Firms’ Reactions to Public Information on Business Practices: Case of Search Advertising*

Justin M. Rao  Andrey Simonov
Microsoft Research  University of Chicago (Booth)

May 20, 2016

Abstract

We use five years of bidding data to examine the reaction of advertisers to widely disseminated press on the lack of effectiveness of brand search advertising (queries that contain the firm’s name) found in a large experiment run by eBay (Blake, Nosko and Tadelis, 2015). We estimate that 11% of firms that did not face competing ads on their brand keywords, matching the case of eBay, discontinued the practice of brand search advertising. In contrast, firms did not react to the information pertaining to the high value and ease of running experiments—we observe no change in the experiment-like variation in advertising levels. Further, while 72% of firms had sharp changes in advertising suitable for estimating causal effects, we find no correlation between firm-level advertising effects and the propensity to advertise in the future. We discuss how a principal-agent problem within the firm would lead to these learning dynamics.

*The views expressed reflect those of authors and not Microsoft or the University of Chicago. We thank Matthew Gentzkow and Matt Goldman for useful comments and suggestions.
1 Introduction

Empirical findings in economics and management science often have normative implications for firm behavior. Indeed, papers that speak directly to the profitability of decisions faced by firms appear regularly in peer-reviewed journals. Particularly relevant findings are amplified by media coverage. Perhaps surprisingly, very little is known about how scientific progress in these areas impacts business practice. Our lack of knowledge can be explained by a few factors. First, it can be difficult to record decisions with the required granularity and quantitative rigor. Second, it is generally even more difficult to link these choices to their direct economic consequences. Third, even when such measurement is possible, the relevant data usually do not enter the academic discourse or do so via one-off arrangements with single firms, making it hard to draw broader conclusions.

We aim to overcome these challenges with a unique combination of detailed data on firm-level decisions with directly measurable consequences and the release of a particularly impactful academic paper that received widespread media attention. Our domain is sponsored search, a popular form of online advertising in which advertisers bid on slots at the top of search engine results. Since an advertiser can be present in both the sponsored links and the “organic” links returned by the search engine, there is the possibility of crowd-out between paid and free clicks. Blake, Nosko and Tadelis (2015, herein BNT) study this crowd-out with a field experiment involving tens of millions of dollars of search advertising. The experimental results show that ads on branded keywords (e.g. “eBay shoes”) had almost no effect on traffic to eBay. Despite the minuscule causal impact, the nominal metrics of these ads—high click-through-rate (CTR) and low cost-per-click (CPC)—made them appear to be strong performers. Based on these results eBay stopped advertising on branded terms, recouping an annual expenditure exceeding fifty million dollars.

BNT’s results call into question the entire enterprise of advertising on own-brand queries. Since the practice is widespread, it is no surprise that the paper received attention in the popular press (e.g. The BBC), search engine marketing blogs (e.g. Search Engine Land) and the business press (e.g. the Harvard Business Review article titled “Did eBay just prove that paid search ads do not work?”). We study the impact of this information disclosure using detailed bidding data covering a five-year period on Bing, focusing our analysis on the roughly one thousand firms for which there is adequate data coverage. We examine firms’
reactions in two dimensions: 1) the propensity to advertise on own-brand queries; 2) the propensity to conduct experiments to measure ad effectiveness.

A natural way a firm could react to BNT’s findings is to judge whether or not they are in a comparable situation as eBay and, if so, simply follow the example of eBay by stopping brand search advertising. A critical part of this comparable situation caveat is that eBay did not face competing advertisers that stood to supplant them in the top slot.1 BNT note that this is a critical consideration and Simonov, Nosko and Rao (2015) subsequently showed that the value of brand ads indeed lies in the ability to prevent a competitor from occupying the top spot.2 Based on this observation, we use a difference-in-differences framework specifying firms that did not regularly face competitors as the treated group and firms that did regularly face competition as the untreated group. We find that treated firms decrease advertising levels on brand keywords by 11% relative to their untreated counterparts. The untreated companies show no break in the pre-existing time-trend, whereas treated companies show a marked downward shift the propensity to advertise. The result is significant beyond the 0.01 level and robust to functional form assumptions, changes in the treatment assignment threshold and various other robustness checks.

The relatively modest impact on the propensity to advertise could potentially be explained by firms evaluating their expenditure, getting positive results (contrary to eBay) and staying the course. This leads to the second reaction we study: do firms adopt active experimentation and increasingly evaluate the ad effectiveness for themselves? The “search pause” methodology described in BNT is incredibly easy to adopt because it can be done through the bidding interface with no additional programming. The ease of experimentation was aptly summarized in the popular press and held up as a contrast to the staggering value it offered eBay. We can identify experimentation in our data by looking for sharp changes in advertising levels—indeed we are easily able to identify the dates of eBay’s experiments. In general, sharp changes could reflect active experimentation, or “natural experiments” due to budget exhaustion, churn in marketing personnel and so forth. Since the value of experimentation might not depend on the presence of competing ads, we compare the frequency of experiment-like variation in advertising levels before and after information disclosure. This produces precise estimates that reveal no significant difference, indicating, perhaps surpris-

1 The ads above the organic links account for the vast majority of search engine revenue.
2 Simonov et al. (2015) estimates that competitors on average steal around 18 percent points of clicks when they are in the top position (first paid link), while they steal only 2-3 percent points when they are in position 2–4. Hence, removing own brand ad would on average lead to the loss of 15 percent points of clicks.
ingly, that firms were not moved to adopt this powerful method. A difference-in-differences approach leads to similar point estimates and the same conclusion.

