Down-to-the-Minute Effects of Super Bowl Advertising on Online Search Behavior

RANDALL A. LEWIS, Google, Inc.
DAVID H. REILEY, Google, Inc.

Online consumer data presents new opportunities for measuring the effects of advertising. We combine search query data with the television commercial schedule for the 2011 Super Bowl to measure the causal impact of TV advertising on consumer search behavior. Examining 46 Super Bowl commercials, we generally find large spikes in search behavior related to the advertiser or product within 15 seconds following the conclusion of the TV commercial. We present results for four categories of Super Bowl advertisers: movies, online services, cars, and other consumer goods. We also consider the economic and statistical feasibility of scaling this analysis to use search data to measure the impact of other TV commercials.

Categories and Subject Descriptors: **H.3.5 [Information Storage and Retrieval]**: Online Information Services; **H.5.4 [Information Interfaces and Presentation]**: Hypertext/Hypermedia; **J.4 [Social and Behavioral Sciences]**: Economics; **J.1 [Administrative Data Processing]**: Marketing.

General Terms: Management, Measurement, Economics, Human Factors

Additional Key Words and Phrases: Advertising effectiveness; television commercials; search queries

ACM Reference Format:

Randall A. Lewis and David H. Reiley, 2013. Down-to-the-Minute Effects of Super Bowl Advertising on Online Search Behavior. *14th ACM Conference on Electronic Commerce*, 9, 4, Article 61 (June 2013). DOI:<http://dx.doi.org/10.1145/0000000.0000000>

1. INTRODUCTION

Television advertising continues to represent an enormous market, estimated at over \$200 billion annually in 2012; for comparison, this is still more than twice the size of the entire market for online advertising [ZenithOptimedia 2012]. Furthermore, consumer behavior exhibits deep connections between the online and offline media worlds, with consumers often using laptops or tablets while watching television, for example. We take advantage of these connections in order to measure the causal impact of TV advertising on consumer searches for the advertised brands. We show that high-frequency search data, readily accessible for all major advertised brands, enable clear measurements of consumer behavior caused by television ads.

Advertisers need to know how effective their advertising campaigns are at engaging consumers and thereby boosting sales. However, measuring the causal effects of advertising has been a very difficult problem with sparse evidence to date. One of the biggest hurdles is linking data on ad exposure and purchase behaviors. Early work [Abraham and Lodish 1990; Lodish, *et al.* 1995a,b; Hu, Lodish, and

Author's addresses: R. Lewis, randall@econinformatics.com, and D. Reiley, david@davidreiley.com. Both at Google, Inc., 1600 Amphitheatre Parkway, Mountain View, CA 94043.

Permission to make digital or hardcopies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credits permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or permissions@acm.org.

© 2010 ACM 1539-9087/2010/03-ART39 \$15.00

DOI:<http://dx.doi.org/10.1145/0000000.0000000>

Krieger 2007] studied the effects of hundreds of television commercial campaigns on purchases by using panels of thousands of customers whose commercial exposure they could manipulate through a split-cable TV signal. Their work, while finding aggregate evidence for TV advertising influencing sales, produced relatively imprecise, statistically inconclusive estimates for any given individual advertising campaign. Our previous work studying online display advertising [Lewis and Reiley 2012a,b; Lewis, Reiley, and Schreiner 2012; Johnson, Lewis, and Reiley 2012] required randomized experiments with sample sizes of over a million customers in order to begin to get statistically significant estimates of the effectiveness of individual online display ad campaigns on both online and in-store purchases by consumers. Detecting the effects of advertising on purchase behavior turns out to be like looking for a needle in a haystack. Even with very large-scale experiments, statistical imprecision limits our ability to make actionable recommendations to advertisers. For example, Lewis and Reiley [2012a] estimated a return on investment in excess of 100% for the advertiser, but the confidence intervals were sufficiently wide to make it difficult to reject the null hypothesis of no effect. Lewis and Rao [2013] expounds on the sources of this restrictive imprecision: sales are generally quite noisy due to their “perfect storm” of high variance and rare occurrence.

Because purchase data can be difficult to link to individuals’ ad exposure, advertisers often work with proxies to measure ad effectiveness, particularly using consumer survey questions measuring recall, awareness, and affinity for the advertised brand. Since these survey responses are rather distant from the purchase outcome, we find it more desirable to work with actual consumer behavior whenever possible. In particular, search-engine queries have become a key part of the purchase process for many consumers—potentially a “leading indicator” of purchase. Unlike surveys, they measure voluntary consumer behavior—a consumer actually takes the time and attention to do a query. As such, search queries represent a potentially valuable proxy for the purchase outcome.

While purchases due to advertising typically are recorded within days or weeks after exposure to the advertising, search data show clear causal effects within minutes or seconds. Examining a shorter period of time after exposure enables us to eliminate a lot of relatively uninformative variance, thereby increasing the signal-to-noise ratio of the causal effects on advertising. Relative to sales, online searches are a more attractive outcome measure for the causal effects of advertising for two reasons: (1) increased precision and (2) ubiquity of data, thanks to the fact that search engines are continuously logging timestamped data on brand-name searches by consumers.

In previous work, Lewis, Rao, and Reiley [2011] and Lewis and Nguyen [2012] made use of web searches as outcomes in measuring the effectiveness of online display advertising. By using higher-frequency search data instead of lower-frequency purchase data, these studies have achieved much more precise (i.e., statistically significant) estimates of causal effects of advertising on consumer behavior. Previous researchers have also used online search data in observational studies of the effects of television. Joo, Wilbur, and Zhu [2012] used publicly released AOL search data (at much lower frequencies than our down-to-the-minute effects) to demonstrate that online searches can also be used to measure the impact of television advertising; Joo, *et al.* [2013] replicates, strengthens, and extends the AOL analysis with similar, but much higher volume Google search data. We have adapted their idea to measure the effects of the highest-profile advertising on television: advertising during the Super Bowl. Our dataset consists of all searches on Yahoo! for brand names advertised during Super Bowl XLV on February 6, 2011. Other

researchers have performed similar analyses using searches during live sporting events. Oldham [2012] performed an analysis for game-related searches (as opposed to advertising-related searches) during the 2012 Super Bowl; Zigmond and Stipp [2010, 2011] studied similar effects for a select group of advertisers during the Olympic Games but used much lower-frequency outcome data.

Our most important contribution to this literature is a demonstration that high-frequency data (periods shorter than a minute) combined with short post-exposure event windows provide an extremely strong case for causal effects, as opposed to mere correlation.¹ We find these results quite striking. These techniques also provide increased statistical significance, with confidence intervals as narrow as 10% of the estimated incremental queries due to advertising.²

Our second contribution is to demonstrate significant heterogeneity in consumer search responses to TV commercials across a broad spectrum of large advertisers during the same event. Some advertisers experience a large volume of brand-related searches in response to their TV commercials while others see little to no change.

