Multi-Channel Attribution: The Blind Spot of Online Advertising*

Working Paper

Vibhanshu Abhishek[1], Stylianos Despotakis[2], and R. Ravi[2]

[1] Heinz College, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213
[2] Tepper School of Business, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213
{vibs, steliosdes, ravi}@cmu.edu

April 28, 2017

Abstract

In this paper, we study the problem of attributing credit for customer acquisition to different components of a digital marketing campaign using an analytical model. We investigate attribution contracts through which an advertiser tries to incentivize two publishers that affect customer acquisition. We situate such contracts in a two-stage marketing funnel, where the publishers should coordinate their efforts to drive conversions.

First, we analyze the popular class of multi-touch contracts where the principal splits the attribution among publishers using fixed weights depending on their position. Our first result shows the following counterintuitive property of optimal multi-touch contracts: higher credit is given to the portion of the funnel where the existing baseline conversion rate is higher. Next, we show that social welfare maximizing contracts can sometimes have even higher conversion rate than optimal multi-touch contracts, highlighting a prisoners' dilemma effect in the equilibrium for the multi-touch contract. While multi-touch attribution is not globally optimal, there are linear contracts that "coordinate the funnel" to achieve optimal revenue. However, such optimal-revenue contracts require knowledge of the baseline conversion rates by the principal. When this information is not available, we propose a new class of 'reinforcement' contracts and show that for a large range of model parameters these contracts yield better revenue than multi-touch.

Keywords: attribution; advertising; Shapley value; multi-touch; game theory; marketing funnel

*All three authors contributed equally to the paper and are listed in alphabetical order. The authors thank Kartik Hosanagar, Ganesh Iyer, and Kannan Srinivasan for comments. The authors are particularly grateful to Amin Sayedi, Kaifu Zhang, and Yi Zhu for their extensive feedback on an earlier draft.

1 Introduction

The last decade has seen a large shift of advertising effort from offline to online channels. Internet advertising revenue is projected to overtake TV advertising for the first time in 2017.[1] The key benefit of the online medium is the accountability provided by the user clicks and the cookie trails left by their visits to various advertising venues. This fine-grained view of the user's journey through the decision funnel along with the specific advertising actions they are exposed to (banner and display ads, video ads, text ads after search, emails) provides a unique opportunity to solve the traditional marketing mix problem in a very user-specific way. The key to arriving at optimal resource allocations across the channels is to determine the response model of how each of the interventions affects the decision-making journey of the customer. This is commonly phrased as an **attribution** problem of how the credit for a digital conversion should be split among the various advertising actions. Attribution allows an advertiser to determine the impact of each ad-type so that the effectiveness of different types of ad activities can be taken into consideration while deciding how to split the advertising budget.

The problem of attribution is not new. It arises in traditional advertising channels like television and print as well, where advertisers have resorted to marketing mix models using aggregate data (Naik et al., 2005; Ansari et al., 1995; Ramaswamy et al., 1993). However, online advertising offer a unique opportunity to address the attribution problem as advertisers have disaggregate individual-level data which were not previously available (Goel, 2014). Disaggregate data offer the possibility of determining the effectiveness of an ad on an individual customer at a specific time. Although better data has improved the accountability and performance in online advertising, several advertisers still use simplistic approaches like last/first touch attribution that might be suboptimal under a variety of conditions (PWC, 2014; Abhishek et al., 2016). The rapidly growing size of the industry and concerns raised by the advertisers have lead to tremendous recent focus on attribution. Companies like Google, Marketo, and Datalogix have designed and offered several new algorithmic attribution techniques in the last few years. Google (2017a) alone offers a variety of such models including last-touch and first-touch (where the last and first publisher gets the full credit respectively), as well as other weighted models including options for weighting based on fixed weights, time decay or position. Google (2017b) also offers an alternate data-driven attribution method that is based on the Shapley value of each publisher.

At the same time, many researchers have proposed empirical models of attribution. At its core, the attribution rule determines the payment received by publishers for showing an advertiser's ad.[2] Most of the existing literature has assumed that the publishers are not strategic (with the exception of Berman (2015)) and that their actions are not affected by the attribution methodology used. However, publishers (e.g. Facebook or Google) have to exert a considerable amount of effort (such as investments in technology) to match an ad impression with the right customer, and this effort level is affected by the incentives. Unfortunately, a

[1] http://www.pwc.com/us/outlook (accessed March 2017).

[2] In this paper, we denote publisher as an entity that is responsible for matching ads to consumers and the delivery of ads. In most cases, this role would be fulfilled by an ad network such ad Double Click of Yahoo Display Network, a search engine like Google or a large publishers like Facebook or Snap. We use the term publisher to be consistent with the prior literature (e.g. Berman (2015))

typical advertiser cannot observe the effort exerted by a single publisher in the funnel while the final conversion is based on the total effort by all publishers in it. This might lead to free-riding where a publisher aims to benefit from the effort exerted by another publisher in the pipeline, and creates an opportunity for moral hazard (Holmstrom, 1982). In addition, the publishers are much better informed about the consumers as they observe them in many different contexts as compared to the advertiser. This information asymmetry prevents the advertiser from estimating the effectiveness of a publisher's marketing action and can lead to adverse selection. As an example of such an inefficiency, Abhishek et al. (2016) show that once display publishers have determined which consumers are most likely to convert, they flood them with ads. This crowds out more effective publishers and drives revenues from them to less effective publishers. Both the moral hazard and adverse selection problems create a misalignment of incentives between the advertiser and the publishers resulting in a loss of efficiency in online advertising.

The misalignment in incentives is manifested in many ways. Ad fraud has become a major concern and publishers charge advertisers for ads that consumers never see. In fact, methodologies like last-touch give credit to publishers even if the consumer would have converted without seeing an ad. There is a substantial increase in low-quality ad inventory, that do not lead to any meaningful outcomes such as conversions (Scott, 2016). A recent Economist (2016) article shows that ad fraud will cost advertisers US$ 7 billion this year and is growing rapidly. Even though advertisers are aware of these issues, they are not able to address them due to the lack of appropriate data and the complexity of the online advertising industry. This complication is exacerbated even further due to the multiple touch-points spanning several publishers that jointly affect the consumer decision-making funnel. The multiplicity of publishers create a lack of accountability (Economist, 2016), which affects the advertisers and ultimately the entire advertising industry adversely. Although advancement in attribution methodologies have made advertising more efficient,[3] they have not eliminated the moral hazard or adverse selection issue completely. One way to eliminate or reduce the moral hazard issue is using newer attribution methodologies, e.g. Berman (2015) presents a Shapley value based attribution scheme that performs better than last-touch attribution when the publishers' ads are strategic complements and the uncertainty in consumer behavior is low.

In this paper, we propose a simplified two-stage model of the purchase funnel to determine the most appropriate attribution methodology. Prior literature (Mulpuru, 2011; Court et al., 2009; Bettman et al., 1998) shows that consumers move through different stages before they purchase a product. Consequently, in our analysis, the two stages considered are awareness and consideration. Incorporating the temporal dynamics is not only a more natural approach to addressing the attribution problem but also leads to interesting new results as opposed to more static models considered in prior literature. We consider two distinct publishers that are responsible for generating ad impressions on behalf of the advertiser and jointly drive consumers towards conversion. In the basic model, we consider that one publisher can create awareness and the other one can drive conversions.[4]

[3] For example, "According to Forrester Research, B2B companies are seeing an average of 15 to 18 percent lift in revenue as a result of implementing a closed-loop attribution system." (`https://www.clickz.com/lift-off-a-new-study-on-attribution-and-revenue/41687/`, accessed March 2017).

[4] In an extension of the basic model, we assume that both publishers compete for both awareness and

After reviewing related work and background in Sections 2 and 3, we define our model in Section 4. In Section 5, we first analyze a linear attribution rule that splits a fraction of the marketing dollar as f and $1-f$ between the two publishers (for a conversion). This rule resembles the commonly used multi-touch attribution rules, such as first- and last-touch. We term the resulting contracts as f-contracts, which in turn determine the efforts exerted by the publishers in the co-production process. One might expect the optimal f to give more credit to the stage where the baseline conversion rate is lower. We show the counter-intuitive result that, all else being equal, an advertiser using an optimal f-contract should give *more* credit to a publisher with higher baseline probability of advancing (down the purchase funnel) in its stage. We arrive at this counter-intuitive result because of the way multi-touch contracts compensate publishers. There is a complementarity between the baseline rate of one stage and the effort exerted by the publisher in the other stage since the final conversion is a co-production process. In the presence of this effect, to provide an incentive for the publisher in this stage to exert optimal effort, the optimal f-contracts gives this publisher more credit if its baseline rate is high.

Next, we show that social welfare maximizing contracts can lead to have even higher conversion rates than optimal multi-touch contracts, highlighting a prisoners' dilemma effect in the equilibrium for the multi-touch contract. Our result shows that the optimal f-contract gives an inefficient equilibrium due to lack of coordination between the publishers. This result is unique in the attribution literature as we show that the general class of f-contracts, which are commonly used in practice, can result in inefficiencies in the effort exerted by the different publishers.[5]

In Section 6, we explore the broader design space of linear contracts. We show that there are optimal linear contracts that can "coordinate the marketing channels" to achieve optimal revenue. However, these optimal contracts suffer from the same problem as optimal multi-touch contracts in that they require full knowledge of the baseline conversion rates by the principal (advertiser). For this reason we propose a new class of 'reinforcement' contracts. These contracts perform significantly better over a wide range of parameters than other multi-touch contracts while not relying on the knowledge of the baselines. Finally, in Section 7, we examine several extensions of our underlying models and show that the main findings still hold under a variety of circumstances.

This paper addresses an important gap in the attribution literature, namely the strategic decision by the publishers in a dynamic purchase funnel. As discussed earlier, incorporation of the dynamic nature of the conversion process leads to new and interesting results. This paper has also several managerial implications about designing of multi-channel advertising contracts under information asymmetry and uncertainty. Advertisers can use the contracts outlined in the paper to increase the effectiveness of multi-channel advertising.

---

conversions.

[5] Given the inefficiency, a related issue is how bad an f-contract can be in terms of social welfare. Using a concept called the Price of Anarchy (Koutsoupias and Papadimitriou, 1999; Roughgarden and Tardos, 2002), we show that this multiplicative ratio is bounded by $\frac{4}{3}$.

2 Literature Review

Recent years have seen a tremendous amount of academic interest in the attribution problem given the importance to the industry. Here, we discuss the different streams of literature that are relevant to our research.

We start off with some of the empirical work on attribution from the marketing literature. Shao and Li (2011) propose two multi-touch attribution models, a bagged logistic regression model and a probabilistic model, and they apply these two approaches to a real-world data-set. Jordan et al. (2011) explain why current attribution methods are inefficient and they find an optimal ad allocation and payment scheme in a model they developed. Dalessandro et al. (2012) propose an attribution methodology based on a casual estimation problem that uses the concept of Shapley Value. Anderl et al. (2013) introduce a graph-based framework for attribution using Markov models and test it in real-world data-sets. Li and Kannan (2014) propose a measurement model for attributing conversions to different channels and find that the relative contributions of these channels are different from those estimated by traditional metrics like last-touch. Xu et al. (2014) develop a multivariate point process that captures the dynamic interactions among ad clicks and find that even though display ads may have low direct effect on conversions, they have also an indirect effect by stimulating subsequent visits through other ad formats. Abhishek et al. (2016) use a hidden Markov model for consumer behavior to find that different channels and types of ads affect consumers in different stages in the purchase funnel.

One of the first analytical papers on attribution was by Berman (2016). Berman uses an analytical model with two publishers and two advertisers that involves externalities between the publishers and uncertainty about consumer visit order. Using his model, he shows that bidding truthfully in ad auctions is not an equilibrium for the advertisers. He also shows that last-touch attribution results in lower profits for advertisers compared to not using attribution, while an attribution based on Shapley value can result in higher profits when conversion rates are low.

Our work differs from Berman's work in that we explicitly model the marketing funnel and we take into account customer's microtrails. We believe that the nature of some websites make them more or less likely to be at a certain point in a customer's trail. For example, when a customer searches for a product he is already interested in, we can assume that he is in the final stage of consideration, and therefore an ad in the search engine is more likely to be the last-touch point before purchase. This affects the behavior of publishers who publish ads for different stages in the funnel, and as a result it is important especially for multi-touch contracts. Another difference is that we model a wider range of attribution rules, where last-touch and Shapley value are special cases, and we examine the advertiser's problem to determine the optimal attribution. However, we assume a functional form for the relation between the effort of the publisher and the increase in conversion rate, while Berman models the process of price setting at each stage more realistically via a second-price auction.

The attribution problem is also related to the team production literature in contract theory. There is a big literature on the topic, but to mention just a few, Holmstrom (1982) studies the problem of moral hazard in team compensation and how it can be dealt with by breaking the budget-balancing constraint. Eswaran and Kotwal (1984) followed by explaining how not balancing the budget can result in a new source of moral hazard. Holmstrom and Milgrom (1987) study compensation schemes for incentivizing agents. Dearden and Lilien (1990) consider a problem in which the firm learns over time. McAfee and McMillan (1991) study the interaction between moral hazard and adverse selection in a team model.

Besides the context, another key difference to our work is that we explore the multiplicative effect of agents' efforts, which is a result of the marketing funnel. In other words, instead of all agents putting efforts together that adds to a total effort, they put efforts in stages in a way that the result of the effort an agent puts in a stage is affected by what happened in other stages. Incorporating this difference, which is a key to the marketing funnel approach, leads to some interesting results.

3 Background

We present some background on the purchase funnel, attribution models, and fairness considerations.

3.1 Purchase Funnel and Common Attribution Rules

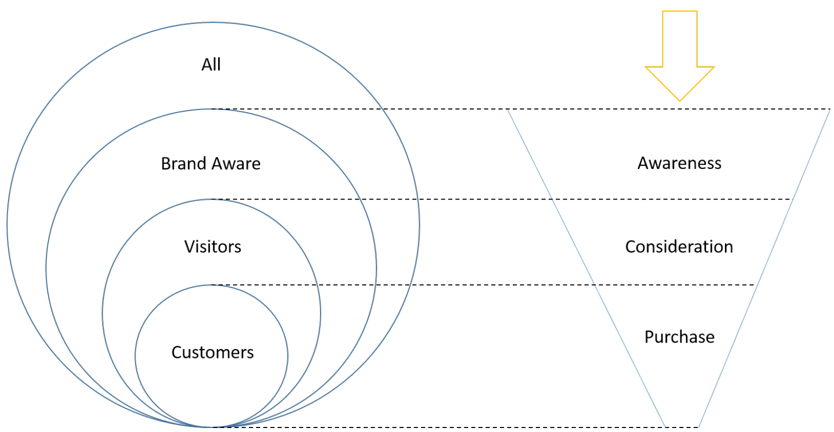

Figure 1: Purchase funnel.