Although experiment-like changes in advertising did not increase, they were nonetheless quite common—72% of the 395 ‘treated’ firms (facing few competitors) had at least one sharp change in advertising levels over the five year period. One possibility is that firms do not appear to react because they were already rationally conducting experiments. In this case, the findings of BNT would not reflect new information for most market participants, and while this runs contrary to the narrative in academic circles and the press, it is nonetheless an important possibility to investigate. We use sharp changes in the firms’ advertising to estimate advertising causal effects at the firm level (we only use data that would be available to advertisers as well). The resulting estimates reveal firm-level heterogeneity in the ad effectiveness and are broadly consistent with the estimates from randomized trials on the same platform (Simonov et al., 2015). In other words, if the sharp changes we observe represent intentional experimentation, some firms got relatively good news from the trials and most got bad news. Strikingly, the magnitude of the ad effect has no predictive power for advertising levels in the last six months of our data, which we use as a hold-out evaluation period. Even firms that possess data to tightly estimate small ad effects do not show a propensity to eliminate own-brand advertising. This stands in contrast to eBay’s reaction to completely shut down brand search advertising. The evidence thus strongly rejects the “already optimizing” hypothesis.

Putting all the pieces together reveals a nuanced picture. First, the negative information on brand search effectiveness led to a reduction in the propensity to advertise on own-branded queries. While this impact is highly significant statistically and economically meaningful, the majority of firms nonetheless continued with business as usual, pointing to substantial “inertia” of business practices. Second, there is no measurable impact on the propensity to conduct experiments. Not only did firms fail to increase experimentation in response to very positive news about the value of experimentation, the insights that can be derived from existing experiment-like variation in ad levels do not appear to impact decisions. This indicates this variation either does not represent intentional experimentation, or there is a principal-agent problem within the firm preventing the gathered evidence from guiding decisions.

Past work has theoretically explored how agency problems can arise when the incentives
of the proximate decision makers diverge from those of the firm as a whole (Scharfstein and Stein, 2000). In our application, it is marketing managers who are tasked with optimizing and evaluating advertising expenditure. While search advertising experiments on own-brand queries are clearly quite valuable at the firm level, at the individual level they appear quite risky and expected to produce “bad news,” making them unattractive from the point of view of the proximate decision maker (by revealing past mistakes, for instance).\(^3\) In contrast, a high-level manager, such as the Chief Financial Officer, has the incentive to learn true advertising effectiveness. Such a manager could read an article in an outlet like the Harvard Business Review, quickly fire off a few tests searches to observe what ads, if any, appear on the firm’s brand queries and, if necessary, order own-brand ads to be shut down. This could be done with the “flip of a switch” and would not require changing the culture of measurement at the firm.

To the best of our knowledge, this is the first paper to quantitatively measure how firms react to scientific findings on business practice. Past work has established that firms within narrowly-defined industries display large differences in productivity per worker that cannot be attributed to differences in capital (Bartelsman and Doms, 2000; Foster et al., 2008)—these differences have been linked to how well firms employ incentive structures, measurement and monitoring in line with established best practices (Bloom and Van Reenen, 2010). So-called best practices, in turn, are often established by within-firm field experiments similar to the one we study in this paper (see Bandiera et al. (2011) for a review).\(^4\) The large and growing body of field experiments of this sort have addressed topics such as how managerial practices can improve supply chain efficiency (Bloom et al., 2013),\(^5\) evaluating performance-based pay (Lazear, 2000; Shearer, 2004), manager incentives (Bandiera et al., 2007), the interaction of social preferences and incentives (Bandiera et al., 2005; Ashraf et al., 2014) and optimally running auctions (Ostrovsky and Schwarz, 2011).

Perhaps the closest related work is in health economics. Evidence suggests that doctors’ choices among competing treatment options are broadly consistent with rational information processing with knowledge spillovers (Chandra and Staiger, 2007). Fiedler (2013) argues that

---

\(^3\)Reporting incentives have been previously studied theoretically. An example is the persuasion game of Shin (1994). The analog here would be a marketing manager opting to continue to report nominal CPC and CTR, not incremental traffic induced by the ad, or selectively report experimental results.

\(^4\)Bandiera, Barankay and Rasul (2011) note that such experiments often have a large impact on profits when the experimental findings are implemented by the firm.

