# Beyond the Last Touch: Attribution in Online Advertising

Preliminary Version

Current version at: http://ron-berman.com/papers/attribution.pdf

Ron Berman[1]

September 20, 2013

[1]University of California, Berkeley; Haas School of Business. email: ron_berman@haas.berkeley.edu
I would like to thank Zsolt Katona, Ganesh Iyer and Shachar Kariv for their dedicated support during this project. I am also indebted to J. Miguel Villas-Boas, Przemyslaw Jeziorski, Ben Hermalin, Haluk Ergin and participants at the MI 12 conference in Israel for helpful comments.

# Beyond the Last Touch: Attribution in Online Advertising


## Abstract

Advertisers who run online advertising campaigns often utilize multiple publishers concurrently to deliver ads. In these campaigns advertisers predominantly compensate publishers based on effort (CPM) or performance (CPA) and a process known as Last-Touch attribution. Using an analytical model of an online campaign we show that CPA schemes cause moral-hazard while existence of a baseline conversion rate by consumers may create adverse selection. The analysis identifies two strategies publishers may use in equilibrium – free-riding on other publishers and exploitation of the baseline conversion rate of consumers.

Our results show that when no attribution is being used CPM compensation is more beneficial to the advertiser than CPA payment as a result of free-riding on other's efforts. When an attribution process is added to the campaign, it creates a contest between the publishers and as a result has potential to improve the advertiser's profits when no baseline exists. Specifically, we show that last-touch attribution can be beneficial for CPA campaigns when the process is not too accurate or when advertising exhibits concavity in its effects on consumers. As the process breaks down for lower noise, however, we develop an attribution method based on the Shapley value that can be beneficial under flexible campaign specifications.

To resolve adverse selection created by the baseline we propose that the advertiser will require publishers to run an experiment as proof of effectiveness. Although this experiment trades-off gaining additional information about the baseline with loss of revenue from reduced advertising, we find that using experimentation and the Shapley value outperforms campaigns using CPM payment or Last-Touch attribution.

Using data from a large scale online campaign we apply the model's insights and show evidence for baseline exploitation. An estimate of the publishers' Shapley value is then used to distinguish effective publishers from the exploiting ones, and can be used to aid advertisers to better optimize their campaigns.

# 1 Introduction

Digital advertising campaigns in the U.S. commanded US $36.6 Billion in revenues during 2012 with an annual growth rate of 19.7% in the past 10 years,[1] surpassing all other media spending except broadcast TV. In many of these online campaigns advertisers choose to deliver ads through multiple publishers with different media technologies (*e.g.* Banners, Videos, etc.) that can reach overlapping target populations.

This paper analyzes the *attribution* process that online advertisers perform to compensate publishers following a campaign in order to elicit efficient advertising. Although this process is commonly used to benchmark publisher performance, when asked about how the publishers compare, advertisers' responses range from "We don't know" to "It looks like publisher X is best, but our intuition says this is wrong." In a recent survey[2], for example, only 26% of advertisers claimed they were able to measure their social media advertising effectiveness while only 37% of advertisers agreed that their facebook advertising is effective. In a time when consumers shift their online attention towards social media, it is surprising to witness such low approval of its effectiveness.

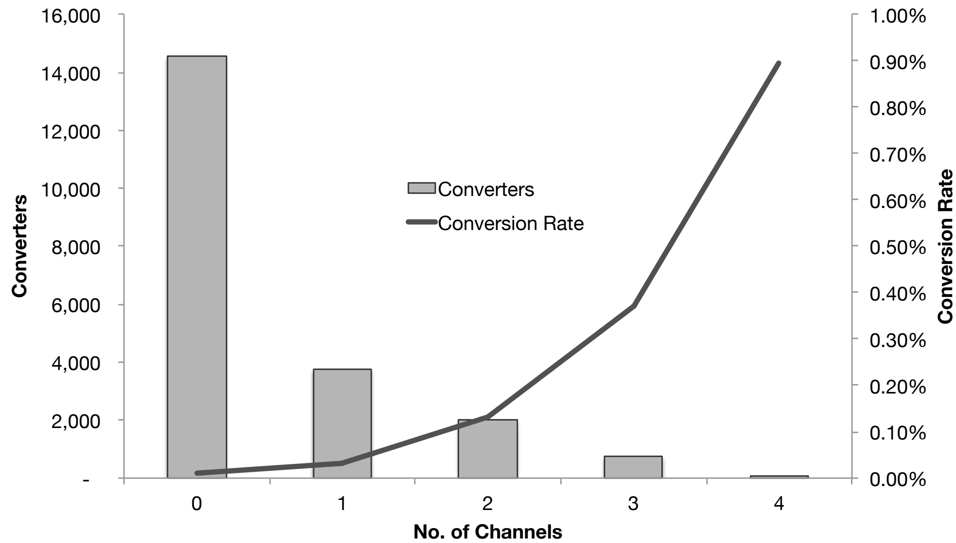
To illustrate the potential difficulties in attribution from multiple publisher usage, Figure 1 depicts the performance of a car rental campaign exposed to more than 13 million online consumers in the UK, when the number of converters[3] and conversion rates are broken down by the number of advertising publishers that consumers were exposed to. As can be seen, a large number of converters were exposed to ads by more than one publisher; it also appears that the conversion rate of consumers increases with the number of publishers they were exposed to.

An important characteristic of such multi-publisher campaigns is that the advertisers do not know a-priori how effective each publisher may be. Such uncertainty may arise, *e.g.*, when publishers can target consumers based on prior information, when using new untested ads or because consumer visit patterns shift over time. Given that online campaigns collect detailed browsing and ad-exposure history from consumers, we ask what obstacles this uncertainty may create to the advertiser's ability to properly mount a campaign.

The first obstacle that the advertiser faces during multi-publisher campaigns is that the ads interact in a non-trivial manner to influence consumers. From the point of view of the advertiser,

[1]Source: 2012 IAB internet advertising revenue report.
[2]Source: "2013 Social Media Marketing Industry Report", www.socialmediaexaminer.com
[3]Converters are car renters in this campaign. Conversion rate is the rate of buyers to total consumers.

Figure 1: Converters and Conversion Rate by Publisher Exposure




getting consumers to respond to advertising constitutes a team effort by the publishers. In such situations a classic result in the economics literature is that publishers can piggyback on the efforts of other publishers, thus creating moral hazard (Holmstrom 1982). If the advertiser tries to base its decisions solely on the measured performance of the campaign, such free-riding may prevent it from correctly compensating publishers to elicit efficient advertising.

A second obstacle an advertiser may face is lack of information about the impact of advertising on different consumers. Since the decision to show ads to consumers is delegated to publishers, the advertiser does not know what factors contributed to the decision to display ads nor does it know the impact of individual ads on consumers. The publishers, on the other hand, have more information about the behavior of consumers and their past actions, especially on targeted websites with which consumers actively interact such as search-engines and social-media networks. Such asymmetry in information about ad effectiveness may create adverse selection – publishers who are ineffective will be able to display ads and claim their effectiveness is high, with the advertiser being unable to measure their true effectiveness.

To address these issues advertisers use contracts that compensate the publishers based on the data collected during a campaign. We commonly observe two types of contracts in the industry: effort based and performance based contracts. In an effort based contract, publishers receive payment based on the number of ads they showed during a campaign. These schemes, commonly known as cost per mille (CPM) are popular for display (banner) advertising, yet their popularity is declining in favor of performance based payments.

Performance based contracts, in contrast, compensate publishers by promising them a share of the observed output of the campaign, *e.g.*, number of clicks, website visits or purchases. The popularity of these contracts, called Cost Per Action (CPA), has been on the rise, prompting the need for an attribution process whose results are used to allocate compensation. Among these methods, the popular *last-touch* method credits conversions to the publisher that was last to show an ad ("touch the consumer") prior to conversion. The rationale behind this method follows traditional sales compensation schemes – the salesperson who "closes the deal" receives the commission.

This paper uses analytical modeling to focus on the impact of different incentive schemes and attribution processes on the decision of publishers to show ads and the resulting profits of the advertisers. Our goal is to develop payment schemes that alleviate the effects of moral-hazard and asymmetric information and yield improved results to the advertiser. To this end Section 3 introduces a model of consumers, two publishers and an advertiser engaged in an advertising campaign. Consumers in our model belong to one of two segments: a baseline and a non-baseline segment. Baseline consumers are not impacted by ads yet purchase products regardless. In contrast, exposure to ads from multiple publishers has a positive impact on the purchase probabilities of non-baseline consumers. Our model allows for a flexible specification of advertising impact, including increasing returns (convex effects) and decreasing returns (concave effects) of multiple ad exposures. The publishers in our model may have private information about whether consumers belong to the baseline and make a choice regarding the number of ads to show to every consumer in each segment. The advertiser, in its turn, designs the payment scheme to be used after the campaign as well as the measurement process that will determine publisher effectiveness.

Section 4 uses a benchmark fixed share compensation scheme to show that moral-hazard is more detrimental to advertiser profits than using effort based compensation. We find that CPM campaigns outperform CPA campaigns for every type of conversion function and under quite general conditions. As ads from multiple publishers affect the same consumer, each publisher experiences an externality from actions by other publishers and can reduce its advertising effort, raising a question about the industry's preference for this method. We give a possible explanation for this behavior by focusing on single publisher campaigns in which CPA may outperform CPM for convex conversion functions.

Since CPA campaigns suffer from under-provision of effort by publishers, we observe that advertisers try to make these campaigns more efficient by employing an attribution process such as last-touch. By adding this process advertisers effectively create a contest among the publishers to receive a commission, and can counteract the effects of free-riding by incentivizing publishers to increase their advertising efforts closer to efficient amounts. We include attribution in our model through a function that allocates the commission among publishers based on the publishers' efforts and performance and has the following four requirements: Efficiency, Symmetry, Pay-to-play and Marginality. To model Last-Touch attribution with these requirements, we notice that publishers are unable to exactly predict whether they will receive attribution for a conversion because of uncertainty about the consumer's behavior in the future. As a result, our model admits last-touch attribution as a noisy contest between the publishers that has these four properties. The magnitude of the noise serves as a measurement of the publisher's ability to predict the impact of showing an additional ad on receiving attribution and depends on the technology employed by the publisher. Our analysis of this noisy process shows that in CPA campaigns with last-touch attribution, publishers increase their equilibrium efforts and yield higher profits to the advertiser when the noise is not too small. When the attribution process is too discriminating or the conversion function too convex, however, no pure strategy equilibrium exists, and publishers are driven to overexert effort. Cases of low noise level can occur, for example, when publishers are sophisticated and can predict future consumer behavior with high accuracy.

The negative properties of last-touch attribution under low noise levels as well as adverse selection[4] has motivated us to search for an alternative attribution method that resolves these issues. The Shapley value is a cooperative game theory solution concept that allocates value among players in a cooperative game, and has the advantage of admitting the four requirements mentioned above along with uniqueness over the space of all conversion functions with the addition of an additivity property. Intuitively, the Shapley value (Shapley 1952) has the economic impact of allocating the average marginal contribution of each publisher as a commission, and this paper proposes its use

[4]These results are presented in Section 6.

