Online Appendix - Beyond the Last Touch: Attribution in Online Advertising

OA.1 Multiple Global Bidders

The analysis of two competing global bidders is complex because of the need to find an asymmetric Bayesian equilibrium with two dimensional strategies for each player. Since we are unable to find a closed-form solution for this problem, we use numerical analysis to find the equilibria bidding strategies and compute firm profits.

Section OA.1.1 details the algorithm we use to numerically analyze the Bayesian equilibria when one global attributing advertiser is competing with a second global advertiser. It should be noted that finding Bayesian Nash equilibria is computationally hard (NP-Complete), which limits the scope of parameter values we are able to numerically test.

Our analysis focuses on cost per click (CPC) auctions with click through rate (CTR) adjusted bids. In these auctions, advertisers pay for an impression only if the consumer clicked on an ad. Their payment is an adjusted bid of the second highest bidder, and this adjustment is done using the expected CTR that the publisher observes. Using this model, a conversion of the consumer will be a click on an ad.

Specifically, when advertiser 1 bids b_{11} and advertiser 2 bids b_{21} at the publisher which is visited first, advertiser 1 will win the auction if $\gamma b_{11} > \gamma b_{21}$ and will pay $\frac{\gamma b_{11}}{\gamma}$ if the consumer clicks on the ad (which happens with probability γ). When bidding at the publisher which is visited second, however, the advertiser will win the auction if $\gamma_1^E b_{12} > \gamma_2^E b_{22}$, and pay $\frac{\gamma_2^E b_{22}}{\gamma_1^E}$ if she wins the auction and the consumer clicked the ad.

γ_1^E is the expected CTR if the consumer is exposed to an ad at the second publisher, but that CTR depends on exposure at the first publisher. Thus, if the first publisher is visited with probability p_1, the CTR at the second publisher is:

$$\gamma_1^E = p_1 \mathbf{I}[\gamma b_{11} > \gamma b_{21}]\gamma(1+d) + p_1 \mathbf{I}[\gamma b_{11} \le \gamma b_{21}]\gamma + (1-p_1)\gamma = \gamma\left(1 + p_1 \mathbf{I}[\gamma b_{11} \le \gamma b_{21}]d\right) \quad (1)$$

Similarly:

$$\gamma_2^E = \gamma\left(1 + p_1(1 - \mathbf{I}[\gamma b_{11} \le \gamma b_{21}])d\right) \quad (2)$$

The externality of the first impression on the second impression now comes into play in two aspects compared to the CPM case. The first impact is on the revenue, exactly as was the case for the CPM case. The second impact, however, is on the cost. Because d is negative, when an advertiser wins the first impression, the second publisher expects a lower CTR from that advertiser, and thus adjusts their bid downward in the auction. That means that advertisers have an incentive to lower their bidding strategies at the first publisher in order to avoid this penalty when competing at the second publisher.

When comparing the equilibrium profits of the attributing advertisers, we ask what the relevant benchmark is for comparison in terms of the second global advertiser. Depending on the choice of attribution (or lack of) of the second global advertiser, the result may shift dramatically. We therefore perform a comparison with two types of global advertisers – a weak advertiser that does not know the state of nature in the campaign and behaves like an NA advertiser, and a strong advertiser that has full information about the state of the world and behaves like an *Opt* advertiser. In reality, global advertisers will have a mix of knowledge distributions, and as they apply more sophisticated attributions algorithms, they will get closer to the *Opt* case.

Figure 1 shows the expected relative profit of each attributing advertiser to *Opt* advertiser when playing against an NA advertiser (left) and when playing against an *Opt* advertiser (right). A striking result emerges when comparing the left to the right panel. The relative profits in the right panel are higher than one, meaning that all attributing advertisers (including the NA advertiser) make higher profits when they have less information or a less efficient attribution scheme than if they had full state information.

This result exemplifies the prisoner's dilemma-like effect of increasingly improving advertising measurement in a competitive market – as more advertisers become more efficient, the competition becomes stronger which results in lowered profits (and higher revenue going to the publishers).

Essentially, less efficient attribution methods allow advertisers to soften the competition in a multi-publisher advertising market.

