app store. Any user who upgraded to this version of the app could potentially be a part of the experiment.

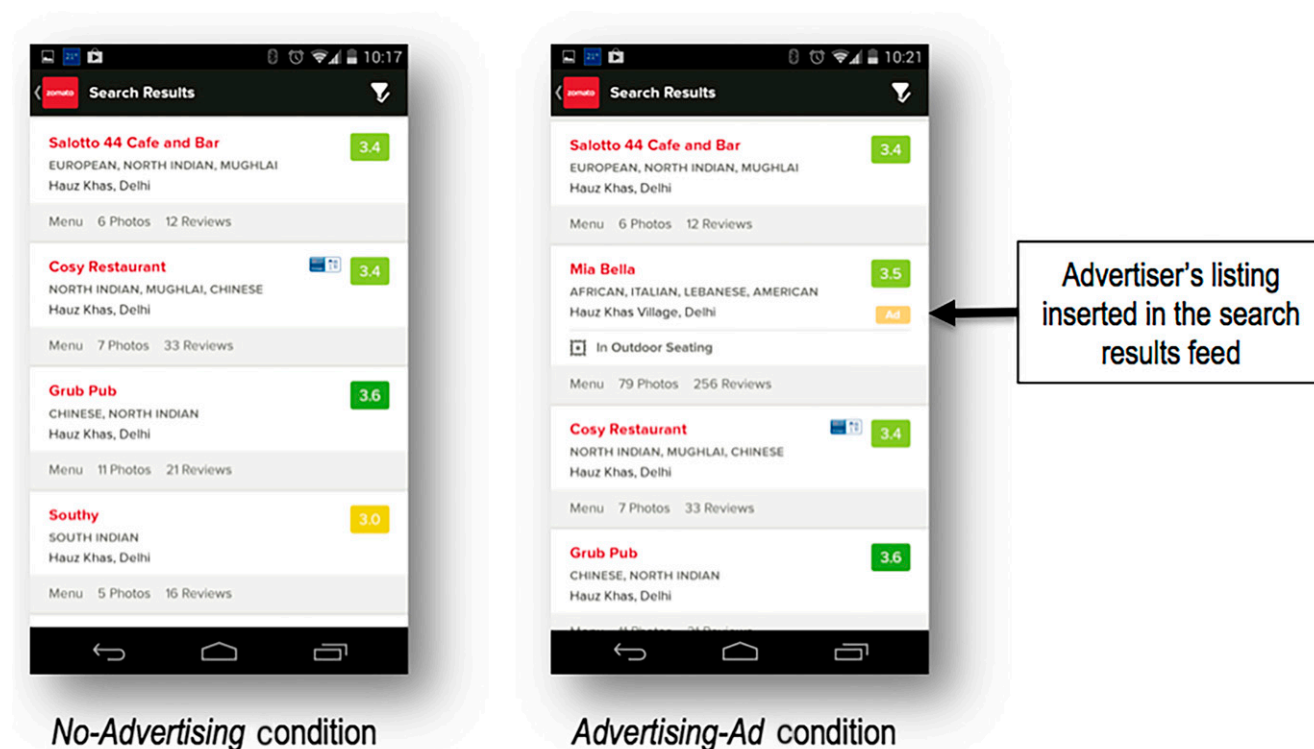
A user in the experiment, identified by a unique login ID, was allocated to one of two experimental conditions. The experimental conditions are as follows:

- No-Advertising (the control group). A user in this condition does not get exposed to any advertising. The user's experience remains unchanged compared with before the experiment.

- Advertising (placing experimental ads). A user in this condition gets the same set of restaurant options in the search results as a user in the No-Advertising condition, and in the same sequence. However, unlike in the No-Advertising condition, the experimental-advertiser restaurants' ads are also placed in the search results. We discuss the selection of the experimental restaurants in detail later in this section.

Figure 4 shows an example of how the search results on a mobile device look like in the No-Advertising and Advertising conditions. The advertised restaurant's listing (Mia Bella, in this example) does not appear in the No-Advertising condition. In the Advertising condition, the advertiser's listing is inserted into the search results feed and appears in the second position in the snapshot. The Advertising condition is comprised of additional subconditions that correspond to the various types of disclosure regimes discussed in the experimental design above. Specifically, in one set of manipulations, we vary the prominence by which native ads are distinguished from organic (nonsponsored) search results. At one extreme, we create a condition in which there is no distinction between ads and organic

**Figure 4.** (Color online) Example of the Search Results Page in the No-Advertising and Advertising Conditions



*Notes.* The advertised restaurant's listing gets inserted in the search results feed if the user is in the Advertising condition. The listings for the other restaurants (organic results) remain the same and are presented in the same order.

content. At the other extreme, ads appear with an added highlight that make them easily distinguishable from nonsponsored content and, arguably, hard to miss. In a second manipulation, we experimentally vary within the Advertising condition, the text used to label the ads. In the baseline condition, we label an advertisement using the word “Ad,” which is commonly used by search platforms such as Google and Yelp. In another condition, we change the label to “Sponsored” (also commonly used by several platforms like Facebook) to understand users’ sensitivity to the manner in which ads are indicated. In total, we implement the following five subconditions within the Advertising condition (examples are shown in Figure 5):

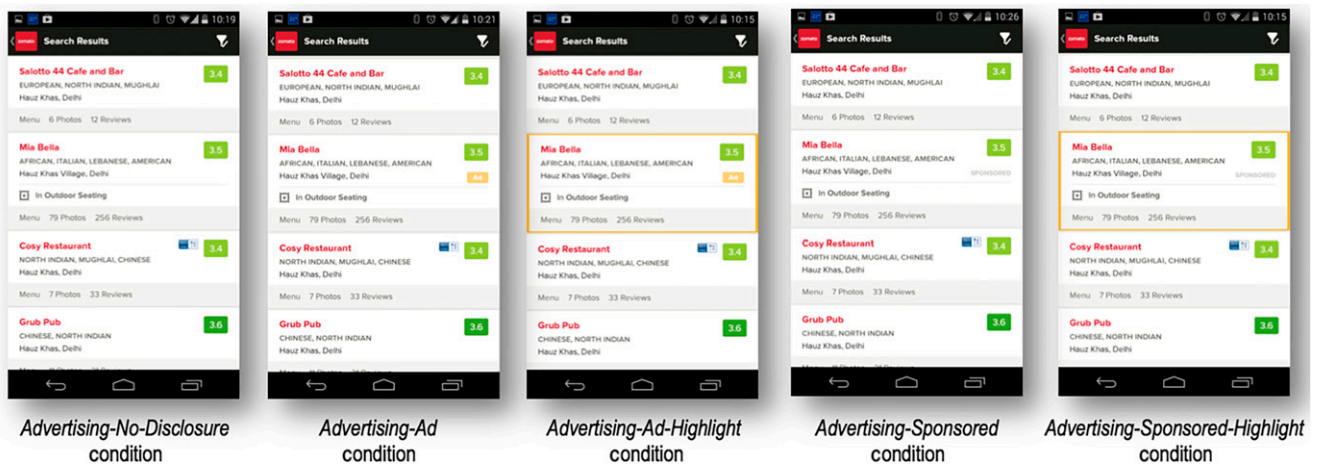
- **Advertising-No-Disclosure.** In this condition, the advertisers appear in the search results, but there is no label or any other sign differentiating the advertisers’ listing from the rest of the search results.
- **Advertising-Ad.** In this condition, the advertisers’ listings appear in the search results with a label reading “Ad.” This label distinguishes the advertised restaurants from the nonadvertised search results in a format that states explicitly that the listing is an ad.
- **Advertising-Ad-Highlight.** In this condition, the ads appear with the label as in the Advertising-Ad condition, along with an added highlight along the boundary of the box containing the advertiser’s listing. This highlight increases the visibility of the ads significantly and makes them more distinguishable from the nonadvertised restaurants.
- **Advertising-Sponsored.** This condition is very similar to Advertising-Ad, except that users in this condition see a label that reads the word “Sponsored” instead of “Ad.”

- **Advertising-Sponsored-Highlight.** This condition is similar to Advertising-Ad-Highlight, except for the text of the label, which reads “Sponsored” instead of “Ad.”

The Advertising-Ad and Advertising-Sponsored subconditions correspond to standard formats of disclosure currently used in the industry. Vis-à-vis the experimental design discussed previously, these represent Typical Disclosure regimes. The Advertising-No-Disclosure subcondition represents the No Disclosure regime. Finally, the Advertising-Ad-Highlight and Advertising-Sponsored-Highlight subconditions represent our execution of Prominent Disclosure regimes. In our discussions with the firm before the experiment, managers at the firm agreed that the highlighted format increased prominence without inducing substantial annoyance.<sup>13</sup> Importantly, this manner of highlighting the ad is not atypical, and is also recommended by the regulator in its guidance to businesses (Federal Trade Commission 2015b).<sup>14</sup> We also implemented a survey to assess the noticeability of this condition, which is described in Appendix A.<sup>15</sup> The survey shows that most respondents indicate that this highlighting increases the visibility of the ad. These indications are consistent with the requirements of the Prominent Disclosure regime discussed in Section 3.2.

Once a user is allocated to one of the experimental conditions, her allocation remains fixed throughout the time period of the experiment (there is no rerandomization over time). Furthermore, advertising in our experiment occurs only on the page showing the search results. There is no advertising on restaurant pages hosted on the mobile platform. Also, there is no change in the restaurant pages across the experimental conditions, regardless of whether the restaurant advertised.<sup>16</sup>

**Figure 5.** (Color online) Example of the Search Results Page in the Five Subconditions of the Advertising Condition



**Notes.** The advertised restaurant’s listing is placed in the search results in the same position across the five conditions. The other restaurants, called organic results, are also the same and presented in the same order. The format of advertising varies. In the Advertising-No-Disclosure condition, ads appear looking exactly as the organic listings. In the other five conditions, there is some distinction between ads and organic listings. In Advertising-Ad-Highlight and Advertising-Sponsored-Highlight, the advertised restaurants’ listings are outlined to make them more conspicuous.



## 5.1. Experimental Advertisers

We pick as advertisers for the experiment those restaurants that intend to advertise to the user on Zomato. Specifically, for any search, the experiment displays ads for restaurants that *would* have been advertised if the search were conducted on the Zomato website, instead of the mobile app. To understand how this works, recall that the Zomato website had a functioning advertising market during our experimental time frame. Consider a user of Zomato's Android mobile app who conducts a search during the experiment. Once this search reaches the Zomato mobile ad server, our experiment retrieves ads that would be shown for the same search on Zomato's website, and shows the listing for those restaurants in that user's search results. If the user is in the No-Advertising condition, or if there is no restaurant that wishes to advertise for the search criteria applied by the user, no ads are displayed.<sup>17</sup>

## 5.2. Experimental Ads Appear Only on Specific Searches

Ads in our experiment appear in search results *only* when the search criterion is based on a (1) location and/or (2) search category such as "home delivery" or "dine out" that captures the context of the search. If a consumer includes a cuisine in the search filter (or any other factor apart from location and search category), she does not see any ads in the search results, regardless of the user's experimental condition, by design. This is also true if a user sorts the search results by a criterion different from the platform's default. For example, if a user sorts by restaurant ratings, no advertising appears on the search results page, regardless of the user's experimental condition. This step reflects the fact that advertising on [Zomato.com](https://www.zomato.com) is sold based on location and search category. Thus, experimental ads are effectively being served only for the searches that advertisers contracted on. Incorporating this aspect into the design has two advantages. First, it makes it easier to implement the experiment (because we did not have to go market by market to pick advertisers). Second, we ensure that the restaurants that appear as ads are consistent with the user's search query and not unrelated to it. For example, we avoid a situation where a user searches for Chinese restaurants and sees a series of organic restaurants that serve Chinese cuisine but ads that may not satisfy this criterion.<sup>18</sup> Furthermore, ads served correspond to functioning, in-market restaurants. This enables us to avoid potentially problematic Hawthorne effects due to experimentation.