In the online world, consumers are exposed daily to a number of different types of ads, e.g. display ads, search ads, affiliate ads and sponsored content. Before a consumer purchases a product, he is influenced by these ads as he moves through his decision making process, which has been commonly captured using the purchase funnel as shown in Figure 1. The purchase funnel captures the progression of individuals from being unaware about the firm to purchasing products and becoming the firm's customers. A fraction of the total population becomes aware of the firm and moves into the state of awareness. Some of these brand aware individuals might be further interested in purchasing products from the firm and move into the next consideration state. Finally, a small fraction of individuals that consider the product will eventually purchase it. Since each of these stages contains fewer number of consumers than the previous one, the progression is typically illustrated as a funnel. Consumers enter through the top of the funnel, pass through the different stages, and some of them convert. Each type of ad that a consumer is exposed to helps his move to a further stage on this path, and some ads are more effective in some stages than others (e.g. display ads in the awareness stage, and search ads in the consideration stage).

For advertisers using both stages of the funnel, it is important to have an attribution rule for eventual conversions. This rule will determine who gets credit every time there is a conversion, conditional on all the ads the user has previously seen. As an example, last-touch attribution is a widely used attribution model, where all the credit is assigned to the ad responsible for the last ad exposure before conversion. Last touch is very popular for its convenience and because it is easy to implement. However it might not be optimal, since it fails to take into account ads used to build awareness and interest to the consumer.

To address this issue, firms have started offering alternative models like first-touch (Google, 2017a), where the credit goes to the first ad the user was exposed to because it was the ad that made him enter the funnel. In even more general multi-touch models, the advertiser determines how to split the credit between all the ads in a consumer's trail. For example, he could give equal weight to all the ads in the trail or give more weight to the first and the last touch points and less in the middle.

3.2 Fairness and Shapley Value

One approach that has been gaining popularity in the advertising literature for deciding payment rules is fairness (Berman, 2015; Dalessandro et al., 2012; Abhishek et al., 2016). Fairness has also been gaining attention in the economics literature since its introduction by Rabin (1993) to examine game-theoretic problems. Defining fairness has been a challenging problem and researchers have used a set of axioms to delineate what is fair (van den Brink, 2002; Lan et al., 2010). Shapley (1953) proposed four such natural axioms and proved that there is a unique rule that satisfies them. We call this rule the *Shapley Value*, and the payoff of each player according to this rule is a weighted sum of his marginal contributions to every subset of players.[6] This approach, also referred to as the incremental approach, has been commonly adopted in the attribution literature (Abhishek et al., 2016; Berman, 2015) and practice (Google, 2017b). Shapley value is a great attribution rule if our goal is to achieve a fair result and if the attribution rule we use does not affect the performance or actions of the players. Now suppose that the effort of each player depends on the attribution rule we use, and our goal is to maximize the total output of the game. Even though Shapley value based attribution meets the fairness axioms, it might not be the most optimal for an advertiser. For example, if a publisher doesn't get enough credit with the Shapley value, he might put less effort in showing ads and the advertiser will end up with fewer conversions overall. It is possible that an alternative attribution scheme might give the publisher higher payouts so that he puts in more effort with a better overall result. How should we then choose an attribution rule that maximizes the total value of the game? Our paper tries to answer this question.

[6]For a formal definition, see Appendix A.3.

4 Model

4.1 Overview

Before purchasing a product, a consumer moves through two stages: awareness and consideration.[7] This model is based on the idea of a conversion funnel which is frequently used in marketing (Mulpuru, 2011; Court et al., 2009; Bettman et al., 1998). Every period, a new consumer arrives in the system and moves to the first stage (awareness).[8] A consumer in the first stage either moves to the second stage (consideration stage) with probability $f(a)$, or leaves the system. The function $f(a)$ has as argument the advertising level a for the first stage. The ads in the first stage create awareness, e.g. display ads or sponsored content. Similarly, a consumer in the second stage either purchases the product with probability $g(c)$, or leaves the system. c is the advertising level for the second stage. Ads in the second stage are more transaction oriented and lead directly to conversion, e.g. search ads. Note that it is not necessary for consumers to see ads in either stage before they purchase as $f(0)$, $g(0)$, the baseline rates at each stage, can be strictly positive.

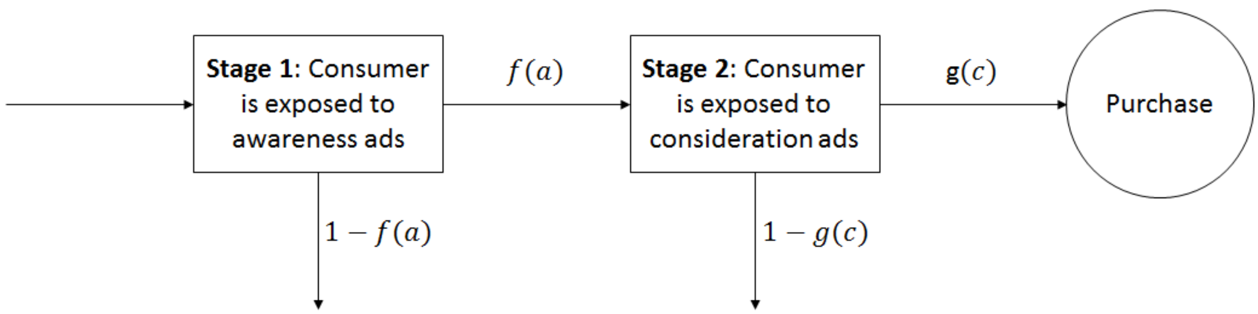

Figure 2: General representation of the model.

4.2 Consumer Model

Awareness Stage

For the customer to purchase a product, it should belong in the customer's consideration set. If the consumer is not aware of the product, then it is unlikely that he will eventually purchase the product. We assume that a fraction of consumers (≥ 0) might be aware of the product even in the absence of advertising. We represent the **baseline rate** of these consumers by $q_0 \in \left[0, \frac{1}{2}\right]$. Some consumers learn about the product because they have seen an ad from Publisher 1. The rate of these consumers is $a \in \left[0, \frac{1}{2}\right]$, where a is a decision variable for Publisher 1. In other words, we assume the functional form $f(a) = a + q_0$ for the function f (see also Figure 3). When a consumer learns about the product, they move to

[7] We analyze two stages for clarity of exposition, but out results continue to hold with more than two stages.

[8] The model can be generalized by including some stochastic waiting time in each stage, but this does not influence the results.

the consideration stage. If a consumer does not learn about the product in this first stage, which happens with probability $1 - q_0 - a$, they leave the system.[9]

We assume the convex functional form of $w \cdot a^2$ for the effort that Publisher 1 has to exert to result in additional conversion a in the first stage. The probability a captures how effective the publisher is in showing the ads to the relevant audience, while the convexity of the functional form models diminishing returns to effort by the publisher.[10]

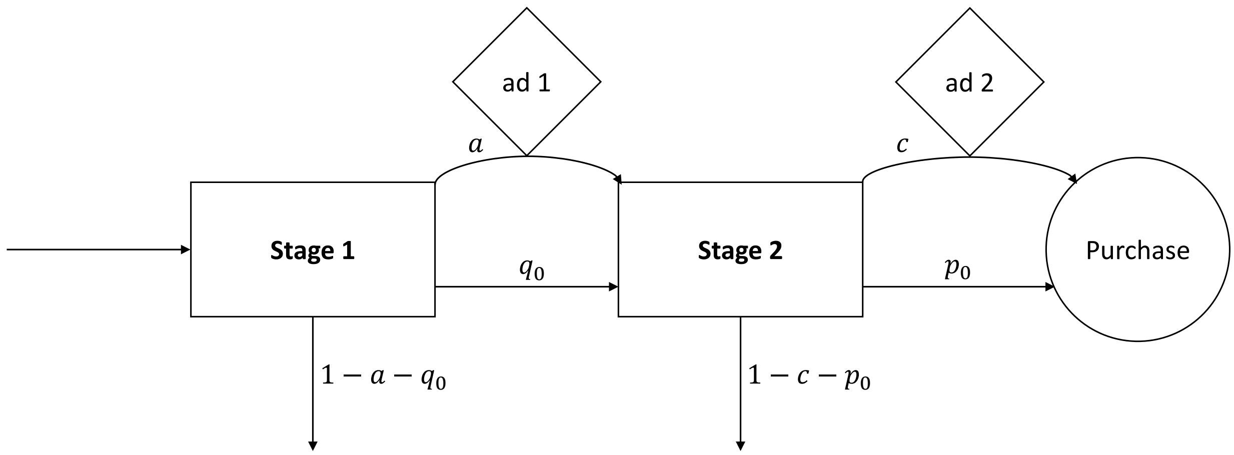

Figure 3: A more detailed representation of the model.

Consideration Stage

After consumers reach the consideration stage and are interested in the product, then we have the second stage of our model. Some consumers in that stage will decide to buy the product after seeing an ad from publisher 2. This will happen with probability $c \in \left[0, \frac{1}{2}\right]$, which is a decision variable for Publisher 2. Some other aware consumers will buy the product without seeing any ad and this will happen with probability $p_0 \in \left[0, \frac{1}{2}\right]$. In other words, we assume that the function g has the form $g(c) = c + p_0$. The probability that a consumer in the second stage will not buy the product in the end is $1 - p_0 - c$.

Publisher 2 can be considered as a website with content related to the product, where aware consumers go to look for more information before they decide if they will buy (e.g. a search engine, a site with reviews or comparisons of similar products). As before, we assume the convex functional form of $v \cdot c^2$ for the effort that Publisher 2 has to exert to result in additional conversion c in the second stage.[11]

[9] In the Appendix A.2 we also consider a more general model of the funnel and show how it can be reduced to the simple model analyzed here.

[10] Note that the probabilities a and q_0 are at most $\frac{1}{2}$, to make sure that $f(a) \leq 1$. In reality, these probabilities are very small, so we don't lose anything by assuming the upper bound of $\frac{1}{2}$. Moreover, to make sure that a will not exceed this bound when we find the equilibria, we assume that w is sufficiently high. We will use the bound $w \geq 1$, which is sufficient for our model.

[11] Similarly to the first stage, we assume that c and p_0 are at most 1/2 and to enforce this upper bound of c in the equilibria, we assume that $v \geq 1$.

4.3 Firm's Problem

In the absence of advertising, a fraction $q_0 \cdot p_0$ of consumers convert. With online advertising, the firm gains the ability to convert more consumers $((q_0 + a) \cdot (p_0 + c))$ by showing them ads. The firm cannot advertise directly, but can use the advertising real estate provided by the publishers to reach its customers.

The firm can observe the consumers who decided to buy the product and all the ads they've seen prior to that (e.g., through the trail captured in browser cookies). It may not know the efforts the publishers put (that resulted in the additional rates a and c in the two stages), or the baseline probabilities q_0, p_0.[12] What the firm can infer by observing a large number of consumers are the percentages of people who followed each of the four possible paths before conversion. For example, it will know that an $a \cdot c$ fraction of consumers have seen an ad from the first publisher in the first stage and an ad from the second publisher in the second stage before they convert. Similarly, it will know that an $a \cdot p_0$ fraction of consumers have seen an ad from the first publisher in the first stage and no ad in the second stage before they convert, and so on (Figure 4).

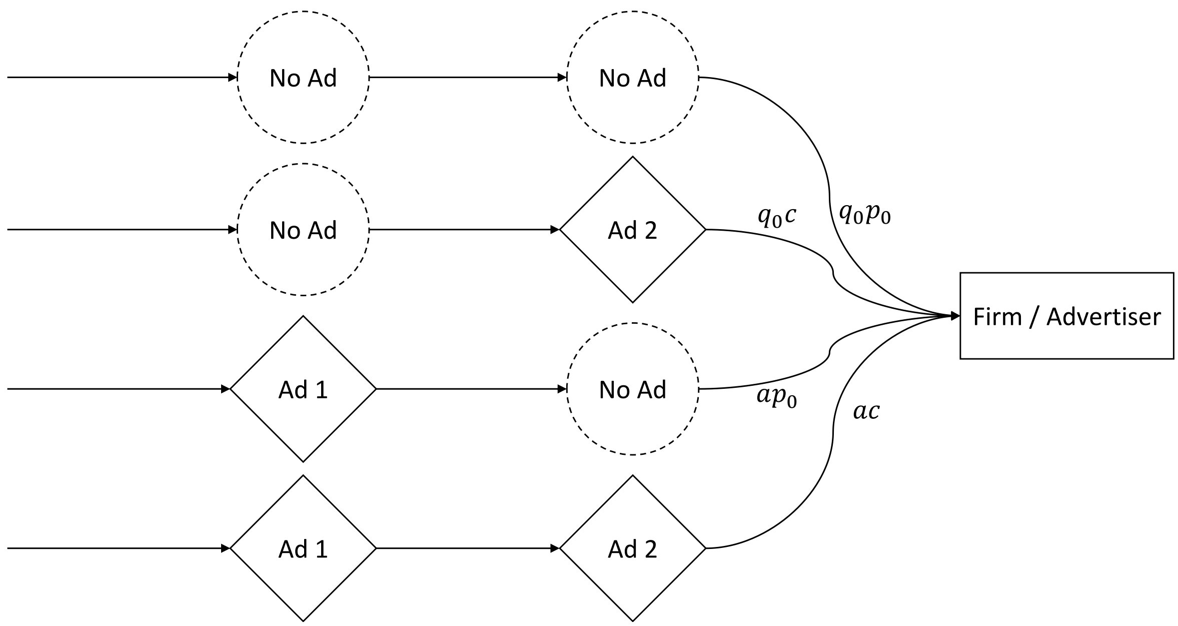

Figure 4: The four types of click-streams the firm can observe.

For every conversion, the firm wants to spend some fixed amount in advertising, which we normalize to \$1.[13] The question that we address is the optimal way to split this dollar between the two publishers in order to maximize the conversion rate $(a + q_0)(c + p_0)$.

[12]The advertiser can potentially observe consumers who do not convert. However, as long as the advertiser does not observe *all* the consumers in the market, it cannot determine p_0, q_0 with certainty.

[13]This limit on the spend can arise from the cost of the competing outside advertising options, e.g. print or television advertising.

4.4 Valid Contracts and Publishers' Problem

Any attribution rule the firm uses will lead to a contract with the publishers. The publishers try to maximize their profits by putting in the optimal amount of effort based on the contract they have with the firm. A contract is defined by two payment functions g_1, g_2 that satisfy the equality

$$g_1(a, q_0, c, p_0) + g_2(a, q_0, c, p_0) = (a + q_0)(c + p_0),$$

where g_1 is the payment to the first publisher and g_2 is the payment to the second publisher. Since $(a + q_0)(c + p_0)$ is the conversion rate and we assume that the firm wants to spend \$1 per conversion, the functions g_1 and g_2 specify how the firm should split their advertising budget between the two publishers.