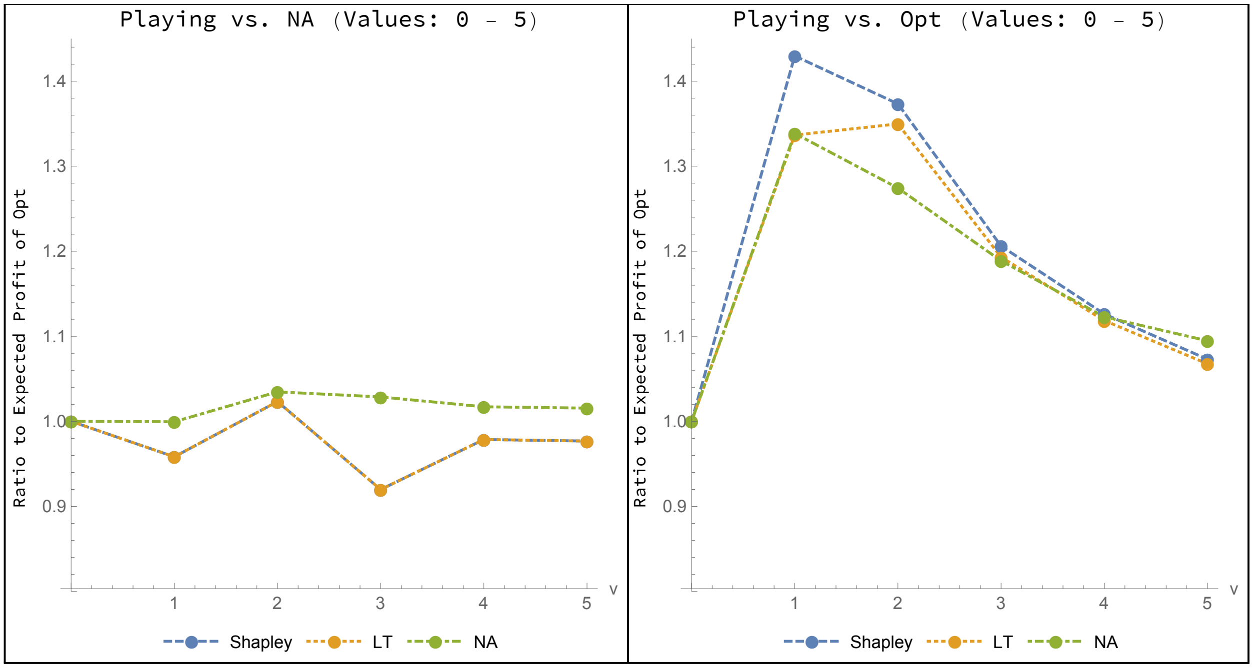

Figure 1: Expected relative profit of the Sh, LT and NA advertisers to the Opt advertiser. $\gamma = 1$, $p_H = 1$, $p_L = 1/2$ and $d = -1/2$.

When focusing on the left panel, we see that advertising profits are no less than 90% of the profits achieved for the Opt player, implying that the attribution methods are quite efficient. We do see however that in this case, the NA player, which is a weaker bidder in equilibrium, makes higher profit.[1] Stronger bidding yields lower profit when using attribution, though the differences are not as significant as when the competing advertiser has full knowledge.

Finally, we compare between the two attribution methods LT and Sh to the NA method. When competing with an NA advertiser, both methods yield similar profit,[2] which is lower than Opt for low and high conversion values. When competing with an Opt advertiser, we see a result that echoes the previous results of our analysis in the CPM case – The profit under the Shapley value is higher for lower values of v, and the profits are higher under NA for higher values of v. We also do not find a case where the LT profits are the highest, but there are cases when the LT profit is higher than the Sh profit, namely for high values of v.

[1] Further details of the equilibrium bids and analysis are provided below.

[2] The profits are not identical, but are very close. We attribute the difference to limited machine precision.

OA.1.1 Algorithm

Finding pure strategy Bayesian-Nash equilibria (BNE) is an NP-Complete problem as shown by Conitzer and Sandholm (2003, 2008), and approximating these equilibria efficiently is still an active research field in computer science and economics. A classic algorithm for finding BNEs is the fictitious play algorithm (FP) due to Brown (1951), also known as an iterated best response algorithm. In fictitious play, each player is initially assigned a pure strategy, and the algorithm proceeds in rounds. In each round players take turns in playing against the empirical mixed strategy of the previous actions of the opposing player.