## 5.3. Advertising in This Setting Is Relevant

It is important to note that on account of (a) the personalization implied by search and (b) the features of the experiment design, ads in the experiment are relevant to the users' context. Unlike TV advertising, or

advertorials and other forms of native advertising, users who are exposed to ads for restaurants are those who are also searching for restaurants. Furthermore, ads are targeted, both by location and/or by intent. Such targeting ensures that advertised options are aligned with the goals of the searching user. Taken together, these aspects imply that ads in our setting are less likely to be annoying than in other contexts. This perspective is useful to keep in mind to interpret the results that follow.

## 6. Empirical Analysis

### 6.1. Setup of the Analysis

We now describe the analysis strategy and the process by which we construct our main analysis data set. Recall that not all searches are eligible for advertising by experiment design. Also, some searches may not be advertised to because no restaurant contracted with Zomato to advertise in response to the search's criteria. The implication of these two features is that not every user who uses the Zomato Android mobile app during our experimental time period is eligible for experimental advertising. To obtain a sample of users who would be affected by experimental ads, we start by focusing on users who engaged in at least one search that is eligible for advertising.

Once we obtain users in this manner, we then restrict attention to the effect of the first ad exposure to a user on her behavior in her first session after downloading the app update.<sup>19</sup> The rationale behind this step is as follows. If the experimental treatment works—that is, the users in different experimental conditions respond to the ads differently—users' search/browsing behavior after the first ad exposure will differ. This difference in search behavior can lead to differences in the exposure to subsequent ads. In other words, all experimental ads after the first one are endogenously exposed. As an extreme example, if ads work in one condition (A) and not in the other one (B), users in condition A who prefer the advertiser may stop browsing after the first ad exposure. Then, users in A who are exposed to the second ad will comprise a selected subpopulation. Hence, the individuals across conditions A and B who are exposed to the second ad are not comparable. Put differently, the experiment design provides equivalent comparison sets for the first ad exposure, but not necessarily for the subsequent ones.

To construct a data set that reflects this, we take the following steps. For every individual user in every condition, we pick the first search that is eligible for an ad exposure and focus on the first ad the user sees, or could have seen, if she is in the control condition. Each such search may have one or multiple first advertisers associated with it. The latter occurs if Zomato has contracted with multiple advertisers for the first ad. We collect the user's behavior with respect to all



advertisers relevant to the first ad and stack across all such users in all conditions.<sup>20</sup> In the final data set so constructed, each observation is a user  $\times$  an advertiser. We have 265,975 such users in our data. In total, there are 622 advertisers whose ads are displayed. Table 1 shows the distribution of the number of users across each experimental condition.

Randomization checks reported in Appendix B show that randomization is induced properly. In other tests of balance, we verify that users in all conditions have the same probability of having a search that would be eligible for an ad exposure (we fail to reject the null hypothesis that users in all conditions have the same probability of conducting a search to which the experiment serves ads,  $p = 0.19$ ). Furthermore, we verify that the characteristics of the first search with ad exposure, such as its date and the advertiser associated with it, are the same in expectation across the six experimental conditions.

## 6.2. Outcome Measures

The two main outcome variables we focus on are (1) whether the user visits the advertiser's page during her session (we refer to this as a *page visit*) and (2) whether the user calls the restaurant during her session by clicking on the call button on the Zomato app (we refer to this as a *call*). Both (1) and (2) are behavioral measures important to the search platform and the advertising restaurants.<sup>21</sup> A *page visit* represents a user considering the advertiser for purchase and gathering information about it. A *call* represents a consumer's interest in purchase through the mobile app. A *call* can occur only after the user visits the advertiser's page.<sup>22</sup> In addition to these measures, we also investigate the channel through which the *page visit* occurs to understand better the user's response to native ads. In particular, we investigate whether the user reaches an advertiser's page by clicking on the advertiser's experimental ad or through her own search effort (i.e., by scrolling down the in-feed listings and clicking on an organic listing). Figure 6 shows a schematic of the user actions and dependent variables we analyze.

## 6.3. First Cut: Means Across Experimental Conditions

We start by presenting average statistics across the six experimental conditions to describe stylized features of user behavior across conditions. Subsequently, we present the hypothesis tests and analysis related to native advertising.

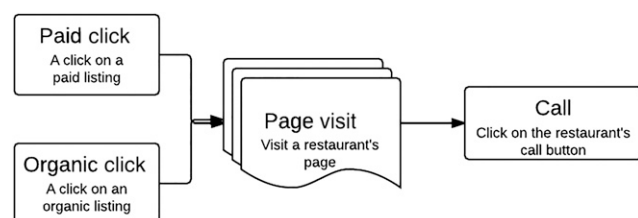
Table 2 presents the means for the outcome measures to compare consumer behavior across experimental conditions. Column (1) shows the probability of the user clicking on an ad during the session. Because no ads are shown in the No-Advertising condition, the probability of an ad click in that condition is zero. Across the other five conditions, the average probability of a user clicking on the experimental ad is 0.50%. We also see that the probability of clicking on an ad does not vary significantly across the conditions that show experimental ads. Column (2) now shows the chance of a consumer visiting the advertiser's page during her session. Because a user can visit the advertiser's page through organic listings, the visit probability in the No-Advertising condition is nonzero and equals 0.83%. For the other conditions, this probability is close to 1.1%. The No-Advertising condition stands out: the variation among the five conditions that show the experimental ad is small relative to their difference from the No-Advertising condition.

We can also examine this table for how well clicks on advertised listings capture incremental user response to advertising. A click on an advertised listing leads to a *page visit*, but this *page visit* need not be incremental—the user may have visited the advertising restaurant's page via a click on the restaurant's organic listing even in the absence of the ad. To assess this, we subtract from the probability of a *page visit* (reported in column (2)) the chance of a *page visit* in the No-Advertising condition and report this difference in column (3). Column (3) thus represents an estimate of the incremental probability of visitation to a restaurant's page from advertising. If all clicks on ads were leading to incremental *page visits*, we would expect the ad-click probabilities reported in column (1) to match the incremental visitation probabilities reported in column (3). On the other hand, if users who

**Table 1.** Distribution of Users Across the Six Experimental Conditions

	Condition	Description	Number of users
1	No-Advertising	Ads do not appear in the search results	44,233
2	Advertising-No-Disclosure	Ads appear but are indistinguishable from organic search results	44,637
3	Advertising-Ad	Ads appear in a typical format: the ad label reads "Ad"	44,333
4	Advertising-Sponsored	Ads appear in a different format: the ad label reads "Sponsored"	44,482
5	Advertising-Ad-Highlight	Ads appear in the typical format, with an additional highlight around the advertised listing to make the ad prominent	44,454
6	Advertising-Sponsored-Highlight	Ads appear in the alternative format, with an additional highlight around the advertised listing to make the ad prominent	43,836
Total			265,975

**Figure 6.** Sequence of User Actions on the App



*Notes.* The figure shows the set of actions a user takes that we report on in our analysis. In a condition with advertising, a user can click on either a paid or an organic listing for a restaurant. Clicking on either listing takes the user to the restaurant's page. The user can choose to call the restaurant by clicking on the "call" button on the restaurant page.

click on the ad would have reached the advertiser's page even in the absence of the ad, we expect the numbers in column (3) to be lower than the numbers in column (1). The table shows the latter is the case: the probabilities in column (3) are consistently lower than the corresponding ones in column (1). This difference suggests that many users who click on the ad would have reached the advertiser's page anyway. Column (4) calculates this difference and shows that about 30% of the users who click on the ads would have reached the advertiser's page using organic listings. These statistics suggest that ad clicks may not be a good measure to judge the incremental effect of advertising.

Column (5) of Table 2 shows the chance of a user calling the advertised restaurant during her session. The numbers in this column are significantly lower than those in column (2), suggesting that a large proportion of users who visit an advertiser's page end up not calling it. Furthermore, it shows that the calling probability is lower in the No-Advertising and Advertising-No-Disclosure conditions relative to the others with disclosure.

Finally, the premise to the debate on native advertising is that advertising in this format works, and benefits advertisers by driving more traffic and as well as more conversion of traffic to sales.<sup>23</sup> We can assess this in our setting: column (2) in Table 2 shows that advertiser's *page visits* increase when its ad is shown, and column (3) shows the same for *calls*. An *F*-test comparing averages in the No-Advertising condition with the other conditions pooled together shows that the changes are statistically significant along both outcomes, *page visits* and *calls* ( $p$ -values < 0.01). In relative terms, showing the ad in one of the five formats increases likelihood of a *page visit* by 41% and likelihood of a *call* by 68%, on average.

#### 6.4. Do Consumers Fail to Notice Native Ads Under Typical Disclosure?

As noted previously, deception may occur if a consumer fails to notice that the product is advertised, perhaps

**Table 2.** Comparisons of Averages Across Experimental Conditions

No.	Experiment condition	Chance of a click on the ad (average) (1)	Chance of visiting the advertiser's page			Chance of calling the advertiser (average) (5)
			Average (2)	Difference in chance of visiting the advertiser's page relative to the No-Advertising condition (3)	[column (1) – column (3)] ÷ column (1) (4)	
1	No-Advertising		0.83% (0.04%)			0.030% (0.006%)
2	Advertising-No-Disclosure	0.47% (0.03%)	1.10% (0.04%)	0.27% (0.05%)	42.6% (10.3%)	0.031% (0.006%)
3	Advertising-Ad	0.51% (0.03%)	1.15% (0.04%)	0.32% (0.05%)	37.3% (9.0%)	0.068% (0.01%)
4	Advertising-Sponsored	0.47% (0.03%)	1.17% (0.04%)	0.34% (0.05%)	27.7% (10.4%)	0.052% (0.008%)
5	Advertising-Ad-Highlight	0.52% (0.03%)	1.14% (0.04%)	0.31% (0.05%)	40.4% (9.4%)	0.047% (0.008%)
6	Advertising-Sponsored-Highlight	0.57% (0.03%)	1.26% (0.04%)	0.42% (0.06%)	26.3% (8.6%)	0.058% (0.009%)

*Notes.* The table reports on comparisons of averages across the six experimental conditions. Standard errors are in parentheses. Column (1) shows the probability of clicking on an ad. Note that because no ads are shown in the No-Advertising condition, the probability of an ad click is zero in that condition. Column (2) shows the chance of a user visiting the advertiser's page during the session. Because a user can visit the advertiser's page through organic listings, the visit probability in the No-Advertising condition is nonzero. Column (3) subtracts from column (2), the chance of a visit in the No-Advertising condition, to show the change in the likelihood of a visit to the advertiser's page relative to the condition in which there are no ads. Numbers in column (3) will be lower than the numbers in column (1) if users who clicked on the ad would have reached the advertiser's page even in the absence of the ad. The table shows this is the case. Column (4) shows that about 30% of the users who click on the ads would have reached the advertiser's page using organic listings. These statistics suggest that ad clicks may not be a good measure to judge the incremental effect of advertising.

because she does not see the ad label, or if she misinterprets the ad label, perhaps because she does not understand that the label used by the platform signifies an ad. We test these two aspects sequentially.