\(^5\)Particularly relevant, the authors conclude that changing incorrect beliefs about important practices is the main driver of the effect.
new treatment adoption is consistent with weighing the evidence in published papers and from a doctor’s personal experience. In contrast, the knowledge spillovers arising from the public disclosure of the BNT findings are present but modest in size and incomplete on some dimensions. The differences between these two setting can help illuminate the root causes of the learning dynamics we observe. In particular, in medicine there is much less of a concern about agency and incentive problems—new techniques can be good for patients, doctors and hospitals alike. Further, doctors are highly trained to process technical information about the efficacy of new treatment options, actively contribute to the academic literature and hospitals are themselves often institutions of higher learning.

As the case of medicine shows, our findings certainly do not extend to situations where principal-agent problems are not present. For example, they would not (necessarily) apply to the adoption of a newly invented technology. However, many advances in business practice are not the result of new inventions, but rather new approaches to solving existing problems. It is commonly argued that agency problems, such as those we highlighted for advertising, are generally present within large firms because the incentives of workers and the firm generally diverge and divisions within a firm often have different incentives from the broader organization (e.g. Scharfstein and Stein (2000) and Lazear (2000)). If so, our results may point to a broader drag on economic efficiency and could be an importance factor in explaining the wide variation in worker productivity in narrowly defined industries.

The paper proceeds as follows: Section 2 describes the economic setting and the data, Section 3 gives results for the propensity to advertise and experiment, Section 4 estimates advertising effects at the firm-level and examines the correlation of these estimates with companies’ post-experimentation behavior, and Section 5 concludes with a broader discussion.

2 Economic Setting and Data Description

In sponsored search, advertisers bid on keywords to have their ad displayed at the top of search engine results. An advertiser specifies keywords that determine which queries a given ad will be entered into a real-time Generalized Second Price auction. The search engine sets parameters such as the reserve price, relevancy requirements for the ad and the maximum number of allowable ads that appear above the organic listings (currently four for Google and
“Brand search” advertising occurs when an advertiser bids on keywords that include a brand name. Figure 1 gives an example for the department store Macy’s advertising on the keyword “macys.” Macy’s paid link occupies the top position on the page and a competing advertisement occupies the second slot. Importantly, Macy’s also appears in the first organic position on the page. All clicks on the organic results are free.

Brand search is an attractive domain to study the effect of the information disclosure for a few reasons. First, BNT report that the potential cost savings are large—eBay management decided to discontinue expenditure of over fifty million dollars per year. Second, the inference problem is challenging because “nominal” metrics of ad performance, such as CTR and CPC, tend to overstate the actual advertising effect (Lewis et al., 2015), creating a large wedge between the reported price (CPC), which is typically very low for the focal brand low due to the high relevance, and real “cost per incremental click.” Since these real metrics are

---

6The pricing rule of the GSP rewards high CTR ads with low CPCs, which makes sense since the opportunity cost of the search engine is impressions, see Edelman et al. (2007) for more details. Simonov et
not directly observable, firms must conduct experiments to estimate them, just as eBay did. Third, published evidence on the effectiveness on own brand keyword was limited until the release of the eBay study.\footnote{A quick web search reveals practitioners guides that warn of the problem of click crowd-out and others that recommend advertising on own-brand keywords. On the academic side, until March 2013, there was only one paper examining the interdependence of paid and organic traffic Yang and Ghose (2010), which used observational data to conclude, counter-intuitively, that clicks on paid links actually increased clicks on organic links (crowd-in).} Finally, at the time BNT was released firms had heterogeneous behavior with respect to own-brand search: roughly half of reasonably large firms that otherwise used search advertising chose to advertise on their own keywords and the other half chose not to (Simonov et al., 2015).

### 2.1 Blake, Nosko and Tadelis (2015) and Media Coverage

Blake, Nosko and Tadelis (2015) reported the results of a large field experiment performed in which eBay shut down their brand keyword advertising for several weeks of spring 2012. The results revealed search ads on queries containing “eBay” had almost no effect on traffic to eBay’s website. The paper was released as a working paper at the beginning of March 2013 and was quickly picked up by the business press. A detailed write-up appeared in *Harvard Business Review* on March 11, 2013, with the provocative headline shown in Figure 2 and was quickly followed by the coverage in the BBC (March 13), *Business Insider* (March 14), various smaller outlets and leading search engine blogs. *The Atlantic* and *The Economist* ran feature stories in April and July 2013 respectively. For the purposes of the information disclosure data, we use March 11, 2013. While the coverage often started with an attention grabbing headline, the articles generally gave a faithful description of the study and reasonable recommendations for business practice. BNT had a large impact in academic circles as well, securing publication in *Econometrica* and quickly racking up citations.

BNT ran their experiment on Bing and Google, meaning the “traces of experimentation” should be observable in our data and indeed they are. Figure 3 shows the frequency which eBay advertised on their own brand keywords over time. Since the authors provided a detailed description of the experiment and subsequent reaction of the company, this Figure does not reveal anything new. It does confirm that prior to Q2 2012 eBay regularly advertised
Figure 2: Harvard Business Review article on the eBay search experiment. The story was published March 11, 2013 (hbr.org/2013/03/did-ebay-just-prove-that-paid).

on their keywords. The experiment is visible just at the time period reported by BNT, after which there was a return to business as usual before the ads were discontinued in 2013, outside a brief period of what appears to be a follow-up experiment. This check validates that our data can capture experimentation behavior and is entirely consistent with the descriptions of an experiment from the advertiser.