Having demonstrated feasibility of measuring TV ad effectiveness in search data, our third contribution is to derive economic limits on the feasibility of making similar measurements for less expensive commercials. We focus on statistical significance as a necessary condition for causal inference. Specifically, we outline the number of lower-budget commercials required to achieve the same statistical significance as the Super Bowl commercial and, alternatively, the lowest-cost commercial whose search lift is statistically distinguishable from zero. We calculate a threshold likely to be met by only the most expensive network and local-market TV commercials.

We use a simple before-after comparison to infer causal effects of the “natural experiments” of the TV commercials. While we would normally prefer to run a randomized experiment (see Lewis and Reiley [2012a,b] for examples), TV broadcasting currently lacks the technology required to control commercial exposure at the level of the individual or geographic market, which renders such experiments impossible. However, in this case, with the ability to observe consumer behavior on very short timescales, we believe the before-after analysis will produce results very similar to those that would be generated by a perfectly randomized experiment.

The time-series graphs we explore tell the entire story. Super Bowl ads affect consumer behavior in a meaningful, measurable way, stimulating brand-related search queries. If one did not know the commercial schedule during the Super Bowl, one could easily tell from the graphs exactly what minute each ad aired on television. The high-frequency nature of the data and the huge changes in search volume at exactly the minute the ad aired both make it unambiguously clear that the spikes are causal effects. In general, one has to be careful about non-causal correlations between advertising and consumer behavior, such as when an advertiser’s advertising and sales both increase during the holidays. However, as we will describe in more detail below, it is much harder to devise a non-causal story explaining why the TV commercial and the spike in searches both take place at exactly the same *minute*.

Several interesting results are worth noting. First, the effects start the very minute that the ad airs online. People are doing searches online while they are

¹ Joo, Wilbur, and Zhu [2012], in their study of the impact of TV advertising on AOL search behavior, focused exclusively on advertising for online brokerages. They imposed a complex set of modeling assumptions in order to identify causal effects; by contrast, our time-series graphs tell our story quite simply, without requiring untestable modeling assumptions.

² We are amused to point out that for effects estimated this precisely ($t=21$), we can reject the null hypothesis of no effect of advertising with a significance level as low as 10^{-100} (one in a googol!).

watching television, as previously pointed out by Joo, *et al.* [2013]. A skeptic might imagine a non-causal explanation, such as people leaving their televisions during the commercial break and going to their computers to do all sorts of activities, and this causes the spike in searches. However, these commercials represent “natural experiments” in the sense that we can use different advertisers as controls for each other’s commercial airings. The fact that “Captain America” searches spike exactly during the *Captain America* commercial, but not during the Doritos commercial, makes it very clear that the commercial caused the consumer search behavior.

Another interesting result is that the effects for movie advertisements are generally much larger than those for consumer goods. We believe this is due to movie commercials stimulating consumers to search for the full movie trailer to watch online. Movie trailers are an online form of product sampling, while physical goods are much less easy for consumers to sample online. Another result is the main exception to the rule that search spikes occur exactly when the commercial airs online: in the case of the Volkswagen commercial, we see a second large spike in views two days before the Super Bowl. This early spike occurred because of Volkswagen’s decision to pre-release the commercial online in hopes of generating “viral” attention. The commercial, featuring a child in a Darth Vader costume, was sufficiently cute and engaging that it inspired a long period of repeated searches, beyond the initial spike, for consumers who wished to watch the commercial again.

For these Super Bowl advertisers, the effects of the ad on purchases are not nearly as measurable as the effects on searches. The lifts in brand-relevant search queries are likely correlated with increases in sales, but not perfectly so. Many of the queries suggest that consumers are searching to view movie trailers or view an amusing commercial again. Many of these searches are likely not leading directly to increased purchases. However, many searches may represent merely the tip of the iceberg in terms of future shopping behavior; the search during the game may be a task simple enough to remain socially acceptable during a football party, but the consumer may do additional research later to learn more about the products or services. Given the fact that we see search lifts for most of the advertisers, even the ones with less amusing ads and products difficult to sample online, we believe that on the whole we are seeing important evidence of TV serving a role in building awareness and leading potential customers to learn more about the advertised brands.

2. RESEARCH DESIGN

We focus on the biggest annual event in TV advertising: the Super Bowl. Super Bowl television commercials are well known for having unrivaled reach and for having invested in high production values.³ These two factors cause high impact, making Super Bowl ads particularly likely to have measurable effects on search behavior. By linking data on the exact timing of each commercial with the exact timing of related search queries, we can observe the impact of the Super Bowl commercials.

We examine data from Super Bowl XLV, held on February 6, 2011. The TV commercial schedule included advertiser, product, time (EST), and duration data for all commercials on the FOX Network from 6:30pm until 10:15pm. This included the post-game show but not the pre-game show. The search data comes from Yahoo! Search, accounting for 14% of relevant United States search events at the time,

³ The Super Bowl commercials examined in this paper can easily be found online by searching for “super bowl ads 2011.” The following link from Advertising Age has freely accessible videos of the ads: <http://adage.com/article/special-report-super-bowl/watch-super-bowl-commercials/148677/>.

according to public sources.⁴ We examine queries related to each advertiser’s brand, defining a query as *related* if either the query includes the product’s name or any link in the page of search results includes the advertiser’s domain name.⁵

For example, the movie *Captain America* would match search page views that either included the phrase “captain america” in the search query or a link to a website with “captainamerica.marvel.com” as part of its URL. This encompasses a large number of unique search queries ranging from “captain america trailer” to “when will we see captain america appear in thor” to “captain america kids halloween costume.”⁶

We obtained search data for a sample of queries related to the Super Bowl festivities and commercials. The Super Bowl teams and entertainers included searches for “Packers,” “Steelers,” “Christina Aguilera,” “Black Eyed Peas,” and “Usher.” The Super Bowl commercials advertised products from four broad categories: movies, cars, internet services, and consumer goods. There were 67 commercials on the schedule, but our sample of searches only covered 46 or 70% of the commercials—21 commercials were inadvertently omitted.⁷ We expect the results for the 46 commercials to be representative of the categories, in spite of the omissions.

Our analysis of the commercials’ impact on related searches is straightforward. We present graphs of the related search volume over time to visualize the impact of the commercials on search behavior. To understand the statistical significance of the spikes in searches that coincide with the commercials, we compute t -statistics for the difference in mean search volume for one hour preceding and one hour following the commercial, for a total of 7200 second-level observations.⁸ Qualitatively and quantitatively, the statistical significance⁹ of the search spikes using this two-hour time window¹⁰ is robust to longer time windows and alternative models such as Poisson regression. We prefer to use the simplest model possible for exposition and share the visually compelling histograms and simply computed t -statistics.