To assess the possibility of deception due to noticeability, we compare consumer behavior in the Advertising-Typical-Disclosure condition with the Advertising-No-Disclosure and the Advertising-Highlight conditions. The Advertising-Typical-Disclosure condition pools the Advertising-Ad and Advertising-Sponsored conditions. The Advertising-Highlight condition pools the Advertising-Typical-Highlight and Advertising-Sponsored-Highlight conditions. We pool the data in these two conditions to increase the power of our test and do not discriminate between formats with different text in the label. We implement the test in a linear regression framework, in which we regress the *page-visit* and *call* outcome variables on an indicator of whether the user is in the Advertising-No-Disclosure condition or the Advertising-Highlight condition. The baseline is the Advertising-Typical-Disclosure condition. So, the coefficients on the Advertising-No-Disclosure and Advertising-Highlight dummies represent the change in the outcome variable from the Advertising-Typical-Disclosure condition.

The estimates are presented in columns (3) and (4) of Table 3. Looking at the regression with *page visit* as the dependent variable (column (3)), we see that the coefficients on the indicators are statistically indistinguishable from zero. The point estimates are also small—the difference in *page visits* in the condition with extra highlight is 3% of the likelihood of visiting the advertiser in the baseline condition. These comparisons suggest that the estimated likelihood of a user visiting the advertiser's page in the baseline condition—in which native advertising is presented in the typical format—is statistically indistinguishable from the same measure in the other two formats.

Looking at the estimates in column (4), we see that the likelihood of an individual calling the advertiser does vary with the format of advertising: the coefficient with respect to the indicator of Advertising-No-Disclosure is statistically significant and negative. This suggests that the likelihood of the individual calling a restaurant is lower when the ads are completely blended among the nonadvertised search results compared with the condition in which ads are labeled in the typical format.<sup>24</sup> On the other hand, we see the coefficient corresponding to the indicator of Advertising-Highlight is statistically indistinguishable from zero.

**6.4.1. Joint Hypothesis Tests.** A joint test using the seemingly unrelated regression (SUR) framework (Zellner 1962) is unable to reject the null hypothesis that user behavior in terms of *calls* and *page visits*

remains the same when a highlight is added to the typical format of advertising ( $p = 0.32$ ). On the other hand, we strongly reject the null hypothesis that user behavior in terms of *calls* and *visits* remains the same when ads are not disclosed using a similar joint test ( $p < 0.01$ ). We are also able to reject the null hypothesis that the change in behavior from adding the highlight to typical ad disclosure is the same as the change in behavior from removing ad disclosure from typical disclosure ( $p = 0.02$ ; jointly testing whether the two coefficients in column (3) of Table 3 are equal and the two coefficients in column (4) are equal). This suggests that consumer behavior under typical disclosure is closer to behavior under prominent disclosure, relative to behavior under no disclosure.

Overall, the above analysis compares user behavior under the typical advertising format with that under two extreme formats, one in which the ads are made more salient and a second in which the ads are completely blended into the content and indistinguishable from the nonadvertised content. It finds no significant change in user behavior when ads are made more prominent. Referencing the discussion in Section 3, this result implies that there is no detectable evidence of material deception in this setting. With respect to Figure 3, it appears this setting corresponds to the situation in the bottom-left panel—no detectable evidence of material deception under the typical disclosure formats, but with no disclosure having the potential to deceive consumers (justifying regulatory interest in the role of disclosure).

## 6.5. Statistical Power Considerations

What magnitude of changes in behavior can this analysis detect? Table 3 reports confidence intervals implied by the regression estimates in Table 3. We specifically focus on the coefficients corresponding to Advertising-Highlight in columns (1) and (2) because they convey the difference in behavior between the typical and prominent disclosure formats. We can interpret this coefficient as representing the change induced by adding the highlight to an ad, on the proportion of the population that visits the advertiser's page or calls it.

- **Absolute change in *page visits*.** The 95% confidence interval on the Advertising-Highlight coefficient corresponding to *page visits* (column (1)) is  $[-0.00042, 0.001144]$ . Thus, we can say with 95% confidence that the induced change in page-visit probability is between  $-0.00042$  and  $+0.001144$ . In absolute value, the confidence interval covers change magnitudes  $< 0.1144\%$ . Therefore, we can rule out with 95% confidence that highlighting the ad changes the page-visitation behavior of more than  $0.1144\%$  of the population. We cannot rule out an increase less than that; if highlighting increases page-visit behavior of less than  $0.1144\%$  of the population, we would not detect it in this test.

**Table 3.** Effect of Native Advertising and Its Various Formats

Independent variables	Five conditions, with and without disclosure		Four conditions varying disclosure text	
	Page visit (1)	Call (2)	Page visit (3)	Call (4)
Advertising-No-Disclosure	−0.00057 (0.00048) [−0.00151, 0.00037]; [−0.1302, 0.03197]	<b>−0.000287 (0.000091)</b> [−0.00047, −0.00011]; [−0.7773, −0.18145]		
Advertising-Highlight	0.00036 (0.00040) [−0.00042, 0.00114]; [−0.03655, 0.09862]	−0.000091 (0.000087) [−0.00026, 0.00008] [−0.4368, 0.1328]		
Advertising-Sponsored or Advertising-Sponsored-Highlight			0.00066 (0.00040) [−0.00012, 0.00144]; [−0.01078, 0.12556]	−0.000010 (0.000087) [−0.00018, 0.00016] [−0.32293, 0.28716]
Intercept	<b>0.0116 (0.00028)</b> [0.01105, 0.01214]; [0.95268, 1.04731]	<b>0.0005987 (0.0000638)</b> [0.00047, 0.00072]; [0.79113, 1.20887]	<b>0.0115 (0.00028)</b> [0.01095, 0.01204] [0.95227, 1.04772]	<b>0.000559 (0.0000617)</b> [0.00044, 0.00068]; [0.78366, 1.21633]
Number of clusters	221,742	221,742	177,105	177,105

*Notes.* The table reports on results from several regressions. In each regression, the dependent variable is a 0/1 indicator of a *page visit* or *call*. The standard errors are in parentheses next to the coefficient estimates and are clustered by user. Additionally, below each coefficient in square brackets, we report (1) 95% confidence interval of coefficient and (2) the same confidence interval scaled by the intercept of the regression to show the effect relative to the baseline. For example, the first confidence interval corresponding to Advertising-No-Disclosure in column (1) equals  $[-0.00057, -1.96 \times 0.00048, -0.00057 + 1.96 \times 0.00048] = [-0.00151, 0.00037]$ , and the second confidence interval equals  $[-\frac{0.00057 - 1.96 \times 0.00048}{0.0116}, -\frac{0.00057 + 1.96 \times 0.00048}{0.0116}] = [-0.1302, 0.03197]$ . The coefficient estimates in bold are statistically significant at the 5% significance level. In terms of fit, the independent variables explain little of the total variation in all specifications with  $R^2 < 10^{-4}$ . Columns (1) and (2) report regress the dependent variable on indicators of Advertising-No-Disclosure and Advertising-Highlight conditions. The Advertising-Typical-Disclosure condition pools the Advertising-Ad and Advertising-Sponsored conditions. The Advertising-Highlight condition pools the Advertising-Ad-Highlight and Advertising-Sponsored-Highlight conditions. The regression is conducted on all the data pooled excluding the No-Advertising condition; therefore, the intercept represents the average outcome in the Advertising-Typical-Disclosure condition, and the coefficients represent the difference between the condition indicated and the Advertising-Typical-Disclosure condition. Columns (3) and (4) report regress the dependent variable on indicators of the user being in the Advertising-Sponsored or the Advertising-Sponsored-Highlight condition. The regression is conducted on the data pooling the four conditions Advertising-Typical-Disclosure, Advertising-Typical-Highlight, Advertising-Sponsored, and Advertising-Sponsored-Highlight; therefore, the intercept represents the average outcome in the Advertising-Ad or Advertising-Ad-Highlight condition, and the coefficients represent the change in behavior when the text “Ad” is replaced by “Sponsored,” as shown in Figure (5).

- **Relative change in *page visits*:** To examine relative effects, it may be useful to scale the confidence interval by an estimate of *page visits* the advertiser gets in typical format, which is the point estimate of the intercept of the regression.<sup>25</sup> This gives a scaled 95% confidence interval  $[-0.0365, 0.0986]$ . Therefore, relative to the *page visits* with a typical format, adding the highlight changes behavior of less than 9.86%, following the same rationale. We cannot rule out a change of a relative magnitude smaller than 9.86%.

- **Absolute change in *calls*:** The regression estimates yield a 95% confidence interval  $[-0.000262, 0.000008]$ , suggesting that in absolute terms, we can say that less than 0.0262% of the population change their calling behavior.

- **Relative change in *calls*:** To see the magnitude in relative terms, we scale by the point estimate of the intercept of the regression, which gives a scaled 95% confidence interval  $[-0.43681, 0.13282]$ . Therefore, we can rule out with 95% confidence that calls changed by 43.681%, relative to the typical format when the highlight is added. We cannot rule out a change in *calls* of a smaller magnitude.

These thresholds outline the range of effects we can and cannot detect statistically.<sup>26</sup> Although we do not detect material deception, our analysis lacks statistical

power to detect changes in user behavior below these thresholds.

**6.5.1. Past Effects as Benchmarks.** We can also benchmark our estimates against past studies. Past studies on sponsorship deception in search ads have been survey based. One of the most comprehensive studies on search ad disclosures is by Franklyn and Hyman (2013), who survey individuals from a general population. They report that only 42% of participants in their first survey understood the difference between sponsored and unsponsored search results, implying 58% ( $= 100 - 42$ ) could potentially be deceived (the corresponding number in the second study is 36%, implying 64%  $= 100 - 36$  could potentially be deceived). In another question in the same survey, only 35% reported they found it easy to distinguish between paid and unpaid search results (implying 65%  $= 100 - 35$  could potentially be deceived). Other survey research report magnitudes in similar ranges. The above analysis shows that we can reject such large proportions of individuals being materially deceived by typical disclosure format in our context. Appendix D reports Monte Carlo power simulations that consider a range of a priori effect sizes and documents the statistical power in our data set.



### 6.6. Do Consumers Misinterpret Native Ads Under Typical Disclosure?

We next examine the data for any change in user behavior when the text in the label indicating advertising is changed from the common format reading “Ad” to “Sponsored” in the way shown in Figure 5. This assesses whether consumers interpret these widely used labels differently when the labels are meant to convey the same fact—that the link is advertised. If consumers interpret these labels differently, their behavior may be sensitive to the labels used to disclose ads.<sup>27</sup>

We perform the analysis by grouping the data into two categories: one in which the label used reads “Ad” (combining users in conditions Advertising-Ad and Advertising-Ad-Highlight) and a second in which the label reads “Sponsored” (combining users in conditions Advertising-Sponsored and Advertising-Sponsored-Highlight). The rationale behind combining the conditions with and without the additional highlight is to increase the power in the test by increasing the number of observations. To conduct the test, we regress the two outcome variables on an indicator of the label using the word “sponsored.”

Columns (5) and (6) in Table 3 show the estimates from the regressions. As before, column (5) shows the results with *page visit* as the dependent variable, and column (6) the results with *calls* as the dependent variable. The coefficients are small and statistically insignificant. A joint test is unable to reject the hypothesis that user behavior in terms of *page visits* and *calls* does not change when the text in the ad label changes from “Ad” to “Sponsored” ( $p$ -value = 0.24). We do not detect sensitivity in the ad’s effect to the two types of text used in the label to signify them. Considering the confidence intervals corresponding to these estimates, shown in columns (5) and (6) of Table 3, we can rule out with 95% confidence that more than 0.1% of the population in our experiment changes behavior because of the text in the label. Relative to the behavior when the ad label says “Ad,” we can say with 95% confidence that *page visits* change by less than 13% and *calls* change by less than 33% when the label is changed to “Sponsored.”