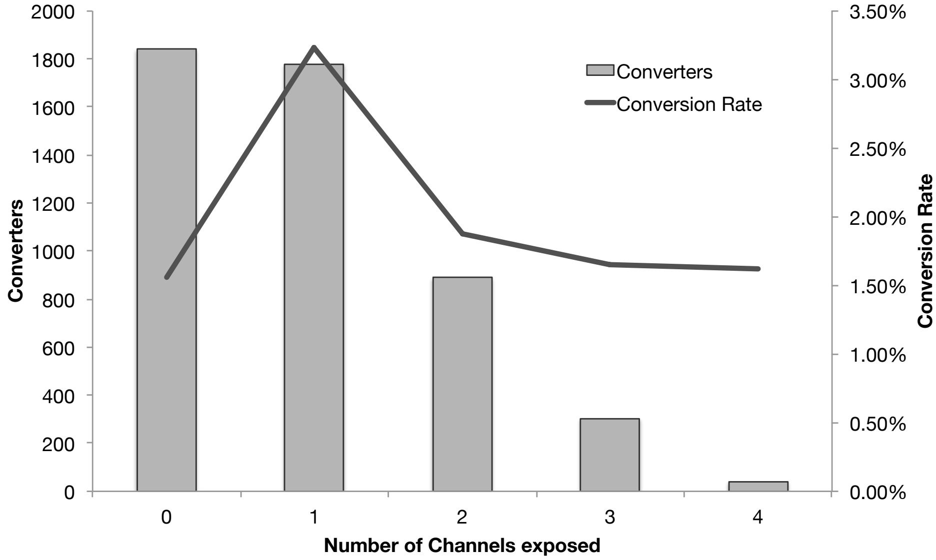
as an improved attribution scheme. In equilibrium we find that the Shapley attribution scheme increases profit for the advertiser compared to regular CPA schemes regardless of the structure of the conversion function, while it improves over last-touch attribution for small noise ranges. Since the calculation of the Shapley value is computationally hard and requires data about subsets of publishers, a question arises whether generating this data by experimentation may be profitable for the advertiser.

Section 6 analyzes the impact of asymmetric information the publisher may have about the baseline conversion rate of consumers and running experiments on consumers. We first show that running an experiment to measure the baseline may control for the uncertainty in the information. The experiment uses a control group which is not exposed to ads to estimate the magnitude of the baseline. Since not showing ads may reduce the revenues of the publisher, we search for conditions under which the optimal sample size is small enough to merit this action. We find that when the population of the campaign is large enough, experimentation is always profitable, and armed with this result, we analyze the strategies publishers choose to use when they can target consumers with high probability of conversion. In equilibrium, we show that publishers in a CPA campaign with last-touch attribution will target baseline consumers in a non-efficient manner yielding less profit than CPM campaigns. Using the Shapley value with the results of the experiment, however, alleviates this problem completely as the value controls for the baseline.

In Section 7 we investigate whether evidence exists for baseline exploitation or publisher free-riding in real campaign data. The data we analyze comes from a car rental campaign in the UK that was exposed to more than 13.4 million consumers. We observe that the budgets allocated to publishers exhibit significant heterogeneity and their estimates of effectiveness are highly varied when using last-touch methods. An estimate of publisher effectiveness when interacting with other publishers, however, gives an indication for baseline exploitation as predicted by our model, and lends credibility to the focus on the baseline in our analysis. Evidence for such exploitation can be gleaned from Figure 2, which describes the conversion behavior of consumers who were exposed to advertising only after visiting the car rental website without purchasing. If we compare the conversion rate of consumers who were exposed to two or more publishers post-visit, it would appear that the advertising had little effect compared to no exposure post-visit.

We posit that the publishers target consumers with high probability of buying in order to be

Figure 2: Converters and Conversion Rate of Visitors by Publisher Exposure




credited with the sale which is a by-product of the attribution method used by advertisers. To try and identify publishers who free-ride on others, we calculate an estimate of average marginal contributions of publishers based on the Shapley value, and use these estimates to compare the performance of publishers to last-touch methods. Calculating this value poses a significant computational burden and part of our contribution is a method to calculate this value that takes into account specific structure of campaign data. The results, which were communicated to the advertiser, show that a few publishers operate at efficient levels, while others target high baseline consumers to game the compensation scheme. We are currently in the process of collecting the information about the changes in behavior of publishers as a result of employing the Shapley value, and the results of this investigation is currently the focus of research. To the best of our knowledge, this is the first large scale application of this theoretical concept appearing in the literature.

The discussion in Section 8 examines the impact of heterogeneity in consumer behavior on publisher behavior and the experimentation mechanism. We conclude with consideration of the managerial implications of proper attribution.

## 2 Industry Description and Related Work

Online advertisers have a choice of multiple ad formats including Search, Display/Banners, Classifieds, Mobile, Digital Video, Lead Generation, Rich Media, Sponsorships and Email. Among these formats, search advertising commands 46% of the online advertising expenditures in the U.S. followed by 21% of spending going to display/banner ads. Mobile advertising, which had virtually no budgets allocated to it in 2009, has grown to 9% of total ad expenditures in 2012. The market is concentrated with the top 10 providers commanding more than 70% of the entire industry revenue.

Although the majority of platforms allow fine-grained information collection during campaigns, the efficacy of these ads remains an open question. Academic work focusing on specific advertising formats has thus grown rapidly with examples including Sherman and Deighton (2001), Dreze and Hussherr (2003) and Manchanda et al. (2006) on banner advertising and Yao and Mela (2011), Rutz and Bucklin (2011) and Ghose and Yang (2009) on search advertising among others. Recent work that employed large scale field experiments by Lambrecht and Tucker (2011) on retargeting advertising, Blake et al. (2013) on search advertising and Lewis and Rao (2012) on banner advertising have found little effectiveness for these campaigns when measured on a broad population. The main finding of these works is that the effects of advertising are moderate at best and require large sample sizes to properly identify. The studies by Lambrecht and Tucker (2011) and Blake et al. (2013) also find heterogenous response to advertising by different customer segments.

When contracting with publishers, advertisers make decisions on the compensation mechanism that will be used to pay the publishers. The two major forms of compensation are performance based payment, sometimes known as Cost Per Action (CPA) and impression based payment known as Cost Per Mille (CPM). Click based pricing, known as Cost Per Click (CPC), is a performance based scheme for the purpose of our discussion. In 2012 performance based pricing took 66% of industry revenue compared to 41% in 2005. The growth has overshadowed impression based models that have declined from 46% to 32% of industry revenue. Part of this shift can be attributed to auction based click pricing pioneered by Google for its search ads. This shift resulted in significant research attention given to ad auction mechanisms from both an empirical and theoretical perspective which is not covered in this study. It is interesting to note that hybrid models based on both performance and impressions commanded only 2% of ad revenues in 2012.

In the past few years, the advertising industry has shown increased interest in improved attribution methods. In a recent survey[5] 54% of advertisers indicated they used a last-touch method, while 42% indicated that being "unsure of how to choose the appropriate method/model of attribution" is an impediment to adopting an attribution method. Research focusing on the advertiser's problem of measuring and compensating multiple publishers is quite recent, however, with the majority focusing on empirical applications to specific campaign formats. Tucker (2012) analyzes the impact of better attribution technology on campaign decisions by advertisers. The paper finds that improved attribution technology lowered the cost per attributed converter. The paper also overviews theoretical predictions about the impact of refined measurement technology on advertising prices and makes an attempt to verify these claims using the campaign data. Kireyev et al. (2013) and Li and Kannan (2013) build specific attribution models for online campaign data using a conversion model of consumers and interaction between publishers. They find that publishers have strong interaction effects between one another which are typically not picked up by traditional measurements.

On the theory side, classic mechanism design research on team compensation closely resembles the problem an advertiser faces. Among the voluminous literature on cooperative production and team compensation the classic work by Holmstrom (1982) analyzes team compensation under moral hazard when team members have no private information. Our contribution is in the fact that the advertiser is a profit maximizing and not a welfare maximizing principal, yet we find similar effects and design mechanisms to solve these issues.

# 3 Model of Advertiser and Publishers

Consider a market with three types of players: an advertiser, two publishers and $N$ homogenous consumers. Our interest is in the analysis of the interplay between the advertiser and publishers through the number of ads shown to consumers and allocation of payment to publishers. We assume advertisers do not have direct access to online consumers, rather they have to invest money and show ads through publishers in order to encourage consumers to purchase their products.

[5]Source: "Marketing Attribution: Valuing the Customer Journey" by EConsultancy and Google.

### 3.1 Consumers

Consumers in the model visit both publishers' sites and are exposed to advertising, resulting in a probabilistic decision to "convert". A conversion is any target action designated by the advertiser as the goal of the campaign that can also be monitored by the advertiser directly. Such goals can be the purchase of a product, a visit to the advertiser's site or a click on an ad.

The response of consumers to advertising depends on the effectiveness of advertising as well as on the propensity of consumers to convert without seeing any ads which we call the baseline conversion rate. The baseline captures the impact of various states of consumers resulting from exogenous factors such as brand preference, frequency of purchase in steady state and effects of offline advertising prior to the campaign. When each publisher $i \in \{1, 2\}$ shows $q_i$ ads, we let $(q_1 + q_2)^\rho$ denote the conversion rate of consumers who have a zero baseline.[6] By denoting the baseline probability of conversion as $s$, the advertiser expects to observe the following conversion rate after the campaign:

$$x(q_1, q_2) = s + (q_1 + q_2)^\rho (1 - s) \tag{1}$$

The values of $\rho$ and $s$ are determined by nature prior to the campaign and are exogenous. To focus on pure strategies of advertising, we assume that $0 < \rho < 2$.[7] The assumption implies that additional advertising has a positive effect on the probability of buying of a consumer, yet allows both increasing and decreasing returns. When $\rho < 1$ the response of consumers to additional advertising has decreasing returns and publishers' ads are substitutes. When $\rho > 1$ publishers' ads are complements.

Finally, we let the baseline $s$ be distributed $s \sim Beta(\alpha, \beta)$ with parameters $\alpha > 0$, $\beta > 0$. The flexible structure will let us understand the impact of various campaign environments on the incentives of advertisers and publishers.

[6] The additivity of advertising effects is not required but simplifies exposition. Asymmetric publisher effectiveness is discussed in Appendix A.

[7] Restricting $\rho < 2$ is sufficient for the existence of profitable pure strategies when costs are quadratic.

### 3.2 Publishers

Publishers in the model make a simultaneous choice about the number of ads $q_i$ to show to each consumer and try to maximize their individual profits. When showing these ads publishers incur a cost resulting from their efforts to attract consumers to their websites. We define the cost of showing $q_i$ ads as $\frac{q_i^2}{2}$. Both publishers have complete information about the values of $\rho$ and $s$, as well as the conversion function $x$ and the cost functions.

At the end of the campaign, each publisher receives a payment $b_i$ from the advertiser that may depend on the amount of ads that were shown and the conversion rate observed by the advertiser. The profit of each publisher $i$ is therefore:

$$u_i = b_i(q_1, q_2, x) - \frac{q_i^2}{2} \tag{2}$$

### 3.3 The Advertiser

The advertiser's goal is to maximize its own profit by choosing the payment contract $b_i$ to use with each publisher prior to the campaign. The structure of the conversion function $x$, as well as the value of $\rho$ are known to the advertiser. Initially, we assume as a benchmark that the baseline $s$ is known to the advertiser, which we normalize to zero without loss of generality. The goal of this assumption, to be relaxed later, is to distinguish the effects of strategic publisher interaction on the advertiser's profit from the effects of additional information the publishers may have about consumers.

Normalizing the revenue from each consumer to 1, the profit of the advertiser is then:

$$\pi = x(q_1, q_2) - b_1(q_1, q_2, x) - b_2(q_1, q_2, x) \tag{3}$$

### 3.4 Types of Contracts - CPM and CPA

The advertising industry primarily uses two types of contracts - performance based contracts (CPA) in which publishers are compensated on the outcome of a campaign, and effort based contracts (CPM) in which publishers receive payment based on the amount of ads they show. As noted in the introduction, hybrid contracts that make use of both types of payments are uncommon. As shown by Zhu and Wilbur (2011), in environments that allow hybrid campaigns, rational publishers expectations will rule out hybrid strategies by advertisers.

CPM contracts (cost per mille or cost per thousand impressions) are effort based contracts in which the advertiser promises each publisher a flat rate payment $p_i^M$ for each ad displayed to the consumers. The resulting payment function $b_i^M(q_i; p_i^M) = q_i p_i^M$ depends only on the number of ads shown by each publisher. The profit of the publisher becomes:

$$u_i = q_i p_i^M - \frac{q_i^2}{2} \tag{4}$$

CPA contracts (cost per action) are performance based contracts. In these contracts the advertiser designates a target action to be carried out by a consumer, upon which time a price $p_i^A$ will be paid to the publishers involved in causing the action. The prices are defined as a share of the revenue $x$, yielding the following publisher profit:

$$u_i = (q_1 + q_2)^\rho p_i^A - \frac{q_i^2}{2} \tag{5}$$

The timing of the game is illustrated in Figure 3. The advertiser first decides on a compensation scheme based on the observed efforts $q_i$, performance $x$ or both. The publishers in turn learn the value of the baseline $s$ and make a decision about how many ads $q_i$ to show to the consumers. Consumers respond to ads and convert according to $x(q_1, q_2)$. Finally, the advertiser observes $q_i$ and $x$, compensates each publisher with $b_i$ and payouts are realized.