3. TV COMMERCIALS’ IMPACT ON ONLINE SEARCH

There are several components to the Super Bowl. The Super Bowl began with the USA national anthem sung by pop-artist Christina Aguilera, followed by the football matchup between the Green Bay Packers and the Pittsburgh Steelers. During the

⁴ <http://blog.compete.com/2011/03/16/february-2011-search-market-share-report/>

⁵ Appendix 2 (available on the authors’ websites) provides the full set of regular expressions used to define *related queries* for each advertiser.

⁶ Appendix 3 provides an even longer list of examples of related queries for *Captain America*.

⁷ We originally extracted search data for an incomplete list of advertisers; by the time we realized our omission, the raw search data had been deleted. The missing advertisers included 2 movie ads (*Fast Five*, *Mars Needs Moms*), 9 car-related ads (Chevy, Chrysler, Mini, Castrol Edge), 2 internet service ads (Career Builder, TheDaily.com), and 8 consumer goods ads (Budweiser, Lipton, Stella Artois, Wendy’s, Verizon).

⁸ A careful econometrician might worry about positive autocorrelation at such short time scales, which (given our implicit assumption of independent observations) could cause us to overstate our statistical significance. On the other hand, we have been agnostic about the shape of the response function, but modeling the shape could easily yield higher significance levels. We present the difference-in-means t -statistics as a simple quantification that captures the qualitative evidence apparent in the figures; this footnote alerts the interested reader to details that may be valuable in future research.

⁹ While the statistical significance is not qualitatively impacted by using longer time windows, we likely underestimate the total impact by omitting any incremental searches beyond one hour. We trade off the omitted search lift with the bias potentially introduced from widening the window around the commercial.

¹⁰ Lewis and Nguyen [2012] use a ten-minute time window in their experimental analysis of online display advertisements. Their results are similar in nature: a significant spike immediately following exposure accounts for most of the statistical significance.

game's half-time intermission, the Black Eyed Peas and Usher performed together in a mini-concert. The game concluded with the Green Bay Packers triumphing over the Steelers with a score of 31 to 25. The TV commercials are shown interspersed during the entire presentation of game play during time-outs, official commercial breaks, and other lulls in game play. In total, the commercials account for 40 out of 225 minutes of the total scheduled game and post-game-show TV time (6:30-10:15pm). Our analysis covers 26 minutes of commercial time spanned by 46 of the 67 total commercials aired during the game.

3.1 The Big Game

Before examining the results for the commercials, we would like to know whether there are signals in the data which can answer basic questions regarding the game:

- Does the winning team get more searches than the losing team?
- Is the timing of the national anthem, half-time, or post-game recap noticeable?
- Are there systematic changes in search behavior during commercial breaks?

In Figure 1, we present histograms of the searches over time. Note that the vertical yellow bars show commercial breaks. First off, we see that searches for the Packers fluctuate over the course of the game. Interestingly, searches for the Packers spike at the end of the game—perhaps indicating that they had just won the Super Bowl. In contrast to the spike, there is a lull in search activity for both the Packers and Steelers at 8:08 PM, contemporaneous with spikes in searches for the half-time show artists, the Black Eyed Peas and Usher. We also see spikes for the Black Eyed Peas, Usher, and Christina Aguilera at the end of the game, presumably coinciding

Fig. 1: Searches for teams and musical artists during Super Bowl LXV.

with post-game show and news outlet coverage of the event. Finally, Christina Aguilera unintentionally omitted a few lines of the national anthem during her a cappella performance. There is an initial spike in search activity surrounding the national anthem, followed by an additional spike in activity after the game had started and the news had spread about her gaffe.

There are very strong signals about the composition of the event—who the actors are as well as when they are performing. But are there any systematic lulls or spikes in search behavior that appears to be correlated with the commercial breaks? Casual inspection¹¹ suggests that search behavior related to the Super Bowl is not systematically correlated with the commercial breaks, leading us to conclude that the interruption of commercials is not changing the intensity of search behavior related to the programming. But what is the impact on searches related to the commercials?

3.2 The Commercials: Movies

Eleven movie commercials aired during the Super Bowl; Figure 2 plots histograms of related search page views for the nine we observed: *Captain America: The First Avenger*, *Cowboys and Aliens*, *Limitless*, *Pirates of the Caribbean*, *Rango*, *Rio*, *Super 8*, *Thor*, and *Transformers 3: Dark Side of the Moon*. The figure also includes the aforementioned yellow bars for commercial breaks and a green bar corresponding to when the movie’s TV ad was aired.

The results are amazingly stark: spikes for each of the movies immediately follow each of the ads. In fact, the spikes begin less than 15 seconds following the end of the TV ad—roughly the time it takes to type “captain america” into a search engine. However, the boost over baseline search behavior persists throughout the remainder of the evening for virtually all of the movies following their Super Bowl ad. There is no doubt that these spikes are clear indications of the causal impact of TV ads on online search behavior; a statistical comparison of the 60 minutes before and after each commercial yields expectedly large t -statistics with *Rio* ($t=9.14$) and *Transformers 3* ($t=42.75$) bounding the movie category.

Note, however, the significant variation in the magnitude of the initial spikes in searches across movies: *Super 8* and *Rio* differ by an order of magnitude (spikes of ~400 searches versus ~40 during 15-second intervals following the commercials). Part of this may be attributable to a decline in viewership toward the end of the game or to a difference in the fundamental appeal of the two movies to the audience. *Super 8*, as a sci-fi thriller, effectively built up tension in their ad that may have piqued the curiosity of viewers ($t=18.94$). *Rio*, as a family movie, may have generated the same level of appeal among children and parents, but they may not have been as likely to search to learn more immediately due to the content of the ad ($t=9.14$). In addition, the method to associate search queries related to the movies may have been more effective for *Super 8* than for *Rio*, even though each has many alternative associations (e.g., Super 8 Hotels and *rio*, which means “river” in many languages).

We see some small but detectable natural spillovers between *Captain America* ($t=32.64$) and *Thor* ($t=17.89$) in terms of search behavior, with a clear spike in queries for each movie following the other movie’s commercial airing (most clearly for *Thor* during the *Captain America* commercial). This overlap might result either from consumers’ mental associations between the two Marvel superheroes or from search

¹¹ Any correlation between search behavior and commercial breaks is much smaller than the correlation with the game or the commercials. We highlight the commercial breaks in all figures to make this inspection easy for every commercial break and for each advertiser.

results for one movie stimulating questions about the other movie. Regardless, the detectable spillovers in search behavior from one movie's ad on searches for another movie are modest. This is in contrast to Lewis and Nguyen [2012] who, using randomized ad-exposure data, find statistically and economically meaningful relative spillovers among advertisers, especially in the auto industry.

Fig. 2: Searches for movies advertised during Super Bowl LXV.

In total, 25 car-related ads were shown; Figure 3 shows histograms of related search page views for 16 of those commercials for Audi, BMW, Hyundai, Kia, Mercedes-Benz, Volkswagen, Bridgestone, Carmax, and Cars.com. As before, the figure includes yellow bars for commercial breaks and a green bar showing the advertiser's airtime.