2.2 Data Description

We start with 87,000 brand names of online retailers taken from the Open Directory Project. We aggregate historical search logs at the daily level from October 07, 2011, to May 31, 2015, which both anonymizes the data and reduces it to a practical size. The resulting sample consists of 85,725 brands searched at least once during this period. We further limit to firms that are (1) searched consistently, (2) are in the top organic position on their own-brand queries, (3) advertise on their brand keywords at least some of the time and (4) firms that “want the traffic” (that is, the brand is not generally sold through a reseller). We thus further reduce the sample by (1) keeping companies which are searched for at least 20 times on each day in our sample (left with 6,258 companies), (2) keeping companies which are in organic search position 1 more than 90% of the times (left with 1,861 companies), (3)
keeping companies which advertise in top advertising position more than 90% of the time on at least one day in the sample (left with 1,234 companies) and (4) keeping companies that get at least a 50% combined CTR from the ad and organic links (left with 1148 companies).

Table 1 gives summary statistics. An average brand gets more than three million searches over the time period we study. There is more than one mainline ad on average, with a standard deviation of 0.9. This implies our data has a lot of variance in a number of ads shown in the advertising slots above the sponsored positions. We further see that there is variation in both percents of the own brand ad in mainline 1 occasions and competitor ads in mainline 2-4 occasions. The percent of search occasions leading to own-brand click is around 75 percent, which is consistent with previous results in Simonov et al. (2015).

3 Estimation and Results

We start by studying the impact this publication had on firms’ propensity to engage in own-brand advertising. Our identification comes from the observation that companies which face
### Table 1: Data Summary

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Median</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposures</td>
<td>3,262,494</td>
<td>575,711</td>
<td>16,969,288</td>
</tr>
<tr>
<td># of mainline ads</td>
<td>1.67</td>
<td>1.41</td>
<td>0.9</td>
</tr>
<tr>
<td>% of times own ad in mainline 1</td>
<td>71.35</td>
<td>81.62</td>
<td>26.2</td>
</tr>
<tr>
<td>% of times competitor’s ad in mainline 1</td>
<td>12.37</td>
<td>6.93</td>
<td>14.82</td>
</tr>
<tr>
<td>% of times competitor’s ad in mainline 2</td>
<td>38.14</td>
<td>30.9</td>
<td>27.53</td>
</tr>
<tr>
<td>% of times competitor’s ad in mainline 3</td>
<td>25.47</td>
<td>13.39</td>
<td>27.4</td>
</tr>
<tr>
<td>% of times competitor’s ad in mainline 4</td>
<td>16.77</td>
<td>5.16</td>
<td>23.67</td>
</tr>
<tr>
<td>% of search with navigation to own brand website</td>
<td>74.60</td>
<td>77.08</td>
<td>8.85</td>
</tr>
<tr>
<td>- through organic position 1</td>
<td>46.84</td>
<td>46.11</td>
<td>14.6</td>
</tr>
<tr>
<td>- through mainline 1 ad</td>
<td>27.76</td>
<td>29.42</td>
<td>14.3</td>
</tr>
<tr>
<td>Number of observations</td>
<td>1,525,692</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of days</td>
<td>1,329</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of companies</td>
<td>1,148</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Means are unweighted

The same level of competition as eBay—having no competitors advertising on their keyword—should be affected by this information, while companies which generally face competing ads are not directly affected. The reason is simple: when a firm faces competing ads, if it kills its own-brand ad, a competitor will supplant it at the top of the page. Simonov et al. (2015) show that in such cases, competitors can siphon off 20% of clicks on average, an order of magnitude higher than the average causal impact of own-brand ads without competitors present. This implies that in this case removing own brand ads would lead to losses of clicks far higher than that reported by BNT, who are careful to highlight the critical importance of competing ads (or the lack thereof in their case).

Figure 4 shows the distribution of firms over the frequency of facing a competitor in slot 2. There are companies for which competitors rarely advertise in Mainline 2 and companies with competitor frequently advertising in Mainline 2. We define companies which face competitor’s ad in Mainline 2 less than 20% of the time as companies with low competition level, and companies which face a competitor in Mainline 2 more than 80% of the time as companies with high competition. The resulting group with the low level of advertising competition (which we refer to as “treatment”) contains 395 companies, and the group with high level of advertising competition (“control” or “untreated”) contains 145 companies.
3.1 Verifying Diff-in-diff Identifying Assumptions

We start with examining the validity the identifying assumptions for our “treatment” and “control” groups. The design is valid provided there was not a confounding factor around the time of information disclosure that differentially (by group) affected the propensity to conduct own-brand keyword advertising. Since we define groups using the level of competition, we examine if the extent of competition changed around the time of disclosure and the time trends in each group leading up to disclosure.