Specifically, suppose there are 2 players, i and j. The players have types θ_i and θ_j drawn from the same prior probability with density $f(\theta)$, and can take actions b_i and b_j from a set of actions $\mathcal{B}$ with size B. The payoffs of player i is denoted $u_i(b_i, b_j|\theta_i, \theta_j)$ and similarly for player j u_j is defined.

Suppose at round t it is player i's turn to play. Denote player j's sequence of previous actions as $b_j(1|\theta_j), \ldots, b_j(t-1|\theta_j)$. The empirical mixed strategy of player j and type θ_j is defined by the distribution $s_j(t-1|\theta_j)$ over actions $b_k \in \mathcal{B}$ with $s_j(t-1|\theta_j) = (p_1(t-1|\theta_j), \ldots, p_B(t-1|\theta_j))$ and $p_k(t-1|\theta_j) = \frac{\sum_{r=1}^{t-1} \mathbb{I}(b_j(r|\theta_j)=b_k)}{t-1}$. Using this definition, player j's empirical mixed strategy is a distribution for each type θ_j.

Given $s_j(t-1|\theta_j)$, player i plays her pure strategy best response:

$$b_i(t|\theta_i) = \arg\max_{b \in \mathcal{B}} \mathbf{E}_{\theta_j}[\sum_{k=1}^{B} u_i(b, b_k|\theta_i, \theta_j) p_k(t-1|\theta_j)]$$

Given player i's play, player j then recomputes the empirical mixed strategy, and best responds to it, and so forth. The algorithm stops when the best responses converge such that $b_i(t|\theta_j) = b_i(t-1|\theta_j)$ and $b_j(t|\theta_j) = b_j(t-1|\theta_j)$. It is easy to show that in this case the resulting convergent actions are a BNE of the game.

Fictitious play is not guaranteed to converge for generic games, and even if it does converge, the convergence time is unknown a-priori in the general case. Moreover, since estimation is run on computers with finite precision (and finite type or strategy spaces) for large games it is probable that the algorithm will not converge even if it would have converged with infinite precision and continuous type and action spaces.

For this reason, the majority of work involving finding BNEs numerically focuses on finding ϵ-BNEs. A pair of mixed (or pure) strategies s_i and s_j is an ϵ-BNE if:

$$\mathbb{E}_{\theta_j}[u_i(s_i, s_j|\theta_i, \theta_j)] \geq \mathbb{E}_{\theta_j}[u_i(s, s_j(\theta_j)|\theta_i, \theta_j)] - \epsilon$$

for every strategy s available to i, and similarly defined for j. Thus if the players can mix over $\mathcal{B}$, the equilibrium will be defined over mixed strategies, and otherwise it will be defined over pure strategies. When computing ϵ-BNEs, the FP algorithm stops when the best response of each player (and type) cannot improve their current payoff by more than ϵ. Our definition uses an ex-interim definition of a BNE (where each player knows their type but not the opponent's type). Alternatives include an ex-ante definition where each player maximizes over their expected type as well. We followed the literature with the most common definition used. Our analysis uses a value of 0.00001 for ϵ.

Our analysis makes use of the FP algorithm to find an ϵ-BNE with a finite action space. In this setting, Rabinovich et al. (2013) show that if the type space is continuous, mixed strategies can be converted to equivalent pure strategies and this fact can be used efficiently find ϵ-BNEs. Despite the great progress made in that paper, the algorithm is still only able to compute the equilibria for games with no more than 5 to 10 bid levels with one dimensional strategies. Since advertisers in our game bid at two publishers, they will have two dimensional strategies, growing the action space exponentially, which will disallow efficient computation of equilibria.

To resolve this issue we focus on computing equilibria in the attribution game with a discrete type and action space. Because we are interested in comparing how different attribution methods perform, we designate one player to be a Benchmark player in all four cases we compare (NA, LT, Sh and Opt), and the second player uses one of the attribution methods analyzed in this game. The benchmark player is fully strategic and best responds to the actions of the attributing player. We analyze two types of benchmark players, to get a sense of the "best" and "worst" case scenario facing an attributing player. In the first scenario, the benchmark player does not have information about the state of the world and bids using the same strategies in every state. Essentially, this is an NA player. In the second scenario, we assume the benchmark player has full information about the state of the world, and bids optimally for each state. Essentially in this scenario the benchmark player is an Opt player.