Valid Contracts. The contracts that the firm can offer to the publishers must be based on the information that the firm knows. In this paper we will focus on the class of linear payment functions with respect to the four products q_0p_0, q_0c, ap_0, ac. This type of contracts have two advantages. First, they don't require the knowledge of the baseline rates p_0, q_0 or the publishers' efforts a, c to be implemented, but only knowledge of the percentages of people who followed each of the four possible paths before conversion (Figure 4).[14] Second, even if the baseline rates p_0, q_0 are known to the advertiser, attribution rules that depend linearly only on the products q_0p_0, q_0c, ap_0, ac can be applied in an online fashion. Every time there is a conversion, the advertiser can attribute one unit of credit among publishers based on the conversion path the customer followed. Note also that the widely used class of multi-touch attribution contracts is a special case of linear contracts.

The profit of the first publisher is given by

$$\pi_1 = g_1(a, q_0, c, p_0) - w \cdot a^2,$$

where $w \cdot a^2$ is the advertising cost in the first stage. Publisher 1 decides a to maximize his profits. Similarly, the profit of the second publisher is given by

$$\pi_2 = g_2(a, q_0, c, p_0) - v \cdot c^2,$$

where $v \cdot c^2$ is the advertising cost in the second stage. Publisher 2 decides c to maximize his profits.

4.5 Benchmarks

Next, we consider two benchmark models to compare with our main model. In both models, we assume that the firm and the publishers are integrated, i.e. the firm controls the advertising efforts. In the first model, the goal for the firm is to maximize the social welfare, i.e. the sum of profits of the two publishers. In the second model, the goal is to maximize the number of conversions.

[14] Note that if the firm can only observe the four products q_0p_0, q_0c, ap_0, ac, it cannot infer the individual efforts. This is because for any solution (q_0, a, p_0, c) that satisfy these products, there are other solutions of the form $\left(q_0x, ax, \frac{p_0}{x}, \frac{c}{x}\right)$, for $x > 0$ that satisfy them too.

Maximizing Social Welfare (Publishers' Optimal) In the first benchmark model, the firm will make all the decisions about advertising, and its goal is to maximize the social welfare. The social welfare is equal to the total revenue minus the total cost for the publishers. Therefore, the optimization problem for the firm is as follows.

$$\begin{aligned} \max_{a,c} \quad & (a+q_0)(c+p_0) - (wa^2 + vc^2) \\ & a, c \geq 0. \end{aligned}$$

Maximizing Conversion Rate (Firm's Optimal) There are two ways we could potentially model this benchmark case. One is to model it as an optimization problem with the conversion rate as the objective and no constraints. In other words, we don't care about the cost of advertising effort, but we want to find the optimal effort level that maximizes the number of conversions. However, this would be unrealistic, simply because we can always achieve a conversion rate of 1 in this optimization problem.

The more appropriate benchmark is achieved by including a constraint on the cost of advertising effort. Note that in the main model, the firm can spend $1 for every conversion, which goes to the two publishers. Therefore, the appropriate benchmark is to determine the optimal conversion rate given that the cost of advertising effort is exactly the number of conversions (multiplied by unit revenue per conversion). In other words, the optimization problem for the firm is as follows.

$$\begin{aligned} \max_{a,c} \quad & (a+q_0)(c+p_0) \\ & (a+q_0)(c+p_0) = wa^2 + vc^2 \\ & a, c \geq 0. \end{aligned}$$

We denote by a^*, c^* the efforts in the optimal solution of this problem.

5 Multi-Touch Attribution

5.1 Definition of f-contract

We start by considering simple contracts that split conversion credit among the touch points of a consumer's trail. The canonical form of such a contract, that we term an f-contract, is summarized in the following table.

Ad in the awareness stage	Ad in the conversion stage	Credit to Publisher 1	Credit to Publisher 2
no	no	$\frac{1}{2}$	$\frac{1}{2}$
no	yes	0	1
yes	no	1	0
yes	yes	f	$1-f$

The parameter $f \in [0,1]$ is some value determined by the firm or externally. Note that in the table above, in the case that the consumer sees no ad, we split the dollar equally to the two publishers. This is an arbitrary choice made to be consistent with the assumption that the firm always pays out a dollar for every conversion. However, since this is just the same constant amount each publisher gets, its value does not affect the equilibrium behavior of the publishers. In other words, we could also assume that in case of no ad, both publishers get 0, without any change in our results.

In an f-contract, the profit of the publishers are as follows,

$$\begin{aligned}\pi_1 &= \frac{1}{2}q_0p_0 + ap_0 + fac - wa^2,\\ \pi_2 &= \frac{1}{2}q_0p_0 + q_0c + (1-f)ac - vc^2. \end{aligned} \tag{1}$$

Some examples of contracts of this form are the last-touch ($f = 0$), where all the credit goes to the last ad a consumer had seen prior to the conversion, and the first-touch ($f = 1$) where all the credit goes to the first ad. It is also interesting to note that for $f = \frac{1}{2}$, we get the Shapley value attribution of this model. In other words, the f-contract captures a wide range of attribution models commonly used in practice (Google, 2017a).

If we solve for the equilibrium under an f-contract, the equilibrium efforts exerted by the publishers are as follows,

$$\begin{aligned}a(f) &= \frac{fq_0 + 2vp_0}{4vw - f(1-f)},\\ c(f) &= \frac{(1-f)p_0 + 2wq_0}{4vw - f(1-f)},\end{aligned}$$

while the conversion rate as a function of f is given by

$$r(f) = \frac{(f^2q_0 + 2v(p_0 + 2wq_0))((1-f)^2p_0 + 2w(q_0 + 2vp_0))}{(4vw - f(1-f))^2}.$$

5.2 Optimal f-contract

We can now investigate the properties of the optimal contract and show how different parameters affect the split of the advertising dollar between Publishers 1 and 2. We study the effect of the baseline conversion rates here and consider the effect of the publisher costs later in Section 7.1. We then investigate the social welfare properties of the equilibrium of the f-contract.

5.2.1 Effect of baseline conversion rates

Our first proposition shows how the optimal contract f^* varies with baseline rates q_0 and p_0. To motivate it, note that if the value of the baseline rate q_0 is high, there is already a high rate of intrinsic conversion in the first stage. This suggests that there is less value in increasing the incentive to the first publisher in the co-production process, and hence could lead to lowering f as q_0 increases. However, we find the opposite to be the case. (All proofs are provided in the appendix.)

Proposition 1. *Let $f^* = f^*(q_0, p_0)$ be the value of f in the optimal f-contract for the firm as a function of q_0 and p_0. Then, f^* is increasing in q_0 and decreasing in p_0.*

This proposition shows that as the baseline rate of the first stage (q_0) increases, the amount we give to the first publisher in the optimal f-contract increases. We arrive at this counter-intuitive result because of the way multi-touch contracts compensate publishers. If the baseline rate in the first stage increases, the number of consumers who don't see an ad in the first stage but see an ad from the second publisher in the second stage increases. This implies that the payoff of the second publisher increases (as shown in Equation 1). Since there is complementarity between q_0 and the effort exerted by Publisher 2, this increases his incentive to put more effort. However, the incentive for the first publisher does not increase due to an increase in q_0. To balance things out and give incentive to both publishers to put more effort to increase the overall conversion rate, we should increase the value of f, i.e. the amount we give to the first publisher. Increasing f gives Publisher 1 more incentive to increase a, leading to an overall increase in the total conversion rate. The argument for the decrease in f^* with an increase in p_0 is similar as the advertiser wants to incentivize Publisher 2 to put more effort.

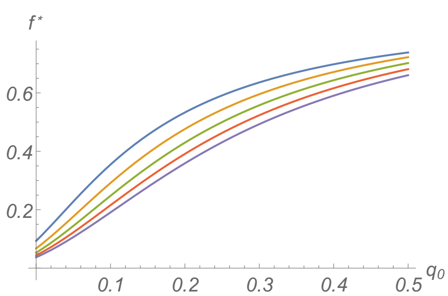

Figure 5: The value of f^* as a function of q_0 for various values of v, $p_0 = 0.25$, and $w = 2$.

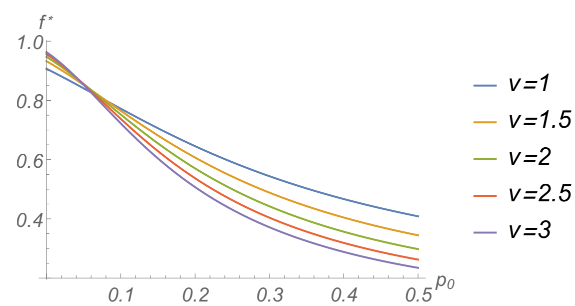

Figure 6: The value of f^* as a function of p_0 for various values of v, $q_0 = 0.25$, and $w = 2$.

Our result indicates that if the level of awareness of a product or brand is high, the family of f-contracts needs to incentivize the publisher creating awareness even more, as the second publisher that leads to conversions is automatically receiving a relatively higher payoff. This suggests the counterintuitive recommendation that well known advertisers should be allocating relatively more resources to informational advertising. On the other hand, products that deliver high intrinsic value to consumers so that consumers are likely to buy them if they are aware of the products represent products with high p_0: advertisers for such products should increase their spending on persuasive forms of advertising (in the second stage). As explained in the previous paragraph, these will best balance the automatic incentive for increase in effort in the other stage in optimally splitting the budget. Figures 5 and 6 illustrate the proposition and show the optimal contract as a function of q_0 and p_0 for some fixed values of the parameters.

In the preceding discussion, we presented the properties of the optimal contract. In reality however, the advertiser may not know the baseline rate of awareness and consideration. However, the following corollary offer insights on how the advertiser should split the advertising dollar if he knows the relationship between the two baseline rates in the first and second stage. If the advertiser has reasons to believe that a larger fraction of consumers are aware of the product but a relatively smaller fraction are likely to convert on their own then he should compensate Publisher 1 more if the cost of advertising is the same across the two publishers.

Corollary 1. *Assume that $w = v$. It holds that*

- *If $q_0 > p_0$, then $f^* > \frac{1}{2}$.*
- *If $q_0 = p_0$, then $f^* = \frac{1}{2}$.*
- *If $q_0 < p_0$, then $f^* < \frac{1}{2}$.*

5.2.2 Social Welfare and Price of Anarchy

How does the optimal f-contract perform compared to a centralized solution that maximizes the social welfare? It is clear that the f-contract will be worse than the first-best or socially optimal solution in terms of the social welfare, but it is not clear if the resulting conversion rate would be higher or lower. The following proposition answers this question. It shows that the first benchmark model, where a central planner maximizes the social welfare, can yield a result where everyone (including the firm) is better off compared to the optimal f-contract.

Proposition 2. *Consider the first benchmark (where we maximize social welfare of the publishers) and the equilibrium in the optimal f-contract. In the first benchmark, the conversion rate is higher. Moreover, sometimes the publishers' payments are higher too.*

Proposition 2 shows that the f-contract gives an inefficient equilibrium, since there can be an alternative solution where everyone is better off. This is a result of the lack of coordination between the publishers. In the centralized solution, both publishers put more effort which turns out to be good for both of them. However, in the f-contract equilibrium we observe a version of the prisoner's dilemma. In particular, when one of the publishers puts a lot of effort, the other one prefers to lower his effort and free-ride. We believe that this result is quite unique in the attribution literature as we show that the general class of f-contracts, which are commonly used in practice, can result in a prisoner's dilemma that adversely affects the efforts exerted by the different publishers.

This result illustrates how multi-channel attribution can be a blind spot for advertisers. Not only they might have limited information that obstructs their view of what the optimal attribution is, but there is also some inherent inefficiency in the widely used multi-touch rules. Thus, even if an advertising firm obtains good estimates of the conversion propensities in the market and determines the best multi-touch, the outcome can still be inefficient. In Section 6, we explore how we can use alternative attribution rules to resolve this problem.

Given the inefficiency, a related issue is how bad an f-contract can be in terms of social welfare. This can be measured by a concept called the Price of Anarchy (Koutsoupias and Papadimitriou, 1999), defined as the maximum value of the ratio between the social welfare in the optimal centralized (first-best) solution and the worst possible social welfare in any equilibrium.[15] The higher the price of anarchy, the more inefficient the equilibrium.[16]

As an example in our setup, if we consider the last-touch contract ($f = 0$), with the publishing costs set as low as possible, i.e. $w = v = 1$, and the baseline rate of the second stage set as low as possible, i.e. $p_0 = 0$, then the social welfare in the first benchmark is $\frac{q_0^2}{3}$, while the social welfare in the equilibrium is $\frac{q_0^2}{4}$. This gives a ratio of $\frac{4}{3}$. In the following proposition, we show that this is actually the worst possible case, i.e. the price of anarchy is equal to $\frac{4}{3}$. In other words, the social welfare in an f-contract is never worse that $\frac{3}{4}$ of the optimal.

Proposition 3. *Let SW_{OPT} be the social welfare in the first benchmark and SW_f be the social welfare in an f-contract equilibrium. It holds that*

$$1 \leq \frac{SW_{OPT}}{SW_f} \leq \frac{4}{3}.$$

Moreover, as $w \to +\infty$ or $v \to +\infty$, the ratio $\frac{SW_{OPT}}{SW_f}$ tends to 1.

The intuition behind why an increase in imbalance of the publishing costs w and v results in the social welfare approaching the optimal (first-best) social welfare, is that the high costs resolve the lack of coordination between the publishers in the prisoner's dilemma described above. As the cost of advertising in a particular stage increases, the publisher of this stage puts lesser and lesser effort. At some point he will put no effort at all in both the centralized solution and the equilibrium. At that point only one publisher exerts effort, which means that there is perfect coordination resulting in the first-best outcome.

Figure 7 shows the values of the ratio $\frac{SW_{OPT}}{SW_{\text{Last-Touch}}}$ for various parameters. We see that last-touch performs the worst in terms of social welfare when the cost v in the second stage is low and when the baseline p_0 in the second stage is low. Figure 8 shows the comparison of the conversion rate in the first benchmark case to the conversion rate under the optimal f-contract. Note that the conversion rate of the optimal f-contract, similarly to the social welfare, performs the worst when the costs (v and w) are low.

In summary, for both the firm and the publishers, multi-touch contracts (even the optimal one) are sub-optimal, while they perform better when there is a high degree of heterogeneity in the advertising costs across the two channels (i.e. $w \gg v$ or $w \ll v$). This indicates that advertising will be more efficient (and profitable) in industries where there is a wide discrepancy in the advertising costs.