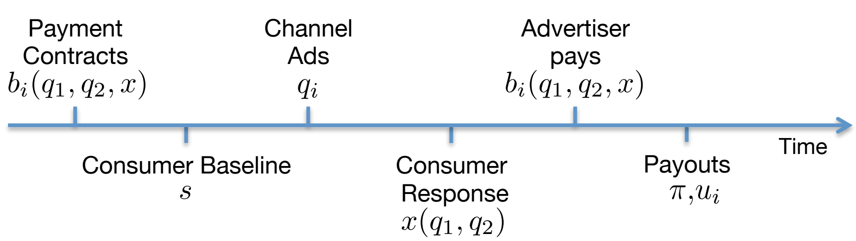



Figure 3: Timing of the Campaign

Several features of the model make the analysis interesting and are considered in the next sections. The first is that the interaction among the publishers is essentially of a team generating conversions. A well known result by Holmstrom (1982) shows that no fixed allocation of output among team members can generate efficient outcomes without breaking the budget. In our model, however, a principal is able to break the budget, yet its goal is profit maximization rather than efficiency. Nonetheless, the externality that one publisher causes on another by showing ads will create moral hazard under a CPA model as will be presented in the next section.

The second feature is that under CPM payment neither the performance of the campaign nor the effect of the baseline enter the utility function of the publishers directly and therefore do not impact a publishers's decision regarding the number of ads to show. Consequently, if the advertiser does not use the performance of the campaign as part of the compensation scheme, adverse selection will arise.

Finally, we note that both the effort of the publishers as well as the output of the campaign are observed by the advertiser. Traditional analysis of team production problems typically assumed one of these is unobservable by the advertiser and cannot be contracted upon. Essentially, CPA campaigns ignore the observable effort while CPM campaigns ignore the observable performance. As we will show, a primary effect of an attribution process is to tie the two together into one compensation scheme.

We now proceed to analyze the symmetric publisher model under CPM and CPA payments. The analysis builds towards the inclusion of an attribution mechanism with a goal of making multi-publisher campaigns more profitable for the advertiser.

# 4 CPM vs. CPA and the Role of Attribution

We start by developing a benchmark that assumes the advertiser is integrated with the publishers. The optimal allocation of ads is found by solving $\max_{q_1,q_2}(q_1+q_2)^\rho - \frac{q_1^2}{2} - \frac{q_2^2}{2}$ yielding

$$q_1^* = q_2^* = \left(\rho \cdot 2^{\rho-1}\right)^{\frac{1}{2-\rho}} \tag{6}$$

which is strictly increasing in $\rho$.

When using CPM based payments, the publisher will choose to show $q_i^M = p_i^M$ ads. Because of symmetry, in equilibrium $q^M = p^M = p_1^M = p_2^M$ and the number of ads displayed is:

$$q^M = p^M = \arg\max_p (2p)^\rho - 2p^2 = \frac{\rho^{\frac{1}{2-\rho}}}{2} \tag{7}$$

In contrast, under a CPA contract, publisher $i$ will choose $q_i$ to solve the first order condition $q_i = \rho(q_i + q_{-i})^{\rho-1} p_i^A$. Invoking symmetry again, we expect $p_1^A = p_2^A$ and $q_1^A = q_2^A$, as a result yielding:

$$q^A = \left(\rho 2^{\rho-1} p^A\right)^{\frac{1}{2-\rho}} \tag{8}$$

We notice that the number of ads displayed in a CPA campaign increases with the price $p^A$ offered to the publishers.

By performing the full analysis and solving for the equilibrium prices $p^M$ and $p^A$ offered by the advertiser we find the following:

**Proposition 1.** *When* $0 < \rho < 2$*:*

- $q^A < q^M < q^*$ *- the level of advertising under CPA is lower than the level under CPM. Both of these are lower than the efficient level of advertising.*
- $\pi^M > \pi^A$ *- the profit of the advertiser is higher when using CPM contracts.*
- *There exists a critical value* $\rho^c$ *with* $0 < \rho^c < 1$ *s.t. for* $\rho < \rho^c$*,* $u^A > u^M$ *and CPA is more profitable for the publishers. When* $\rho > \rho^c$*,* $u^M > u^A$ *and CPM is more profitable for the publishers.*

Proposition 1 shows that using CPA causes the publishers to free-ride and not provide enough effort to generate sales in the campaign. The intuition is that the externality each publisher receives from the other publisher gives an incentive to lower efforts, which consequently lowers total output of the campaign. Under CPM payment, however, publishers do not experience this externality and cannot piggyback on efforts by other publishers. By properly choosing a price for an impression, the advertiser can then incentivize the publishers to show a higher number of ads.

In terms of profits, we observe that advertisers should always prefer to use CPM contracts when multiple publishers are involved in a campaign. This counter-intuitive result stems from the fact that the resulting under-provision of effort overcomes the gains from cooperation by the publishers even when complementarities exist.

The final part of Proposition 1 gives one explanation to the market observation that campaigns predominantly use CPA schemes. When the publishers have market power to determine the payment scheme, *e.g.* the case of Google in the search market, the publishers should prefer a CPA based payment when $\rho$ is small, *i.e.*, when publishers are extreme substitutes. In this case, the possibility for free-riding is at its extreme, and even minute changes in efforts by competing publishers increase the profits of each publisher significantly. For example, if consumers are extremely prone to advertising and a single ad is enough to influence them to convert, any publisher that shows an ad following the first one immediately receives "free" commission. If a search engine which typically arrives later in the buying process of a consumer, is aware of that, it will prefer to use CPA payment to free-ride on previous publisher advertising.

A question that arises is about the motivation of advertisers, in contrast to publishers, to prefer CPA campaigns over CPM ones. The following corollary shows that when advertisers do not take into account the interaction between the publishers, CPA campaigns are also profitable for the advertiser.

**Corollary 1.** *When there is one publisher in a campaign and* $0 < \rho < 2$*:*

- $q^A > q^M$ *iff* $\rho > 1$*: the publisher shows more ads under CPA payment.*
- $\pi^A > \pi^M$ *iff* $\rho > 1$*: more revenue and more profit is generated for the advertiser when using CPA payment and advertising has increasing returns (*$\rho > 1$*).*

Corollary 1 reverses some of the results of Proposition 1 for the case of one publisher campaigns. Since free-riding is not possible in these campaigns, we find that CPA campaigns better coordinate the publisher and the advertiser when ads have increasing marginal returns, while CPM campaigns are more efficient for decreasing marginal returns.

### 4.1 The Role of Attribution

An attribution process in a CPA campaign allocates the price $p^A$ among the participating publishers in a non-fixed method. We model the attribution process as a two-dimensional function $f(q_1, q_2, x) = (f_1, f_2)$ that allocates a share of a conversion to each of the players respectively. When publishers are symmetric and the baseline is zero, candidates for effective attribution functions will exhibit the following properties:

- Efficiency - The process will attribute all conversions to the two publishers: $f_1 + f_2 = 1$.

- Symmetry - If both publishers exhibit the same effort ($q_1 = q_2$) then they will receive equal attribution: $f_1(q, q, x) = f_2(q, q, x) = \frac{1}{2}$.
- Pay to play (Null Player) - Publishers have to invest to get credit. When a publisher does not show any ads, it will receive zero attribution: $f_i(q_i = 0, q_{-i}, x) = 0$.
- Marginality - Publishers who contribute more to the conversion process should receive higher attribution: if $q_1 > q_2$ then $f_1 \geq f_2$.

Although these properties are straightforward, they limit the set of possible functions that can be used for attribution. We also assume that $f(\cdot)$ is continuously differentiable on each of its variables.

The profit of each publisher in a CPA campaign can now be written as:

$$u_i^A = f_i(q_i, q_{-i}, x)x(q_1, q_2)p^A - \frac{q_i^2}{2} \tag{9}$$

An initial observation is that the process creates a contest between the two publishers for credit. Once ads have been shown, the investment has been sunk yet credit depends on delayed attribution. It is well known (see, *e.g.*, Sisak (2009) and Konrad (2007)) that contests will elicit the agents to overexert effort in equilibrium compared to a non-contest situation. As a result the attribution process can be used to incentivize the publishers to increase their efforts and show a number of ads closer to the integrated market levels.

In the next section we analyze the impact of the commonly used last-touch attribution method, and compare it to a new method based on the Shapley value we developed to attribute performance in online campaigns.

# 5 Last-Touch and Shapley Value Attribution

Advertiser surveys report that last-touch attribution is the most widely used process in the industry. This process gives 100% of the credit for conversion to the last ad displayed to a consumer before conversion. From the point of view of the publisher, if the consumer visits both publisher sites, last-touch attribution creates a noisy contest in which the publisher cannot fully predict whether it will receive credit by showing a specific impression. Even if the publisher can predict the equilibrium behavior of the other publisher and expect the number of ads shown by the other publisher, it has little knowledge of the timing of these ads, and in addition it cannot fully predict the timing of a consumer purchase.

Consequently, we model the process as a noisy contest. The noise in the contest models the uncertainty the publisher has about whether a consumer is about to purchase the product or not, and whether they will visit the site again in the future. We let $\varepsilon_i$ denote the uncertainty of publisher $i$ with respect to its ability to win the attribution process. When publisher $i$ shows $q_i$ ads it will receive credit only if $Pr(q_1\varepsilon_1 > q_2\varepsilon_2)$. In a static model this captures the effect of showing an additional ad by the publisher. By assuming that $\varepsilon_i$ are uniformly i.i.d on $[1, d]$ for $d > 1$, we can define the last-touch attribution function as following:

$$f_i^{LT}(q_i, q_{-i}) = Pr(q_i\varepsilon_i > q_{-i}\varepsilon_{-i}) = \int_1^d G\left(\frac{q_i}{q_{-i}}\varepsilon\right) g(\varepsilon)d\varepsilon \tag{10}$$

when $G(\cdot)$ is the CDF of the uniform distribution on $[1, d]$ and $g(\cdot)$ its PDF.

The value of $d$ measures the amount of uncertainty the publishers have about the consumer's behavior in terms of future visits and purchases, and will be the focus of our analysis of Last-Touch attribution. Higher values of $d$, for example, can model consumers who visit both publishers with very high frequency, allowing both of them to show many ads to the consumer. Lower values of $d$ make the contest extremely discriminating, having a "winner-take-all" effect on the process. In such cases, the publishers can time their ads exactly to be the last ones to be shown, and as a result compete fiercely for attribution. A natural extension which is left for future work is to allow asymmetric values of $d$ among the publishers. This will allow modeling of publishers who have an advantage in timing their advertising to receive credit, although their ads may have the same effectiveness.

Two noticeable properties of last-touch attribution are due discussion. The first is that the more ads a publisher will show, the higher probability it has of being the last one to show an ad before a consumer's purchase. Last-touch attribution therefore has the Marginality property described above. It also trivially has the 3 other properties. The second property is that last-touch attribution makes use of the conversion rate only in a trivial manner. The credit given to the publisher only depends on the number of ads shown to a consumer and whether the consumer had converted. It does not depend on the actual conversion rate of the consumer and therefore ignores the value of $x$.