In our notation, we will denote $\pi^*_{LT,NA}$ as the equilibrium profit of a last-touch attributing player playing against an NA global player, and $\pi^*_{LT,Opt}$ will denote the profit of a last-touch attributing player playing against an optimal player.

The second choice we made are parameter values for γ, d, p_H and p_L. Estimating the ϵ-BNEs using finite actions and types creates a tradeoff between accuracy, running time and space considerations. The more refined the action and type spaces, the closer the resulting equilibrium will be to the continuous space equilibria, but this comes at a cost of exponential growth in required memory (and possibly runtime) of the algorithm. We would like to be able to set bid values such that the advertiser can choose to bid in increments which represent the increments in conversion values it faces in the campaign. The smallest increment in value for an advertiser will come from the effect of the externality and will be $\gamma \cdot p_H \cdot p_L \cdot d \cdot v$. Thus, we let bids increase in increments of $\gamma \cdot p_H \cdot p_L \cdot d \cdot v$, or have a resolution of $\frac{1}{\gamma \cdot p_H \cdot p_L \cdot d \cdot v}$ per unit value. This means that the smaller each factor in the multiplication is, the higher resolution for bid values we would require. We picked the values $\gamma = 1$, $d = -1/2$, $p_H = 1$ and $p_L = 1/2$, that generate a resolution of 4 bid increments per unit value. When setting the total number of values for a conversion to n_v and the bid resolution to r, the size of the action space available to an advertiser is $(n_v \cdot r - 1)^2$. For 6 values and a resolution of 4 bids that means an action space of 529 possible actions. The fully mixed strategy of a player in this case will have 6×529, or $3,174$ elements.

OA.1.2 Equilibrium Bid Analysis

To give more intuition for the results we analyze the equilibrium bids of the attributing global player when they play vs. an NA and an *Opt* advertiser.

Figure 2 shows the equilibrium bidding strategies of the global attributing advertiser when they compete with an NA global advertiser, and Figure 3 shows the strategies in equilibrium when competing with an *Opt* advertiser. The bids for states $2H$ and $2L$ are symmetric.

As we can see, when competing with an NA advertiser, attribution causes stronger bidding at the first publisher and weaker at the second compared to the bidding of the *Opt* advertiser. The NA advertiser bids less than their value at both publishers, but the bids are quite close to the click value. We also see that both attribution schemes we analyze yield very close equilibrium bidding strategies, which is why we do not see too much difference between their profits in Figure 1. When looking at the competition with the *Opt* player, however, we see that the Shapley value encourages stronger bidding when the first publisher visited has visit propensity p_L, and that both attribution schemes bid weaker than *Opt* and NA for how values at the second publisher. This explains why we see higher profits for the Shapley value at lower values, and higher profits for the NA scheme in higher values, as also found in the analysis when competing with two local advertiser.

OA.2 Comparison to Models that Use Aggregate Data

When individual ad exposure data was not available, the advertisers could not tell how their ad expenditures translated to ad exposures. This is similar to standard offline advertising campaigns (Newspaper and TV), where there wasn't usually accurate data about which individual consumers were exposed to what ads that covers the entire market.[3]

In such cases, for example, media mix models were used to estimate the effectiveness and impact of ad spending, but these effects were only estimated in aggregate and on average over all consumers.[4] The advantage of multi-touch attribution is that the additional data about ad exposure and ad timing may be used as additional information in the estimation and optimization processes. This should allow, for example, to properly compensate strategic publishers that may have freedom in how ad impressions are allocated and maximize their own profit.

In the classic media mix setup, it is assumed that the advertising channels are not strategic and the only change in the market are the budgets allocated to each channel. These classic marketing response and optimization problems usually assume ad prices are exogenous and that competitors do not respond to changes in budgets by a focal advertiser. A second difference between our setup and the standard media mix setup is that we take into account the resulting equilibrium of ad allocation and ad prices given the results of the measurement process by the advertiser.

To estimate the benefit of multi-touch attribution we would like to perform a similar analysis to the one done in Section 5.3 and find the highest profit possible when employing an aggregate data model in the measurement and optimization process.

Finding the best model is a complex problem (similarly to finding the best attribution solution),

[3]For example, most TV household advertising measurement is done on a sample, and not on the entire target population.