Fig. 3: Searches for automobile brands advertised during Super Bowl LXV.

Similar to the movie commercials, all of the auto manufacturers' ads generated meaningful spikes in search behavior immediately. The statistical significance for this category was strongest for Volkswagen ($t=12.42$) and Cars.com ($t=22.63$). There are a few other noteworthy facts as well. Lewis and Nguyen [2012] find that for an Acura display ad, significant spillovers are generated for similar brands, vehicles, and sales outlets. Careful inspection suggests that Audi's searches are somewhat boosted immediately following commercials for BMW and Mercedes-Benz, and BMW's searches are also boosted during the Mercedes-Benz commercial. This (weakly) suggests that similar commercials occurring later in the event may remind viewers of competitors' commercials that have already been shown. The evidence of spillovers is not uniform; the Audi commercial did not appear to generate any lift in search behavior for BMW or Mercedes-Benz. Perhaps the most noteworthy boost occurred for Cars.com—not from its own ad, but from the 2-minute Chrysler 200 ad shown at 9pm. This highly specific coincidence is too great to attribute the spike in searches to any other credible cause. This is consistent with Lewis and Nguyen's findings that advertising for a product can stimulate searches for related products, brands, and services.

Volkswagen showed two commercials: a cute Passat ad featuring a young boy in a Darth Vader costume and a Beetle ad featuring an animated beetle running around like a racecar.¹² Volkswagen pre-released the Darth Vader commercial online in advance of the Super Bowl, generating 18 million views even before the game began [Dreier 2011].¹³ The Beetle ad generated a huge spike around 9:45PM. We also see a sustained massive increase in searches at 8:33PM, which is puzzling because it does not correspond to our records of a commercial airtime. Perhaps there was a featured mention of the commercial at that point in halftime, or perhaps a celebrity with many Twitter followers tweeted about it at that time, causing many retweets (and searches). We know the Passat commercial generated heavy online interest, with over 57 million views on YouTube as of April 2013.

3.4 The Commercials: Internet Services

Eleven commercials for internet services aired during the Super Bowl; Figure 4 plots graphs of related queries for nine commercials for GoDaddy.com, Teleflora.com, Salesforce, E-Trade, HomeAway, and Groupon.com. The figure also includes yellow bars for commercial breaks and a green bar corresponding to the ad airtime.

All commercials for internet services provide great examples of large impacts (t -statistics range from 14.83 to 25.02) as one might expect: internet services are naturally found on the Internet, usually via navigational search. This begs the question—how much direct traffic are these advertisers receiving in addition to the navigational traffic via search? If the effects of television advertising are very long-lived, then the number of incremental searches we measure could generate large effects on revenue. For HomeAway, a relatively unknown firm seeking brand awareness for its vacation-rental matching market, we estimate that there were 3,000 incremental searches that evening just on Yahoo! Search. Across all search engines, the total could easily be 20,000 incremental searches; the number of incremental visitors that evening could easily be as high as 30,000 if some consumers navigated directly to HomeAway.com without a search engine. Given the

¹² <http://www.sbnation.com/2011-super-bowl/2011/2/7/1979815/super-bowl-commercials-2011-volkswagen-score-big-with-beetle-literally-volkswagen-2011-beetle>

¹³ See Figure 6, below, for the related increase in query volume several days before the Super Bowl.

Fig. 4: Searches for Internet services advertised during Super Bowl LXV.

approximately \$3 million price of the advertising, this represents an acquisition cost on the order of \$100 per customer gained that evening. While hardly precise, this back-of-the-envelope calculation indicates that the price per incremental customer is high but not unreasonable. The costs are much lower, of course, if the effects on consumer behavior are more long-lived than just the duration of the game.

3.5 The Commercials: Other Consumer Goods

In total, 20 ads for other consumer goods were shown; Figure 5 shows histograms of related search page views for 12 of those commercials for Doritos, Pepsi, Motorola Xoom, Coca-Cola, Snickers, Best Buy, and Skechers. The figure also includes yellow bars for commercial breaks and a green bar showing the advertiser's airtime.

There are two patterns in consumer goods: durables and consumables. The durables, Motorola Xoom ($t=20.95$) and Skechers ($t=10.01$), show strong spikes in searches similar to movies, cars, and internet services. Consumers can easily research these products online. However, the consumables like Doritos, Pepsi, Coca-Cola, and Snickers are very difficult to experience other than by eating or drinking

Fig. 5: Searches for other consumer goods advertised during Super Bowl XLV.

the product. In terms of search behavior, all four consumable products show some impact from the TV commercials ($t < 7$) with Doritos ($t = 27.35$) being the extreme outlier. The commercials for Doritos were entertaining and included mention of special websites that were created for the Super Bowl. Consumers likely wanted to see the commercial again or visit the advertiser's special Super Bowl websites.

4. PROSPECTS FOR MEASUREMENT OF OTHER TV AD CAMPAIGNS

We have demonstrated the feasibility of measuring causal effects of TV commercial exposure on online search activity. We did so using the most favorable conditions possible: expensively produced, popular advertisements simultaneously reaching

more than 100 million people during a 30-second time window. We now examine the prospects for extending this technique to measure the effects of the other 99.9% of annual spending on TV advertising.¹⁴

First, we consider the problem of causality versus correlation that plagues research on advertising effectiveness. We have argued that the causal effects of the ads are quite clear for many advertisers because of the large increases in search volume starting the very minute that the ad aired on television. In measuring the effects of other television campaigns, we may not be so lucky, because the signal strength may be much lower and we may therefore need to look at a longer time window in order to detect statistically significant effects. In that case, causal inference becomes trickier, because of the problem of establishing a credible counterfactual baseline for the number of searches that would have taken place during the relevant time period in the absence of the advertising. Establishing this counterfactual baseline is most credible when search activity is stable over longer periods of time.

Figure 6 provides several interesting examples of search activity for an eight-day period ending on Super Bowl Sunday, allowing us to compare activity during the

Fig. 6: Eight days of searches for several advertisers.

¹⁴ The global budget for television advertising is around \$200 billion [ZenithOptimedia 2012]. The 40 minutes of Super Bowl commercials collectively cost only \$240 million in 2011.

game with activity in a prior “control baseline” period. The figure shows search activity for queries about the Steelers, Packers, and Black Eyed Peas. We see a huge boost in game-day activity for the Steelers and Packers relative to the preceding week, but the boost begins long before game time. We would not want to measure the causal effects of the TV broadcast by comparing queries on Super Bowl Sunday with queries on the preceding seven days, because we know (from the increase in consumer searches before game time) that there is also a large increase in query volume not caused by anything specific to the broadcast. In this case, the question would be, “Did the advertising have a causal impact, or do consumer searches exhibit unusual patterns on the evening of Super Bowl Sunday for other reasons?” The advertiser for the movie *Limitless* may feel comfortable concluding that the advertising caused the lift in game-day searches, given the stability in search behavior leading up to the game. However, would Best Buy feel comfortable concluding that their Super Bowl commercial caused a significant decrease in game-day searches for their products? While it is clear in Figure 6 that online searches for Best Buy’s consumer electronics systematically dropped (relative to the previous Sunday) when one-third of the country started watching the game, that is not a causal effect of their Super Bowl commercial. Only with high-frequency data can we see the patterns clearly.