### 6.7. Heterogeneity in Effects

We now present additional analysis to better describe the effects of disclosure and to gain insight into how the effects operate. We use data on user and advertiser characteristics to examine systematic heterogeneity in effects.

What factors would affect the likelihood of users correctly spotting ads? Noticing an ad and interpreting its significance depends on the user’s understanding of the format and text used for disclosure, which may vary with the user’s experience with the platform or with similar media. A person’s ability to spot the ads also depends on his or her physiological visual perception

capabilities. This capability can vary with factors such as a person’s health or age (e.g., Johnson 2013).

We do not have data related to Zomato users’ visual perception, but we do have access to measures related to their experience with Zomato, which we use for this analysis. One difficulty is that it is a priori unclear how the effect of the ad format would vary with experience. For example, it is possible that users who are new to the platform are less comfortable with its user interface, so they are more likely to misinterpret the disclosure used to present the ads than experienced users. On the other hand, it is also possible that users who are new are more cautious relative to experienced users, and are more likely to spot ads in their search results. Given this uncertainty, we empirically examine whether the effect of advertising and its disclosure in our context varies with a user’s experience with the platform without taking an a priori stance on the direction of the effects.

We also investigate whether the effects vary systematically with advertiser characteristics, such as its rating, price, and popularity on Zomato. Even though these advertiser characteristics may not directly change the effect of advertising disclosure, these characteristics may be associated with different segments of the user population. Therefore, including these advertiser characteristics in the analysis enables exploring whether the effects are especially large for some user segments, which may signify existence of material deception for subsegments of the population.<sup>28</sup>

#### 6.7.1. Measures of User and Advertiser Characteristics.

To create a measure of experience, we use data on the date a user created her login on Zomato. The difference between the date of the session and the date of login creation is the *age* of the user on Zomato, which we use as a proxy measure for the user’s experience with the platform.<sup>29</sup> For each advertising restaurant, we observe the average rating it receive on Zomato, the number of times it is rated, and an index of the average price of food at the restaurant, all at baseline. To standardize these characteristics within a market, we categorize them into deciles, so each characteristic varies from 1 to 10 (10 being the highest).

To assess heterogeneity, we estimate the following two regressions:

$$y_{ir} = \beta_1 \text{Advertising-Highlight}_i \times Z_{ir} + \beta_2 \text{Advertising-No-Disclosure}_i \times Z_{ir} + \gamma Z_{ir} + \delta_1 \text{Advertising-No-Disclosure}_i + \delta_2 \text{Advertising-Highlight}_i + \epsilon_{ir} \quad (1)$$

$$y_{ir} = \text{Advertising-Sponsored}_i \times Z_{ir} + \gamma Z_{ir} + \delta \text{Advertising-Sponsored}_i + \epsilon_{ir}. \quad (2)$$

Each observation in the regressions corresponds to a user  $i$  and advertiser restaurant  $r$ . The term  $Z$  is a vector

of user experience and three restaurant characteristics. Both models are estimated with two dependent variables: (a) *whether i visited r's page*, or (b) *whether i called r*. The coefficients in regression (1) represent effects relative to typical disclosure. The coefficients in regression (2) represent effects of disclosing using the “Sponsored” label relative to the “Ad” label.

Table 4 shows the results. User and restaurant characteristics are found to be correlated with the likelihood of the user visiting an advertiser's page or calling the advertiser. Conditional on other factors, more experienced users are more likely to visit the advertiser's page. Higher-rated advertisers are more likely to be called and have their pages visited. More expensive advertising restaurants are less likely to be called though more likely to have their pages visited. Looking at the interaction terms in columns (1) and (2), we see the change in *calls* from not disclosing the ads varies with advertiser characteristics. When ads are not disclosed, *calls* to higher-rated restaurants decrease, whereas *calls* to that restaurants rated more times increase. The interaction terms with Advertising-Highlight are not significant. Therefore, we detect no evidence of systematic heterogeneity in the effect of highlighting the disclosure. Finally, the interaction terms in columns (3) and (4) are statistically insignificant, showing no evidence of systematic heterogeneity in the effect of the text in the label.

**6.7.2. Takeaways.** We can summarize these findings as follows:

1. In the previous section, we found that adding a highlight to the typical disclosure leads to no detectable change in *page visits* and *calls*. The current analysis shows that the spread around the average does not significantly vary with observed user and advertiser characteristics.
2. Removing the ad disclosure reduces users' *calls* to advertisers on average. This change is driven by advertisers with high ratings, and by those that have been rated fewer times.
3. The text in the ad label does not change average *page visits* or *calls*, and the spread around the average does not vary significantly with observed characteristics.

## 7. Are Consumers Tricked into Clicking on Native Advertising and Converting?

One of the criticisms of native advertising is that by blending ads into nonadvertised content, it makes unsuspecting individuals *click* on the ads and eventually purchase the advertised products. By facilitating clicks on listings, concerns have also been expressed that native advertising might work as a “misleading door opener” (see Federal Trade Commission 2015a, p. 7), that leads the consumer to an option they would otherwise not have considered. For this to lead to purchases, the implicitly maintained viewpoint is that consumers also do not

meaningfully and deliberately make subsequent choices once they click on the advertising messages, either because they are inertial and remain “stuck with the click” or because they face high search costs that reduce their propensity to search further. In this section, we analyze *how* ads work in our empirical context, focusing specifically on the role of clicks on ads (or *ad clicks*) to assess these viewpoints empirically. Specifically, we use the data to explore three questions:

1. Are incremental *calls* from disclosure associated with ad clicks?
2. Is there significant consumer search after ad clicks?
3. If incremental *calls* are not associated with ad clicks, how do ads work?

### 7.1. Are Incremental Calls from Disclosure Associated with Ad Clicks?

The cuts of the data relevant to this section are organized in Table 5. In the first panel of the table, we ask whether disclosure changes the propensity to click on ads. To do this, we compare (1) the likelihood of a user clicking on an ad when the advertiser's listing is placed in the same format as the organic listings and completely blended with the nonadvertised search results (i.e., the Advertising-No-Disclosure condition) with (2) that when the same ad is placed in a format that allows consumers to recognize ads (i.e., the remaining four subconditions in the Advertising condition). A comparison of the numbers across columns (1) and (2) of the first row in Table 5 shows that more users click on the ad when the ad is distinguishable from nonadvertised search results (0.52% versus 0.47%). However, this difference is small in magnitude compared with the change in *calls*, and is marginally statistically significant (at 90% confidence level). Now, recall from Section 6.2 that *calls do* increase quantitatively and statistically significantly with disclosure. Putting these two facts together suggests two things. First, reduced disclosure does not increase clicks by much, at least in this context. Second, the incremental *calls* arising from disclosure do not seem to be produced by users clicking on the ads.

We can see more suggestive evidence of the latter point if we split the probability of an individual calling a restaurant separately by whether she clicks on the restaurant's ad or not. Row (2) of Table 5 shows the chance of a user calling the advertiser conditional on clicking on its ad. Row (3) shows the corresponding probability for users who arrive at the advertiser's page through their own search. These estimates show that the probability of calling is higher when the users arrive at the page through their own search. This is more evident for the users who see the ads with some form of distinction from the organic listings (column (2)). Interestingly, across columns for both rows (2) and (3), we see that the probability of a user calling the advertiser is

**Table 4.** Heterogeneity in the Effect of Various Native Advertising Formats

	Five conditions, with/without disclosure		Four conditions with disclosure text “Ad” or “Sponsored”	
	Page visit (1)	Call (2)	Page visit (3)	Call (4)
Experience: User’s experience with Zomato	<b>0.000013 (0.000002)</b>	0.000001 (0.000001)	<b>0.000013 (0.000002)</b>	0.000001 (0.000001)
Rating: Advertiser’s rating on Zomato	<b>0.001472 (0.000157)</b>	<b>0.000133 (0.000038)</b>	<b>0.001513 (0.000157)</b>	<b>0.000079 (0.000036)</b>
Price: Advertiser’s price index	<b>0.000687 (0.000146)</b>	<b>−0.000096 (0.000036)</b>	<b>0.000770 (0.000143)</b>	<b>−0.000077 (0.000034)</b>
Num. ratings: Number of times advertiser is rated on Zomato	−0.000143 (0.000179)	−0.000001 (0.000040)	−0.000163 (0.000176)	0.000043 (0.000033)
Advertising-No-Disclosure	−0.000056 (0.002457)	−0.000160 (0.000436)		
Advertising-No-Disclosure × Experience	−0.000002 (0.000004)	−0.000001 (0.000001)		
Advertising-No-Disclosure × Rating	−0.000436 (0.000265)	<b>−0.000217 (0.000062)</b>		
Advertising-No-Disclosure × Price	0.000045 (0.000248)	0.000073 (0.000048)		
Advertising-No-Disclosure × Num. ratings	0.000291 (0.000303)	<b>0.00012 (0.00006)</b>		
Advertising-Highlight	−0.001344 (0.002052)	−0.000173 (0.000415)		
Advertising-Highlight × Experience	0.000006 (0.000004)	0.000000 (0.000001)		
Advertising-Highlight × Rating	−0.000020 (0.000224)	−0.000055 (0.000050)		
Advertising-Highlight × Price	0.000108 (0.000208)	0.000038 (0.000048)		
Advertising-Highlight × Num. ratings	0.000095 (0.000254)	0.000022 (0.000049)		
Advertising-Sponsored (with and without highlight)			0.000533 (0.002052)	0.000150 (0.000415)
Advertising-Sponsored × Experience			0.000004 (0.000004)	0.0000001 (0.000001)
Advertising-Sponsored × Rating			−0.000102 (0.000224)	0.000054 (0.000050)
Advertising-Sponsored × Price			−0.000059 (0.000208)	−0.000001 (0.000048)
Advertising-Sponsored × Num. ratings			0.000135 (0.000254)	−0.000066 (0.000049)
Intercept	<b>−0.004247 (0.001464)</b>	0.000314 (0.000324)	<b>−0.005189 (0.001404)</b>	0.000152 (0.000276)
Number of clusters	213,707	213,707	170,645	170,645

Notes. The table reports coefficients and standard errors clustered by user in parentheses from several regression. Columns (1) and (2) show estimates of the regression Equation (1), with dependent measures as indicators of *page visits* and *calls*, respectively. Columns (3) and (4) show corresponding estimates of the regression Equation (2). The estimates in bold are statistically significant at the 5% significance level. In terms of fit, the focal interaction terms explain little of the total variation in all specifications with partial  $R^2 < 10^{-4}$ .

higher when she sees the ads with disclosure. The difference is particularly large and significant among individuals who did not click on the experimental ads.<sup>30</sup>

Is there significant consumer search after ad clicks? We define the propensity to continue to search by an indicator of whether the user visited other restaurant page(s) during the same session subsequent to an ad click. Row (5) of Table 5 shows the likelihood of continued search of a user who arrives on an advertiser’s page after an ad click. Row (6) shows the likelihood of continued search for individuals who arrive at the page without clicking on an ad. Looking down the columns, we see that the probabilities of continued search are high for both groups. On average, about 85% of the individuals who arrive at the advertiser’s page after clicking on its ad continue to explore other options (row (5)). This suggests considerable search after ad clicks,

and low levels of inertia and stickiness with the clicked option. Finally, the difference in the propensity to search is the same irrespective of disclosure. [The differences are not significant across the two groups of users in columns (1) and (2).] This shows that individuals who click on the ad are more likely to continue other options irrespective of whether ads are distinguishable from other content, suggesting that consumer search does not conclude after an ad click.