[4]There are models that try to connect aggregate spending data with a sample of individual exposure data, and these are close to an attribution model.

but we are able to give an upper bound on how well the best model will perform using theoretical results from team sharing problems. Nandeibam (2002) analyzes sharing rules in teams when the production function of the team is common knowledge and has no uncertainty, and when the effort of the team members is unobserved by the principal. In the case of advertising campaigns, the team members are the publishers, and the fact that their effort is unobserved is equivalent to not being able to observe the resulting ad allocation from a budget given to a publisher. Because a bid at a publisher has a one-to-one mapping to a budget spend, we can use the bids in our model to reflect the actions a publisher takes in the team.

The results of the analysis in Nandeibam (2002) show that when there is no market uncertainty, the optimal sharing rule, which is equivalent to the attribution function in our model, is a function which is linear in the output of the team (or the campaign). In other words, the famous result in Holmstrom (1982) shows that a team sharing rule cannot induce the optimal allocation, while Nandeibam (2002) finds the best result that can be achieved and the sharing rules that achieve them.

We can use this result as an upper bound on the best aggregate model possible because it assumes that the output function of the team is common knowledge. In an advertising campaign the output function is unknown at the campaign design stage, which means this best case scenario cannot be implemented under general conditions.

Using the results of this analysis and a numerical analysis[5] we are able to show that the ratio $\mathbb{E}[\frac{\pi^*_{MM}}{\pi^*_{Opt}}]$ has a minimum of 0.75. Comparing this value to the previous values 0.875 for last-touch and Shapley value attribution allows us to put an estimate on the value of attribution methods in increasing the profits of advertisers. This increase in value can be attributed to the additional data available to the advertisers, and particularly, to the data on timing and ad exposure post campaign which are available online.

OA.2.1 Details of Analysis

The following proposition, due to Nandeibam (2002) allows us to find the maximal profit of the global advertiser without finding the optimal aggregate data model explicitly.[6] We mention the most relevant details and assumptions for our analysis.

[5]See section OA.2.1 of the Appendix for details.

[6]Nandeibam (2002) also details how this mechanism may be found, but in our case it doesn't have a closed form.

Proposition OA.2.1 (Nandeibam (2002) Proposition 3)**.** *Let $U^i(m, a_i)$ be a team's member utility function, with m their payment in monetary terms and a_i their action. Define U^i_j, $j \in \{1, 2\}$ as the derivative of U^i with respect to variable j. Let $f(a_1, \ldots, a_n)$ be the production function of the team members.*

Assume:

- *(A1) f is strictly increasing and concave.*
- *(A2) U^i is concave, strictly increasing in income and strictly decreasing in action.*

Let (a, p) be an outcome (a the vector of actions, p the vector of payments) with $a_i > 0$ for all i.

Then (a, p) is implementable non-cooperatively if and only if (a, p) satisfies:

$$\sum_{i=1}^{n} \left[\frac{U^i_2(p_i, a_i)}{U^i_1(p_i, a_i) f_i(a)} \right] = 1 \tag{3}$$

Nandeibam (2002) further defines a non-cooperative implementation if there is a sharing rule $s(f(a))$ such that $p_i = s_i(f(a))$ and (a, p) is a Nash-equilibrium of the team.

In our campaign model, the production function is the revenue function $r(b_1, b_2)$, the utility functions of the team players are $u_i(m, b_1) = m - c_i(b_i)$ and the payment sharing rule s are the attribution functions $a_1(r(b_1, b_2))$ and $a_2(r(b_1, b_2))$ (without using additional data such as timing or ad exposure).

Assumptions (A1) and (A2) hold for these functions, and as a result, we can use the condition in Equation 3 to solve the following constrained maximization problem:

$$\pi^*_{AD}(state) = \max_{b_1, b_2} \pi(b_1, b_2 | state) \text{ s.t. (3) holds.} \tag{4}$$

Equation 3 translates in our model to the following condition:

$$\frac{\frac{b_2}{b_1 dp_H + \gamma} + \frac{b_1}{b_2 dp_L + \gamma}}{v} = 1$$

This problem assumes that the profit function $\pi(b_1, b_2)$ is known prior to the campaign (so a contract can be signed based on the outcome), which is of course not the case in our model. We therefore conclude the result from this analysis will serve as an upper bound on the possible profit achieved from an aggregate data model.