It is useful to examine the equilibrium best response of the publishers in a CPA campaign in order to understand the impact of last-touch attribution on the quantities of ads being displayed. Recall that when no attribution is used, the publisher will display $q$ ads according to the solution of:

$$(2q)^{\rho-1}\rho p^A = q \tag{11}$$

When using last-touch attribution, a publisher faces a winner-take-all contest which increases its marginal revenue when receiving credit for the conversion, even if the conversion rate remains the same. In a CPA campaign the first order condition in a symmetric equilibrium becomes:

$$(2q)^{\rho-1}\left(2f_1'(1)+\frac{1}{2}\rho\right)p^A = q \tag{12}$$

where $f'(1)$ is the marginal increase in the share of attribution when showing an additional ad when $q_1 = q_2$. Comparing equations (11) and (12) we see that if $\left(2f_1'(1)+\frac{1}{2}\rho\right) > \rho$, then the publisher faces a higher marginal revenue for the same amount of effort. As a result it will have an incentive to increase its effort in equilibrium when the conversion function is concave compared to the case when no attribution was used. Gershkov et al. (2009) show conditions under which such a tournament can achieve Pareto-optimal allocation when symmetric team members use a contest to allocate the revenue among themselves. Whether this contest is sufficient to compensate for free-riding in online campaigns remains yet to be seen.

To answer this question we are required to perform the full analysis that considers the price $p^A$ offered by the advertiser in equilibrium. In addition, the accuracy of the attribution process which depends on the magnitude of the noise $d$ has an impact and may yield exaggerated effort by each publisher. Finally, the curvature of the conversion function $x$ that depends on the parameter $\rho$ may also influence the efficiency of last-touch attribution.

When performing the complete analysis for both CPA and CPM campaigns, we find the following:

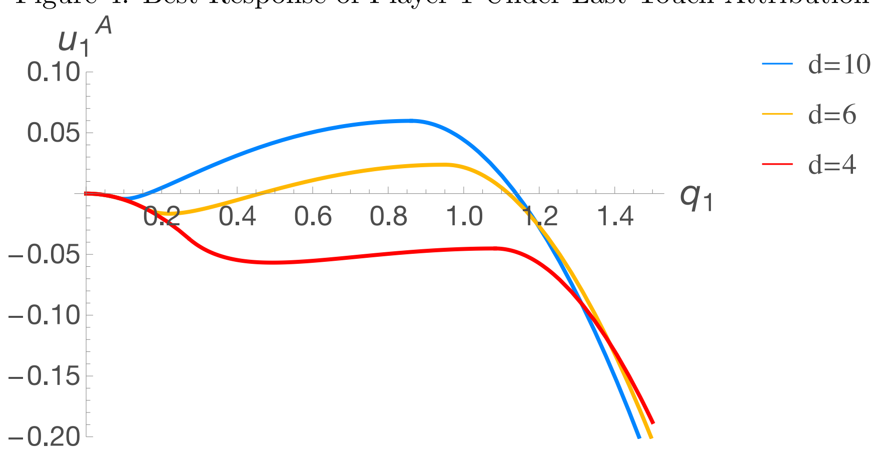
**Proposition 2.** *When $0 < \rho < 2$ and last-touch attribution is being used:*

- *In a CPA campaign a symmetric pure strategy equilibrium exists for $0 < \rho < 2 - \frac{4}{d-1}$. In this equilibrium $q^{A-LT} = \left(\frac{\rho}{2}2^{\rho-1}(\frac{d+1}{d-1}+\frac{\rho}{2})\right)^{\frac{1}{2-\rho}}$.*

- *For any noise level $d$, $q^A < q^M < q^{A-LT}$.*

Proposition 2 shows surprising findings about the impact of last-touch attribution on different campaign types. The contest among the publishers has a symmetric pure strategy equilibrium in a CPA campaign when $\rho$ is low enough or when the noise $d$ is high enough. In these cases, more advertising is being shown in equilibrium compared to regular CPM and CPA campaigns, and more revenue will be generated by the campaign. As a result, the advertiser may make higher profit compared to the case of no attribution as well as for the case of CPM campaigns with no attribution. To understand the impact of low noise, we focus on the case of $d < 3$. In this case, the contest is too discriminating and the effort required from the publishers in equilibrium is too high to make positive profit, and publishers would prefer not to participate. Figure 4 illustrates the best-response of publisher 1 to publisher's 2 equilibrium strategy to give intuition for this result. When the noise becomes small and the contest too discriminating, the best-response function loses the property of having a maximum point which yields positive profit as a result of too strong competition for attribution.

Figure 4: Best Response of Player 1 Under Last-Touch Attribution




Publisher's 1 best response to publisher's 2 strategy of showing $q^{A-LT}$ ads when $\rho = 1$.

Finally, a comparison of the profits the advertiser makes with and without last-touch attribution yields the following result:

**Corollary 2.** *When $0 < \rho < 2 - \frac{4}{d-1}$, $\pi^{A-LT} > \pi^M > \pi^A$ and the advertiser makes higher profit under last-touch attribution.*

### 5.1 The Shapley Value as an Attribution Scheme

The Shapley value (Shapley 1952) is a cooperative game theory solution concept that allocates value among players in a cooperative game. A cooperative game is defined by a characteristic function $x(q_1, \ldots, q_M)$ that assigns for each coalition of players and their contribution $q_i$ the value they created. For a set of $M$ publishers, the Shapley value is defined as following:[8]

$$\phi_i(x) = \sum_{S \subset (M \setminus i)} \frac{|S|!(|M| - |S| - 1)!}{|M|!}(x_{S \cup i} - x_S) \tag{13}$$

where $M$ is the set of publishers and $x$ is the set of conversion rates for different subsets of publishers.

The value has the four properties mentioned in the previous section: Efficiency, Symmetry, Null Player and Marginality.[9] In addition, it is the unique allocation function that has these properties with the addition of an additivity property over the space of cooperative games defined by the conversion function $x(\cdot)$. For the case of two publishers $M = 2$ the Shapley value reduces to:

$$\phi_1 = \tfrac{x(q1+q2) - x(q_2) + x(q_1) - 0}{2} \quad \phi_2 = \tfrac{x(q1+q2) - x(q_1) + x(q_2) - 0}{2} \tag{14}$$

Using the Shapley value has the benefit of directly using the marginal contribution of the publishers to compensate them. In addition, the process's accuracy does not depend on exogenous noise and yields a pure strategy equilibrium for all values of $\rho$.

In a CPA campaign, the profit of a publisher will become: $u_i^{A-S} = \phi_i p^{A-S} - \frac{q_i^2}{2}$.

Solving for the symmetric equilibrium strategies and profits of the advertiser and publishers yield the following result:

**Proposition 3.**

*When $0 < \rho < 2$, using the Shapley value for attribution yields $q^{A-S} = \left(\frac{\rho^2}{4}(2^{\rho-1} + 1)\right)^{\frac{1}{2-\rho}}$.*

*For $\rho < 2 - \frac{4}{d-1}$, $q^A < q^{A-S} < q^{A-LT}$.*

*The profit of the advertiser is higher under Shapley value than under Last-Touch attribution iff $q^{A-S} > q^{A-LT}$,* i.e. *$d < \frac{4}{2-\rho} + 1$.*

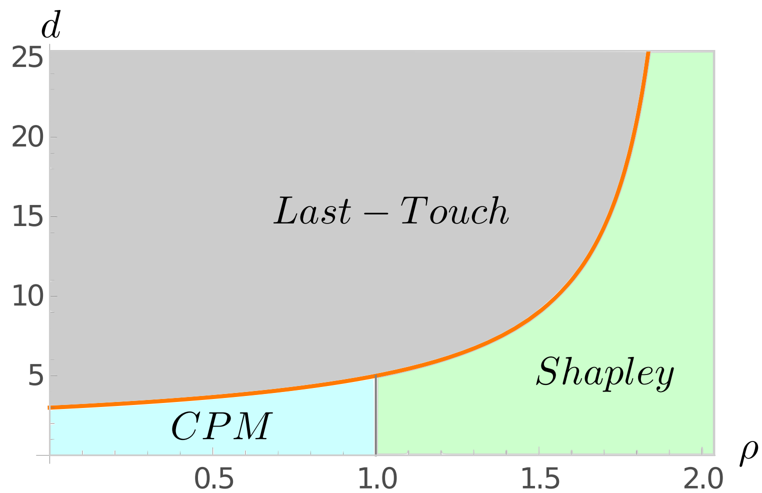
[8] This is a continuous version of the value.

[9] Some of these properties can be shown to be derived from others.

*The profit of the publisher is higher under the Shapley value attribution than under regular CPM pricing iff $\rho > 1$.*

Proposition 3 is a major result of this paper, showing that the Shapley value can be more profitable when publishers are complements. Contrary to Last-touch attribution, a symmetric pure strategy equilibrium exists for any value of $\rho$, including very convex functions. When considering lower values of $\rho$ for which Last-Touch attribution improves the efficiency of the campaign, we see that when the noise level $d$ is low enough, the Shapley value will yield better results for the advertiser if $\rho > 1$, while CPM will be better when $\rho < 1$. Figure 5 depicts for which values of $\rho$ and $d$ is each attribution and compensation scheme more profitable.

Figure 5: Profitability of Each Compensation Scheme




Values of $\rho$ and $d$ for which each compensation scheme is more profitable for the advertiser.

The intuition behind this result can be illustrated best for extreme values of $\rho$. When $\rho < 1$ and is extremely low, the initial ads have the most impact on the consumer. As a result, there will be significant free-riding which Last-touch is best suited to solve, while the marginal increase that the Shapley value allocates is not too high. When $\rho > 1$, however, if the noise is low enough, the publishers will be inclined to show too many ads because of the low uncertainty about their success of being the last one to show an ad. In essence, the competition is too strong and overcompensates for free-riding. The Shapley value in this case is better suited to incentivize the players as the marginal increase between two symmetric publishers to one is highest with a convex function.

To make use of the Shapley value in an empirical application, it is required that the advertiser can observe the conversion rates of consumers who were exposed to publisher 1 solely, publisher 2 solely and to both of them together. In addition, when a baseline is present, it cannot be assumed that not being exposed to ads yields no conversions.

The next section discusses the baseline and the use of experimentation to generate the data required to calculate the Shapley value.

## 6 Baselines and Experiments

In this Section we relax the assumption that the baseline $s = 0$ and examine its impact on the performance of the attribution schemes, and methods to fix this impact. When the baseline is non-zero, the advertiser cannot discern from conversions whether they were caused by advertising effects or simply because consumers had other reasons for converting. As publishers have more information about consumers reaching their sites, this private information may cause adverse selection - publishers can target consumers with high baselines to receive credit for those conversions.

Specifically, if we consider again equation (12) the first order condition of an advertiser showing $q$ ads to all consumers now becomes:

$$\left[(2q)^{\rho-1}\left(2f_1'(1)+\frac{1}{2}\rho\right)(1-s)+f'(1)s\right]p^A = q \tag{15}$$

In the extreme case of $s = 1$, the publishers will elect to show advertising to baseline consumers and be attributed credit.

To understand how experimentation may be beneficial for the advertiser in light of this problem, we analyze a model with a single publisher, but now assume the baseline is non-zero and known to the publisher. We also assume $\rho = 1$, and recall that $s$ is distributed $Beta(\alpha, \beta)$. Thus, if all consumers are exposed to $q$ ads, the expected observed number of converters will be $N\left(s + q(1-s)\right)$. We note however that if non-baseline consumers are not exposed to ads at all, the advertiser would still expect to observe $N\left(s + q(1-s)\right)$ converters.

When the advertiser is integrated with the publisher and can target specific consumers, it can choose to show $q_b$ ads to baseline consumers and $q$ ads to the non-baseline consumers. If the cost of showing $q$ ads to a consumer is $\frac{q^2}{2}$ the firm's profit from advertising is:

$$\pi(q, q_b; s) = N\left(s + q(1-s) - s\frac{q_b^2}{2} - (1-s)\frac{q^2}{2}\right) \tag{16}$$

The insight gained from this specification is that when consumers have a high baseline, the advertiser has a smaller population to affect with its ads, as consumers in the baseline would convert anyway.

It is obvious that when the advertiser can target consumers exactly, it has no reason to show ads to baseline consumers, and therefore will set $q_b = 0$. The allocation of ads that maximizes the advertiser's profit under full information is then $q^* = 1$ and $q_b^* = 0$, while the total number of ads shown will be $N(1-s)$. We call this strategy the *optimal* strategy and note that the number of ads to show decreases in the magnitude of the baseline. The profit achieved under the optimal strategy is $\pi^{max} = N\frac{\mu+1}{2}$ when $\mu$ is the expectation of $s$.[10] This profit increases with $\alpha$, and decreases with $\beta$. This means that when higher baselines are more probable in terms of mass above the expectation, a higher profit is expected.