If incremental *calls* are not associated with ad clicks, how do ads work? If the incremental *calls* are not being driven by ad clicks, the alternative explanation is that there is a *direct effect* whereby users exposed to the ad update their evaluation of the advertised product and call the advertiser without clicking on its ad, instead reaching the advertiser’s page through their own search effort. In this alternative path, ads work through

exposure alone, affecting purchases without inducing clicks on them; so ad clicks do not matter for purchase outcomes. Establishing such a direct effect is nontrivial, however. A naive approach would be to compare the *call* rates of users in the treated group who call conditional on not clicking on ads to those of users in the control group who call conditional on not clicking on the ad. This naive approach compares subpopulations of users defined by conditioning on not clicking on an ad, which is an endogenous posttreatment outcome. Any such conditioning induces a potential selection bias in the comparison. Below, we describe a test of the direct effect that circumvents this problem.

## 7.2. A Test for the Direct Effect of Ads

The test compares likelihood of a user calling and not clicking on the ad in the control group with the likelihood of a user calling and not clicking on the ad in the treatment group.<sup>31</sup> We argue that “No Direct effect”  $\Rightarrow$  “the proportion of users who call without clicking the ad in treatment group will be  $<$  the proportion of users who call without clicking on the ad in the control group.” We show that the latter statement is not true in our data, implying the existence of a direct effect.

**7.2.1. Rationale Behind the Test.** To understand the test, suppose there are  $N$  users in the control group, and  $n_0$  of them call the focal restaurant without clicking on the ad. Now consider the treatment group, also with  $N$  users. Figure 7 depicts these visually. Consider the

subgroup of users in the treatment group who would have called the advertiser and not clicked on the ad *if* they were instead in the control group. Let us call this subgroup  $S$ . Subgroup  $S$  is expected to be of size  $n_0$  because of randomization. There are two possible scenarios of how the treatment affects the behavior of this subgroup. Because “*call*” is a binary outcome, these two scenarios cover all possibilities:

*Scenario 1.* The treatment has no effect on the call behavior of subgroup  $S$ , that is, all these users in  $S$  call the advertiser.

*Scenario 2.* The treatment has a negative effect on the call behavior of subgroup  $S$ , that is, some of the users in  $S$  do not call the advertiser.

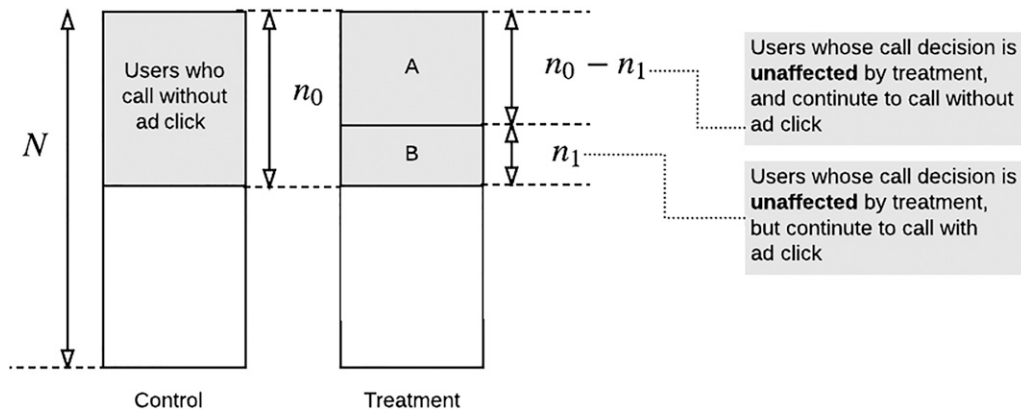
First, consider Scenario 1. This is depicted in the right panel of Figure 7, where the  $n_0$  users who would have called the restaurant in the control group continue to call the restaurant in the treatment group. Out of these  $n_0$  users, a subset  $n_1$  happen to click on the ad, and the remaining  $n_0 - n_1$  call the restaurant but do not click on the ad. These groups are referred to as “B” and “A,” respectively, in the figure.<sup>32</sup> Now, if the treatment has no direct effect, then none of the remaining  $N - n_0$  individuals call without clicking on the ad. Therefore, the proportion of users in the treatment group who call without ad clicking is  $\frac{n_0 - n_1}{N}$ , which will always be less than the proportion of users in the control group who call without ad clicking,  $\frac{n_0}{N}$ . On the other hand, if the treatment has a direct effect, then some of the remaining  $N - n_0$  users call without ad clicking. Let us

**Table 5.** Exploring Search, Clicks, and the Mechanism Driving Ad Response

	(1) Ads placed with no indication (Advertising-No-Disclosure condition)		(2) Ads placed with some indication (four conditions pooled)		$H_0$ : means in columns (1) and (2) are equal ( $p$ -value)
	Mean	Standard error	Mean	Standard error	
(1) Percentage of users clicking on ads	0.47%	0.03%	0.52%	0.01%	0.07
(2) Probability that an advertised restaurant is called for users who click on its ad	1.74%	0.70%	2.76%	0.42%	0.21
(3) Probability that an advertised restaurant is called for users who visit its page but do not click on its ad	2.62%	0.71%	4.94%	0.48%	<0.01
(4) $p$ -value testing the null hypothesis that means in rows (2) and (3) are equal	0.49		<0.01		
(5) Probability of continuing to search, conditional on visiting an advertiser’s page via an ad click	85.79%	1.88%	84.75%	0.92%	0.61
(6) Probability of continuing to search, conditional on visiting an advertiser’s page for users who did not click on ads	74.92%	1.64%	72.55%	0.83%	0.21
(7) $p$ -value testing the null hypothesis that means in rows (5) and (6) are equal	<0.01		<0.01		

*Notes.* The table presents averages, standard errors, and hypothesis tests to examine differences between users who do or do not click on experimental native ads. Observations in the data are categorized into two groups: (1) individuals in the condition that show experimental ads with no distinction from the rest of the nonadvertised content (Advertising-No-Disclosure) and (2) individuals in any one of the four conditions that show native ads in a format that is distinguishable from the rest of the content (the four conditions included are Advertising-Ad, Advertising-Sponsored, Advertising-Ad-Highlight, Advertising-Sponsored-Highlight). Column (3) shows the  $p$ -values from testing the null hypothesis that the means in columns (1) and (2) are equal.



**Figure 7.** Testing for a Direct Effect of Ad Exposure Without Ad Clicks: Rationale

say there are  $n_2$  such users. Then the proportion of users who call without ad clicking in the treatment group,  $\frac{n_0 - n_1 + n_2}{N}$ , can be higher compared with the control, if  $n_2$  is large enough. Thus, this difference in *call* rates from non-ad clickers in the treatment and control groups forms a test statistic for the existence of the direct effect.

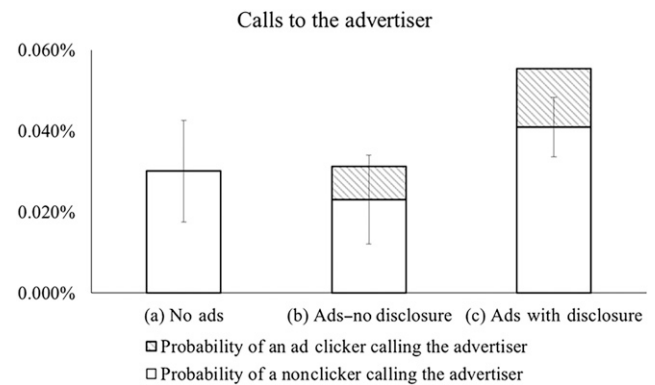
Now consider Scenario 2. In this scenario, the treatment causes some of the users,  $x$  in number, who would have called the advertiser without clicking on the ad in the control condition to not call the advertiser. Among the remaining  $n_0 - x$  who continue to call,  $n_1$  happen to click on the ad. Now, if the treatment has no direct effect, then none of the remaining  $N - n_0$  individuals call without clicking on the ad. Therefore, the proportion of users in the treatment group who call without ad clicking is  $\frac{n_0 - x - n_1}{N}$ , which will always be less than the proportion of users in the control group who call without ad clicking,  $\frac{n_0}{N}$ .

This test avoids the selection bias of the naive approach because it uses an unconditional average as the test statistic (the proportion of the treatment group as a whole that calls without ad clicking). The naive approach uses a conditional average as the test statistic (the proportion of the treatment group that call conditional on not ad clicking). The conditional comparison is subject to selection, the unconditional is not.

**7.2.2. Results from the Test.** Figure 8 shows the probability of calling the advertiser for three groups of users: (a) those in the experimental condition with no ads, (b) those in the condition with nondisclosed ads, and (c) those in the four conditions with ads disclosed. Our test focuses on the probability of calling without ad clicking, which is represented by the white, non-shaded portions of the bars. We test for both a direct effect of advertising (comparison of panels (c) and (a)) and a direct effect of disclosure (comparison of panels (c) and (b)). As discussed above, if there was no direct effect of advertising, we would expect the white portion of the bar in (c) to be smaller than that in (a), which

is rejected in our data ( $p = 0.04$ ). This establishes a direct effect of advertising. Similarly, if there was no direct effect of ad disclosure, we would expect the white portion of the bar in (c) to be smaller than that in (b), which is also rejected in our data ( $p < 0.01$ ). This establishes a direct effect of disclosure.

**7.2.3. Takeaways.** We draw two conclusions from the above analysis. First, clicks on the ads do not seem the only channel by which the causal effect of advertising functions: even non-ad clickers change their calling behavior when ads are placed and disclosed. Mere exposure to ads has a direct effect—individuals view the ads in their search results, update their impression of the advertiser, and continue with their search

**Figure 8.** Calls by Whether the Individual Clicked on the Ad

**Notes.** The chart shows the average probability of observing a *call* to the advertiser. This average is estimated for individuals in three groups: (a) those in the condition with no ads, (b) those in the condition with nondisclosed ads, and (c) those in the four conditions with ads disclosed. The shaded portion represents outcomes for individuals who clicked on the ad; the white portion represents outcomes for individuals who did not click on the ad. Obviously, the shaded portion is absent for condition with no ads. Because the chart is used to make inferences using averages for non-ad clickers, we show 95% confidence intervals for averages for non-ad clickers. Confidence intervals for ad clickers are not shown to avoid clutter.

without clicking on the ad. Eventually, if they decide to pick the advertised option, consumers reach it through their own organic search.<sup>33</sup> A second implication is that users in our data are sophisticated consumers of advertising. Users get exposed to advertising, process the information it conveys to them, and act on it in a deliberate manner after considered search. Users in our context do not exhibit stickiness; most of them continue searching after visiting a restaurant's page. We find no detectable evidence of these users being tricked into purchase by clicking on native ads.