6 Beyond Multi-touch Contracts

In the previous section we showed that the general class of f-contracts are generally sub-optimal. We now consider alternative payment functions and compare them with each other in terms of the conversion rate in the equilibrium.

[15] In our case the equilibrium is unique, so we don't have to worry about determining the worst.

[16] For a prominent example where this concept is used in computer science, see Roughgarden and Tardos (2002).

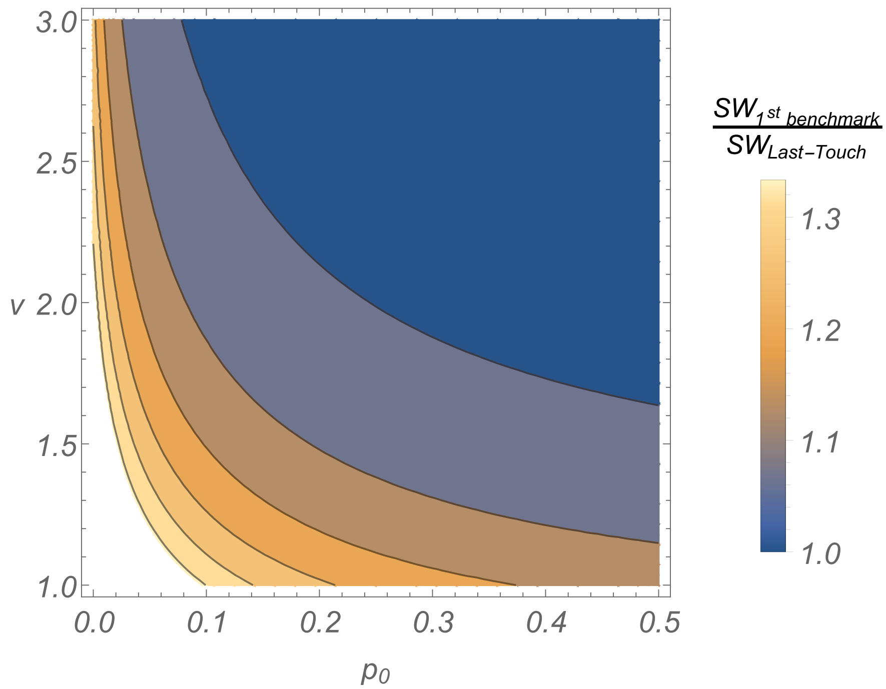

Figure 7: Social welfare in the first benchmark (publishers' optimal) over the social welfare in the last-touch contract as a function of p_0 and v, for $q_0 = 0.25$, and $w = 2$.

6.1 Optimal Contract

One may wonder if there is a valid contract (in the sense of Section 4.4) that achieves the optimal conversion rate (the one achieved in the second benchmark by a central planner). The following proposition shows that this is actually possible with a contract that is linear w.r.t. the observed products q_0p_0, q_0c, ap_0, ac.

Proposition 4. *There is a contract that achieve the optimal conversion rate, given by the following payment functions.*

$$g_1(a, q_0, c, p_0) = \frac{1}{2}q_0p_0 + sap_0 + tq_0c + fac$$
$$g_2(a, q_0, c, p_0) = \frac{1}{2}q_0p_0 + (1-s)ap_0 + (1-t)q_0c + (1-f)ac$$

The values of s, t, f are defined below.

$$s = 1 + \frac{q_0p_0 + 2v(c^*)^2}{2a^*p_0}$$
$$t = -\frac{q_0p_0 + 2w(a^*)^2}{2q_0c^*}$$
$$f = \frac{1}{2} + \frac{-a^*p_0 + q_0c^* + 3w(a^*)^2 - 3v(c^*)^2}{2a^*c^*}$$

Note that a^*, c^* in the above proposition are the optimal values of the optimization problem in the second benchmark (Section 4.5).

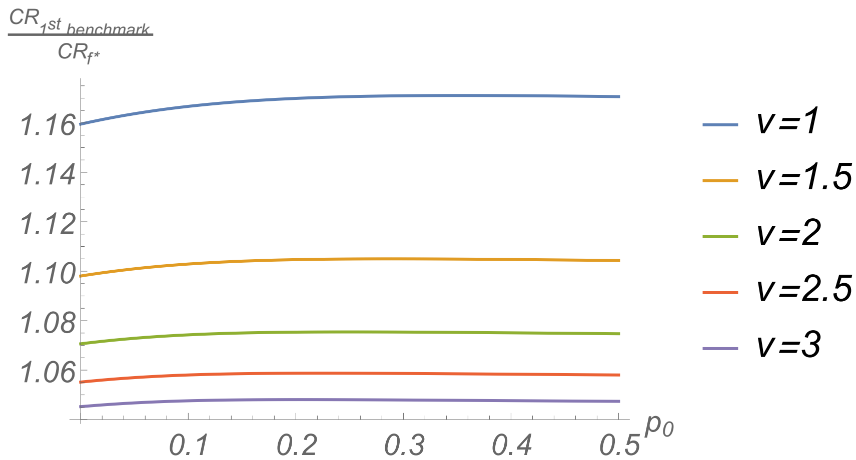

Figure 8: Conversion rate in the second benchmark (firm's optimal) over the conversion rate in the optimal f-contract as a function of p_0 for various values of v, $q_0 = 0.25$, and $w = 2$.

In order to understand why the contract in Proposition 4 achieves the optimal rate, it is useful to compare it to the payments for the publishers in Equation 1. As mentioned earlier, any f-contract suffers from a free-riding problem. Both publishers want the *other* publisher to exert effort that leads to a prisoner's dilemma. In the payoff outlined here, the publishers are punished for the effort exerted by the other publisher (both t and $1-s$ are negative). This payment scheme disincentivizes both publishers from free-riding. By appropriately choosing s, t and f, the advertiser is able to achieve perfect coordination between the publishers. In equilibrium the publishers' profits are zero (their outside option).[17]

A drawback of the aforementioned contract is that it is not clear how an advertiser can implement this contract if he has incomplete information about q_0 and p_0, which are required to determine the values of a^* and c^*.

Even though the preceding contract might not be implementable, it provides us the intuition that contracts that penalize publishers for the effort of the other publisher, instead of just compensating them for their efforts, tend to work better. The penalty due to under-performance can increase the efficiency of the system, because it gives incentive to both publishers to put more effort instead of free-riding and depending on the effort of the other publisher. In the following subsection, we define one such contract. We call it a reinforcement contract[18] and show that it performs better than any multi-touch contract in the majority of the cases. Moreover, the firm does not need to know the exact values of the baseline probabilities to implement it, which makes it more practical.

6.2 Reinforcement Contracts

Perhaps the simplest form of a valid contract is given by the payment function $g_1 = \frac{1}{2}(a + q_0)(c + p_0)$, which we call *equal-split*. According to this payment, both publishers get the

[17]We can extend the model such that the publishers have a non-zero outside option, but the qualitative results do not change.

[18]The name is motivated by the similarity to rewarding good performance and penalizing bad performance in reinforcement learning.

same credit for every conversion. The reinforcement or (r)-contract is a generalization of the equal-split ($r = 0$) and it is given by the following payment function.

$$g_1 = \left(\frac{1}{2} + \left(\frac{a}{a+q_0} - \frac{c}{c+p_0}\right) r\right)(a+q_0)(c+p_0), \text{ for } r \in \mathbb{R}.$$

It is evident from the preceding equation that an (r)-contract not only compensates publishers for the effort they put, but it also penalizes them for the effort the other publisher puts. The following table shows the canonical form of the corresponding attribution rule.

Ad in the awareness stage	Ad in the conversion stage	Credit to Publisher 1	Credit to Publisher 2
no	no	$\frac{1}{2}$	$\frac{1}{2}$
no	yes	$\frac{1}{2} - r$	$\frac{1}{2} + r$
yes	no	$\frac{1}{2} + r$	$\frac{1}{2} - r$
yes	yes	$\frac{1}{2}$	$\frac{1}{2}$

Note that r can take values above $\frac{1}{2}$, i.e. this attribution rule has the unique property of being able to assign negative credit to channels that do not appear in conversion paths.[19] Thus, instead of focusing only on the touch points like the multi-touch attribution rules do, it pays attention to non-touch points as well. This change of focus can help resolve the commitment problem illustrated in Section 5.2.2. The following proposition will help us determine the optimal (r)-contract.

Lemma 1. *The conversion rate under an (r)-contract is a convex function of r.*

Lemma 1 tells us that the optimal (r)-contract is achieved for the maximum possible value of r that keeps both publishers into the game, i.e. both publishers have non-negative payoff by putting some effort in the equilibrium. This is useful, because in a practical situation the firm can determine the optimal (r)-contract by increasing the value of r until one of the publishers drops out. In other words, implementing the (r)-contract reduces to a simple search problem in a repeated dynamic environment.

A practical way to implement a type of reinforcement contract like the (r)-contract is to use an attribution method that assigns negative credit to publishers who do not appear in a conversion path. This will incentivize publishers to put more effort into being part of conversion paths, and as a result the conversion rates will increase.

In Figure 9, we can see a comparison of performance of the different contracts we considered in this paper. As we can see, the optimal (r)-contract performs significantly better than any other contract that does not require the knowledge of the baselines in the majority of the cases (except for very small values of q_0, where an f-contract is better). The chosen values of the parameters for drawing the plot are arbitrary and the picture looks very similar for different values as well. Note also that in the symmetric case, i.e. when $w = v$ and $q_0 = p_0$, the optimal (r)-contract achieves the first best conversion rate.

[19] In an f-contract, even if we allow f to take negative values, it is never optimal to do so. As we can see in the proof of Proposition 1, it is always $0 \le f^* \le 1$.

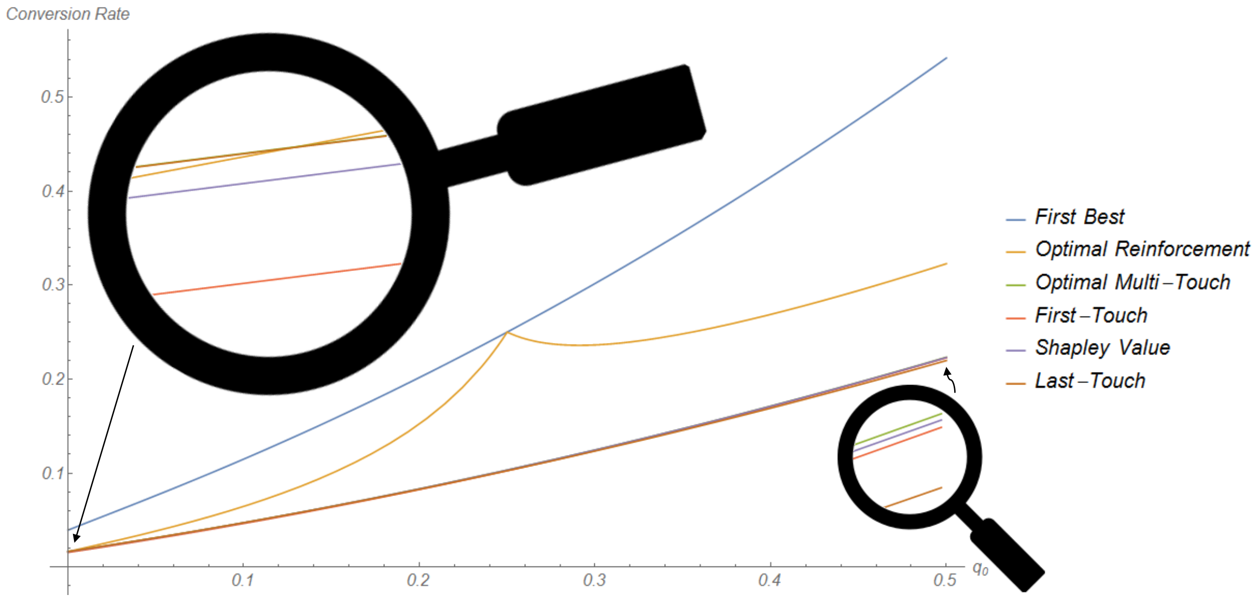

Figure 9: Conversion rate for different contracts as a function of q_0 (for $w = v = 2$ and $p_0 = 0.25$). The first best is the one that maximizes conversion rate. The reinforcement contract is second best for a wide range of parameter values.

7 Extensions

In this section, we first discuss how the optimal f-contract behaves as a function of the cost parameters of the publishers, and how a firm can choose the optimal f-contract under limited uncertain information about baseline conversion rates. Then we examine the effect of competition on the advertising game.

7.1 Effect of Cost Parameters of Publishers in Optimal f-contracts

We now return to the popular f-contracts and explore how the advertising dollar might be split between the two publishers based on the difficulty of advertising in the different stages. The advertising cost can be driven by several factors, e.g. the reach of a publisher, its ability to target the right set of consumers, the intensity of advertising by competitors, or type of the product being sold.

Recall that w represents the cost of advertising effort in the first stage, and v the corresponding value for the second.

Proposition 5. *There is a threshold $\overline{w} \in [1, +\infty]$ such that f^* is decreasing in w for $w \in [1, \overline{w}]$ and increasing in w for $w \in [\overline{w}, +\infty)$.*[20] *Similarly, there is a threshold $\overline{v} \in [1, +\infty]$ such that f^* is increasing in v for $v \in [1, \overline{v}]$ and decreasing in v for $v \in [\overline{v}, +\infty)$.*

Interestingly, f^* is not monotone with respect to w. It is easy to see that when the cost of advertising in the first stage increases, Publisher 1 would reduce the amount of effort. As the

[20] $\overline{w}$ is given implicitly by the solution to the equation $\overline{w} = \frac{p_0(1-f^*(\overline{w}))}{2q_0}$.

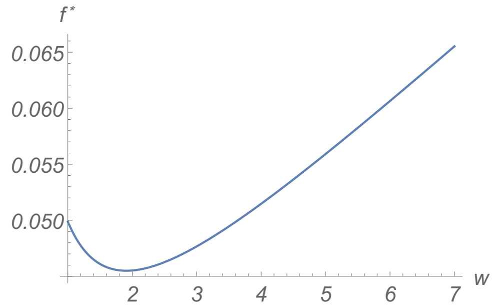

Figure 10: The value of f^* as a function of w, for $q_0 = 0.1$, $p_0 = 0.4$, and $v = 10$. The monotonicity changes at $\overline{w} = 1.909$.

result, the advertiser compensates Publisher 1 by increasing f^* to induce more effort as w increases. However, we find that when w is sufficiently small, f^* decreases when w increases. In order to understand this counterintuitive finding, it is important to note that even though Publisher 1 is affected directly due to the increase in advertising cost, Publisher 2 is also affected indirectly as w increases. This results in Publisher 2 exerting less effort because his payoff is reduced due to the multiplicative factor of the two efforts in the payoff function presented in Equation 1. Consequently, the advertiser needs to incentivize both publishers to increase the effort. When w is relatively small, it is easier to increase the effort exerted by Publisher 1 as compared to Publisher 2. Under this condition, it is more important to increase the fraction paid to Publisher 2 to increase the overall conversion rate, which leads to a decrease in f^* with w. Figure 10 shows the variation in f^* with w.