Figure 5 plots the frequency of competitor ad in Mainline 2 for groups of companies with high and low competition levels. We fit a local polynomial\(^8\) to assess the degree of change, which is given in Panel A. The frequency of competitors advertising in Mainline 2 in “treatment” case is decreasing in the middle of 2012, but is stable at the time of publication, which made clear in panel (b), which models just the differences in the propensity of groups to face competition. Further, the absolute changes over time in both cases are minimal.

\(^8\)LOESS of second degree
The parallel time-trend assumption is examined in Figure 6, with the difference in the series given in the panel (b). Examining the period before information disclosure (the region left of the vertical line), the time trends are both increasing at nearly the same rate. Note the level differences are expected—firms facing competition are more likely to engage in own-brand advertising as the returns are higher in this case. Finally, the lack of a meaningful pre-period time trend is shown clearly panel (b), which plots the differences and reveals a slight downward trend, the impact of which is entirely swamped by the treatment effect we find later.

Overall the identifying assumptions for the difference-in-differences estimation hold nearly exactly and where they do not, the small divergence is not large enough to generate significant treatment effects.

3.2 Effect on Advertising Levels

Figure 6 plots the frequency of own-brand advertising in slot 1 for companies with low competition (blue) and companies with a lot of competition (red). The frequency of advertising in the top slot increases in both groups before the information is revealed, but the low level
of competition group flattens out after the publication. The plot on the right presents the difference between groups with low and high levels of competition, and non-parametrically illustrates a significant treatment effect that takes roughly one year to fully take hold.

Figure 6: Own advertisement level for “treatment” and “control” companies: average level

![Graph showing advertisement levels for treatment and control companies over time.](image)

Fraction of times own brand ad is shown in Mainline 1 for companies facing and not facing competitors in Mainlines 2-4. Daily data

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>-0.1351</td>
<td>-0.1351</td>
<td>-0.3944</td>
<td>-0.3944</td>
</tr>
<tr>
<td></td>
<td>(0.0281)</td>
<td>(0.0282)</td>
<td>(0.0144)</td>
<td>(0.0144)</td>
</tr>
<tr>
<td>After</td>
<td>0.0926</td>
<td>0.1175</td>
<td>0.0926</td>
<td>0.1175</td>
</tr>
<tr>
<td></td>
<td>(0.0188)</td>
<td>(0.0256)</td>
<td>(0.0188)</td>
<td>(0.0255)</td>
</tr>
<tr>
<td>Treatment * After</td>
<td>-0.0679</td>
<td>-0.0679</td>
<td>-0.0678</td>
<td>-0.0679</td>
</tr>
<tr>
<td></td>
<td>(0.0237)</td>
<td>(0.0237)</td>
<td>(0.0237)</td>
<td>(0.0237)</td>
</tr>
<tr>
<td>Query FE</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Day FE</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
</tbody>
</table>

Robust standard errors clustered on query level.
All coefficients presented are significant at 1% level

We use a difference-in-differences estimator to provide formal tests of this effect, incorporating various controls for seasonality and query fixed effects. The results are presented in Table 2. The estimated impact is 0.067 percent points (averaged over the entire post-period) is very stable across specifications. This estimate is significant beyond the 0.01 level. Given
that average level of advertising for companies not facing competition is 0.626, this decrease corresponds to 10.8% change in advertising levels.

This change could come from two qualitatively different reactions, which at the extremes are given by: 1) all companies reduce their advertising level by 11% 2) 11% of companies completely turn off their advertising. Indeed this distinction offers a robustness check for the results. If decreases in the level of advertisement are due to the effect of information, then we would expect the latter pattern—the ads are simply turned off, just as eBay did. On the contrary, if pattern (1) is observed, then it is difficult to reconcile as a reaction to the news on DNT.

Figure 7: Own advertisement level for “treatment” and “control” companies: fraction of companies advertising less than 20% of the time

![Graph](image)

(a) Levels

(b) Difference

The fraction of companies with own brand ad in Mainline 1 < 20% of the time. Daily data

We investigate these possibilities by measuring the number of occasions that companies advertise on their own query on less than 20% of searches in the week (advertising “turned off”) in both groups. If companies indeed stop advertising, we should find that there are more companies without competition that move to the no-advertising state after the publication compared to the companies which face competition. Figure 7 presents levels and difference of the share of companies not advertising on their own keywords. The share of companies not advertising on their own keywords is decreasing in both groups before the publication. After the publication, among companies which do not face competition this quantity stabilizes, while among companies which do face competition the share not advertising keeps declining.
The difference between treated and untreated firms is given in the right panel of Figure 7, which shows the clear relative uptick in not advertising for the treated firms.