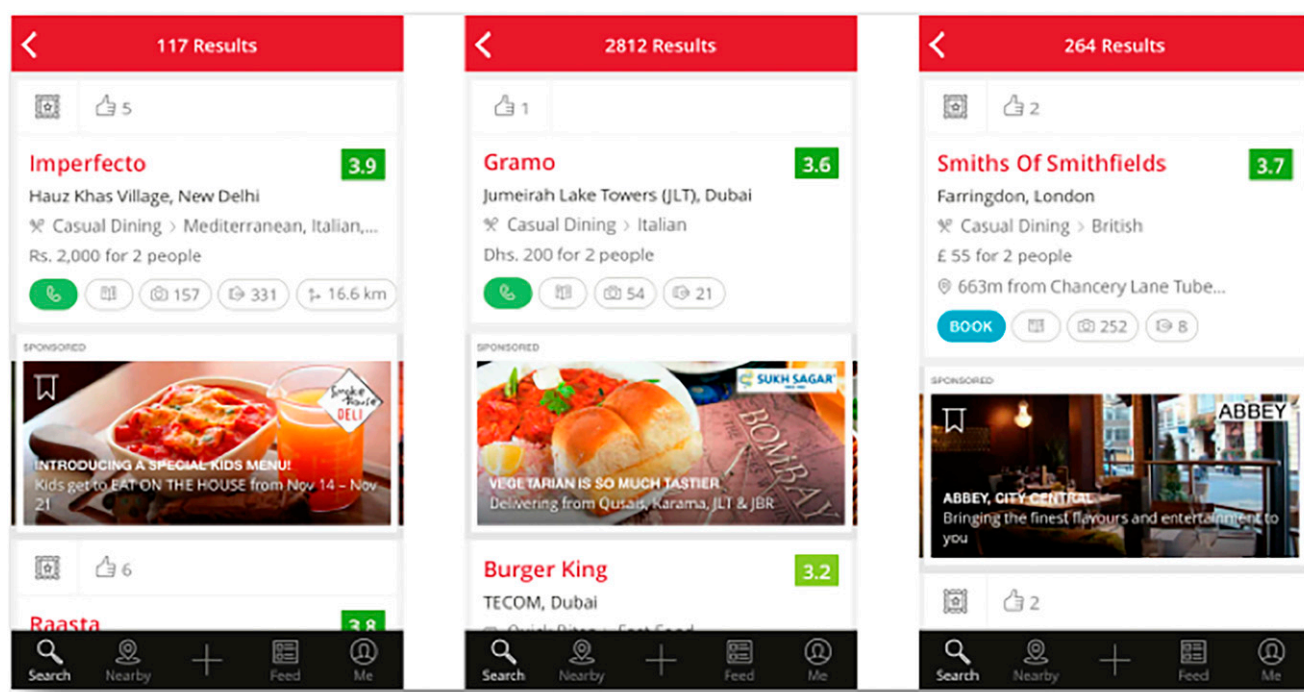
## 8. Conclusions

An assessment of whether native advertising materially deceives consumers and tricks them into changing their actions related to paid-search ads is presented. The assessment utilizes a field experiment implemented on a restaurant search platform. The novelty of the experiment is in presenting a way to assess the possibility of deception via revealed preference arguments. The experimental design randomizes users into conditions in which the sponsorship of a listing by an advertiser is disclosed prominently or not at all. Deception is said to have occurred if actions under typical standards of disclosure seen in the marketplace look closer to those in the counterfactual world in which

advertisement status is not revealed. An advantage of the setup is that user actions related to both the advertisement and the product being advertised are tracked, to assess the material nature of any possible effects from the interventions.

The empirical results do not provide evidence for substantive deception: user actions under typical disclosure look similar to those under prominent disclosure. We do detect a significant scope for deception because consumer behavior changes when disclosure is removed. Furthermore, users in our data seem to be sophisticated consumers of advertising, making restaurant choices deliberately after substantial search and exploration, and do not seem to be tricked by native formatting into purchasing products. Overall, the results inform the recent debate about native advertising, as well as the more long-standing one about separation between advertising and content in media. The following qualifications to our findings are important. Although we do not detect deception, our analysis lacks statistical power to detect changes in user behavior below the thresholds discussed in Section 6.3. We conjecture our findings regarding consumer sophistication may be relevant to other media settings. However, we caution against plainly extrapolating the findings to other contexts, especially ones in which

**Figure 9.** (Color online) Screenshot of Postexperiment In-Feed Advertising Introduced into the Mobile App



**Notes.** The figure shows a screenshot of the ad disclosure format introduced into the Zomato mobile platform after the experiment. Ads are prominently displayed in the search results page with a picture and logo of the advertising restaurant along with a label noting “Sponsored.”

consumers are less knowledgeable, and find media browsing and searching more effortful.

### 8.1 Epilogue

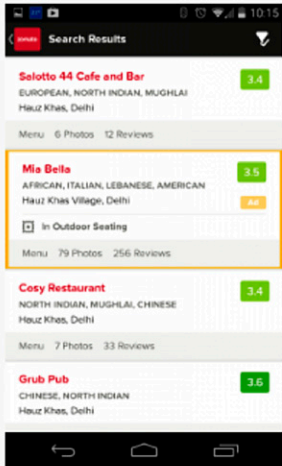
Following the experiment, Zomato launched mobile advertising on its Android app in November 2014 (see Russell 2014). The ads used the “Sponsored” label format with a picture and logo of the advertiser inserted into the search results to clearly and prominently make advertising salient to searching users. This decision was driven in part by the fact that no adverse user reaction or drop in *calls* to restaurants was seen from disclosure as part of this experiment. Figure 9 presents a screenshot of the new format implemented.

## Appendix A. Survey for Assessing Prominence of the Prominent Disclosure Condition

**Figure A.1.** (Color online) Screenshot from Survey: Assessing Noticeability of the Prominent Disclosure Condition Used in the Experiment

Please read the information below and answer questions about an app used to search for restaurants.

Are you able to spot the ad in the screen below? Specifically, is it clear *which one* of the links in the screen is an ad?



Yes, I can spot the ad

No, I cannot spot the ad

### Acknowledgments

The authors thank Dan Edelman, Puneet Manchanda, Ginger Jin; conference discussants Yesim Orhun and Brad Shapiro; and seminar participants at Harvard Business School, MIT Sloan, the University of Michigan, the 2016 Big Data Marketing Analytics Conference, MKS-FTC 2016 Conference, the 2016 Great-China Conference on Mobile Big Data Marketing, and 2015 and 2017 Marketing Science Conferences for very helpful comments and suggestions. The authors thank Shuo Xie and Charles Zhang for excellent research assistance. The authors declare they have no commercial relationship with Zomato, and the research was conducted with no restrictions on the publishability of the findings from the field experiment. The views discussed here represent those of the authors and not of Stanford University or Zomato. The usual disclaimer applies.

We conducted a survey to assess the assumption maintained in the experiment that highlighting the ad in the Prominent Disclosure condition makes it clear and conspicuous to users.<sup>34</sup> We implemented the survey on 200 Amazon Mechanical Turk subjects through Qualtrics. Subjects were prescreened for ownership of a smartphone and for having used a smartphone for doing a product search. The survey showed consumers a screenshot from the Prominent Disclosure condition and asked the respondents whether they could spot the ad. Additionally, the survey also showed the respondents a screenshot from the Typical Disclosure condition and the same screenshot from the Prominent Disclosure condition to test whether they found the ad more recognizable with the highlight. Figures A.1 and A.2 show screenshots from the survey. We found the following:

- Responses of 196 out of 200 subjects (98%) indicated they could spot the ad in the highlighted condition. This is consistent with the maintained assumption that the highlighted condition makes ads clearly visible.<sup>35</sup>

- Responses of 185 of the 200 subjects (92.5%) indicated they would find it easier to spot ads in the highlighted condition, relative to the condition without the highlight. Specifically, 72% indicated that the highlight would “very likely” make spotting the ad easier; 20.50% indicated “likely,” 4.50% indicated “unlikely,” and 3% indicated “very unlikely.”

- Responses of 47 of the 200 subjects (23.5%) indicated that they were “very likely” to spot the ad in the condition without a highlight; 61.5% responded “likely,” and the rest “unlikely” (14%) or “very unlikely” (1%).

- Responses of 166 of the 200 subjects (83%) indicated that they were “very likely” to spot the ad in the condition with a highlight; 12.5% responded “likely,” and the rest “unlikely” (3.5%) or “very unlikely” (1%).

Overall, the survey data are supportive of the assumption that adding the highlight makes the ad noticeable. However, we are cautious about these findings because these data may be biased due to demand effects or respondent acquiescing.

**Figure A.2.** (Color online) Screenshot from Survey: Comparing Noticeability of the Typical and Prominent Disclosure Conditions

Please see two presentations of the identical list of restaurants below. Let's call the picture on the left (A) and the picture on the right (B).

(A)

(B)

We want to know which screen (A) or (B) is clearer in terms of showing the ad. The ad is the second link in both the pictures.

	Very likely	Likely	Unlikely	Very unlikely
While using an app like (A) will you be able to spot an ad?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
While using an app like (B) will you be able to spot an ad?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Will it be easier to spot an ad in (B) relative to (A)?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

<<
>>



## Appendix B. Randomization Checks

Table B.1 shows randomization checks across the six experimental conditions in the experiment. We see that randomization is induced properly.

**Table B.1.** Randomization Checks

User characteristic	Variable	Averages: experiment condition number						<i>p</i> -value of test ( $H_0$ : equal means across conditions)
		1	2	3	4	5	6	
Past engagement with Zomato	<i>Number of reviews</i>	0.76	0.77	0.76	0.76	0.78	0.75	0.95
	<i>Number of ratings</i>	1.70	1.68	1.62	1.68	1.70	1.68	0.72
	<i>Days since signed up</i>	285.76	286.16	285.48	285.64	282.48	285.97	0.34
Search activity prior to the experiment	<i>Number of restaurant clicks</i>	2.71	2.71	2.81	2.79	2.77	2.75	0.71
	<i>Number of menus seen</i>	1.31	1.33	1.38	1.35	1.35	1.35	0.66
	<i>Number of calls made</i>	0.14	0.14	0.13	0.14	0.14	0.14	0.89
	<i>Number of searches</i>	1.36	1.34	1.34	1.37	1.33	1.34	0.49
	<i>Number of sessions</i>	0.48	0.48	0.48	0.49	0.49	0.48	0.84
Characteristics of restaurants visited in the past	<i>Average rating of the restaurants</i>	3.50	3.50	3.49	3.48	3.49	3.49	0.24
	<i>Average price (indexed) of restaurants</i>	13,985.91	14,568.83	14,165.84	15,120.68	14,083.27	14,094.49	0.85
	<i>Popularity in terms of number of reviews</i>	334.83	340.72	341.63	330.97	337.60	334.90	0.41

*Note.* This table tests whether the baseline characteristics of users in the six experimental conditions are statistically significantly different from each other, to assess whether randomization is induced correctly.

## Appendix C. Counts of Page Visits and Calls

Table C.1 shows the counts of *page visits* and *calls* received by advertisers within the focal session in the six experimental conditions.