Turning to the case of a firm with uncertainty about $s$, one approach the firm may choose is to maximize the expected profit over $s$ by showing a number of ads $q$ to all consumers independent of the baseline. This *expected* strategy solves:

$$\max_q \mathbb{E}_s[\pi(q, q; s)] \tag{17}$$

The achieved profit in this case can serve as a lower bound $\pi^{min}$ on profit the firm can achieve in the worst case. Any additional information is expected to increase this profit; if it does not, the firm can opt to choose the expected strategy.

The following result compares the expected strategy with the optimal one:

**Lemma 1.** *Let* $q^E = \arg\max_q \pi(q, q; s)$. *Then:*

- *The firm will choose to show* $q^E = 1 - \mu$ *ads when using the expected strategy.*
- *The firms profit,* $\pi^{min}$ *is lower than* $\pi^{max}$ *by* $\frac{N}{2}\left(\mu - \mu^2\right)$.

[10] $\mu = \frac{\alpha}{\alpha+\beta}$

Lemma 1 posits that the number of ads displayed using this strategy treats the market as if $s$ equals its expected value. As a result, the achieved profit increases with the expected value of $s$. When this strategy is the only one available the value of full information to the firm is highest when the expected baseline is close to 1/2.

The most common strategy that firms employ in practice, however, is to learn the value of $s$ through experimentation. The firm can decide to not show ads to $n < N$ consumers and observe the number of converters in the sample. This information is then used to update the firm's belief about $s$ and maximize $q$. We call this strategy the *learning* strategy.

When the firm observes $k$ converters in the sample it will base the number of ads to show on this updated belief (DeGroot 1970). The expected profit of the firm in this case is:

$$n\mathbb{E}_s[x(q=0;s)] + (N-n)\mathbb{E}_s\left[\mathbb{E}_{k|s}\left[\max_q \pi(q;s)|s\right]\right] \tag{18}$$

The caveat here is that by designating consumers as the sample set, the firm forfeits potential added profit from showing ads to these consumers. We are interested to know when this strategy is profitable, and also how much can be gained from using it and under what conditions.

Let $n^*$ denote the optimal sample size that maximizes (18) given the distribution of $s$. As the distribution of the observed converters $k$ is $Bin(n,s)$, the posterior $s|k$ is distributed $Beta(\alpha + k, \beta + n - k)$. Using Lemma 1, the optimal number of ads to show when observing $k$ converters becomes $q^*(\mu(k))$ when $\mu(k) = \mathbb{E}_s[s|k] = \frac{\alpha+k}{\alpha+\beta+n}$. A comparative statics analysis of the optimal sample size $n^*$ shows the following behavior:

**Lemma 2.** *The optimal sample size $n^*$:*

- *Is positive when the population $N$ is larger than $\frac{\beta(\alpha+\beta)(1+\alpha+\beta)}{\alpha} = \beta\frac{1-\mu}{\sigma^2}$.*
- *Increases with $N$ and decreases with $\beta$.*
- *Decreases in $\alpha$ when $\alpha$ is large .*

Lemma 2 shows that unless the distribution of $s$ is heavily skewed towards 0 by having a large $\beta$ parameter, even with small populations some experimentation can be useful. On the flip side, when the distribution is heavily skewed towards 1 with very large $\alpha$, the high probability baseline makes it less valuable to experiment, and the optimal sample size decreases.

Having set conditions for the optimal size of the sample during experiments, we now revisit our question: when is it profitable for the firm to learn compared to choosing an expected strategy. Our finding is that for a large enough population $N$, it is always more profitable to learn than to use an expected strategy:

**Proposition 4.** *When $N > \beta\frac{1-\mu}{\sigma^2}$, learning yields more profit than the expected strategy.*

To exemplify this result, if $\alpha = \beta = 1$, then the baseline $s$ is distributed uniformly over $[0, 1]$. In this case, it is enough for the population to be larger than 6 consumers for experimentation to be profitable. [11]

## 6.1 Baseline Exploitation

When the advertiser is not integrated with the publisher, the publisher has a choice of which consumers to target and how many ads to show to each segment. We can solve for the behavior of an advertiser under CPM and CPA pricing in this special case without attribution to get the following result:

**Proposition 5.**

- *Under CPM the publisher will show $q^M = \frac{1-\mu}{2}$ ads to each consumer in both segments.*
- *Under CPA, the publisher will show $q^A = \frac{2\mu-1}{2\mu-2}$ when $\mu < \frac{1}{2}$. The ads will be shown to consumers only in the non-baseline segment. When $\mu > \frac{1}{2}$, the advertiser will opt to not use CPA at all.*
- *Under CPA the publisher will show a total number of ads which is higher than the efficient number $q^*$, as well as higher than $q^M$, for every value of $s$.*
- *The profit of the advertiser under CPM is higher than under CPA for any value of $\mu$.*

Proposition 5 exposes two seemingly contradicting results. Since under CPM payment the publisher is paid for the amount of ads it shows, it will opt to show both $q > 0$ and $q_b > 0$ ads. Given the same price and cost for each ad displayed, it will show exactly the same amount to both segments, which will be lower than the efficient amount of ads to show. Specifically, when

[11] This result assumes $n$ is continuous. As $n$ is discrete, the actual $n^*$ is slightly larger than this bound to allow for discrete sizes of samples.

$\mu$ is high, *i.e.*, the expectation of the baseline is high, the publisher will lower its effort as the advertiser would have wanted. Under CPA, however, the publisher will use an efficient allocation of ads in terms of targeting and will not show ads to the baseline population. Since the publisher gets a commission from the baseline as well, however, it experiences lower effective cost for each commission payment, and as a result will show too many ads compared to the optimal amount. The apparent contradiction may be that although the publisher now allocates its ads correctly under CPA compared to CPM, the profit of the advertiser is still higher under CPM payment for low baseline values. The intuition is that CPM allows the advertiser to internalize the strategy of the publisher and control it through the price, while in CPA the advertiser will need to trade-off effective ads for ineffective exploitation of the baseline if it lowers the price paid per conversion.

Adding Last-Touch attribution to the CPA process will only exacerbate the issue. If the publisher will show a different number of ads in each segments, the advertiser can infer which segment may be the baseline one and not compensate the publisher for it. The publisher, as a result, will opt to show the same number of ads to all consumers, and the number of ads shown will now depend on the size of the baseline population $s$. The result will be too many ads shown by a CPA publisher to the entire population, and reduced profit to the advertiser.

Using the Shapley value, in contrast, will allocate revenue to the publisher only for non-baseline consumers, as the Shapley value will control for the observed baseline through experimentation. When solving for the total profit of the advertiser including the cost of experimentation, it can be shown that Shapley value attribution in a CPA campaign reaches a higher profit than CPM campaigns.

We thus advocate moving to an attribution process based on the Shapley value considering the adverse effects of the baseline. The next section discusses a preliminary analysis of data from an online campaign using Last-Touch attribution to detect whether baseline exploitation is indeed occurring.

# 7 An Application to Online Campaigns

This section applies the insights from Sections 4, 5 and 6 to data from a large scale advertising campaign for car rental in UK.

The campaign was run during April and May 2013 and its total budget exceeded US $\$65,000$

while utilizing 8 different online publishers. These publishers include two online magazines, two display (banner) ad networks, two travel search websites, an online travel agency and a media exchange network. During the campaign more than 13.4 million online consumers[12] were exposed to more than 40.4 million ads.

The summary of the campaign results in Table 1 shows that the campaign more than quadrupled conversion rates for the exposed population.

| Ad Exposure | Population | Converters | Conversion Rate |
|---|---|---|---|
| Exposed | 13,448,433 | 6,030 | 0.045% |
| Not Exposed | 144,745,194 | 15,087 | 0.010% |
| Total | 158,193,627 | 21,117 | 0.013% |

Table 1: Performance of Car Rental Campaign in the UK

To associate the return of the campaign the advertiser computed last-touch attribution for the publishers based on the last ad they displayed to consumers. Table 2 shows the attributed performance alongside the average cost per attributed conversion. We see that the allocation of budgets correlates with the attributed performance of the publishers, while the cost per conversion can be explained by different average sales through each publisher and quantity discounts.[13]

| Publisher No. | Type | Attribution | Budget ($) | Cost per Converter ($) |
|---|---|---|---|---|
| 1 | Online Magazine | 386 | 8,300 | 21.50 |
| 2 | Travel Agency | 218 | 8,000.02 | 36.69 |
| 3 | Travel Magazine | 40 | 6,000 | 150 |
| 4 | Display Network | 168 | | |
| 5 | Travel Search | 50 | | |
| 6 | Display Network | 1,330 | 13,200 | 9.92 |
| 7 | Travel Search | 69 | | |
| 8 | Media Exchange/Retargeting | 3,769 | 33,200 | 8.80 |
| Total | | 6,030 | 68,700 | 11.39 |

Table 2: Last Touch Attribution for Car Rental Campaign

We observed that in order to achieve high profits, the advertiser needs to be able to condition payment on estimates of the baseline as well as on the marginal increase of each publisher over the sets of other publishers. This result extends to the case of many publishers, where for a set of publishers $M$ the advertiser will need to observe and estimate $2^{|M|}$ measurements.

[12] An online consumer is measured by a unique cookie file on a computer.

[13] Publisher number 3 targets business travelers and yields more profit per attributed conversion.

| Publishers | logdiff coef | se |
|---|---|---|
| 1 | -0.657 | (0.849) |
| 2 | -2.175*** | (0.693) |
| 3 | -1.960*** | (0.703) |
| 4 | -0.986 | (0.751) |
| 5 | -1.559** | (0.691) |
| 6 | -1.689** | (0.744) |
| 7 | -0.588 | (0.748) |
| 8 | -0.539 | (0.813) |
| $R^2$ | 0.650 | |
| Observations | 88 | |

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3: Logit Estimates of Publisher Effectiveness

Even small campaigns utilizing 7 publishers require more than 100 of these estimates to be used and reported. Current industry practices do not allow for such elaborate reporting resulting in advertisers using statistics of these values. The common practice is to report one value per publisher with the implicit assumption that if a publisher's attribution value is higher, so is its effectiveness.

## 7.1 Evidence of Baseline Exploitation and Detection of Free-Riding

Section 6 shows that publishers can target high baseline consumers to deceive the advertiser regarding their true effectiveness. To test the hypothesis that publishers target high baseline consumers, Table 3 shows the results of the logit estimates on the market share differences of each publisher combination in our data.[14] The estimate shows that no publisher adds a statistically significant increase in utility for consumers compared to the baseline. More surprising is the result that a few publishers seem to decrease the response of consumers, thus supporting our hypothesis.

Section 5 predicts that using a last-touch method will lead publishers to strategically increase the number of ads shown, while attempting to free-ride on others. If publishers were not attempting to game the last-touch method, we would expect to see their marginal contribution estimates be close to their last-touch attribution in equilibrium. An issue that arises with using marginal estimates from the data, however, is that the timing of ads being displayed is endogenous and depends on a

[14]The estimation technique is described in Appendix C.

decision by the consumer to visit a publisher and by the publisher to display the ad. The advertiser does not observe and cannot control for this order, which might raise an issue with using ad view data as created by random experimentation.

The use of the Shapley value, however, gives equal probability to the order of appearance of a publisher when a few publishers show ads to the same consumers. The effect is a randomization of order of arrival of ads when multiple ads are observed by the same consumer. Because of this fact, using the Shapley value as is to estimate marginal contributions will be flawed when not every order of arrival is possible. For example, the baseline effect needs to be treated separately while special publishers such as retargeting publishers and search publishers that can only show ads based on specific events need to be accounted for. An additional hurdle to using the Shapley value is the computation time required as it is exponential in the size of the input.

We developed a modified Shapley value estimation procedure to handle these issues. The computational issues are addressed by using specific structure of the advertising campaign data and will be described in Berman (2013).