Table 3: DiD regression on occasions of not advertising

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>0.829</td>
<td>0.833</td>
<td>4.4732</td>
<td>4.514</td>
</tr>
<tr>
<td></td>
<td>(0.1871)</td>
<td>(0.1877)</td>
<td>(0.1215)</td>
<td>(0.124)</td>
</tr>
<tr>
<td>After</td>
<td>-0.9077</td>
<td>-1.0149</td>
<td>-1.3261</td>
<td>-1.5468</td>
</tr>
<tr>
<td></td>
<td>(0.1879)</td>
<td>(0.2149)</td>
<td>(0.2743)</td>
<td>(0.3389)</td>
</tr>
<tr>
<td>Treatment * After</td>
<td>0.7755</td>
<td>0.7736</td>
<td>1.0751</td>
<td>1.0768</td>
</tr>
<tr>
<td></td>
<td>(0.2038)</td>
<td>(0.2042)</td>
<td>(0.3125)</td>
<td>(0.3165)</td>
</tr>
<tr>
<td>Query FE</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Day FE</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
</tbody>
</table>

Robust standard errors, clustered on query level.
All coefficients presented are significant at 1% level
Link function is logit

To formally test the significance of this visual evidence we use a difference-in-differences logistic regression with “not advertising” (=1) as the binary outcome variable. The results are presented in Table 3. The effect for treated firms is significant effect beyond on the 0.01 level. Based on specification (4), the corresponding average marginal effect of information on probability of not to advertise is 10.4%. This indicates that nearly all of the advertising level effect we previously observed is driven by companies discontinuing own-brand ads, consistent with an informational story.

3.3 Effect on Experimentation Levels

In the previous, section we showed that companies reacted to new information by reducing advertising levels, but the number of firms reacting was rather modest given the strength of the information. One natural concern a firm could have is that there is heterogeneity in ad effectiveness across companies, so it would be unwise to take eBay’s results as rep-

9For robustness, we also check another measure of the level of advertisement. Figure 13 in the Appendix plots the fraction of companies with own brand ad in Mainline 1 > 90% of the time for “treatment” and “control” group. Results are the same.
resentative of their returns on brand keywords.\textsuperscript{10} Given the ease of implementing BNT’s experimental method—the baseline version can be done by just pausing the ad through the bidding interface with no programming required—the rational response would then be to copy the protocol and learn one’s own ad effect. We now examine if firms indeed reacted in this fashion.

As shown in Figure 3, intentional experimentation is characterized by large, sharp changes in advertising levels. These changes, however, do not necessarily represent experiments, as they could be due to budget shortfalls, change in personnel that handles the ad buying and so forth. Accordingly, this “experiment-like” variation is a necessary but not sufficient condition for intentional experimentation. We start by looking at the amount of experiment-like changes in advertising before and after the experiment for companies with high and low level of advertising competition. If companies similar to eBay started to experiment more, we expect the total amount of experimentation to change. In this subsection we focus on weekly level data, to avoid variation in the advertising level due to weekdays and weekends. We define experiment-like changes as a 50 percentage point difference in advertising levels week-on-week (our results are not sensitive to the precise definition).

Figure 8: Fraction of companies changing their advertisement level by more than 50%

\textsuperscript{10}Simonov et al. (2015) show that it is indeed true that effect of companies ad on its own brand query is different for different types of companies; in particular, the effect is larger for companies with less brand capital.
Figure 8 plots the weekly propensity of experiment-like variation over time. The average propensity is 2% per week, and the time trends reveal that there is no detectable increase in the experiment-like changes after BNT was released. A difference-in-differences specification confirms there is no significant differences for the treatment group relative the untreated group and a simple before-and-after estimator reveals no significant differences for either group.

Although we do not observe an uptick in the experiment-like behavior, large changes in advertising week-on-week are not entirely uncommon (2% of weeks, or about once-a-year per firm). Out of 395 companies that do not face competition, more than 71% have changed their advertisement level by more than 50% at least once in the observed period. One reason a firm may adjust their advertisement levels is the existence of demand shocks (e.g. holiday shopping period or product releases). If expected demand changes affect the level of advertising, we should find a strong correlation between the change in the number of total query exposures in a given week and change in the advertising rates. The correlation between percent change in exposures and change in advertising rates that we find explains less than 0.1% of variation of changes of advertising levels. This allows us to conclude that demand factors are not the primary reason for the changes in advertisement levels that we observe.\textsuperscript{11}

The results here indicate that there was no experimentation uptick but that a majority of firms nonetheless had large changes in advertising levels suitable for estimating causal effects. This either reflects widespread experimentation prior to BNT or natural experiments due to the supply-side mechanisms (e.g. running of out budget). Although the narrative of coverage of BNT and the impact it has had in academic circles was around the innovative nature of the experiment and findings, it’s possible, albeit unlikely, that most firms had been conducting these experiments all along. This would explain the muted reaction in terms of advertising levels and the lack of a reaction in terms of experimentation levels. In the next section we examine if these sharp changes in the advertising levels were indeed active experimentation by computing the implied advertising effectiveness and examining if firms changed their advertising levels in response.

\textsuperscript{11}Recall that all results for this section are based on weekly data. If we use daily data, results would be different in this case: we find that demand changes explain around 2% of all advertising level changes in daily data. This is explained by some firms treating weekdays to weekends (lower query volume) differently.
4 Inferences from Experiment-like Variation

If the experiment-like variation we observe indeed represent intentional experiments, rational firms that find large effects should decide to keep advertising and those that find small effects of advertising should stop advertising. Here we examine this relationship. We start by computing advertising effects for companies which have the requisite variation and then examine their post-experimentation behavior in light of these estimates.