**Table C.1.** Counts of *Calls* and *Page Visits*

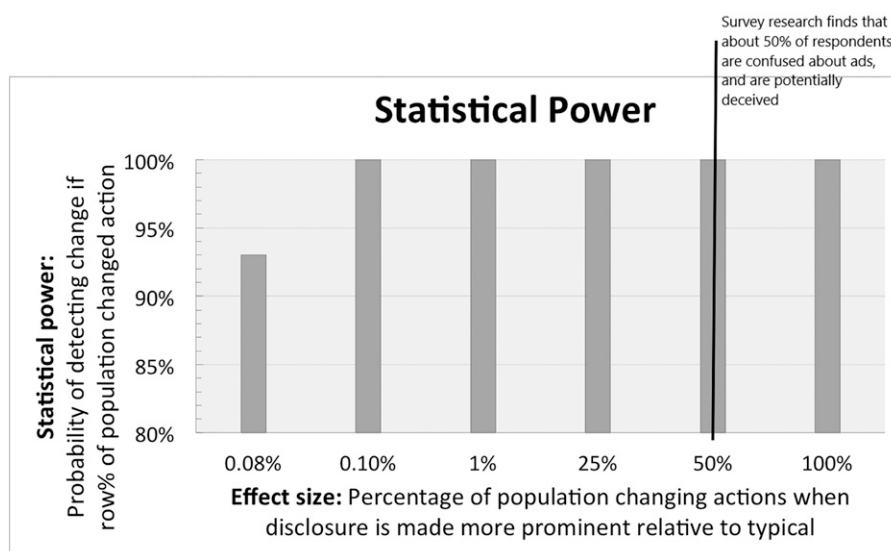
	Number of <i>calls</i>	Number of <i>page visits</i>
No-Advertising	22	605
Advertising-No-Disclosure	23	814
Advertising-Ad	50	845
Advertising-Sponsored	32	837
Advertising-Ad-Highlight	38	863
Advertising-Sponsored-Highlight	42	907
Total	207	4,871

## Appendix D. Monte Carlo Power Simulations

This section reports on Monte Carlo simulations to assess the statistical power of the experiment to detect deception in our setting. We consider a broad range of effect sizes and ask whether our experiment would be able to detect a change in consumer behavior of that magnitude. Specifically, given an effect size, we assess the ability of the experiment and the sample to reject the null of equality between the typical and prominent conditions. We compute bootstrapped estimates of power by simulation. For each effect size, we bootstrap 10,000 samples from the data, run the SUR regression reported in Section 6.2 for each bootstrapped sample, collect the 10,000 *p*-values from the tests, and report as power the

proportion of these *p*-values that are less than 5%.<sup>36</sup> Figure D.1 plots the statistical power against a range of effect sizes. The *x*-axis of the figure is in absolute terms (percentage of the population). One could scale it with any benchmark statistic to assess relative effects. Looking at Figure D.1, we see that if 0.1% of the population was being deceived (0.1% change their decision and 99.9% do not change their decision), we would have detected it at the 5% significance level with probability close to 1. The data rule out deception of the magnitudes one might extrapolate from survey research: for our setting, we have close to 100% power to conclude that 50% of the population did not change their decisions when ads with typical disclosure were highlighted.

**Figure D.1.** Statistical Power Varying with the Effect: Percentage of the Population Changing Decisions Because of the Highlight



## Endnotes

<sup>1</sup> The Interactive Advertising Bureau, the main trade group of the digital advertising industry in the United States, classifies digital native ads into six types: in-feed units, paid search units, recommendation widgets, promoted listings, IAB standard ads with native elements, and custom ad units (Interactive Advertising Bureau 2013).

<sup>2</sup> The focus on materiality is a fundamental aspect of regulation of deceptive practices by the FTC. The FTC's mandate is to regulate unfair methods of competition when "it shall appear to the Commission that a proceeding by it in respect thereof would be to the interest of the public" (Federal Trade Commission Act of 1914, § 5(b) <https://www.law.cornell.edu/uscode/text/15/45>). The focus of the mandate on the public interest motivated the incorporation of the materiality concept into the commission's deception standard (Richards and Preston 1992). As the *Harvard Law Review* (1967) elaborates, "There is a twofold justification for requiring the Commission to go beyond showing a falsehood and show how it could affect buyers' actions. First, it conserves public resources by focusing Commission attention on only those advertising falsehoods which lead to production-distorting economic behavior or other significant harm. Second, advertisers have a legitimate claim to freedom from regulation except where their activities threaten to injure others."

<sup>3</sup> The FTC bases the materiality definition on whether the disclosure affects consumers' actions, without a stand on whether the disclosure benefits the advertiser and/or consumer. Unlike the separate "deceit" standard, which focuses on damages, for the "deception" standard under consideration here, what matters is only whether the object is other than what is represented and whether that affects behavior (Richards and Preston 1992). The focus on is on accurate information provision so that market participant's behavior is not distorted. This paper uses the term "deception" in the strictly legal sense of its usage in the U.S. justice system for cases related to the deception standard.

<sup>4</sup> While evaluating practices as deceptive, the FTC may presume materiality, or require evidence for it, depending on the circumstances. The commission has the legal authority to presume materiality for several types of information, including express claims; omitted information the seller knew, or should have known, which ordinary consumers would need to evaluate the product or service; claims the seller knew, or should have known, were false; implied claims, where there is proof the seller intended to convey them;

information involving health, safety, or other areas with which the reasonable consumer would be concerned; and information pertaining to the product's central characteristics, for example, involving the product's purpose, safety, efficacy, cost, durability, performance, warranties, or quality, or regarding findings by another agency concerning the product. See Federal Trade Commission (1983) and Richards and Preston (1992) for discussions.

<sup>5</sup> The exact choice of a prominent disclosure condition that satisfies these criteria depends on the context. To draw conclusions from this research design, all parties to the debate would need to agree on this choice. We discuss these considerations in detail in Section 3.2.

<sup>6</sup> Note these contrasts are not due to the differences in the type of consumers or advertising restaurants, or the listing position or other content attributes of the ads. These factors are held fixed in the comparisons because of randomization.

<sup>7</sup> What is relevant for the legal standard on deception is whether the deception *changes* consumer behavior per se, irrespective of the direction of the change and whether the change benefits the advertiser. For more discussion, see, for instance, Richards and Preston (1992).

<sup>8</sup> Clearly, assessing that consumers are equally "misled" between the Typical and Prominent Disclosure regimes would also imply that there is no evidence of material deception (i.e.,  $\mathcal{B} \Rightarrow \mathcal{A}$ ). But we are proceeding here under the assumption this test is harder to implement precisely because it requires asking consumers whether they are misled, which comes with a host of problems mentioned above. Philosophically, we are giving tests based on revealed preference more weight than those based on stated preferences.

<sup>9</sup> To preview our results, this is the case we find in the empirical setting in the analysis later in this paper.

<sup>10</sup> In our application, we focus on detecting statistically significant changes in behavior and present the confidence intervals that serve as bounds for the changes. A threshold effect size will enable us to check whether the threshold lies within the confidence interval or not.

<sup>11</sup> For example, Craswell (1997) notes that in cases about false claims in advertising, courts have often used a threshold of 15%–20%; using a copy test, a challenged ad is judged as making a false claim if the proportion of consumers believing the false claim is higher than the threshold.

<sup>12</sup> Because user-level demographics are not collected reliably on the Zomato app, we utilize comparative profiling information from

Alexa.com, a widely used provider of internet analytics. Alexa's browsing data come from multiple sources, including the users of their browser toolbar. The demographic data are self-reported. We have access to data on demographic comparisons only.

<sup>13</sup> The firm has an interest in avoiding formats that induce consumer annoyance, and has some past experience in being able to judge and screen out such formats.

<sup>14</sup> The Federal Trade Commission (2015b) says, "Any background shading used to differentiate native ads from non-advertising content should be sufficiently saturated for consumers to notice it. Advertisers also should consider using visual cues in addition to background shading, such as a prominent border that sets off native ads from surrounding content, in case consumers cannot see color differences."

<sup>15</sup> To be clear, the purpose of the survey is to assess the responders' opinions on the prominence of our execution of prominent disclosure. The survey is not designed to evaluate the central question of material deception.

<sup>16</sup> Situations in which the same restaurant appears as an ad and in the organic results on the same page are rare in the data. Specifically, an advertiser and its organic listing appears together in 5.2% of all searches in our data. Because of randomization, this number is, on average, the same across the three experimental conditions ( $p = 0.50$ ). Furthermore, removing such cases does not change our findings.

<sup>17</sup> A page on Zomato's website may be able to display more ads (has more ad slots) than the mobile app. In such cases, Zomato's proprietary algorithm decides which restaurants are to be advertised on the mobile app. The algorithm's criteria is uniformly applied across the experimental conditions, so differences across conditions do not reflect the effect of these placements.

<sup>18</sup> This would be specifically problematic in the Advertising-No-Disclosure condition.

<sup>19</sup> A session comprises a sequence of actions on the mobile app. A session is said to have ended when there is continuous inactivity for three hours or more. The goal in constructing sessions in this manner is to collect together a series of actions on the app that map to a distinct purchase occasion.

<sup>20</sup> Suppose on the day when a search occurred, Zomato contracted with two restaurants,  $r_1$  and  $r_2$ , on the same advertising position. In such situations Zomato displays one of the two restaurants in that position, rotating between the two. In this case, we study the effect of the user's experimental condition on her behavior with respect to both  $r_1$  and  $r_2$ . Note that for the purpose of estimating the precision of our estimates, we cluster the standard errors by individual, to keep the effective number of observations the same as the number of users.

<sup>21</sup> This is related to the question of what constitutes relevant actions. In general, materiality pertains to whether actions related either to the advertisement and/or the product being advertised are affected. There is precedence suggesting regulators consider intermediate measures of consumer behavior, and not just the purchase behavior to judge materiality. This suggests that *page visits* may be a measure of consumer behavior that is worth considering, in addition to *calls* (the proxy for purchase intent).

<sup>22</sup> Sahni and Nair (2016) document that *calls* form a proxy for purchase intent in the experimental markets. It is possible that some consumers purchase without calling; therefore, the measure is not perfect.

<sup>23</sup> Articulating the considerations that make a paid-search listing to be classified as native, "The Native Advertising Playbook" from the Interactive Advertising Bureau (2013) notes, "While the content and format of organic search engine results varies depending on the search engine and the platform through which the service is being accessed (desktop, mobile, tablet, etc.), there is one definitive principle that defines an ad as native: native ads in search must present their content in a format and layout that is readily available to organic search engine results."

<sup>24</sup> This finding is important and has implications for how we think about advertising effects. A companion paper, Sahni and Nair (2016), investigates this question in detail.

<sup>25</sup> We picked this benchmark to illustrate the confidence intervals on the suggestion of a reviewer. The reader could use other benchmarks to make inferences in a similar manner.

<sup>26</sup> Alternately, one could ask, what effect sizes can an analysis with our sample size detect, given a particular statistical power threshold? To assess this, we focus on the change from adding a highlight to typical disclosure. With 50% power (one minus probability of a type 2 error) our analysis can detect a 0.0009 change in the probability of a *page visit* to the advertiser and 0.0002 change in the probability of a *call*, at the conventional 95% confidence level. The corresponding numbers for 80% power are 0.0014 for *page visits* and 0.0003 for *calls*.

<sup>27</sup> For instance, Edelman and Gilchrist (2012, p. 77) say, "The word 'sponsored' creates considerable ambiguity. For one, 'sponsored' uses the passive voice—leaving unstated the subject of the sentence, i.e., who exactly did the sponsoring. If a user searches for 'Priceline' at Google and sees a link to Expedia, who is the 'sponsor' of that link? Sophisticated users are likely to understand that 'sponsored' means 'advertisement.' But the disclosure does not specifically inform anyone who does not already know."

<sup>28</sup> For example, if making the ads prominent increases *calls* to low-priced restaurants, it may signify that lower-income individuals who see cheaper restaurants are materially affected by typical advertising disclosure.

<sup>29</sup> We did the same analysis adding an indicator of whether the user rated any restaurants on Zomato, which is a proxy for engagement with the platform. The inference from the analysis did not change. We did not include this variable in this section because it is highly correlated with our measure of user experience.

<sup>30</sup> Even though this finding is indicative of some ad-disclosure effect operating without ad clicks, this comparison is not conclusive because it compares individuals who are self-selected into these comparison groups, causing a potential selection bias. We present a test that circumvents the selection problem later in the section.