Figure 6: Last Touch Attribution vs. Shapley Value

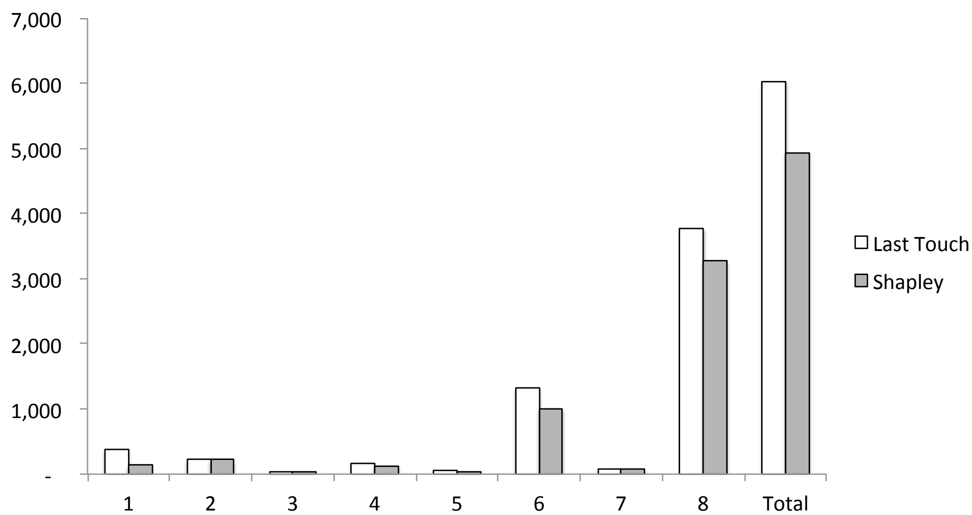



Figure 6 compares the results from a last-touch attribution process to the Shapley value estimation.

More than $1,000$ converters were reallocated to the baseline. In addition, a few publishers lost significant shares of their previously attributed contributions, showing evidence of baseline exploitation. Using these attribution measures the advertiser has reallocated its budgets and significantly lowered its cost per converter. We are currently collecting the data on the behavior of the publishers given this change in attribution method, to be analyzed in the future.

## 8 Conclusion

As multi-publisher campaigns become more common and many new publisher forms appear in the market, attribution becomes an important process for large advertisers. The more publishers are added to a campaign, however, the more complex and prone to errors the process becomes. Our two-publisher model has identified two issues that are detrimental to the process – free-riding among team members and baseline exploitation. This measurement issue arises because the data does not allow us to disentangle the effect of each publisher accurately and using statistics to estimate this effect gives rise to free riding. Thus, setting an attribution mechanism that does not take into account the equilibrium behavior of publishers will give rise to moral hazard even when the actions of the publishers are fully observable. On the other hand, if the performance of the campaign is not explicitly used in the compensation scheme through an attribution mechanism, adverse selection cannot be mitigated and ineffective publishers will be able to impersonate as effective ones.

The method of last-touch attribution, as we have showed, has the potential to make CPA campaigns more efficient than CPM campaigns under some conditions. In contrast, attribution based on the Shapley value yields well behaved pure strategy equilibria that increase profits over last-touch attribution when the noise is not too small. Adding experimentation as a requirement to the contract does not lower the profits of the advertiser too much, and allows for collection of the information required to calculate the Shapley value, as well as estimating the magnitude of the baseline.

The analysis of the model and the data has assumed homogenous consumers. If the population has significant heterogeneity, which is observed by the publishers but not by the advertiser, the marginal estimates will be biased downwards, as the publishers will be able to truly target consumers they can influence. Another issue that arises from the analysis is that publishers may have access to exclusive customers who cannot be touched by other publishers.

Exclusivity can be handled well by our model as a direct extension. In those campaigns where a publisher has access to a large exclusive population, it may be beneficial to switch from CPM to CPA campaigns, or vice versa, depending on the overlap of other populations with other publishers.

To handle the heterogeneity of the baseline and consumers, we propose two solutions. To understand whether the baseline estimation affects the results significantly, we can compare the Shapley Value estimates with and without the baseline. In addition, the data include characteristics of consumers which can be used to estimate the baseline heterogeneity, and control for it when estimating the Shapley Value. Propensity Score Matching is a technique that will allow matching sets of consumers who have seen ads to similar consumers who have not seen ads and estimate the baseline for each set. One issue with this approach is that consumer data may include thousands of parameters per consumer including demographics, past behavior, purchase history and other information. Our tests have shown that using regularized regression as a dimensionality reduction technique performs well in this setting, and work is underway to implement it with a matching technique.

This study has strong managerial implications in that it identifies the source of the attribution issue that advertisers face. Advertisers today believe that if they improve their measurement mechanism campaigns will become more efficient. This conclusion is only correct if the incentive scheme based on this measurement is aligned with the advertiser's goal. If it is not, like last-touch methods, the resulting performance will be mediocre at best.

A key message of this paper is that performance based incentive schemes require a good attribution method to alleviate moral hazard issues. The observations that proper estimates of marginal contributions as well as a proof based mechanism can solve these issues when employed together creates a path for solving this complex problem and providing advertisers with better performing campaigns.

## References

Berman, Ron. 2013. An application of the shapley value for online advertising campaigns. *Work in Progress* .

Blake, Thomas, Chris Nosko, Steven Tadelis. 2013. Consumer heterogeneity and paid search effectiveness: A large scale field experiment. *NBER Working Paper* 1–26.

DeGroot, Morris H. 1970. Optimal statistical decisions .

Dreze, Xavier, François-Xavier Hussherr. 2003. Internet advertising: Is anybody watching? *Journal of interactive marketing* **17**(4) 8–23.

Gershkov, Alex, Jianpei Li, Paul Schweinzer. 2009. Efficient tournaments within teams. *The RAND Journal of Economics* **40**(1) 103–119.

Ghose, Anindya, Sha Yang. 2009. An empirical analysis of search engine advertising: Sponsored search in electronic markets. *Management Science* **55**(10) 1605–1622.

Holmstrom, Bengt. 1982. Moral hazard in teams. *The Bell Journal of Economics* 324–340.

Kireyev, Pavel, Koen Pauwels, Sunil Gupta. 2013. Do display ads influence search? attribution and dynamics in online advertising. *Working Paper* .

Konrad, Kai A. 2007. Strategy in contests: An introduction. Tech. rep., Discussion papers//WZB, Wissenschaftszentrum Berlin für Sozialforschung, Schwerpunkt Märkte und Politik, Abteilung Marktprozesse und Steuerung.

Lambrecht, Anja, Catherine Tucker. 2011. When does retargeting work? timing information specificity. *Timing Information Specificity (Dec 02, 2011)* .

Lewis, Randall A, Justin M Rao. 2012. On the near impossibility of measuring advertising effectiveness. Tech. rep., Working paper.

Li, Alice, P.K. Kannan. 2013. Modeling the conversion path of online customers. *Working Paper* .

Manchanda, Puneet, Jean-Pierre Dubé, Khim Yong Goh, Pradeep K Chintagunta. 2006. The effect of banner advertising on internet purchasing. *Journal of Marketing Research* 98–108.

McAfee, R. Preston, John McMillan. 1991. Optimal contracts for teams. *International Economic Review* 561–577.

Rutz, Oliver J., Randolph E Bucklin. 2011. From generic to branded: A model of spillover in paid search advertising. *Journal of Marketing Research* **48**(1) 87–102.

Shapley, Lloyd S. 1952. A value for n-person games .

Sherman, Lee, John Deighton. 2001. Banner advertising: Measuring effectiveness and optimizing placement. *Journal of Interactive Marketing* **15**(2) 60–64.

Sisak, Dana. 2009. Multiple-prize contests–the optimal allocation of prizes. *Journal of Economic Surveys* **23**(1) 82–114.

Tucker, Catherine. 2012. The implications of improved attribution and measurability for online advertising markets .

Yao, Song, Carl F. Mela. 2011. A dynamic model of sponsored search advertising. *Marketing Science* **30**(3) 447–468.

Zhu, Yi, Kenneth C Wilbur. 2011. Hybrid advertising auctions. *Marketing Science* **30**(2) 249–273.

# A Extension - Asymmetric Publishers

We briefly overview the modeling of asymmetric effectiveness of publishers and results about the impact on campaign effectiveness.

When publishers are asymmetric the advertiser may want to compensate them differently depending on their contribution to the conversion process. If we assume the advertiser has full knowledge of the effectiveness level of each publisher, we can treat publisher one's effectiveness as fixed, and use the relative performance of publisher two as influencing its costs. Specifically, we let the cost of publisher two be $\frac{q_2^2}{2\theta}$.

When $\theta = 1$, we are back at the symmetric case. When $\theta < 1$, for example, publisher one is more effective as its costs of generating a unit of contribution to conversion are lower. [15]

Solving for the decision of the publishers and the advertiser under CPM and CPA contracts yields the following results:

**Proposition 6.** *When publishers are asymmetric:*

- *Under a CPM contract the same price* $p^M = \left(\frac{\rho(\theta+1)^{\rho-1}}{2}\right)^{\frac{1}{2-\rho}}$ *per impression will be offered to both publishers.*
- *Under a CPA contract, if* $\theta < 1$*, the advertiser will contract only with publisher one. If* $\theta > 1$*, the advertiser will only contract with publisher two.*

We observe that asymmetry of the publishers creates starkly different incentives for the advertiser and the publishers. Under CPM campaigns having more effective publishers in the campaign increases the price offered by the advertiser to all publishers. As a result publisher one will benefit when a better publisher joins the campaign yet will suffer when a worse one joins.

Performance based campaigns using CPA, in contrast, make the advertiser exclude the worst performing publisher from showing ads. The intuition is that because conversions are generated by symmetric "production" input units of both publishers, the advertiser may just as well buy all of the input from the publisher who has the lowest cost of providing them. The only case when it is optimal for the advertiser to make use of both publishers is when $\theta = 1$ and they are symmetric.

[15] It should be noted that this specification is equivalent to specifying the costs as being equal while the conversion function being $x(q_1, \zeta q_2)$ for some value $\zeta$.

Using a single publisher is significantly less efficient when two are available to the publisher. Adding an attribution process creates an opportunity for this shut-out publisher to compensate for its lower effectiveness with effort. The resulting asymmetric equilibrium is currently under investigation to understand the ramifications of the attribution process on such a campaign.

### A.1 Asymmetric Information with Asymmetric Publishers

When the publishers may be asymmetric yet their relative asymmetry is unknown to the advertiser, the problem exhibits adverse selection. The mechanism design literature has dealt with similar scenarios when either moral hazard is present, *i.e.*, the effort of publishers is unobserved, or with a scenario when both moral hazard and adverse selection are present. A novel result by McAfee and McMillan (1991) has developed second-best mechanisms for the case of team production when agents are complements in production.

We extend this result to the case of substitute production and note that publishers can be seen as contributing a measure of output we call *efficiency units* to the performance of the campaign $x$ (McAfee and McMillan 1991). This measure of input to the advertising process is not observed by the advertiser, but the mechanism will elicit optimal choice of efficiency units in equilibrium.

Let $y_i = \theta_i q_i$ be the output of publisher $i$ measured in efficiency units, and let $y = (y_1, y_2)$, $\theta = (\theta_1, \theta_2)$ be the vectors of efficiency units and effectiveness of publishers. We assume $\theta_i \sim U[0, 1]$. Then the expected observed performance will be $x(y) = (y_1 + y_2)^\rho$, and the cost of each publisher will be $c(y_i, \theta_i) = \frac{y_i^2}{2\theta_i}$.

We denote by $\gamma(y_i, \theta_i) = c(y_i, \theta_i) - (1 - \theta_i)\frac{\partial c}{\partial \theta}(y_i, \theta_i)$. This is the virtual cost as perceived by the advertiser resulting from the actual cost of the publisher and the cost of inducing the publisher to reveal its true effectiveness.