4.1 Firm-level Advertising Effects

Our dependent measure is total traffic to the brand’s website from the sponsored and organic links. Figure 9 presents the histogram of probabilities of getting to brand’s website for companies that do not face competition. On average, 80% of clicks go to own brand website.

Figure 9: Histogram of probability to get a click on own brand query across companies

![Figure 9](image.png)

Weekly data

Figure 10 presents similar histograms for occasions when companies in the sample advertise more than 90% in a given week and when they advertise less than 10%. The distribution
is shifted to the right during advertising spells and this difference is significant (Kolmogorov-Smirnov, p-value = 2.58e-12). During advertising spells, the average probability of a click is 81.5%, and on occasion without advertising it is 78%, implying a 3.5 percent point difference between advertising and no advertising cases. This simple comparison reveals a correlation between advertising and probability to get clicked, which could be due to the causal effect of the ad or endogenous timing of advertising. Simonov et al. (2015) estimate the causal effects of own-brand ads on Bing using full randomized experiments and find an average effect of advertising on probability to be clicked of 1.4-2 percent points depending on firm size. This indicates that our simple comparison overstates the true effect of advertising by about 100%.

Figure 10: Histogram of probability to get a click on own brand query across companies, by level of advertising

Given this bias, we impose additional restrictions to better identify ad effects with experiment-like variation. To do so, we use only weeks on the border of the large changes in advertising, before and after ads are turned on/off, to reduce any co-varying factors which change over time. We emphasize that this is not as good as fully-randomized experiment as there can be different demand shocks in two subsequent weeks, but we will be able to compare
these estimates to the ground truth from fully randomized trials on the same platform. To control for differences in levels across the “experiments” we use changes in advertising levels and probability of click. This gives us 1081 experimentation occasions for 282 companies.

Table 4: Effect of advertising on probability to get a click

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advertising Level</td>
<td>0.0295</td>
<td>0.0297</td>
</tr>
<tr>
<td></td>
<td>(0.0022)</td>
<td>(0.0028)</td>
</tr>
<tr>
<td>Query FE</td>
<td>N</td>
<td>Y</td>
</tr>
</tbody>
</table>

Robust standard errors, clustered on query level
All coefficients presented are significant at 1% level

Table 4 gives the results. On average, advertising increases the probability of getting a click by 2.95 percent points. This is a bit higher than results in Simonov et al. (2015), which could be due confounds, or to the fact that a set of companies used are different.\textsuperscript{12} We examine the heterogeneity of advertisement effects across companies by running the regressions at the firm level. Since the majority of companies experiment only once or twice, the power of this estimation is limited. However, it is still fair to treat this estimate as a firm’s best estimate of its advertising effectiveness, as we have been careful to use only data available to firms as well.

Figure 11 presents the distribution of company-specific advertising effects. The distribution is skewed to the right, and a battery of statistical test reject that it is Gaussian (i.e. it is not what we’d expect due to sampling variation alone).\textsuperscript{13} This implies that advertising effect coefficients contain information about ad effectiveness, and companies with higher estimated coefficient should on average have higher effectiveness for advertisement. In other words, some companies got good news and some got bad news.

\textsuperscript{12}We know there is heterogeneity across companies from Simonov et al. (2015)
\textsuperscript{13}To conduct these tests, we normalize the distribution by mean and standard deviation, and perform a series of test: Shapiro-Wilk, Jarque-Bera, D’Agostino, Lilliefors, etc.
Figure 11: Histogram of advertising effects on probability to get a click, across companies
4.2 Post-Experiment Behavior

If companies are using inferences from the experiments (rationally), then they should be more likely to stop advertising when advertisement effect is small, and to keep advertising when the effect is large. On average, the experimental estimates are small relative to the putative ad effects given by nominal metrics (e.g. CTR or CPC), so we might also expect that most firms with experiment-like variation to discontinue advertising as eBay did.

We separate last five months in our data and use it as the post-experimentation period and use estimates from experiment-like variation in prior periods to predict advertising levels in the holdout period. Figure 12 presents the relationship of estimated effects (y-axis) and their decisions to advertise in the post-experimentation period. Companies which do not advertise in the post-experimentation period are on the left and companies which advertise are on the right. There is no detectable correlation with estimated advertising effect, which can be seen visually and is confirmed with formal tests.

Figure 12: Relationship between estimated effects for companies and their decision in the post-experimentation period
We can thus deduce that firms are not acting on the inferences that can be made from the experiment-like variation we observe, making it unlikely they are true experiments. As a check of this explanation, we re-run the analysis using only cases where companies did not change the advertising level for at least three weeks before and after the sharp change. Example of such behavior is presented in Appendix (Figure 15). This behavior is more consistent with experimentation of eBay (Figure 3). We find similar results as above in this sample of 97 companies.\textsuperscript{14}

We cannot rule out the possibility that the proximate decision makers are conducting the right analyses, but the results are not faithfully communicated to upper management. This could be because there is an incentive to increase paid traffic generally (or digital ad budgets). If aggregate reports featuring CPC are used, including low CPC paid traffic from own-brand queries decrease the global average. Framed in this way, it is clear that classical principal agent problems could be at play.