<sup>31</sup> Depending on the objective, one could study (a) the direct effect of placing the ad with disclosure or (b) the direct effect of disclosure alone. For (a), the treatment group would be the group that gets ad exposure with disclosure, and the control group would be the group that does not see the ad. So clicking on the ad is not possible in the control group. For (b), the control group would be the group that sees the ad without disclosure. For this group, clicking on the ad is possible because the ads are placed.

<sup>32</sup> To be clear, the treatment has not affected the call behavior of the A or B group, but has affected the ad-click behavior of the B group.

<sup>33</sup> The inadequacy of ad clicks to assess the efficacy of online banner ads has been pointed out in other academic research. Gordon et al. (2016) document that a large proportion of consumers who convert following exposure to Facebook ads do not click on them. Lewis and Reiley (2014) document a similar phenomenon for banner ads on Yahoo!.

<sup>34</sup> A survey of this sort may be useful even if it does not establish in an iron-clad way that the Prominent Disclosure condition has no deception. Suppose regulators are willing to commit to the notion that prominent disclosure is sufficient for their objectives given its noticeability as established in the survey. At the very least, then, we can assess whether the typical disclosure is more deceptive than an alternative that regulators are comfortable with. We thank an anonymous reviewer for pointing this out.

<sup>35</sup> Three out of the four respondents who said they could not spot the ad with the highlight appears to have picked the wrong option (in the sense that they intended to say *yes* but selected *no*). Specifically, they responded to the second question that they would be likely to spot the

ad in the highlighted condition, which is inconsistent with their response to the first question. One respondent responded negatively to all questions.

<sup>36</sup> Here, we elaborate on the procedure. We estimate the probability of rejecting the null that additional highlighting has no effect, assuming the alternative (that additional highlighting has an effect) is true. To do this, we block sample at the user level with replacement two data sets for each effect size  $x$ : (a) We block sample at the user level with replacement from conditions Advertising-Ad and Advertising-Sponsored pooled together. (b) We change the decision of  $x\%$  of the population and then block sample at the user level with replacement from conditions Advertising-Ad and Advertising-Sponsored pooled together. How do we simulate a change in individual decision? A decision is a duple of two indicators (*visit*, *call*), which can take three values: (0, 0), (1, 0), or (1, 1) (not four, because *call* cannot occur without a *visit*). In this simulation, we assume that a decision can change to any of the other two possibilities with probability one-half. For example, if an individual visited but did not call, decision (1, 0), then a changed decision could be (1, 1) or (0, 0) with the same probability. Sample (a) is a sample of the behavior of users when ads are disclosed in a typical manner. Sample (b) is a sample of the behavior when  $x\%$  of the same population changes behavior. Then we test the null: *calls* and *page visits* are the same in (a) and (b) using an SUR regression and get a  $p$ -value, which is compared with the conventional value of 0.05. We repeat this procedure for many draws and many values of  $x$  to obtain Figure D.1.

## References

- Aribarg A, Schwartz EM (2017) Consumer responses to native advertising. Working paper, Ross School of Business, University of Michigan, Ann Arbor.
- Bakshi A (2015) Why and how to regulate native advertising in online news publications. *J. Media Law Ethics* 4(3–4):4–47.
- Beales H, Craswell R, Salop SC (1981) The efficient regulation of consumer information. *J. Law Econom.* 24(3):491–539.
- Blake T, Nosko C, Tadelis S (2015) Consumer heterogeneity and paid search effectiveness: A large scale field experiment. *Econometrica* 83(1):155–174.
- Burke RR, DeSarbo WS, Oliver RL, Robertson TS (1988) Deception by implication: An experimental investigation. *J. Consumer Res.* 14(4):483–494.
- Carlson M (2014) When news sites go native: Redefining the advertising–Editorial divide in response to native advertising. *Journalism: Theory Practice Criticism* 16(7):1–17.
- Coddington M (2015) The wall becomes a curtain: Revisiting journalism’s news-business boundary. Carlson M, Lewis SC, eds. *Boundaries of Journalism: Professionalism, Practices and Participation* (Routledge, New York), 67–82.
- Coffee P (2016) FTC slams Lord and Taylor for not disclosing paid social posts and native ads. *AdWeek* (March 15), <http://www.adweek.com/brand-marketing/ftc-slams-lord-taylor-deceiving-customers-not-disclosing-its-native-ads-170229>.
- Cohen D (2016) Native advertising dominates Facebook audience Network (study). *AdWeek* (April 5), <http://www.adweek.com/digital/native-advertising-facebook-audience-network-study/>.
- Craswell R (1997) “Compared to what?” The use of control ads in deceptive advertising litigation. *Antitrust Law J.* 65(3):757–791.
- Edelman B, Gilchrist D (2012) Advertising disclosures: Measuring labeling alternatives in internet search engines. *Inform. Econom. Policy* 24(1):75–89.
- Engle M (2013) Updated search engine letter from FTC. Accessed February 1, 2016, <https://www.ftc.gov/sites/default/files/attachments/press-releases/ftc-consumer-protection-staff-updates-agencys-guidance-search-engine-industry-on-need-distinguish/130625searchenginegeneralletter.pdf>.
- Federal Trade Commission (1983) FTC policy statement on deception. Accessed February 1, 2016, <https://www.ftc.gov/public-statements/1983/10/ftc-policy-statement-deception>.
- Federal Trade Commission (2013) FTC consumer protection staff updates agency’s guidance to search engine industry on the need to distinguish between advertisements and search results. Accessed February 1, 2016, [www.ftc.gov/news-events/press-releases/2013/06/ftc-consumer-protection-staff-updates-agencys-guidance-search](http://www.ftc.gov/news-events/press-releases/2013/06/ftc-consumer-protection-staff-updates-agencys-guidance-search).
- Federal Trade Commission (2015a) Commission enforcement policy statement on deceptively formatted advertisements. Accessed February 1, 2016, <https://www.ftc.gov/public-statements/2015/12/commission-enforcement-policy-statement-deceptively-formatted>.
- Federal Trade Commission (2015b) Native advertising: A guide for businesses. Accessed February 1, 2016, <https://www.ftc.gov/tips-advice/business-center/guidance/native-advertising-guide-businesses>.
- Franklyn D, Hyman D (2013) Trademarks as search engine keywords: Much ado about something? *Harvard J. Law Tech.* 26(2).
- Goldfarb A, Tucker C (2011) Search engine advertising: Channel substitution when pricing ads to context. *Management Sci.* 57(3):458–470.
- Gordon B, Zettelmeyer F, Bhargava N, Chapsky D (2016) A comparison of approaches to advertising measurement: Evidence from Big field experiments at Facebook. Working paper, Kellogg School of Management, Northwestern University, Evanston, IL.
- Harvard Law Review (1967) Developments in the law—Deceptive advertising. *Harvard Law Rev.* 80(5):1005–1155.
- Hoetzel M (2015) Spending on native advertising is soaring as marketers and digital media publishers realize the benefits. *Business Insider* (May 20), <http://www.businessinsider.com/spending-on-native-ads-will-soar-as-publishers-and-advertisers-take-notice-2014-11>.
- Hoofnagle C, Meleshinsky E (2015) Native advertising and endorsement: Schema, source-based misleadingness, and omission of material facts. Working paper, University of California, Berkeley, Berkeley.
- Interactive Advertising Bureau (2013) The native advertising Playbook: Six native ad categories, six marketplace considerations, and IAB recommended disclosure principles. Accessed February 1, 2016, <http://www.iab.com/guidelines/native-advertising/>.
- Interactive Advertising Bureau (2015) IAB concerned about FTC guidance on native advertising. Accessed February 1, 2016, <http://www.iab.com/news/iab-concerned-about-ftc-guidance-on-native-advertising/>.
- Jansen B, Brown A, Resnick M (2007) Factors relating to the decision to click on a sponsored link. *Decision Support Systems* 44(1):46–59.
- Jin GZ, Kato A (2006) Price, quality, and reputation: Evidence from an online field experiment. *RAND J. Econom.* 37(4):983–1005.
- Johnson J (2013) Human visual perception: Impacts our ability to distinguish online content from advertising. Presentation, Blurred Lines: Advertising or Content?—An FTC Workshop on Native Advertising, Federal Trade Commission, December 4, Washington, DC.
- Lewis RA, Reiley DH (2014) Online ads and offline sales: Measuring the effect of retail advertising via a controlled experiment on Yahoo! *Quant. Marketing Econom.* 12(3):235–266.
- Miller JC (1983) FTC policy statement on deception. Accessed February 1, 2018, [https://www.ftc.gov/system/files/documents/public\\_statements/410531/831014deceptionstmt.pdf](https://www.ftc.gov/system/files/documents/public_statements/410531/831014deceptionstmt.pdf).
- Mitchell M (2015) State of the news media 2015. Accessed April 8, 2019, <http://www.journalism.org/2015/04/29/state-of-the-news-media-2015/>.
- Narayanan S, Kalyanam K (2015) Position effects in search advertising and their moderators: A regression discontinuity approach. *Marketing Sci.* 34(3):388–407.



- Owen DK, Plyler JE (1991) The role of empirical evidence in the federal regulation of advertising. *J. Public Policy Marketing* 10(1):1–14.
- Pappalardo JK (1997) The role of consumer research in evaluating deception: An economist's perspective. *Antitrust Law J.* 65(3): 793–812.
- Rao A, Wang E (2017) Demand for healthy products: False claims in advertising. *J. Market Res.* 54(6):968–989.
- Richards JI, Preston IL (1992) Proving and disproving materiality of deceptive advertising claims. *J. Public Policy Marketing* 11(2): 45–56.
- Russell J (2014) Restaurant guide Zomato launches mobile ads and prepares to move into payments. *TechCrunch* (November 28), <http://techcrunch.com/2014/11/28/restaurant-guide-zomato-launches-mobile-ads-and-prepares-to-move-into-payments-for-an-announcement>.
- Russo JE, Metcalf BL, Stephens D (1981) Identifying misleading advertising. *J. Consumer Res.* 8(2):119–131.
- Sahni N (2015) Effect of temporal spacing between advertising exposures: Evidence from online field experiments. *Quant. Marketing Econom.* 13(3):203–247.
- Sahni N, Nair H (2016) Does advertising serve as a signal? Evidence from field experiments in mobile search. Working paper, Stanford Graduate School of Business, Stanford University, Stanford, CA.
- Sebastian M (2014) *Wall Street Journal* adopts native ads, tactic its editor has said can lead to Faustian pacts. *AdAge* (March 10), <https://adage.com/article/media/wall-street-journal-introducing-native-ads-site/292044/>
- Sonderman J, Tran M (2013) Understanding the rise of sponsored content. Accessed June 28, 2019, <https://www.americanpressinstitute.org/publications/reports/white-papers/understanding-rise-sponsored-content/>.
- Swant M (2016) Publishers are largely not following the FTC's native ad guidelines. *AdWeek* (April 8), <http://www.adweek.com/digital/publishers-are-largely-not-following-ftcs-native-ad-guidelines-170705/>.
- Tucker C (2012) Social advertising. Working paper, Sloan School of Management, Massachusetts Institute of Technology, Cambridge.
- Tutaj K, van Reijmersdal EA (2012) Effects of online advertising format and persuasion knowledge on audience reactions. *J. Marketing Comm.* 18(1):5–18.
- Wojdyski BW, Evans NJ (2015) Going native: Effects of disclosure position and language on the recognition and evaluation of online native advertising. *J. Advertising* 45(2):157–168.
- Zellner A (1962) An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *J. Amer. Statist. Assoc.* 57(298):348–368.