Further, consider the plots for Teleflora.com and VW. Both advertisers have abnormal spikes during the days leading up to Super Bowl. Without understanding what caused those differences, you would be left with the question on game day, “What was idiosyncratic about the days leading up to the Super Bowl? Was it the Super Bowl ad or one of a variety of other idiosyncratic causes that led to the game-day search lift?” Perhaps a popular blog, news article, or television show featured their service, causing the significant increase in related searches. It could also be that another large advertising campaign took place on some other medium that day, leading to the boost in search activity.¹⁵

A before-versus-after approach using highly temporally granular event data is viable for measuring search behavior following a TV commercial, at least for Super Bowl ads. However, this same method performs very poorly when applied to online media. Lewis, Rao, and Reiley [2011] coined the phrase “activity bias” in documenting that that exposure to online advertising can be temporally correlated with many outcomes of interest, such as online searches. Activity bias results in correlations that overstate true causal effects of advertising. Consider Figures 7-9 (borrowed from Lewis [2012]), which show search behavior temporally adjacent to ad exposure. In Figure 7, we indicate how many users searched for the name of a retailer after exposure to that retailer’s online ad campaign. We might naturally conclude that the advertising induced the large spike following exposure.

However, Figure 8 shows a similar comparison for these same users—but this time using searches including the control term “Craigslist” to cross-validate. Surprisingly (or perhaps not), we find a similar spike in searches including

¹⁵ As another example, supposing Teleflora.com had carried out a large search-advertising campaign that day and, as a result, the Teleflora domain name appeared in results for many more search queries, then we would discover a mechanical increase in “related searches” as a result. This example highlights the supply-and-demand nature of our definition of “related searches” as an outcome measure: while users can demand information about an advertiser through search, advertisers can also make themselves more visible on the search platform by supplying more search advertising through higher bids in the search-ad auctions. (They can also raise their placement in the auction by making improvements to their website or search ad quality scores, though this can be a much slower process.)

“Craigslist” immediately following exposure. Did the unrelated display ads cause a large increase in searches for “Craigslist?” The answer is “no.”

Figure 9 shows the comparison graph for a randomized control group for whom we suppressed the retailer’s online ad. As we now might expect, there is a natural lift in search behavior following ad exposure resulting from browsing patterns. However, there is also a causal lift in searches—and it is reasonably large, leading us to conclude that the ads do impact consumer behavior, but not as much as one might have concluded without a randomized experiment.¹⁶ A simple before-after comparison likely works much better for temporally concentrated spikes in behavior on a secondary medium where the spikes are large relative to the baseline behavior. In this paper, we find spikes in behavior on online search while the users are being influenced by the secondary medium, TV.

Clearly, establishing unequivocal causality using such methods is tenuous. Our confidence in our causal inference in this paper is bolstered by the fact that each advertiser in Figures 2-5 serves as a “control group” for each other advertiser. If we had found spikes for one advertiser during another advertiser’s commercial, that would have weakened our confidence in the causal inference, just as the “Craigslist” cross-validation did in Figures 7 and 8. The granularity and instantaneity of searching behavior and the exact timing and nationwide reach of the Super Bowl ads makes causality credible in this situation.

5. LIMITATIONS AND FUTURE WORK

Are these results generalizable to other advertisers and the other 99.9%¹⁷ of

Fig. 7: Searches for a retailer's brand increase following ad exposure.

Fig. 8: Searches for "craigslist" also increase following ad exposure.

Fig. 9: The difference between control and treatment groups shows the true causal search lift from the retailer's ad.

¹⁶ Johnson, Lewis, and Reiley [2013] present results for online and offline sales for this campaign.

¹⁷ The 40 minutes of Super Bowl commercials collectively cost ~\$240 million in 2011, representing ~0.1% of the \$200 billion in global TV spending [ZenithOptimedia 2012].

annual spending on TV advertising? The Super Bowl is an unrepresentative and live¹⁸ television program, the ad creatives are exceptional, the audience is more than one-third of the US population, and some people even watch the Super Bowl more to see the commercials than to see the game! On the other hand, social gatherings during the Super Bowl may prevent viewers from searching for content related to the commercials as much as they might normally do when watching TV at home in less social settings.

First, using Yahoo! Search data limited us to only 14% of online search activity in the United States during the Super Bowl. Unifying aggregated time series from Google, Bing, and Yahoo would yield 94% of search activity during the relevant time period. This would enhance the ability to detect such effects by a factor of seven—it would take seven Super Bowl-sized experiments using just Yahoo! Search data to reach the same level of statistical precision that would be obtained using the combined data from the three search market leaders.

Second, we have only considered search activity. Certainly social media tools such as Twitter, Facebook, and blogs would tell a broader story of the impact of the advertising on behavior. Other user data such as site visitation and purchase behavior following the commercial could provide a more holistic perspective regarding the impact of the Super Bowl ad.

Third, a typical commercial has a smaller impact but still suffers from the same size of baseline noise,¹⁹ and the t -statistic of the ad’s impact should be proportional to the cost.²⁰ Hence, to achieve the same level of statistical precision for ads costing 1/20th of a Super Bowl ad (~\$150,000), an advertiser would have to buy 400 ads, spending 20 times the cost of a single Super Bowl ad, or ~\$60 million.²¹ Even relaxing the required statistical significance from our Super Bowl ad’s median of $t=15$ to a much weaker test of whether the ad has a non-zero impact (with an expectation of $t=3$), we only simplify the problem by a factor of 25: we would still need to buy 16 ads with a total cost of \$2.4 million. Further, a test that provides sufficiently informative precision for a confidence interval of $\pm 33\%$ on the estimated effect would require $t=6$, quadrupling the number of necessary commercials to 64 and the budget to \$9.6 million.²² This observation may play an important role in explaining why other

¹⁸ Because we synchronize the commercial airtime to the search behavior, digital video recorder (DVR) use will tend to postpone any searches caused by the ads by people who watch the program later, reducing our detectable signal and causing us to underestimate the total effect. DVRs also allow some viewers to skip commercials, which could also reduce our signal. Of course, live events like the Super Bowl tend to have much lower DVR use. The search lifts we measure should be interpreted as the total immediate effects from live and slightly-delayed DVR viewers.

¹⁹ This statement assumes the advertiser is held constant. Smaller advertisers may have less baseline noise; see Appendix 1: Calculations (available on the authors’ websites) for a related discussion.