The following corollary shows how the different advertising costs affect f^*. A firm which uses multi-touch attribution for its campaign should give more credit to the advertising channel with the highest publishing cost, all else being equal.

Corollary 2. *Assume that $q_0 = p_0$. It holds that*

- *If $w > v$, then $f^* > \frac{1}{2}$.*
- *If $w = v$, then $f^* = \frac{1}{2}$.*
- *If $w < v$, then $f^* < \frac{1}{2}$.*

7.2 Implementing the f-contract Under Uncertainty

To find the optimal f-contract, the firm needs to know the baseline rates q_0, p_0. In this section, we propose an approach that can be used by the firm to determine the optimal f under asymmetric information. The following proposition shows that using very little information the firm can derive the optimal (in expectation) f-contract.

Proposition 6. *Assume that the firm does not know the values of q_0, p_0, but it has some information about the distributions from which they are drawn. Then, the firm can calculate the value f^* of the optimal (in expectation) f-contract, by only using the moments* $\mathrm{E}[q_0 p_0]$, $\mathrm{E}[q_0^2]$, *and* $\mathrm{E}[p_0^2]$.

This proposition shows that the advertiser does not need too much information to derive f^*. As long as he knows the second moments of the baseline rates , he can find the optimal split between the two publishers. Now, suppose that a firm knows that both the publishing costs and the expected baselines rates are the same in the two stages, the following corollary sheds light on how the advertiser should set f^*.

Corollary 3. *Let f^* be the value of the optimal (in expectation) f-contract for the firm. If $v = w$ and $\mathrm{E}[q_0] = \mathrm{E}[p_0]$, then*

- *If* $\mathrm{Var}[q_0] = \mathrm{Var}[p_0]$, *then* $f^* = \frac{1}{2}$.
- *If* $\mathrm{Var}[q_0] > \mathrm{Var}[p_0]$, *then* $f^* > \frac{1}{2}$.
- *If* $\mathrm{Var}[q_0] < \mathrm{Var}[p_0]$, *then* $f^* < \frac{1}{2}$.

Corollary 3 shows that a firm which uses multi-touch attribution for its campaign should give more credit to the advertising channel with higher uncertainty about the baseline rate, all else being equal.

7.3 Competition in Each Stage

In this extension, we assume that both publishers can show ads in both stages of the game. In other words, there is competition not only across stages but also in each stage. In this variation, we want to see how increased competition between the publishers will affect our results.

As in the original model, there is a continuum of consumers of mass 1 moving as in the diagram of Figure 2. The variable a now will be a vector (a_1, a_2), where a_1 is a decision variable for the first publisher, with cost $w_1 a_1^2$, and a_2 is a decision variable for the second publisher, with cost $w_2 a_2^2$. Similarly, the variable c will be a vector (c_1, c_2), where c_1 is a decision variable for the first publisher, with cost $v_1 c_1^2$, and c_2 is a decision variable for the second publisher, with cost $v_2 c_2^2$. For the functions f and g, we assume the functional forms $f(a) = a_1 + a_2 + q_0$ and $g(c) = c_1 + c_2 + p_0$, where q_0 and p_0 are the baseline probabilities.

In each stage, we assume that a consumer can see at most one ad. The effort a_1 represents the probability that during the awareness stage he will see an ad from the first publisher, a_2 is the probability that he will see an ad from the second publisher, and q_0 is the probability that he will move to the conversion stage without seeing an ad. Similarly, the effort c_1 represents the probability that during the conversion stage he will see an ad from the first publisher, c_2 is the probability that he will see an ad from the second publisher, and p_0 is the probability that he will decide to buy the product without seeing an ad during the conversion stage.

As before, we assume that the baseline probabilities q_0, p_0 cannot be more than $\frac{1}{2}$ and that the cost parameters v_1, v_2, w_1, w_2 are sufficiently large (so that $f(a), g(c) \in [0, 1]$ in the equilibrium).

For every conversion, the firm wants to spend \$1 in advertising. The question is what is the optimal way to split this dollar between the two publishers in order to maximize the number of conversions, which is $f(a)g(c)$.

The equivalent of an f-contract in this variation is summarized in the following table.

Ad in the awareness stage	Ad in the conversion stage	Credit to Publisher 1	Credit to Publisher 2
0	0	$\frac{1}{2}$	$\frac{1}{2}$
0	1	1	0
0	2	0	1
1	0	1	0
1	1	1	0
1	2	f	$1-f$
2	0	0	1
2	1	$1-f$	f
2	2	0	1

A 0 in the first two columns means no ad, a 1 means ad from the first publisher and a 2 means ad from the second publisher. The parameter $f \in [0,1]$ is some value determined by the firm or externally.

As in the original model, some examples of contracts of this form are the last-touch ($f=0$), where all the credit goes to the last ad the consumer had seen prior to the conversion, and the first-touch ($f=1$) where all the credit goes to the first ad, which made the consumer aware of the product. The following lemma will help us determine the optimal f-contract in this extended model.

Lemma 2. *Let $r(f)$ be the conversion rate in an f-contract as a function of f. Then, $r(f)$ is concave in $(0,1)$.*

Since $r(f)$ is concave, we know that the optimal value f^* is either the single root of the equation $r'(f)=0$ in $(0,1)$ (when it exists), or 0, or 1. More specifically, if $r'(f)<0$ in $(0,1)$, then $f^*=0$. If $r'(f)>0$ in $(0,1)$, then $f^*=1$. If there is a root f such that $r'(f)=0$, then it is unique and f^* is equal to this root.

To show the robustness of our results in this extended model, we consider the following numerical examples.

Figures 11 and 12 are the analogs of Proposition 1. As we can see, f^* is increasing in q_0 and decreasing in p_0.

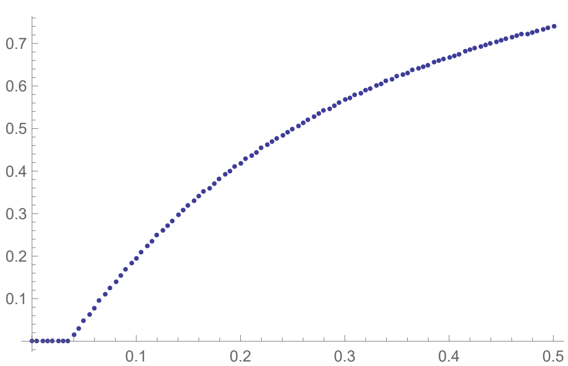

Figure 11: Optimal f for different values of q_0, for $v_1=v_2=w_1=w_2=2$, $p_0=\frac{1}{4}$.

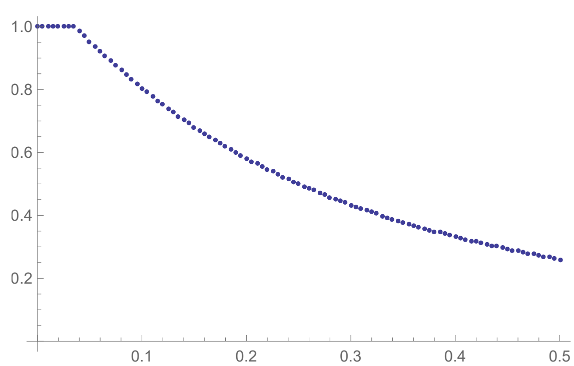

Figure 12: Optimal f for different values of p_0, for $v_1=v_2=w_1=w_2=2$, $q_0=\frac{1}{4}$.

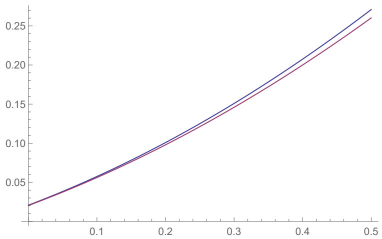

Figure 13: Social Welfare in the equilibrium (purple) and in the first best (blue) for different values of q_0, for $v_1 = v_2 = w_1 = w_2 = 2$, $p_0 = \frac{1}{4}$, $f = 0$.

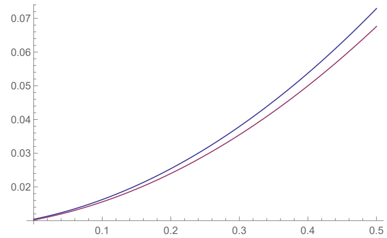

Figure 14: First publisher's profit in the equilibrium (purple) and in the first best (blue) for different values of q_0, for $v_1 = v_2 = w_1 = w_2 = 2$, $p_0 = \frac{1}{4}$, $f = 0$.

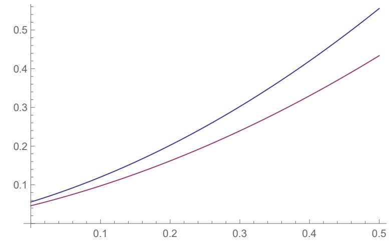

Figure 15: Firm's revenue in the equilibrium (purple) and in the first best (blue) for different values of q_0, for $v_1 = v_2 = w_1 = w_2 = 2$, $p_0 = \frac{1}{4}$, $f = 0$.

Figures 13, 14, and 15 are the analogs of Proposition 2. As we can see, in the first best the conversion rate is higher, and sometimes publisher's profits are higher too.

Figures 16 and 17 are the analogs of Proposition 3. In this case, the upper bound of the ratio is even tighter at $\frac{512}{507}$.

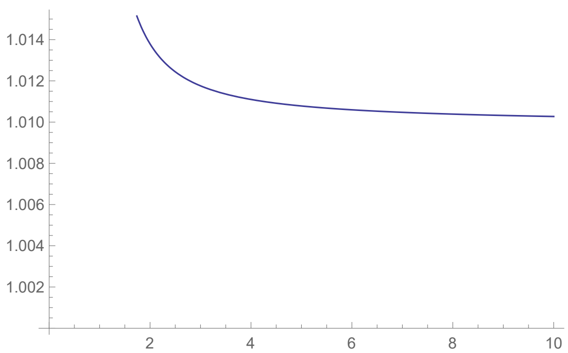

Figure 16: Ratio of Social welfare in the first best over the Social welfare in the equilibrium for different values of a, for $q_0 = p_0 = \frac{1}{4}$, $f = 0$, $v_1 = w_2 = lo = 2$, $v_2 = w_1 = hi$, $hi = a \cdot lo$. The limit as $a \to \infty$ is $\frac{512}{507} = 1.00986$.

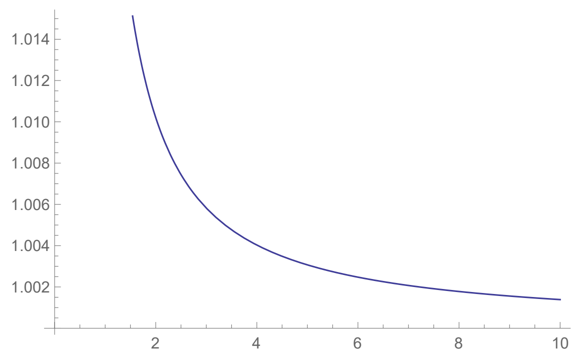

Figure 17: Ratio of Social welfare in the first best over the Social welfare in the equilibrium for different values of a, for $q_0 = p_0 = \frac{1}{4}$, $f = 0$, $v_1 = v_2 = lo = 2$, $w_1 = w_2 = hi$, $hi = a \cdot lo$. The limit as $a \to \infty$ is 1.

Finally, Figures 18 and 19 are the analogs of Proposition 6 and Corollary 3. The firm can determine f^* by using the moments $\mathrm{E}[q_0 p_0]$, $\mathrm{E}[q_0^2]$, and $\mathrm{E}[p_0^2]$. Moreover, if q_0 and p_0 have the same mean, but q_0 has higher variance, then $f^* > \frac{1}{2}$.

8 Conclusion

The online advertising industry has continued to grow rapidly in the last two decades. While it offers several advantages over traditional advertising, the information asymmetry and

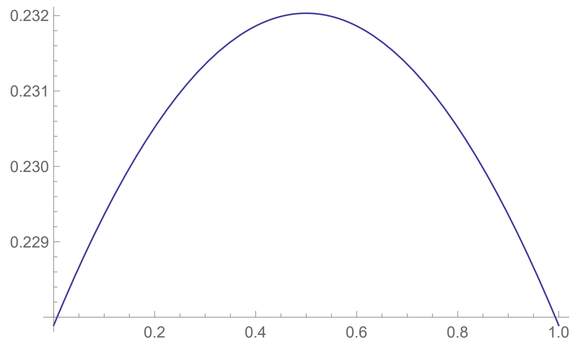

Figure 18: Firm's revenue for different values of f when $\mathrm{E}[p_0] = \mathrm{E}[q_0] = \frac{1}{4}$ and $\mathrm{E}[p_0^2] = \mathrm{E}[q_0^2] = \frac{1}{12}$, for $v_1 = w_1 = v_2 = w_2 = 2$. The optimal value is for $f = \frac{1}{2}$.

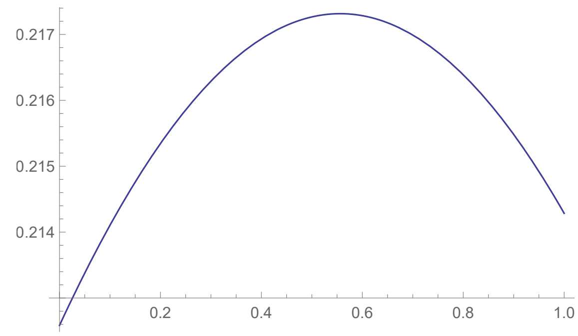

Figure 19: Firm's revenue for different values of f when $\mathrm{E}[p_0] = \mathrm{E}[q_0] = \frac{1}{4}$ and $\mathrm{E}[p_0^2] = \frac{1}{16}$, $\mathrm{E}[q_0^2] = \frac{1}{12}$, for $v_1 = w_1 = v_2 = w_2 = 2$. The optimal value is for $f = 0.556179$.

misalignment of incentives between the advertisers and publishers poses a threat to this industry in form of ad fraud, substandard ad inventory, and sub-optimal effort by advertisers. In this paper, we formalize this problem in the context where consumers move through two stages before they purchase and show how standard contracts, also known as multi-touch contracts, might not lead to the most effective outcome for advertisers *and* publishers. Multi-touch contracts can result in the advertiser receiving lower return on investment and the publishers generating smaller revenues in equilibrium. The inefficiency in multi-touch attribution arises due to a prisoner's dilemma where the dominant strategy of publishers in both stages is to exert less effort in showing ads. We also show that our results are in line with real-world observations. Furthermore, motivated by optimal contracts, we introduce reinforcement contracts, that penalize publishers if they exert relatively less effort than the other publisher, and show how they can reduce the prisoner's dilemma and lead to higher profits for the advertiser.