5 Discussion and Conclusion

Although learning by firms is central to many aspects of economic efficiency, quantitative evidence on when and how they learn and why they do not in some cases, is thin. One reason for this shortage is the steep data requirements for a convincing, broad analysis. We overcome this hurdle with detailed data on firms’ choices and the fortuitous timing of a particularly impactful academic paper, which received as much attention in the popular press as an economics paper can ever hope to, containing specific and actionable information relevant to these choices. Based on these factors, we might have expected a strong response by firms, yet our results reveal a general sluggishness to react. While we do find a significant propensity to reduce own-brand advertising for firms directly impacted by the information, only a minority of firms did so. Further, we find no evidence that firms adopted the powerful (and easy) method of experimentation the paper advocated, nor do they appear to respond to inferences they could make using sharp changes in historical advertising levels (these are either natural or intentional experiments). The fact that we find a response in terms of ad

\textsuperscript{14} Figure 14 in the Appendix shows the relationship between estimates of advertising effect of probability to get a click and post-experimentation decision. Also, Figure 15 corresponds to the company for which advertising effect is estimated to be negative. As we can see, company keeps advertising in post-experimentation period.
levels but not experimentation point to the difficulty in moving beliefs about causal inference more broadly.

There are two related, and not mutually exclusive, mechanisms that could explain our findings. The first is that the relevant personnel are not paying much attention to business journals, business articles in the popular press, search engine management blogs and other sources of information to help optimize choices. The second is that risk aversion, the risk of punishment for past behavior and related principal-agent problems hinder the evolution of best practices at the firm. Although the policy implications to address these two mechanisms differ, they are both a product of incomplete incentives and is unclear which is “worse.” If the relevant decision makers did not come across actionable information that received widespread attention in the media, in addition to the buzz in marketing circles and technical blogs, then this is quite concerning since most advances do not receive this kind of attention and are not as straightforward to implement. On the other hand, if employees do in fact have up-to-date information, but are not incentivized to act on it, this may speak to deeper problems within the firm.

It is important to keep mind that the business practices we study are in a domain, advertising, for which past work suggests principal-agent problems may be particularly large. The effects of advertising are hard to measure and nominal metrics can distort the real impact of campaigns (Lewis and Rao, 2015; Lewis et al., 2015). Marketing managers are often assigned to a certain class of media, e.g. television or digital, and may face negative personal consequences if disappointing effectiveness results are found in their domain. Viewed in this light, BNT represented good news for a firm advertising on own-brand keywords—it may be able to save millions of dollars just as eBay did. But it probably represented bad news for the people making these decisions “on the ground” because past expenditure could be revealed as wasteful, budgets cut may follow, loss of positions and so forth. Seen from this angle, new information is risky, which also helps explain why we do not observe increases in experimentation.

It is certainly not the case that all business practices are subject to agency problems of this nature. Medicine is perhaps at the other extreme. Doctors actively contribute to the academic literature and published advances generally benefit doctors, hospitals and patients alike. Indeed, there are “evidenced based medicine” journals that have the mission to ensure the new methods are quickly adopted (Haynes et al., 2002). In addition, doctors are also
highly trained to process technical information about the efficacy of new treatment options and evidence suggests that their choices among competing treatments are broadly consistent with rational information processing with knowledge spillovers (Chandra and Staiger, 2007; Fiedler, 2013). While our findings do not extend to such cases where the incentive structures are entirely different, other authors have argued that the agency problems we discuss for advertising are generally present within large firms (e.g. Scharfstein and Stein (2000)). If so, our results may point to a broader drag on economic efficiency and could be an important factor in the wide variation in worker productivity observed in narrowly defined industries.

References


6 Appendix

6.1 Press Coverage of Blake, Nosko and Tadelis (2015)

Major press includes:


*Business Insider*, 3/14/2013 EBay Slams Google Ads As A Waste Of Money

*BBC*, 3/13/2013. Google advertising value questioned by eBay.

*The Economist*, 7/13/2013. Simple tests can overstate the impact of search advertising.

*The Atlantic*, 4/13/2014. A dangerous question: Does internet advertising work at all?
6.2 Probability to advertise more than 90% of the time

Figure 13: Own advertisement level for “treatment” and “control” companies: fraction of companies advertising more than 90% of the time

(a) Levels
Fraction of companies with own brand ad in Mainline 1 > 90% of the time

(b) Difference
6.3 Estimated effects versus post-experimentation behavior

Figure 14: Relationship between estimated effects for companies and their decision in the post-experimentation period. Done for companies with similar experimentation patters as in the case of eBay.
6.4 Example of a company that experiments

In Figure 15 we show an unnamed firm’s advertising levels on own-brand keywords. This firm displayed a pattern of experimentation in the end of 2011 and beginning of 2012. Based on this experimentation, we find that advertising effect small and statistically insignificant. Nevertheless, this firm continued to advertise after the “experimentation” period.

Figure 15: Company X advertising frequency over time