²⁰ This holds under mild economic assumptions regarding efficient markets and ad effectiveness—that the rate of return on advertising investment is equilibrated across media. Given the difficulty of estimating ad effectiveness, this assumption may not hold. However, it is a convenient and logical starting point for evaluating marginal investments in advertising. For online display advertising, Lewis [2010] shows that click-through rates decline only modestly with a large number of impressions, suggesting that an increase in impressions may lead to a proportional increase in clicks.

²¹ See Appendix 1 for more details.

²² This assumes nationwide advertising. Geographic or other audience targeted TV advertising, if coupled with query selection to filter searches by the targeting, can reduce the disadvantage from linear to square-root, making the cost equivalent to a Super Bowl ad. Additional filtering technology, such as knowing who was aware of the TV commercial by knowing who was in the room or within earshot, could further enhance the precision of the estimates. But such technology could be used for both Super Bowl ads and other TV shows’ ads—maintaining Super Bowl ads’ relative statistical advantage from concentrating ad spending.

researchers [Joo, Wilbur, and Zhu 2012; Joo, *et al.* 2013] have found it relatively more difficult to detect the impact of TV advertising on search behavior.

Future research may investigate how brand-related search effects vary by audience or by spend, holding the advertiser, audience, and other factors constant. This analysis could be used by advertisers to compare the relative impact of an advertiser's TV spend among various creatives, channels, programs, and audiences. These comparisons could be done using search queries related to the advertiser's brand or competitors' brands, as Lewis and Nguyen [2012] did to evaluate the competitive spillovers from display ads. On the station's side, such measurements could provide a signal for how engaged audiences are with TV shows. Searches related to the show or commercials could provide a scalable form of non-survey, passive feedback to measure overall engagement. Further, differences in the intensity of search behaviors could be a proxy of relative impact of one TV ad versus another, providing a way to quantify audience attention, interest, or impact across shows. This could help TV stations be more proactive at efficiently matching their TV commercial inventory to the most responsive audiences for each advertiser. While not as directly relevant to profits as consumer purchases, these search measurements provide a valuable signal of audience engagement that is more scalable and derived from more active consumer behavior than surveys and, hence, could be used to more effectively allocate investments in TV advertising.

6. CONCLUSION

Online search queries create a new opportunity for advertisers to evaluate the effectiveness of their TV commercials. We evaluated the impact of Super Bowl commercials on users' brand-related search behavior on Yahoo! Search immediately following 46 commercials and find statistically powerful results for most advertisers with t -statistics generally ranging between 10 and 30. The magnitude and statistical significance of the effects widely varied across advertisers, in much the same way as click-through and conversion rates vary across advertisers' online display ads.

Many brand advertisers may find using product- and brand-related searches to be a valuable new tool for assessing advertising effectiveness, a signal that can be both meaningful and statistically significant. Our results indicate that such advertisers may include movie studios, auto manufacturers, and producers of consumer goods. In contrast, direct-response advertisers will not likely benefit much from using search to evaluate their TV ads—online site visits or call-center volume will likely give clearer signals of ad effectiveness. In these cases, search queries will at best provide rich insights regarding what concepts their commercials cause viewers to think about, including competitors' products and other brand-irrelevant ideas stimulated by the commercial's creative. All advertisers may benefit from this rich information to understand both positive and negative effects of their creatives. More searches may or may not be a signal of a good commercial. For example, the commercial may have failed to communicate important details like the date of release or pricing; in these cases, incremental queries may include the word "price," providing a signal that consumers need more information. A more informative commercial could help consumers make a more immediate mental note or decision to buy.²³

Using online searches in conjunction with TV commercial exposure provides an economically and statistically powerful opportunity for advertisers to gain granular

²³ However, see Mayzlin and Shin [2011] for a strategic, information-economic reason why advertisers might choose deliberately to make their advertising uninformative about product characteristics.

insights about the effects of their TV spend on consumer behavior. In presenting a simple initial investigation of the feasibility of this emerging opportunity, this paper conclusively demonstrates that there are many opportunities for search data to be effectively leveraged by advertisers to understand and enhance the value of their TV spend.

ACKNOWLEDGMENTS

The authors thank Iwan Sakran for help with the search data, Yahoo! Inc. for research support, and Ken Wilbur for the Super Bowl ad schedule. The authors also thank Michael Schwarz, Preston McAfee, Justin Rao, and many others for feedback. This research was undertaken while the authors were at Yahoo! Research. Both authors are now employed by Google, Inc.