Our research has several managerial implications. One important findings of our research is that advertisers should spend a relatively larger fraction of their advertising budget on the stage of the purchase funnel that they believe has a higher baseline conversion rate. E.g. if an advertiser knows that the level of brand (or product) awareness is higher than the baseline conversion probability in the consideration stage, they should spend more on brand advertising, even if the exact levels are unknown. Secondly, in industries where the cost to create awareness is similar to the cost of showing consumers ads to drive conversion, the advertisers should engage with publishers that can show ads in both stages of the funnel. This will reduce the prisoner's dilemma across the stages and help in better coordination of the incentives between the advertiser and the publisher. Furthermore, the results presented in this paper underscore the need for better transparency in online advertising. The online advertising ecosystem is becoming extremely complicated with many participants, often with misaligned incentives. This has led to an increase in data fragmentation, which increases the information asymmetry. This can lead to online advertising becoming a less effective medium and reduce advertisers' desire to move their advertising budget online. Sharing information across all the participants resolves this problem and can increase the effectiveness of online advertising, making both advertisers and publishers better off. Finally, if the advertiser has enough market power, it should carefully consider reinforcement contracts for attribution to increase the returns from advertising. This will prevent the publishers from getting stuck in the unfavorable prisoner's dilemma equilibrium.

Our research has a few limitations and presents directions for future work. Our model focuses on a single advertiser as he allocates credit between competing publishers. In that sense, our advertiser has market power to dictate the attribution rule. A possible extension is to consider how competition between different advertisers can affect the attribution process, where the market power of one advertiser would be considerably reduced. Our insights can be used by each advertiser in the stage of budget allocation among different advertising channels, but advertisers' competition could give some extra insights on patterns of overall spending.

The reinforcement attribution contracts we suggest as improvement over multi-touch contracts have some good theoretical properties and are more practical than optimal contracts, but there might still be some challenges in their implementation. The key insight from them is that assigning negative credit to non-appearances of advertising channels in conversion paths can be beneficial for the advertiser, because it increases competition between the publishers and it incentivizes them to put more effort. However, an advertiser who wants to implement such an attribution contract should also take into account the quality of the ads. A low-quality ad that appears in a lot of conversion paths should not be incorrectly considered as important for conversions. To do this properly, an advertiser needs to have sufficient data about consumer paths that did not lead to a conversion. A future research direction, therefore, is to dive more into the intricacies regarding making reinforcement contracts more practical.

Finally, one important result of our research is to show that an attribution based on Shapley value is often not optimal for advertisers. Shapley value uses the marginal contribution of each publisher to the conversion rate in order to provide a fair attribution, but it does not take into account the publishers' incentives. As a result, we have shown that there are alternative attribution schemes with better results for the advertiser. Another future research direction is to extend the framework of this paper and provide a general formulation of an optimal scheme for ad attribution.

References

V. Abhishek, P. Fader, and K. Hosanagar. Media exposure through the funnel: A model of multi-stage attribution. *SSRN*, 2016. 2, 3, 5, 7

E. Anderl, I. Becker, F. V. Wangenheim, and J. H. Schumann. Putting attribution to work: A graph-based framework for attribution modeling in managerial practice. *SSRN*, 2343077, October 2013. 5

A. Ansari, K. Bawa, and A. Ghosh. A nested logit model of brand choice incorporating variety-seeking and marketing-mix variables. *Marketing Letters*, 6(3):199–210, 1995. 2

R. Berman. Beyond the last touch: Attribution in online advertising. *Working paper*, March 2015. 2, 3, 7

R. Berman. Beyond the last touch: Attribution in online advertising. August 2016. Available at SSRN: https://ssrn.com/abstract=2384211 or http://dx.doi.org/10.2139/ssrn.2384211. 5

J. R. Bettman, M. F. Luce, and J. W. Payne. Constructive consumer choice processes. *Journal of Consumer Research*, 25(3):187–217, December 1998. 3, 8

D. Court, D. Elzinga, S. Mulder, and O. J. Vetvik. The consumer decision journey. *McKinsey Quarterly*, June 2009. 3, 8

B. Dalessandro, O. Stitelman, C. Perlich, and F. Provost. Causally motivated attribution for online advertising. *Proceedings of the Sixth International Workshop on Data Mining for Online Advertising and Internet Economy*, pages 7:1–7:9, 2012. 5, 7

J. A. Dearden and G. L. Lilien. On optimal salesforce compensation in the presence of production learning effects. *International Journal of Research in Marketing*, 7(2-3):179–188, 1990. 6

Economist. Invisible ads, phantom readers. May 2016. 3

eMarketer. Digital ad spending benchmarks by industry. *Executive Summary*, May 2015.

M. Eswaran and A. Kotwal. The moral hazard of budget-breaking. *The Rand Journal of Economics*, pages 578–581, 1984. 5

V. Goel. In advertising, it's about who gets credit for the sale. *New York Times*, December 2014. URL http://bits.blogs.nytimes.com/2014/12/26/in-advertising-its-about-who-gets-credit-for-the-sale/. 2

Google. About the default attribution models, 2017a. URL https://support.google.com/analytics/answer/1665189?hl=en&ref_topic=3205717. Accessed: March 2017. 2, 7, 13

Google. Data-driven attribution methodology, 2017b. URL https://support.google.com/analytics/answer/3191594. Accessed: March 2017. 2, 7

B. Holmstrom. Moral hazard in teams. *The Bell Journal of Economics*, pages 324–340, 1982. 3, 5

B. Holmstrom and P. Milgrom. Aggregation and linearity in the provision of intertemporal incentives. *Econometrica: Journal of the Econometric Society*, pages 303–328, 1987. 6

M. Irvine. Google adwords benchmarks for your industry. http://www.wordstream.com/blog/ws/2016/02/29/google-adwords-industry-benchmarks, 2016. Accessed: March 2017.

P. Jordan, M. Mahdian, S. Vassilvitskii, and E. Vee. The multiple attribution problem in pay-per-conversion advertising. In *Proceedings of the 4th international conference on Algorithmic game theory*, SAGT'11, pages 31–43, 2011. 5

E. Koutsoupias and C. Papadimitriou. Worst-case equilibria. In *Annual Symposium on Theoretical Aspects of Computer Science*, pages 404–413. Springer, 1999. 4, 15

T. Lan, D. Kao, M. Chiang, and A. Sabharwal. An axiomatic theory of fairness in network resource allocation. In *INFOCOM, 2010 Proceedings IEEE*, pages 1–9, March 2010. 7

H. Li and P. Kannan. Attributing conversions in a multichannel online marketing environment: An empirical model and a field experiment. *Journal of Marketing Research*, 51(1): 40–56, 2014. 5

R. P. McAfee and J. McMillan. Optimal contracts for teams. *International Economic Review*, pages 561–577, 1991. 6

S. Mulpuru. The purchase path of online buyers. *Forrester Report*, March 2011. 3, 8

P. A. Naik, K. Raman, and R. S. Winer. Planning marketing-mix strategies in the presence of interaction effects. *Marketing Science*, 24(1):25–34, Jan. 2005. 2

PWC. Cross channel attribution is needed to drive marketing effectiveness. Technical report, 2014. 2

M. Rabin. Incorporating fairness into game theory and economics. *The American Economic Review*, 83(5):1281–1302, 1993. 7

V. Ramaswamy, W. S. Desarbo, D. J. Reibstein, and W. T. Robinson. An empirical pooling approach for estimating marketing mix elasticities with pims data. *Marketing Science*, 12 (1):103–124, 1993. 2

T. Roughgarden and É. Tardos. How bad is selfish routing? *Journal of the ACM (JACM)*, 49(2):236–259, 2002. 4, 16

S. Scott. The $8.2 billion adtech fraud problem that everyone is ignoring. *Techcrunch*, January 2016. 3

X. Shao and L. Li. Data-driven multi-touch attribution models. KDD '11, pages 258–264, 2011. 5

L. S. Shapley. *A Value for n-Person Games*. Princeton University Press, 1953. 7, 39

R. van den Brink. An axiomatization of the shapley value using a fairness property. *International Journal of Game Theory*, 30(3):309–319, 2002. 7

L. Xu, J. A. Duan, and A. Whinston. Path to purchase: A mutually exciting point process model for online advertising and conversion. *Management Science*, 60(6):1392–1412, 2014. 5

A Appendix

A.1 Analyses and Proofs

Proof of Proposition 1. The payoff of the first publisher in the f-contract is $\frac{1}{2}q_0p_0 + ap_0 + fac - wa^2$, while the payoff of the second publisher is $\frac{1}{2}q_0p_0 + q_0c + (1-f)ac - vc^2$. This means that the equilibrium efforts are $a = \frac{fq_0+2vp_0}{4vw-f(1-f)}$ and $c = \frac{(1-f)p_0+2wq_0}{4vw-f(1-f)}$. Therefore, the conversion rate in the equilibrium is

$$r(f) = \frac{(f^2q_0 + 2v(p_0 + 2wq_0))((1-f)^2p_0 + 2w(q_0 + 2vp_0))}{(4vw - f(1-f))^2}.$$

It holds that $r'(f) =$

$$\frac{-4f^3(vp_0^2 + wq_0^2) + 12f^2vp_0^2 - 4f(4v^2wp_0^2 + 3vp_0^2 + 8vwp_0q_0 + 4vw^2q_0^2) + 4(p_0^2v + 4vwp_0q_0 + 4vw^2q_0^2)}{(4vw - f(1-f))^3}.$$

The discriminant of the cubic formula in the numerator is

$$\Delta = -256v^2w^2(2(vp_0^2 + wq_0^2) + p_0q_0)^3(2(vp_0^2 + wq_0^2)(27 + 16vw) + 9(3 + 16vw)p_0q_0) \leq 0,$$

which means that the equation $r'(f) = 0$ has a single real root. It's also $r'(0) = \frac{p_0^2v+4vwp_0q_0+4vw^2q_0^2}{16v^3w^3} \geq 0$ and $r'(1) = -\frac{(q_0+2p_0v)^2}{16v^3w^2} \leq 0$. Therefore, the single root is in the interval $[0, 1]$. That root is the value of f^*. The result is a long expression, so to write it down we use the following notation. Let

$$\begin{aligned}
&x = wq_0^2, y = vp_0^2, z = q_0p_0, H = 2(x+y) + z \\
&A = H\left(H(16xy + 27z^2) + 128xyz\right), \\
&B = \sqrt[3]{xyH\left(9(x-y)zH + \sqrt{3}(x+y)\sqrt{A}\right)}, \\
&n = \sqrt[3]{12}, m = 2\sqrt[3]{18}.
\end{aligned}$$

Then

$$f^* = \frac{y}{x+y} + \frac{nB^2 - mxyH(H+2z)}{6(x+y)zB}.$$

Now, we need to prove that f^* is increasing in q_0 and decreasing in p_0. We'll start by proving that it is increasing in q_0. Since $r'(f^*) = 0$, it holds that

$$-a(f^*)^3 + b(f^*)^2 - cf^* + d = 0,$$

where $a = 4(vp_0^2 + wq_0^2), b = 12vp_0^2, c = 4(4v^2wp_0^2 + 3vp_0^2 + 8vwp_0q_0 + 4vw^2q_0^2)$, and $d = 4(p_0^2v + 4vwp_0q_0 + 4vw^2q_0^2)$. By taking partial derivatives with respect to q_0 and solving for $\frac{\partial f^*}{\partial q_0}$, we get

$$\frac{\partial f^*}{\partial q_0} = \frac{-\frac{\partial a}{\partial q_0}(f^*)^3 + \frac{\partial b}{\partial q_0}(f^*)^2 - \frac{\partial c}{\partial q_0}f^* + \frac{\partial d}{\partial q_0}}{3a(f^*)^2 - 2bf^* + c}$$

The denominator is equal to

$$
\begin{aligned}
&12(vp_0^2 + wq_0^2)(f^*)^2 - 24vp_0^2 f^* + 4(4v^2wp_0^2 + 3vp_0^2 + 8vwp_0q_0 + 4vw^2q_0^2) = \\
&12vp_0^2(f^* - 1)^2 + 12wq_0^2(f^*)^2 + 16vw(vp_0^2 + 2p_0q_0 + wq_0^2) \geq 0,
\end{aligned}
$$

therefore, it is enough to prove that the numerator is positive. The numerator is equal to

$$
8w(-q_0(f^*)^3 - 4vw(p_0 + wq_0)f^* + 2vw(p_0 + 2wq_0^2)),
$$

thus we need to prove that $-q_0(f^*)^3 - 4vw(p_0 + wq_0)f^* + 2vw(p_0 + 2wq_0^2) \geq 0$. We know that $1 = \frac{a(f^*)^3 - b(f^*)^2 + cf^*}{d}$, so it is enough to prove that

$$
-q_0(f^*)^3 - 4vw(p_0 + wq_0)f^* + 2vw(p_0 + 2wq_0^2)\frac{a(f^*)^3 - b(f^*)^2 + cf^*}{d} \geq 0,
$$

or equivalently

$$
\frac{fp_0\left((2vp_0 - q_0)(f^*)^2 - 6vp_0 f^* + 4vw(q_0 + 2vp_0) + 2p_0v\right)}{p_0 + 2wq_0} \geq 0.
$$

It is

$$
\begin{aligned}
&(2vp_0 - q_0)(f^*)^2 - 6vp_0 f^* + 4vw(q_0 + 2vp_0) + 2p_0v = \\
&vp_0\left(2(f^*)^2 - 6f^* + 2 + 8vw\right) + q_0\left(-(f^*)^2 + 4vw\right) \geq \\
&vp_0\left(2(f^*)^2 - 6f^* + 2 + 8\right) + q_0\left(-(f^*)^2 + 4\right) = \\
&2vp_0\left(\left(f^* - \frac{3}{2}\right)^2 + \frac{11}{4}\right) + q_0\left(2 - f^*\right)\left(2 + f^*\right) \geq 0.
\end{aligned}
$$

Therefore, it holds that $\frac{\partial f^*}{\partial q_0} \geq 0$, which means that f^* is increasing in q_0. Because of symmetry, $1 - f^*$ is increasing in p_0, which means that f^* is decreasing in p_0. □

Proof of Corollary 1. Let $w = v$. If $q_0 = p_0$, it holds that $wq_0^2 = vp_0^2$, therefore using the notation of the proof of Proposition 1, $x = y$. This means that

$$
A = (4x + z)((4x + z)(16x^2 + 27z^2) + 128x^2z) = (4x + z)(4x + 3z)^3
$$

and

$$
B = \sqrt[3]{2x^3(4x + z)\sqrt{3}\sqrt{A}} = x\sqrt[3]{2\sqrt{3}}\sqrt{(4x + z)(4x + 3z)}.
$$

Thus,

$$
nB^2 - mxyH(H + 2z) = 2x^2\sqrt[3]{18}(4x + z)(4x + 3z) - 2\sqrt[3]{18}x^2(4x + z)(4x + 3z) = 0,
$$

which means that $f^* = \frac{y}{x+y} = \frac{1}{2}$.