We focus on a direct mechanism where publishers report their effectiveness which we denote as $\hat{\theta}_i$. By the revelation principle any incentive compatible mechanism can be mimicked by such a direct mechanism which is truthful in equilibrium. As the assumption that publishers can send a message about their effectiveness $\theta_i$ departs from reality, we later show how this assumption translates to a world where publishers can only make a choice about the number of ads to show.

The scheme the advertiser will offer to the publishers is $\{\hat{\theta}, x, b_1(x, \hat{\theta}), b_2(x, \hat{\theta})\}$, where $x$ is the observed performance and $b_i$ are the payments offered to publishers based on the observed output and reported effectiveness. During a campaign, publishers will report their types $\hat{\theta}_i$, and after the output $x$ is determined, they will receive the payment $b_i(x, \hat{\theta})$.

Define the payments $b_i$ as following when publishers report types $\hat{\theta}_i$ and performance $x$ is observed:

$$b_i(x,\theta) = \alpha_i(\theta)\left[x - x(y^*(\theta))\right] + c(y^*(\theta),\theta_i) - \int_0^{\theta_i} \frac{\partial c}{\partial \theta}(y_i^*(s,\theta_{-i}),s)ds \tag{19}$$

In the above specification:

$$\alpha_i(\theta) = \frac{\frac{\partial c}{\partial y}(y^*(\theta),\theta_i)}{\frac{\partial x}{\partial y_i}(y^*(\theta))} \quad \text{and} \quad y^*(\theta) = \arg\max_y x(y) - \gamma(y_1,\theta_1) - \gamma(y_2,\theta_2) \tag{20}$$

To understand the intuition behind the definition of these payment functions, we first note that $y^*$ finds the optimal allocation of efficiency units the advertiser would like to employ if the cost of advertising amounted to the virtual cost. The advertiser needs to consider the virtual cost since the mechanism is required to incentivize high effectiveness publishers to truthfully report their type and not try to impersonate lower effectiveness publishers. The advertiser then calculates a desired performance level for the campaign given publishers' reports, and pays publishers only if the output matches or exceeds this level.

The payment gives each publisher a share $\alpha_i$ of $x - x(y^*(\theta))$, which is the difference in performance from the expected optimal output given the publishers reports. In addition, the publisher is paid the expected optimal costs of showing ads $c(y^*(\theta),\theta_i)$ corrected for the expected information rent.

We can now prove the following result that shows the payments in (19) yield the optimal result for the advertiser:

**Proposition 7.** *When* $\frac{\partial^2 x}{\partial y_i y_j}(y^*(\theta)) \cdot \frac{\partial y_j^*(\theta)}{\partial \theta_i} \geq 0$ *the payments in* (19) *yield optimal profit to the advertiser. They are incentive compatible and individually rational for publishers. In equilibrium, publishers will choose to generate* $y^*(\theta)$ *efficiency units.*

Proposition 7 shows that when publishers are substitutes in production and when the equilibrium allocations are substitutes, then the linear contract is optimal, extending McAfee and McMillan (1991) for the case of substitution among the publishers.

The intuition behind this result is subtle. When the other publishers $-i$ are of higher effectiveness they will produce more output in equilibrium. The resulting externality on publisher $i$'s profit will then be stronger and as a result it will decide to increase its own output to compensate and redeem its share of the profits. In equilibrium these effects cause the publisher to increase its output with its type, which is a result similar to the standard monotonicity result in single agent mechanism design.

The optimal mechanism allows the advertiser to efficiently screen among publishers at the cost of giving positive rent to very effective publishers. The payment scheme is built as a sum of two separate payments: payment for performance and payment for effort of displaying ads. Using the ratio of the marginal cost to the marginal productivity of the publisher as the share of performance given to the publisher, the advertiser is able to align the incentives of the publisher at the margin. In equilibrium the most effective publishers will show the full information (first-best) number of ads, but will also receive the highest share of profit from the advertiser. The publishers with the lowest effectiveness will be excluded from the campaign, and will receive zero profits. An interesting aspect of this payment scheme is that publishers receive less expected rent when compared to the two standard CPA and CPM schemes, which is a result of using a combination of reported types and observed performance of the campaign.

The optimal mechanism yields improved results compared to standard compensation schemes at the cost of requiring the assumption that publishers can report their effectiveness. In an advertising campaign, however, publishers can only choose the effort they spend in terms of number of ads they show. To remedy this technical issue we employ the taxation principle to transform the direct mechanism into an indirect mechanism where publishers choose the number of ads to show and the output they will achieve. Based on the observed performance and effort, the advertiser will pay $b_i$ in the following way:

**Lemma 3.** *Let $b_i(x, q) = b_i(x, \hat{\theta})$ when $q = (y_i^*(\hat{\theta})/\hat{\theta}_i)$ and $b_i = 0$ otherwise. Then $b_i(x, q)$ yields the same equilibrium as the mechanism in* (19).

Although this result is standard in mechanism design, it typically requires the assumption that publishers can be punished with arbitrary severity when they do not choose output that the advertiser would prefer. We are able to show that in our case, not paying anything sufficient for the mechanism to still be a truthful equilibrium.

The caveat, however, is that the resulting mechanism is highly non-linear in the effort of publishers. The monotonicity of publisher effort also does not hold for many specifications, which gives rise to multiple equilibria of the indirect mechanism. Another issue that arises is the effect of the baseline conversion rate of consumers. This was not considered previously and will prove to be detrimental to these mechanisms.

As it is highly unlikely that advertisers can implement such a mechanism in the reality, we choose to develop a simpler mechanism that holds potential for achieving profits that are closer to the full information (first-best) profits.

# B Appendix – Proofs

*Proof of Proposition 1.* To find $p^A$, notice that the profit of the advertiser is $(2q^A)^\rho(1-2p)$. Since $q^A \sim p^{\frac{1}{2-\rho}}$, we can drop the constants and solve for $p^A = \arg\max_p p^{\frac{\rho}{2-\rho}}(1-2p)$, yielding $p^A = \frac{\rho}{4}$, and $q^A = \frac{\left(\frac{\rho^2}{2}\right)^{\frac{1}{2-\rho}}}{2}$. The second order condition of each agent is:

$$\rho(\rho-1)\left(q_i + q^A\right)^{\rho-2} p^A - 1 < 0 \tag{21}$$

For $\rho <= 1$ it always holds, while for $1 < \rho < 2$ if holds if $q_i > \left(\frac{\rho^2(\rho-1)}{4}\right)^{\frac{1}{2-\rho}} - q^A$ after plugging $p^A$ and collecting terms. The right hand side is negative if $2^{-\rho-1}(\rho-1) < 1$, which holds for every $1 < \rho < 2$

To show that $q^* > q^M$, we notice that $\frac{q^*}{q^M} = 2^{\frac{1}{2-\rho}} > 1$ for $0 < \rho < 2$. Similarly, $\frac{q^M}{q^A} = \left(\frac{2}{\rho}\right)^{\frac{1}{2-\rho}} > 1$ for $0 < \rho < 2$, which proves part 1 of the proposition.

To prove part 2, since $q^M > q^A$, the total revenue generated by the advertiser $x(q_1, q_2)$ is always larger under CPM. The share of profit given to the publisher under CPM is $\frac{2(q^M)^2}{(2q^M)^\rho} = \frac{\rho}{2}$. This is the same share $p^A$ given under a CPA contract. As a result, since revenues are strictly larger and the same share is given, profits under CPM are larger.

To prove part 3, the difference in profit $u^A - u^M$ of the publisher has a numeric root on $[0, 2]$ at $\rho^c = 0.618185$. The function has a unique extremum in this range at $\rho = 0.246608$, which is a local maximum, and the the difference is zero at $\rho = 0$. Thus, it is positive below $\rho^c$ and negative above $\rho^c$ proving part 3. $\square$

*Proof of Corollary 1.* In the single publisher case, $q^M = p^M$ similarly to before, and solving the advertiser optimization problem yields $q^M = \left(\frac{\rho}{2}\right)^{\frac{1}{2-\rho}}$. Under CPA, $q^A = \left(\frac{\rho^2}{2}\right)^{\frac{1}{2-\rho}}$. We immediately see that $q^A > q^M \iff \rho > 1$.

The share of revenue given as payment to the publishers equals $\frac{\rho}{2}$ in both cases. As a result, when $\rho > 1$, $\pi^A > \pi^M$, and vice versa when $\rho < 1$. □

*Proof of Proposition 2 and Corollary 2.* For completeness, we specify the resulting distribution function, $f_1(\frac{q_1}{q_2})$:

$$f_1(\frac{q_1}{q_2}) = \begin{cases} 1 & q_1 \geq dq_2 \\ -\frac{d^2q_2^2 - 2((d-1)d+1)q_1q_2 + q_1^2}{2(d-1)^2q_1q_2} & q_2 < q_1 < dq_2 \\ \frac{(q_2-dq_1)^2}{2(d-1)^2q_1q_2} & \frac{q_2}{d} < q_1 < q_2 \\ \frac{1}{2} & q_1 = q_2 \\ 0 & q_1 \leq \frac{q_2}{d} \end{cases} \tag{22}$$

In a symmetric equilibrium, $f'(1) = \frac{1}{2}\frac{d+1}{d-1}$. Solving for the symmetric equilibrium allocation of the publishers under CPA reaches $q^{A-LT} = \left(2^{\rho-1}p\left[\frac{d+1}{d-1} + \frac{\rho}{2}\right]\right)^{\frac{1}{2-\rho}}$. Plugging into the advertiser optimization problem, we find that $p^A = \frac{\rho}{2}$,yielding the values of $q^A$ specified in the proposition. The SOC at the symmetric point is only negative for $\rho < \frac{3}{2} + \frac{1}{2}\sqrt{\frac{7+25d}{d-1}}$. The profit is positive for $\rho < 2 - \frac{4}{d-1} < \frac{3}{2} + \frac{1}{2}\sqrt{\frac{7+25d}{d-1}}$, proving this is an equilibrium.

A simple comparison to $q^*$ and among $q^{A-LT}$ and $q^M$ yields the conditions in the proposition and finalizes the proof.

Since $q^{A-LT} > q^M > q^A$ and the share of revenues given to the publishers under each scheme is equal, the profit under last touch attribution is higher. □

*Proof of Proposition 3.* The first order condition the publisher faces in a symmetric equilibrium is: $\frac{\rho(2q)^{\rho-1} + \rho q^{\rho-1}}{2}p^A = q$. The solution, after calculating the equilibrium share offered by the principal is: $q^{A-S} = \left(\frac{\rho^2}{4}(2^{\rho-1}+1)\right)^{\frac{1}{2-\rho}}$.

The second orders are negative at $q_1 = 0$ and at $q_1 \to \infty$, while the third order is always negative between these two, implying the second order condition holds. To prove part 2, we recall that $q^{A-LT} = \left(2^{\rho-1}\frac{\rho}{2}\left[\frac{d+1}{d-1} + \frac{\rho}{2}\right]\right)^{\frac{1}{2-\rho}}$ and $q^A = \left(\frac{\rho^2}{2}\right)^{\frac{1}{2-\rho}}$. In this case, $q^{A-S} > q^A$ iff $\frac{1}{2}(2^\rho + 1) > 1$

which holds for every $0 < \rho < 2$. $q^{A-S} > q^{A-LT}$ always when $\rho < 2 - \frac{4}{d-1}$. Comparing to the CPM quantity, $q^{A-S} > q^M$ iff $\rho > 1$.

Finally, since the share of revenue given by the advertiser to the publishers is $\frac{\rho}{2}$, which is equal to the share given under regular CPA campaigns and under last-touch attribution, we find that profit is higher for Shapley value attribution when $q^{A-S}$ is highest. □

*Proof of Lemma 1.* Maximizing the expectation w.r.t to $s$ yield $q^E = 1 - \mu$.

Plugging in $q^E$ yields the profit $\pi_{min} = \frac{N}{2}\left((\mu-1)^2 + 2\mu\right)$, which is smaller than $\pi^{max}$ by $\frac{N}{2}\left(\mu - \mu^2\right)$. □

*Proof of Lemma 2.* The proof was built for $\theta \in [0,1]$ assuming the effectiveness of advertising is $\theta q$. In the text $\theta = 1$.