REFERENCES

- ABRAHAM, MAGID, AND LEONARD M. LODISH, 1990. Getting the Most out of Advertising and Promotion. *Harvard Business Review*, 68, 3, 50-60.
- DREIER, TROY, 2011. Volkswagen's Mini-Darth Vader Ad: Behind the Screens. StreamingMedia.com, <http://www.streamingmedia.com/Articles/Editorial/Featured-Articles/Volkswagens-Mini-Darth-Vader-Ad-Behind-the-Screens-74862.aspx>.
- HU, Y., L. M. LODISH, AND A. M. KRIEGER. 2007. An Analysis of Real World TV Advertising Tests: a 15-Year Update. *Journal of Advertising Research*, 47, 3, 341-353.
- JOHNSON, GARRETT A., RANDALL A. LEWIS, AND DAVID H. REILEY. 2012. Add More Ads? Experimentally Measuring Incremental Purchases Due to Increased Frequency of Online Display Advertising. Working paper, Yahoo! Research.
- JOO, MINGYU, KENNETH C. WILBUR, BO COWGILL, AND YI ZHU. 2013. Television Advertising and Online Search, forthcoming, *Management Science*.
- JOO, MINGYU, KENNETH C. WILBUR, AND YI ZHU. 2012. Effects of Television Advertising on Internet Search, SSRN working paper. <http://ssrn.com/abstract=1720713>.
- LEWIS, RANDALL A. 2012. Ghost Ads: Free Experimentation at Scale. Working paper, Yahoo! Research.
- LEWIS, RANDALL A. AND DAN T. NGUYEN, 2012. "A Samsung Ad and the iPad: Display Advertising's Competitive Spillovers to Online Search," Working paper, Yahoo! Research.
- LEWIS, RANDALL A. AND JUSTIN M. RAO, 2011. On the Near Impossibility of Measuring the Returns to Advertising, Working paper, Yahoo! Research, available at http://www.justinmrao.com/lewis_rao_nearimpossibility.pdf.
- LEWIS, RANDALL A., JUSTIN M. RAO, AND DAVID H. REILEY. 2011. Here, There, and Everywhere: Correlated Online Behaviors Can Lead to Overestimates of the Effects of Advertising, In *Proceedings of the 20th ACM International World Wide Web Conference [WWW'11] (2011)*. Hyderabad, India, 157-166.
- LEWIS, RANDALL A. AND DAVID H. REILEY. 2012a. Advertising Effectively Influences Older Users: A Yahoo! Experiment Measuring Retail Sales. *Review of Industrial Organization*, forthcoming.
- LEWIS, RANDALL A. AND DAVID H. REILEY. 2012b. Does Retail Advertising Work? Measuring the Effects of Advertising on Sales via a Controlled Experiment on Yahoo! Working paper, Yahoo! Research, available at <http://www.davidreiley.com/papers/DoesRetailAdvertisingWork.pdf>.
- LEWIS, RANDALL A., DAVID H. REILEY, AND TAYLOR A. SCHREINER. 2012. Ad Attributes and Attribution: Large-Scale Field Experiments Measure Online Customer Acquisition. Working paper, Yahoo! Research, available at <http://www.davidreiley.com/papers/AAA.pdf>.
- LODISH, L. M., ABRAHAM, M., KALMENSON, S., LIVELSBERGER, J., LUBETKIN, B., RICHARDSON, B., AND M. E. STEVENS. 1995a. How T.V. Advertising Works: a Meta-Analysis of 389 Real World Split Cable T.V. Advertising Experiments. *Journal of Marketing Research*, 32, 2, 125-139.
- LODISH, L. M., ABRAHAM, M., KALMENSON, S., LIVELSBERGER, J., LUBETKIN, B., RICHARDSON, B., AND M. E. STEVENS. 1995b. A Summary of Fifty-Five In-Market Experiments of the Long-Term Effect of TV Advertising. *Marketing Science*, 14, 3, 133-140.
- MAYZLIN, D. AND J. SHIN. 2011. Uninformative Advertising as an Invitation to Search. *Marketing Science*, 30, 4, 666-685.
- OLDHAM, JEFFREY. 2012. Super Bowl XLVI: Mobile, Manning and Madonna. *Official Google Blog*, February 6, 2012. <http://googleblog.blogspot.com/2012/02/super-bowl-xlvi-mobile-manning-and.html>.
- STIPP, HORST AND DAN ZIGMOND. 2010. When Viewers Become Searchers: Measuring the Impact of Television on Internet Search Queries. *Advertising Research Foundation: Rethink*.
- STIPP, HORST AND DAN ZIGMOND. 2011. Vision Statement: Multitaskers May Be Advertisers' Best Audience, *Harvard Business Review*, January.
- ZENITHOPTIMEDIA, 2012. ZenithOptimedia Releases New Ad Forecasts: Global Advertising Continues to Grow Despite Eurozone Fears, June 8, <http://www.zenithoptimedia.com/zenith/zenithoptimedia-releases-new-ad-forecasts-global-advertising-continues-to-grow-despite-eurozone-fears/>.

Online Appendix to: Down-to-the-Minute Effects of Super Bowl Advertising on Online Search Behavior

RANDALL A. LEWIS, Google, Inc.

DAVID H. REILEY, Google, Inc.

APPENDIX 1: CALCULATIONS

The key to our success in detecting large search lifts due to Super Bowl advertising is the concentration of ad spending against a large audience reached at a single point in time combined with temporally granular search data. We consider the statistical problem. Let $C = \$3$ million be the cost of the commercial; assume that the expected ad effect is proportional to the cost: ad impact $= \alpha \cdot C$ for some α . This is reasonable for an advertiser's marginal spending for which their return on investment (ROI) should be close to the cost of capital. Let σ be the baseline standard error of an estimate of a single commercial's impact. We observe that the t -statistic has both components:

$$t = \alpha \cdot \frac{C}{\sigma}$$

where $\alpha \cdot C$ and σ are in units of the outcome of interest.

Now, observe the t -statistic when we split the budget C into N commercials. The signal remains the same—we still spend the same amount of money and expect the same ROI—but the baseline variance is now scaled up by the number of commercials, or equivalently, the standard deviation is scaled up by the square-root of N :

$$t = \alpha \cdot \frac{C}{\sqrt{N} \cdot \sigma}.$$

Alternatively, if we consider the t -statistic for each commercial, we divide the expected signal of each commercial by N :

$$t = \alpha \cdot \frac{C/N}{\sigma}.$$

Intuitively, each commercial is an observation—but here we have made each observation $1/N$ less informative. As a result, to achieve the same t -statistic as before, we will need N^2 of the less informative observations.

For example, if we estimate a t -statistic of 15 for a given commercial's search lift, we should expect to find a t -statistic of $15/N$ if we split a commercial's budget into N less expensive nationwide ads. Consider a $1/20$ ad buy of \$150,000. We would expect a t -statistic of roughly $15/20 = 0.75$ from a single commercial. In order to be confident in detecting statistical significance, we need an expected t -statistic of 3. This would require running $4^2 = 16$ commercials at a cost of $\$150,000 \cdot 16 = \2.4 million. Even by spending the Super Bowl's ad budget of \$3 million, we only achieve an expected t -statistic of $\sqrt{20} \cdot 0.75 = 3.5$ rather than 15 under such dilution. We would need to spend 20 times the Super Bowl budget, or \$60 million, to achieve the same level of statistical certainty about the effects of that spending.

Before concluding, we consider one additional setting. Suppose our budget was instead split among mutually exclusive geographic or audience segments. Let S be the number of segments. Now consider the effect of a single commercial:

$$t = \alpha \cdot \frac{C/S}{\sigma/\sqrt{S}}.$$

Author's addresses: R. Lewis, randall@econinformatix.com, and D. Reiley, david@davidreiley.com, Google, Inc., 1600 Amphitheatre Parkway, Mountain View, CA 94043.

DOI: <http://dx.doi.org/10.1145/0000000.0000000>

Here we have divided both the signal and the noise among the segments. If the commercials and our outcomes can be accurately divided, we actually do not suffer as much as above where the noise was unaffected by splitting up the budget among N commercials. The simplified expression for segmentation follows:

$$t = \alpha \cdot \frac{C}{\sqrt{S} \cdot \sigma}.$$

This expression for a single segmented ad is identical to the expression for spending the whole budget C on N nationwide ads. Therefore, we can achieve the same level of statistical precision afforded by a Super Bowl ad by segmenting. This is quite intuitive: if we ran a Super Bowl ad and observed segment-level and nationwide data, we should expect the same level of statistical significance from both outcomes.

Super Bowl commercials are perfect examples of market-level concentrated exposure in TV—economically significant ad expenditure that produces a detectable effect against a large share of individuals for whom we can observe behaviors [Lewis 2012]. We can use the equations above in conjunction with t -statistics from the Super Bowl to extrapolate the statistical power of measuring brand-related search lift for other TV commercials. We already answered the question, “How many \$150,000 commercials does it take to achieve a Super-Bowl-sized t -statistic?” Now we ask, “What is the smallest commercial for which we should expect a statistically significant search lift?” We again consider nationwide and segmented commercials.