From Proposition 1, we know that f^* is increasing in q_0, therefore if $q_0 > p_0$, $f^* > \frac{1}{2}$, and if $q_0 < p_0$, $f^* < \frac{1}{2}$. □

Proof of Proposition 2. The optimal efforts (in terms of social welfare) are $a_{OPT} = \frac{q_0+2vp_0}{4vw-1}$ and $c_{OPT} = \frac{p_0+2wq_0}{4vw-1}$. In an f-contract, the equilibrium efforts are $a_f = \frac{fq_0+2vp_0}{4vw-f(1-f)}$ and $c_f = \frac{(1-f)p_0+2wq_0}{4vw-f(1-f)}$. It holds that $a_{OPT} \geq a_f$ and $c_{OPT} \geq c_f$, which means that the conversion rate is higher in the first benchmark. This is true for every f-contract, therefore for the optimal f^*-contract as well.

The first publisher's payment in the first best solution (under an f-contract) is

$$p_{1,OPT} = \frac{q_0 p_0}{2} + \frac{(2p_0 v + q_0)(p_0(f + 2vw - 1) + (2f-1)q_0 w)}{(4vw-1)^2},$$

while in the equilibrium of an f-contract is

$$p_{1,f} = \frac{q_0 p_0}{2} + \frac{w(fq_0 + 2p_0 v)^2}{(4vw - f(1-f))^2}.$$

We want to prove that sometimes $p_{1,OPT} \geq p_{1,f^*}$. More specifically, we'll show that this is true for the symmetric case where $q_0 = p_0$ and $v = w$. In that case, it is $f^* = \frac{1}{2}$ and the inequality $p_{1,OPT} \geq p_{1,f^*}$ is equivalent to

$$\frac{1}{2(2w-1)} \geq \frac{4w}{(4w-1)^2},$$

which is true, since it is equivalent to $16w^2 - 8w + 1 \geq 16w^2 - 8w$. □

Proof of Proposition 3. The inequality $\frac{SW_{OPT}}{SW_f} \leq \frac{4}{3}$ is equivalent to

$$a' q_0 p_0 + b' q_0^2 + c' p_0^2 \leq 0,$$

where

$$\begin{aligned}
a' &= 4(1-f)^2 f^2 + 32(1-f) f v^2 w^2 - 4((1-f)^2 f^2 + 8(1-f)f - 4)vw - 64v^3w^3, \\
b' &= w\left(\left(3(1-f)^2 + 4\right) f^2 + 8(1-f)(2-f)vw - 16v^2w^2\right), \\
c' &= v\left(\left(3f^2 + 4\right)(1-f)^2 + 8f(1+f)vw - 16v^2w^2\right).
\end{aligned}$$

Therefore, it is enough to prove that $a' \leq 0, b' \leq 0$, and $c' \leq 0$.

The expression $(3f^2+4)(1-f)^2 + 8f(1+f)vw - 16v^2w^2$ is a second degree polynomial with respect to vw with largest root

$$\frac{f(1+f) + 2\sqrt{1 - (1-f)f\,(2+f^2)}}{4} \leq 1.$$

Since $vw \geq 1$, i.e. larger than the larger root, and the coefficient of v^2w^2 is negative, the value of the polynomial is non-positive. This means that $c' \leq 0$. Similarly, $b' \leq 0$ (we just replace f with $1-f$).

For a', we have that

$$a' + 48 = 4(1-vw)\left(4(3+4vw) + (4vw - (1-f)f)^2\right) \leq 0,$$

which means that $a' < 0$, and we are done.

The upper bound of the ratio is achieved for $(p_0, f, v, w) = (0, 0, 1, 1)$ or $(q_0, f, v, w) = (0, 1, 1, 1)$.

It remains to show that as $w \to +\infty$ or $v \to +\infty$, the ratio $\frac{SW_{OPT}}{SW_f}$ tends to 1. It holds that $\frac{SW_{OPT}}{SW_f} =$

$$\frac{(4vw - f(1-f))^2 \left(vp_0^2 + wq_0(4vp_0 + q_0)\right)}{(4vw-1)\left(w\left(f^2q_0(8vp_0+q_0) - 8fvp_0q_0 + 4vp_0(vp_0+q_0)\right) + (1-f)^2p_0\left(f^2q_0 + vp_0\right) + 4vw^2q_0(4vp_0+q_0)\right)}.$$

Both the numerator and the denominator are third degree polynomials in w. The coefficients of w^3 in both of these polynomials are equal to $16v^2q_0(4vp_0 + q_0)$. Therefore, as $w \to +\infty$, the ratio goes to $\frac{16v^2q_0(4vp_0+q_0)}{16v^2q_0(4vp_0+q_0)} = 1$. Similarly, as $v \to +\infty$, the ratio goes to 1. □

Proof of Proposition 4. Let a^*, c^* be the efforts in the optimal solution (second benchmark). We define the payments functions

$$g_1(a, q_0, c, p_0) = \frac{1}{2}q_0p_0 + sap_0 + tq_0c + fac$$

and

$$g_2(a, q_0, c, p_0) = \frac{1}{2}q_0p_0 + (1-s)ap_0 + (1-t)q_0c + (1-f)ac,$$

where

$$\begin{aligned} s &= 1 + \frac{q_0p_0 + 2v(c^*)^2}{2a^*p_0}, \\ t &= -\frac{q_0p_0 + 2w(a^*)^2}{2q_0c^*}, \\ f &= \frac{1}{2} + \frac{-a^*p_0 + q_0c^* + 3w(a^*)^2 - 3v(c^*)^2}{2a^*c^*}. \end{aligned}$$

Since a^*, c^* are the optimal efforts, we know that they satisfy the equality

$$(a^* + q_0)(c^* + p_0) = w(a^*)^2 + v(c^*)^2.$$

It is

$$\left.\frac{\partial(g_1(a, q_0, c^*, p_0) - wa^2)}{\partial a}\right|_{a=a^*} = sp_0 - 2wa^* + fc^* = \frac{(a^*+q_0)(c^*+p_0) - w(a^*)^2 - v(c^*)^2}{2a^*} = 0.$$

Similarly,

$$\left.\frac{\partial(g_2(a^*, q_0, c, p_0) - vc^2)}{\partial c}\right|_{c=c^*} = (1-t)q_0 - 2vc^* + (1-f)a^* = \frac{(a^*+q_0)(c^*+p_0) - w(a^*)^2 - v(c^*)^2}{2c^*} = 0.$$

This means that (a^*, c^*) is the equilibrium under the contracts g_1, g_2. To complete the proof, we need to verify that the payoffs of the two publishers in the equilibrium are non-negative. For the payoffs, we have that

$$g_1(a^*, q_0, c^*, p_0) - w(a^*)^2 = \frac{(a^*+q_0)(c^*+p_0) - w(a^*)^2 - v(c^*)^2}{2} = 0$$

and

$$g_2(a^*, q_0, c^*, p_0) - v(c^*)^2 = \frac{(a^* + q_0)(c^* + p_0) - w(a^*)^2 - v(c^*)^2}{2} = 0.$$

□

Proof of Lemma 1. The conversion rate under an (r)-contract is equal to

$$\left(\frac{(2r+1)(q_0 + 4p_0 v)}{16vw - 1} + q_0\right)\left(\frac{(2r+1)(p_0 + 4q_0 w)}{16vw - 1} + p_0\right),$$

which is a convex function of r. Furthermore, the conversion rate is maximized when the participation constraint for one of the publisher is binding. □

Proof of Proposition 5. Using the same notation as in the proof of Proposition 1, we have that

$$\frac{\partial f^*}{\partial w} = \frac{-\frac{\partial a}{\partial w}(f^*)^3 + \frac{\partial b}{\partial w}(f^*)^2 - \frac{\partial c}{\partial w} f^* + \frac{\partial d}{\partial w}}{3a(f^*)^2 - 2bf^* + c}$$

and we know that the denominator is positive, therefore we need to see when the numerator is positive. Let N be the numerator, then it holds that

$$wN = wN - (-a(f^*)^3 + b(f^*)^2 - cf^* + d) = 4(1 - f^*)v(4w^2 q_0^2 - (1 - f^*)^2 p_0^2).$$

Therefore, the numerator is positive iff $2wq_0 > (1 - f^*)p_0$ and negative if $2wq_0 < (1 - f^*)p_0$. This means that when f^* is bellow the line $1 - \frac{2wq_0}{p_0}$, f^* is decreasing in w, and when it's above that line, it's increasing in w. From this and the fact that f^* is continuous, we conclude that f^* can cross the line $1 - \frac{2wq_0}{p_0}$ in at most one point. If that point exists, then $\overline{w}$ is the root of the equation $f^* = 1 - \frac{2wq_0}{p_0}$ with respect to w. If that point doesn't exist, then either f^* is always increasing, which means $\overline{w} = 1$, or always decreasing, which means $\overline{w} = +\infty$.

The result for $\overline{v}$ now follows because of symmetry. □

Proof of Corollary 2. Let $q_0 = p_0$. Similarly to the proof of Corollary 1, if $w = v$, it is $x = y$ and therefore $f^* = \frac{1}{2}$.

For the other two parts, we need the fact that if $q_0 = p_0$, then f^* is increasing in w. This doesn't come immediately from Proposition 5, but we'll show that it is true. Using the notation of the proof of Proposition 5, we have that

$$\begin{aligned} wN &= 4(1 - f^*)v(4w^2 q_0^2 - (1 - f^*)^2 p_0^2) = \\ &= 4(1 - f^*)v(4w^2 - (1 - f^*)^2) q_0^2 \geq \\ &\geq 4(1 - f^*)v(4 - (1 - f^*)^2) q_0^2 = \\ &= 4(1 - f^*)v(1 + f^*)(3 - f^*) q_0^2 \geq 0. \end{aligned}$$

Therefore, $\frac{\partial f^*}{\partial w} \geq 0$ and the result follows. □

Proof of Proposition 6. For the expected conversion rate, we have that

$$\mathrm{E}[r(f)] = \frac{\mathrm{E}[p_0^2](2(1-f)^2 v + 8v^2 w) + \mathrm{E}[q_0^2](2f^2 w + 8vw^2) + \mathrm{E}[p_0 q_0]((1-f)^2 f^2 + 4vw + 4(1-f)^2 vw + 4f^2 vw + 16v^2 w^2)}{(4vw - f(1-f))^2}.$$

In other words, it depends only on $\mathrm{E}[q_0 p_0]$, $\mathrm{E}[q_0^2]$, and $\mathrm{E}[p_0^2]$. □

Proof of Corollary 3. Let $w = v$ and $\mathrm{E}[q_0] = \mathrm{E}[p_0]$. Using the notation of the proof of Proposition 1 with the difference that now $x = w\,\mathrm{E}[q_0^2], y = v\,\mathrm{E}[p_0^2]$, and $z = \mathrm{E}[q_0 p_0]$ (instead of wq_0^2, vp_0^2, and $q_0 p_0$), we know that f^* is the single real root of the polynomial $-af^3 + bf^2 - cf + d$, where $a = 4(x+y), b = 12y, c = 4(4vwy + 3y + 8vwz + 4vwx)$, and $d = 4(y + 4vwz + 4vwx)$.

If $\mathrm{Var}[q_0] = \mathrm{Var}[p_0]$, then that implies $\mathrm{E}[q_0^2] = \mathrm{E}[p_0^2]$, which means that $x = y$. For $f = \frac{1}{2}$, we get $-af^3 + bf^2 - cf + d = \frac{1}{2}(x-y)(16vw-1) = 0$. This means that $\frac{1}{2}$ is the root, i.e. $f^* = \frac{1}{2}$.

If $\mathrm{Var}[q_0] > \mathrm{Var}[p_0]$, then $x > y$. The coefficient of f^3 in the polynomial above is negative and since it has a single real root, the polynomial is positive for f smaller than the root and negative for f larger the root. For $f = \frac{1}{2}$, we get $-af^3+bf^2-cf+d = \frac{1}{2}(x-y)(16vw-1) > 0$. This means that for the root f^*, it holds that $f^* > \frac{1}{2}$.

Similarly, if $\mathrm{Var}[q_0] < \mathrm{Var}[p_0]$, then $x < y$. Therefore, for $f = \frac{1}{2}$, we get $-af^3 + bf^2 - cf + d = \frac{1}{2}(x-y)(16vw-1) < 0$, which means that $f^* < \frac{1}{2}$. ☐

Proof of Lemma 2. Here is the sketch of the proof in 11 steps:

- To prove that $r(f)$ is concave, it is enough to prove that $r'(f)$ is decreasing in $(0,1)$.
- To prove that $r'(f)$ is decreasing, it is enough to prove that $r''(f)$ is negative in $(0,1)$.
- We can write $r''(f)$ as $h(f)g(f)$.
- We can prove that $h(f)$ is positive in $(0,1)$. Therefore, it is enough to prove that $g(f)$ is negative in $(0,1)$.
- To prove that $g(f)$ is negative in $(0,1)$, it is enough to prove that $g(f)$ is convex in $(0,1)$ and that $g(0) < 0$ and $g(1) < 0$.
- We can prove that $g(0) < 0$ and $g(1) < 0$, so it remains to prove that $g(f)$ is convex in $(0,1)$.
- To prove that $g(f)$ is convex in $(0,1)$, it is enough to prove that $g''(f)$ is positive in $(0,1)$.
- To prove that $g''(f)$ is positive in $(0,1)$, we can first prove that $g''(f)$ is convex, and then that $g''(m) > 0$, where m is the point in $(0,1)$ where g'' attains its minimum.
- To prove that $g''(f)$ is convex, we can prove that $g^{(4)}(f)$ is positive in $(0,1)$.
- To find m, we will solve the equation $g^{(3)}(f) = 0$, which will give a single root.
- Finally, we can prove that $g''(m) > 0$.

☐

A.2 A More General Model and Equivalence

In this section, we justify some simplifications we made in our model in order to make it easier to analyze. We start with a general model that captures all the important things that we want to model, and then we see how we can transition from the general model to our model. In the end, we prove an equivalence result, where we show that for any set of parameter values of the general model and any attribution rule, there are parameter values in our model that give the same equilibrium behavior and the same conversion rate. In other words, we don't lose much with the simplifications and at the same time, we make the problem more tractable.