The total profit from experimenting is:

$$\frac{\beta\theta^2\left(2\alpha + \beta^2 + \beta(\alpha + N + 1) + N\right) - 2\beta\theta^2\sqrt{\alpha(\beta+1)(\alpha+\beta+N)} + 2\alpha N(\alpha+\beta+1)}{2(\alpha+\beta)(\alpha+\beta+1)} \tag{23}$$

The second order with respect to $n$ is $-\frac{\alpha\beta\theta^2(\alpha+\beta+N)}{(\alpha+\beta)(\alpha+\beta+1)(\alpha+\beta+n)^3}$ and is negative for all $\alpha > 0$,$\beta > 0$ and $N > 0$. At $n = 0$, the first order is positive when $N > \frac{\beta(\alpha+\beta)(1+\alpha+\beta)}{\alpha}$ implying the optimal sample size is positive.

The solution to the first order condition is $n^* = \sqrt{\frac{\alpha(N+\alpha+\beta)}{1+\beta}} - (\alpha+\beta)$ which is independent of $\theta$. Finally, calculation of the change with respect to $\alpha$, $\beta$ and $N$ yield the conditions stated in the lemma. □

*Proof of Proposition 4.* The difference in profit from $\pi^{min}$ is

$$\frac{\beta\theta^2\left((\alpha+\beta)\left(\alpha(\beta+2) + \beta^2 + \beta - 2\sqrt{\alpha(\beta+1)(\alpha+\beta+N)}\right) + \alpha N\right)}{2(\alpha+\beta)^2(\alpha+\beta+1)} \tag{24}$$

Whenever $N > \beta\frac{1-\mu}{\sigma^2}$, the firm can achieve this difference by Lemma 2. It can be verified that this difference is positive for all $\alpha > 0$ and $\beta > 0$, and is increasing in $\theta$. Specifically, it is positive for $\theta = 1$ when $N$ is large enough. □

*Proof of Proposition 5.* In a CPM campaign, the publisher can choose to show $q_b$ ads to baseline consumers and $q$ ads to non-baseline consumers. The profit of the publisher is:

$$u = N\left[(q_b s + q(1-s))p^M - s\frac{q_b^2}{2} - (1-s)\frac{q^2}{2}\right] \tag{25}$$

Maximizing the profit of the publisher yields $q_b = q = p^M$. Plugging into the advertiser's profit and maximizing over the expectation of $s$ yields $p = \frac{1-\mu}{2}$, resulting in an advertiser profit of $N\left[\mu + \frac{(1-\mu)^2}{4}\right]$.

Performing a similar exercise for a CPA campaign, the publisher will opt not to show ads to baseline consumers, as it receives commission for their conversions regardless of showing them ads. Maximizing the publisher's profit yields $q_b = 0$ and $q = p^A$, which yields $p = \frac{1-2\mu}{2(1-\mu)}$ when plugged into the advertiser's profit and maximized. This value is higher than 0 only for $\mu < \frac{1}{2}$, and for $\mu > 1/2$ the advertiser will prefer not to use a CPA campaign. Comparing $q^M$ to $q^A$ and $q^*$ yields the second part of the proposition. The profit of the advertiser is then $N\frac{1}{1-\mu}$ which is lower than the CPM profit for any $\mu < \frac{1}{2}$, concluding the proof. □

*Proof of Proposition 7.* The following lemma from McAfee and McMillan (1991) establishes conditions for payments offered by the advertiser to maximize its profit, and applies to our model:

**Lemma 4** (McAfee and McMillan (1991) Lemma 1)**.** *Suppose the payment functions $b_i$ satisfy*

$$\mathbb{E}_{x,\theta_{-i}}[b_i(\theta)]|_{\theta_i=0} = 0 \tag{26}$$

*and evoke in equilibrium outputs $y^*(\theta)$. Then the payments maximize the advertiser profits subject to publisher individual rationality and incentive compatibility.*

Theorem 2 of McAfee and McMillan (1991) shows that the payments in (19) yield an optimal mechanism under the conditions that agents are complements in production and when $\frac{\partial y_i^*(\theta)}{\partial z_j} \geq 0$ for $j \neq i$. These conditions do not apply in our case when publishers are substitutes in production.

We therefore proceed to show that the linear mechanism is optimal also under substitute production. To prove these payments yield an optimal truthful mechanism, we first prove the following:

**Lemma 5.** *The optimal allocation of efficiency units $y^*(\theta)$ is unique, positive for positive effectiveness and increases with self-reported type and decreases with reported types of other publishers:*

- $\frac{\partial y_i^*}{\partial \theta_i}(\hat{\theta}_i, \theta_{-i}) \geq 0$

- $\frac{\partial y_j^*}{\partial \theta_i}(\hat{\theta}_i, \theta_{-i}) \leq 0$ *for* $j \neq i$

*Proof.* Let $\pi(y, \theta) = x(y) - \gamma_1(y_1, \theta_1) - \gamma_2(y_2, \theta_2)$.

The first order conditions are:

$$1 - \frac{2 - \theta_i y_i}{\theta_i^3} - y_{-i} = 0 \tag{27}$$

With solutions:

$$y_i^* = \frac{\theta_i^3 \left(\theta_{-i}^3 + \theta_{-i} - 2\right)}{\theta_i^3 \theta_{-i}^3 - \theta_i \theta_{-i} + 2\theta_i + 2\theta_{-i} - 4} \tag{28}$$

These are positive for $1 \geq \theta_i > 0$.

The first principal minors of $\pi$ are zero and the second is positive. As a result $\pi$ is negative semidefinitie, and $y^*$ is the unique maximum.

Using the implicit function theorem:

$$\frac{\partial y_i^*}{\partial \theta_i} = \frac{2(\theta_i - 3)(2 - \theta_{-i}) y_1}{\theta_i \left(\theta_i^3 \theta_{-i}^3 - \theta_i \theta_{-i} + 2\theta_i + 2\theta_{-i} - 4\right)} \geq 0 \tag{29}$$

$$\frac{\partial y_i^*}{\partial \theta_{-i}} = \frac{2\theta_1^3 (\theta_2 - 3) y_2}{\theta_2 \left(\theta_1^3 \theta_2^3 - \theta_1 \theta_2 + 2\theta_1 + 2\theta_2 - 4\right)} \leq 0 \tag{30}$$

□

The expected profit of a publisher with type $\theta_i$ reporting $\hat{\theta}_i$ is

$$u_i(\hat{\theta}_i, \theta_i, y_i) = \alpha_i(\hat{\theta}_i, \theta_{-i}) \left[x(y_i, y_{-i}^*(\hat{\theta}_i, \theta_{-i})) - x(y^*(\hat{\theta}_i, \theta_{-i}))\right] \tag{31}$$

$$+ c(y^*(\hat{\theta}_i, \theta_{-i}), \hat{\theta}_i) - \int_0^{\hat{\theta}_i} \frac{\partial c}{\partial \theta}(y_i^*(s, \theta_{-i}), s) ds - c(y_i, \theta_i) \tag{32}$$

The optimal allocation of efficiency units is $y_i^* = \frac{\theta_i^3 \left(\theta_{-i}^3 + \theta_{-i} - 2\right)}{\theta_i^3 \theta_{-i}^3 - \theta_i \theta_{-i} + 2\theta_i + 2\theta_{-i} - 4}$.

We note that $u_i|_{\hat{\theta}_i = 0} = 0$ since $\alpha_i = 0$ in this case, which is the first sufficient condition of Lemma 4.

The publisher will then choose to show $y_i$ ads that solve the first and second order conditions:

$$\frac{\partial u_i}{\partial y_i} = \alpha_i(\hat{\theta}_i, \theta_{-i})\frac{\partial x}{\partial y_i} - \frac{\partial c}{\partial y} = 0 \tag{33}$$

$$\frac{\partial^2 u_i}{\partial y_i^2} = \alpha_i(\hat{\theta}_i, \theta_{-i})\frac{\partial^2 x}{\partial y_i^2} - \frac{\partial^2 c}{\partial y^2} = -\frac{\partial^2 c}{\partial y^2} < 0 \tag{34}$$

The publisher will therefore choose $\hat{y}_i$ s.t.

$$\alpha_i(\hat{\theta}_i, \theta_{-i}) = \frac{-\frac{\partial c}{\partial y}(\hat{y}_i, \hat{\theta}_i)}{\frac{\partial x}{\partial y_i}(\hat{y}_i, y^*_{-i}(\hat{\theta}_i, \theta_{-i}))} \tag{35}$$

We note that $\hat{y}_i|_{\hat{\theta}_i=\theta_i} = y_i^*$, which is the second sufficient condition of Lemma 4.

We therefore need to prove that the payments $b_i$ are incentive compatible for the publishers.

Denote $x_i = \frac{\partial x}{\partial y_i}$ and $c_y = \frac{\partial c}{\partial y}$.

Differentiating (35) with respect to $\hat{\theta}_i$ yields:

$$\frac{\partial \hat{y}_i}{\partial \hat{\theta}_i} = \frac{\frac{\partial \alpha_i(\hat{\theta}_i, \theta_{-i})}{\partial \hat{\theta}_i} + \frac{c_y}{x_i^2} x_{ij} \frac{\partial y_j^*}{\partial \theta_i}(\hat{\theta}_i, \theta_{-i})}{\frac{c_{yy}}{x_i}} \geq 0 \tag{36}$$

This inequality holds since $x_{ij}\frac{\partial y_j^*}{\partial \theta_i}(\hat{\theta}_i, \theta_{-i}) \geq 0$ by Lemma 5 and $\frac{\partial \alpha_i(\hat{\theta}_i, \theta_{-i})}{\partial \hat{\theta}_i} \geq 0$ by Theorem 3 of McAfee and McMillan (1991).

The remainder of the proof follows the proof of Theorem 2 in McAfee and McMillan (1991), p. 574. Using the fact that $\frac{\partial \hat{y}_i}{\partial \hat{\theta}_i} \geq 0$ is then sufficient to prove incentive compatibility. $\square$

*Proof of Lemma 3.* Let $q_i^* = \frac{y^*}{\theta_i}$. Then $\frac{\partial q_i^*}{\partial \theta_i} > 0$. As showing any number of ads not in $q_i^*$ will yield zero revenue with positive costs, the publisher will prefer to show ads in $q_i^*$ only.

Since $q_i^*$ and $y_i^*$ are both monotonically increasing in $\theta_i$, choosing to show the optimal number of ads such that $\hat{\theta}_i = \theta_i$ is an equilibrium strategy for the publisher.

Finally, $b_i(x, q)$ is well defined. Suppose there are $\theta_1^1 \neq \theta_1^2$ s.t. $q_i^*(\theta_1^1) = q_i^*(\theta_1^2)$ yet $b_i(x, \theta_1^1) > b_i(x, \theta_1^2)$. Then because the utility of the publisher increases with the payment, the publisher would prefer to claim its type is $\theta_1^1$ when its true type is $\theta_1^2$. This contradicts the truthfulness of the direct mechanism. Hence $b_i(x, \theta_1^1) = b_i(x, \theta_1^2)$. $\square$

## C Estimation of Publisher Effectiveness

We let $x_i = 1$ denote exposure by consumers to ads from publisher $i \in N$, and specify the following discrete choice model: The utility of a converting consumer $j$ exposed to a subset of ads $I \subseteq 2^N$ is specified as $u_{jI} = s + \sum_{i \in I} b_i x_i + \epsilon_{jI}$, with $s$ the basic utility of consumers in the baseline. A consumer converts if $u_{jI} > s + \epsilon_{j\emptyset}$.

If we assume the $\epsilon_{jI}$ are distributed i.i.d. extreme value, we expect to see the population conversion rate $y_I = \frac{e^{\sum_{i \in I} b_i x_i}}{1+e^{\sum_{i \in I} b_i x_i}}$. For each subset $I$ observed in the data we have exact values for this conversion rate, as well as the total population value who were exposed to ads. We therefore do not need to make assumptions about the total population size or estimate it from the data.

We then have

$$\ln y_I - \ln (1 - y_I) = \sum_{i \in I} b_i x_i \tag{37}$$

which is estimated using OLS.