For nationwide commercials, we still face the same baseline variability in searches. We again use $t=3$ as our requirement for reliably expecting a significant search lift and $t=15$ as our expectation for the Super Bowl commercial’s statistical significance. We compute the relative costs using the ratio of t -statistics where their numerators are proportional to cost and the denominators are the same for the single nationwide commercial (denoted with the subscript 1NC) and the Super Bowl commercial (denoted with the subscript SB):

$$\begin{aligned} \frac{t_{1NC}}{t_{SB}} &= \frac{C_{1NC}}{C} \\ C_{1NC} &= \$3 \text{ million} \cdot \frac{3}{15} = \$600,000. \end{aligned}$$

Thus, a \$600,000 nationwide commercial is the least expensive commercial for which we can reliably detect a search lift for the typical Super Bowl advertiser.

Segmentation is defined as the ability to filter the search queries to the particular TV audience. For segmented commercials, the baseline variability in searches is reduced by the square-root of the segmentation factor, S . This is a result of the impact of the TV commercial’s impact being focused on a smaller audience, which naturally generates a smaller cumulative variance. Again, we take the ratio of t -statistics (denoting the segmented commercial with 1SC):

$$\frac{t_{1SC}}{t_{SB}} = \frac{C_{1SC}/\sqrt{S}}{C}.$$

We know that $C_{1SC} = C/S$ which simplifies the expression to

$$\frac{3}{15} = \frac{(C/S)/\sqrt{S}}{C} \Rightarrow S = 25.$$

Therefore, for a segmented buy, we should be able to detect a TV commercial that has a Super Bowl level of per-person spending (2-3 cents) as long as it costs at least \$3 million / 25 = \$120,000. However, few other commercials achieve a Super-Bowl-sized local or segment reach of 1/3 for a single commercial. Adjusting for reach, r , introduces a variance scaling factor in the denominator:

$$\sigma_S = \sigma / \sqrt{S \cdot r^{1/3}}$$

which adjusts the minimum-detectable cost by a factor of $1/3 / r$: $\$120,000 / 3r$. This is still a large spend for a single commercial, but this minimum-detectable cost can be further reduced if we can identify who is or is not watching the show that the commercial airs on. The improvement gained by focusing on a narrower segment versus measuring the outcome at the national level can be extended by performing the analysis only on those individuals who were likely to have seen the ad. Given that a typical commercial reaches less than 10% of the population, other ways to exclude the 90% of non-viewers from the analysis can theoretically reduce the minimum-detectable cost by a factor of at least 3.

Understanding the practical structure and theoretical limitations of detecting the impact of TV advertising on search behavior not only illustrates the difficulty of the problem, as Super Bowl ads are atypical, but also outlines the feasibility set and opportunities to improve the signal. Online search behavior as a signal of TV commercial effectiveness can be further enhanced by advertisers and technologists.

Advertisers can design commercials to stimulate a viewer's motivation to search online. Orders of magnitude of difference in detectability are observed across products: for example, compare Doritos' much larger search lift with Pepsi's. This could easily be done by highlighting the Pepsi's broadly appealing web presence. This way of strengthening the search signal could be especially useful for advertisers who want to measure attentiveness to the TV commercial across placements.

Technologists can construct better aggregations of commercial-related queries by efficiently extracting only the data for impacted queries. This can both increase the signal by capturing affected queries that were missed in this research and decrease the noise by omitting unaffected queries that were erroneously captured. Along these same lines, only including very significantly affected queries and ignoring less significantly affected queries can also improve detectability as not every signal is worth the noise it carries along when included.

Finally, while the number of incremental searches impacts the detectability of the search lifts, the baseline search variability is the other half of the equation. Large, well-known Super Bowl advertisers may tend to have greater baseline search variability—perhaps in proportion to their size. Thus, in line with Lewis and Rao [2013], there are both affordability and detectability limitations on detecting the effects of TV commercials on search behavior. Smaller advertisers who can afford less expensive commercials may be able to detect meaningful effects from those whereas the large advertisers cannot, due to their more volatile search baseline. Thus, we note that the bounds computed here are for Super-Bowl-scale advertisers, not for all advertisers. Lewis and Rao [2013] show that the detectability of ad effects increases in the absolute cost of advertising media but decreases in the percentage of firm revenues. Future research can investigate how detectability changes with firm size.

APPENDIX 2: DEFINITION OF RELATED QUERIES

A search page view is defined as related if either the query or any search or ad link's URL matches one or more of the following regular expressions:

.*\audiusa\.com.*	.*\doritos\.com.*	.*\salesforce\.com.*	.*\pirates of the.*
.*\audi\.com.*	.*\fritolay\.com.*	.*\skechers\.com.*	.*\rango movie.*
.*\bestbuy\.com.*	.*\doritoslatenight\.com.*	.*\mars\.com.*	.*\rio movie.*
.*\bmwusa\.com.*	.*\doritoschangethegame\.com.*	.*\teleflora\.com.*	.*\rio the movie.*
.*\bmw\.com.*		.*\thor\.marvel\.com.*	.*\super 8 movie.*
.*\bridgestone\.com.*	.*\crashthesuperbowl\.com.*	.*\volkswagen\.com.*	.*\super8 movie.*
.*\bridgestonetire\.com.*	.*\etradel\.com.*	.*\vw\.com.*	.*\thor movie.*
.*\captainamerica\.marvel\.com.*	.*\godaddy\.com.*	.*\transformersmovie\.com.*	.*\thor 2011.*
	.*\groupon\.com.*	.*\rangomovie\.com.*	.*\packers.*
.*\carmax\.com.*	.*\homeaway\.com.*	.*\rio-themovie\.com.*	.*\steelers.*
.*\cars\.com.*	.*\hyundaiusa\.com.*	.*\super8-movie\.com.*	.*\christina aguilera.*
.*\thecoca-colacompany\.com.*	.*\kia\.com.*	.*\captain america.*	.*\black eyed peas.*
.*\coca-cola\.com.*	.*\mbusa\.com.*	.*\cowboys and aliens.*	.*\usher.*
.*\cowboysandaliensmovie\.com.*	.*\mercedes-benz\.com.*	.*\cowboys & aliens.*	
	.*\motorola\.com.*	.*\cowboys-and-aliens.*	
	.*\pepsi\.com.*	.*\limitless.*	

APPENDIX 3: EXAMPLES OF RELATED QUERIES

Below we find some examples of related queries on January 30, 2011 for Captain America listed in decreasing frequency of appearance:

captain america	who will play captain america	captain america poster 2011
captain america trailer 2011	the first avenger captain america 2011	captain america cmoic
captain america movie	trailers	...
captain america trailer	captain america in thor	captain america kids halloween
captain america movie trailer	...	costume
captain america the first avenger	iron man finds captain america	captain america of vietnam
...	wii games captain america	green lantern "captain america"
when will we see captain america	captain america teaser trailer	who is playing captain america
appear in thor	captain america cycling jersey	play captain america games
new captain america movie	is there a female eivalant to captain america	captain america arrest florida
captain america super bowl		