A.2.1 The General Model

Every consumer who enters the funnel is exposed to an awareness ad with probability r_1. If a consumer does not see an awareness ad, then the probability that he will move to the next stage in the funnel is b_1 (baseline). If a consumer is exposed to an awareness ad, but he was not going to consider the product without the ad, then he moves to the next stage with probability e_1. We assume that r_1 and b_1 are fixed, while e_1 is a decision variable for the first publisher (effort).

The reason we assume that r_1 is fixed is because this can be something pre-determined and it is also easily measured. What is not observable by a firm is the effort the publisher puts in showing ads. The effort will affect how effective the ads will be, e.g. it will show how good targeting the publisher does. Higher effort comes with a cost for the publisher given by the convex function $c_1 e_1^2$.

Summarizing the above, there are eight types of consumers as shown in the following table.

	Exposed to ad	Not exposed to ad
Not in the baseline	$(1-b_1)r_1e_1$ $(1-b_1)r_1(1-e_1)$	0 $(1-b_1)(1-r_1)$
In the baseline	b_1r_1 0	$b_1(1-r_1)$ 0

In each cell the upper part is the probability that the consumer will move to the consideration stage (second stage in the funnel), while the lower part is the probability that he will leave from the system (exit the funnel).

As we can see, the fraction of people who move from the first stage to the second without seeing an ad is

$$b_1(1-r_1),$$

while the fraction of people who move from the first stage to the second after seeing an ad is

$$b_1r_1 + (1-b_1)r_1e_1.$$

Similarly, every consumer who enters the consideration stage in the funnel is exposed to a consideration ad with probability r_2. If a consumer in the second stage does not see a consideration ad, then the probability that he will purchase the product is b_2 (baseline probability). If a consumer is exposed to a consideration ad, but he was not going to purchase the product without the ad, then he purchases with probability e_2. We assume that r_2 and b_2 are fixed, while e_2 is a decision variable for the second publisher (effort). The cost for the effort e_2 is given by the convex function $c_2 e_2^2$

The fraction of people who move from the second stage to the purchase without seeing an ad is

$$b_2(1 - r_2),$$

while the fraction of people moving from the second stage to the purchase after seeing an ad is

$$b_2 r_2 + (1 - b_2) r_2 e_2.$$

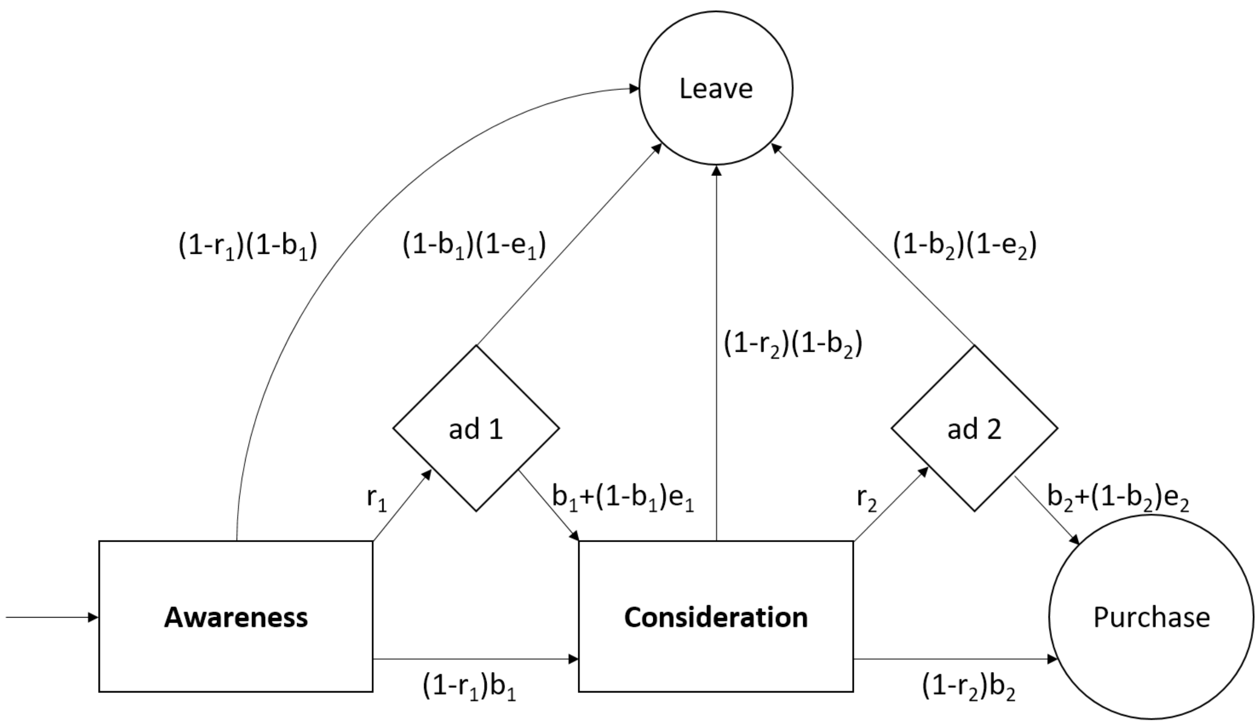

Figure 20: General model.

The firm can see all the consumers who purchased the product in its website and only them. It can also see, for each consumer who purchased, which ads they were exposed to prior to the purchase.

In other words, the firm can observe that a fraction $b_1(1-r_1)b_2(1-r_2)$ of consumers will purchase the product without seeing any ad, a fraction $(b_1 r_1 + (1 - b_1) r_1 e_1) b_2 (1 - r_2)$ will purchase after seeing an ad only in the first stage, a fraction $b_1(1-r_1)(b_2 r_2 + (1-b_2) r_2 e_2)$ will purchase after seeing an ad only in the second stage, and a fraction $(b_1 r_1 + (1-b_1) r_1 e_1)(b_2 r_2 + (1 - b_2) r_2 e_2)$ will purchase after seeing ads in both stages. So the firm can determine these four quantities even though it doesn't know the individual efforts e_1, e_2 or the baselines b_1, b_2.

A.2.2 Transition from the general model to our model

To simplify the notation and make the analysis a bit easier, we make the following changes to the general model:

- We define the constant $q_0 = b_1(1 - r_1)$, which we will call baseline probability (even though the actual baseline is b_1). This is the probability that a consumer in the first stage will move to the second without seeing any ad.
- We define the decision variable $a = b_1 r_1 + (1 - b_1) r_1 e_1$. This is the probability that a consumer in the first stage will see an ad and then move to the second stage. So now we'll assume that the first publisher will have to decide a instead of deciding e_1. The cost of e_1 was $c_1 e_1^2$, which makes the cost of a something of the form $\xi(a - \psi)^2$, for constants ξ, ψ. This means that the first publisher will choose an a such that $a \geq \psi$. We will simplify this cost further to wa^2 for some constant w. This is without loss of generality because we can always approximate the cost of the original model by adjusting w, and make sure that $a \geq \psi$ in equilibrium. Below we prove this claim formally.
- Similarly, we define the constant $p_0 = b_2(1 - r_2)$; the baseline probability for the transition from the consideration stage to purchase.
- Finally, we define the decision variable $c = b_2 r_2 + (1 - b_2) r_2 e_2$ for the second publisher with an associated cost vc^2.

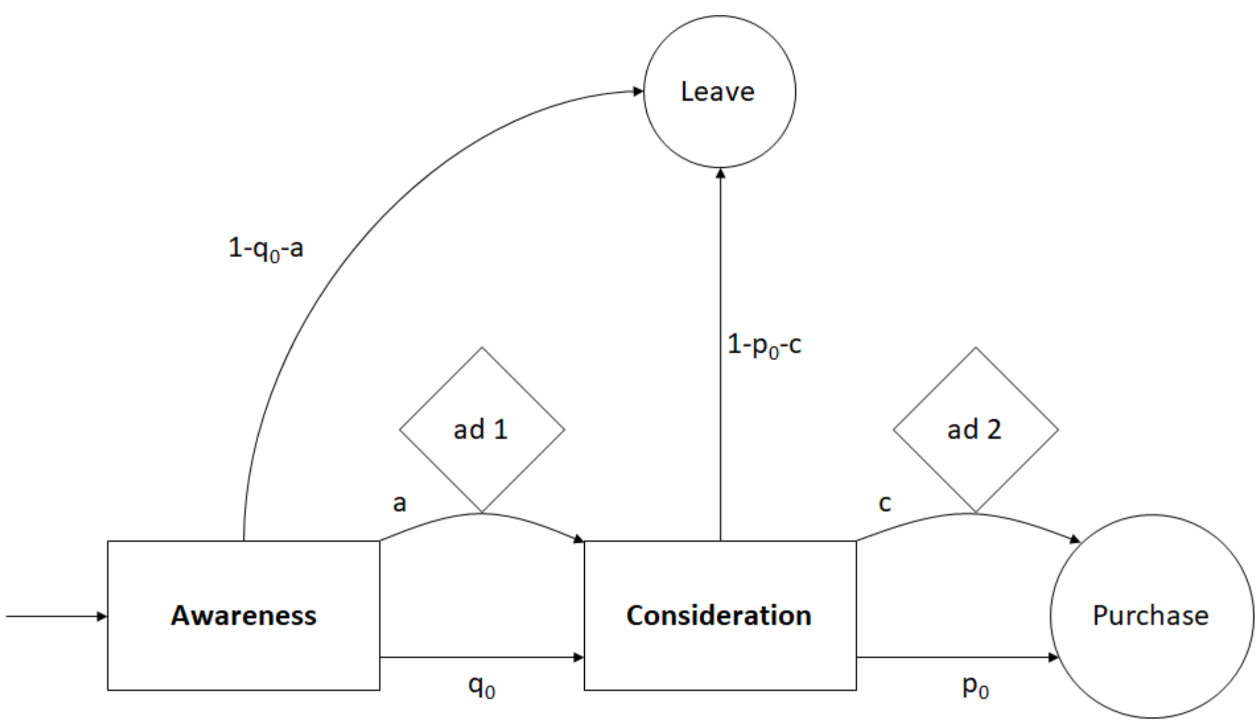

Figure 21: Simplified model.

In this version of the model, the firm can determine the quantities q_0p_0, ap_0, q_0c, and ac, even though it doesn't know the individual values of q_0, p_0, a, and c.

Claim 1. *The two models are equivalent.*

Proof. We will prove that for any equilibrium (e_1^*, e_2^*) in the general model, there are w, v such that there is an equilibrium (a^*, c^*) in the simplified model that satisfies $a^* = b_1r_1 + (1-b_1)r_1e_1^*$ and $c^* = b_2r_2 + (1-b_2)r_2e_2^*$.

We fix the values of $b_1, b_2, r_1, r_2, c_1, c_2$, and let $p_1(e_1, e_2)$ and $p_2(e_1, e_2)$ be the payment functions in the general model for the first and the second publisher respectively. Let (e_1^*, e_2^*) be an equilibrium of the general model under these payment functions.

We define $q_0 = b_1(1-r_1)$ and $p_0 = b_2(1-r_2)$. We also define the functions $g(e_1) = b_1r_1 + (1-b_1)r_1e_1$ and $h(e_2) = b_2r_2 + (1-b_2)r_2e_2$, and let $a' = g(e_1^*)$ and $c' = h(e_2^*)$. The payment functions in the simplified model will then be $p_1(g^{-1}(a), h^{-1}(c))$ and $p_2(g^{-1}(a), h^{-1}(c))$ for the first and the second publisher respectively. We set

$$w = \left.\frac{\partial p_1(g^{-1}(a), h^{-1}(c'))}{\partial a}\right|_{a=a'} \cdot \frac{1}{2a'}$$

and

$$v = \left.\frac{\partial p_2(g^{-1}(a'), h^{-1}(c))}{\partial c}\right|_{c=c'} \cdot \frac{1}{2c'}.$$

For these w and v, it holds that

$$\left.\frac{\partial (p_1(g^{-1}(a), h^{-1}(c')) - wa^2)}{\partial a}\right|_{a=a'} = 0$$

and

$$\left.\frac{\partial (p_2(g^{-1}(a'), h^{-1}(c)) - vc^2)}{\partial c}\right|_{c=c'} = 0.$$

Notice that $p_1(g^{-1}(a), h^{-1}(c')) - wa^2$ is the utility of the first publisher and $p_2(g^{-1}(a'), h^{-1}(c)) - vc^2$ is the utility of the second publisher in the simplified model. In other words, (a', c') is an equilibrium in the simplified model. $\square$

Example 1. Let's consider a $\frac{1}{2}$-contract and let $b_1 = \frac{1}{3}, b_2 = \frac{1}{4}, r_1 = \frac{1}{5}, r_2 = \frac{1}{6}, c_1 = 2, c_2 = 3$. The equilibrium in the general model is $(e_1^*, e_2^*) = (0.00765193, 0.00626063)$. For $w = 1.69573$ and $v = 3.53964$, we get the equilibrium $(a^*, c^*) = (0.0676869, 0.0424492)$ in the simplified model, and it holds that $a^* = b_1r_1 + (1-b_1)r_1e_1^*$ and $c^* = b_2r_2 + (1-b_2)r_2e_2^*$.

A.3 Definition of Shapley Value

Let N be a set of players and let $v : 2^N \to \mathbb{R}$ be a function such that for every subset $S \subseteq N$ of players, $v(S)$ gives the total payoff the members of S will get by working together. For a pair (v, N), an attribution rule is a function $\phi_i(v)$ that gives the payoff of player i.

Axiom 1. *Symmetry*: If two players are equivalent, then they should have the same payoff. Two players i, j are equivalent if their contribution to every subset of other players is the same, or mathematically if $v(S \cup \{i\}) = v(S \cup \{j\})$ for every $S \subseteq N \setminus \{i, j\}$.

Axiom 2. *Null player*: The payoff of a null player should be 0. A player i is called null if he doesn't contribute anything to any subset of other players, or mathematically if $v(S \cup \{i\}) = v(S)$ for every $S \subseteq N \setminus \{i\}$.

Axiom 3. *Additivity*: The sum of payoffs that a player gets for two different games should be equal to the payoff he gets if we consider the two games as one big game. Or mathematically, $\phi_i(v + w) = \phi_i(v) + \phi_i(w)$ for every player $i \in N$ and any two functions $v, w : 2^N \to \mathbb{R}$.

Axiom 4. *Efficiency*: The total payoff is distributed among all the players. Or mathematically, $v(N) = \sum_{i \in N} \phi_i(v)$.

Shapley (1953) proved that there is a unique rule that satisfies these four axioms. We call this rule the *Shapley Value*, and it is given by the following formula

$$\phi_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(n - |S| - 1)!}{n!}(v(S \cup \{i\}) - v(S)).$$