The result of the optimization $\pi^*_{AD}(state)$ does not have a tractable closed form, which is why we use numerical optimization to estimate the ratio $\mathbb{E}[\frac{\pi^*_{AD}}{\pi^*_{Opt}}]$, which is then also minimized numerically.

Essentially, we are finding the MinMax of the problem numerically. The global minimization over 5 parameters does not converge in a realistic time-frame, but a local minimization finds a ratio of 0.75 when $v = 1$ at the parameter values $\gamma \approx 0.5$, $p_H \approx 0.666$, $p_L \approx 0.664$ and $d \approx 0$. Thus we can conclude that this can serve as an upper bound on the ratio when using an aggregate data model.

OA.3 Fixed Externality Share (FES) Attribution

In this Section we will analyze a generic attribution mechanism we call Fixed Externality Share (FES) Attribution that generalizes the last-touch and Shapley values method into a family of attribution methods.

The definition of the Shapley value attributes the externality from winning both ad auctions equally to both publishers, while last-touch attribution allocates the externality to the last publisher visited by the consumers. In fixed externality share attribution, a share α of the externality is allocated to the first publisher visited in a sequence of visits, and a share $1 - \alpha$ is allocated to the second publisher visited, when $0 < \alpha < 1$.

When publisher 1 is visited first, for example, the FES attribution functions are:

$$A_1^{FES} = p_1\gamma v_G F_1(b_{G1}) + \alpha p_1 p_2 d\gamma v_G F_1(b_{G1})F_2(b_{G2}) \tag{5}$$

$$A_2^{FES} = p_2\gamma v_G F_2(b_{G1}) + (1-\alpha) p_1 p_2 d\gamma v_G F_1(b_{G1})F_2(b_{G2}) \tag{6}$$

It is easy to show that the Shapley value is part of this family with $\alpha = 1/2$. Using this definition, we can find the equilibrium bids and profits, and then maximize the resulting expected profit with respect to α. This will give us the optimal FES scheme.

Finding the equilibrium in closed form is not possible for all parameter ranges. We therefore limit the parameter space to $p_L = p_H(1 - p_H)$ and $\frac{d+1}{d+2} \leq \alpha \leq \frac{1}{d+2}$ when $d < 0$. In this restricted space the profit functions are smooth without kinks and we can fully analyze the resulting bids and profits.

Using these assumptions, we prove the following:

Proposition OA.3.1. *When using FES attribution:*

- *Last-touch attribution is equivalent to FES attribution with* $\alpha = \frac{1}{d+2}$.
- *FES attribution will never reach the optimal profit.*
- *When* $-1 < d < 0$, $p_L = p_H(1 - p_H)$ *and* $\frac{d+1}{d+2} \leq \alpha \leq \frac{1}{d+2}$*: Last-touch is never optimal, and the Shapley value is the unique optimal FES attribution scheme.*

Proposition OA.3.1 shows that Last-touch attribution is a part of the FES family (this is true for the entire parameter space), and that any attribution scheme in this family will not be able to achieve the optimal profit. The last item shows that in the limited parameter space we analyze, which generally restricts the campaigns to low values of v, last-touch is never optimal while the Shapley value is the uniquely optimal attribution scheme.

This result gives more credibility to the focus on the Shapley value as an attribution scheme, and opens an interesting avenue for future research into the optimality of FES schemes under general conditions.

OA.3.1 Proof of Proposition OA.3.1

To prove the first item we compare the equilibrium bids under Last-touch attribution to the equilibrium bids under FES attribution.

We denote by b_{1,p_1} the bid at the publisher visited first with propensity p_1. We note this is not publisher number 1 necessarily:

$$b_{1,p_1} = \frac{\gamma v(p_2 v(\alpha(d+2) - 1) + 1)}{p_H p_L v^2 \left((1-2\alpha)^2 + (\alpha - 1)\alpha d^2 + (1-2\alpha)^2 d\right) + 1} \quad (7)$$

$$b_{2,p_2} = \frac{\gamma v(p_1 v(-2\alpha - \alpha d + d + 1) + 1)}{p_H p_L v^2 \left((1-2\alpha)^2 + (\alpha - 1)\alpha d^2 + (1-2\alpha)^2 d\right) + 1} \quad (8)$$

We then equate these bids to the last-touch bids and solve for α yielding $\alpha_{LT} = \frac{1}{d+2}$.

To prove the second item we equate the FES bids to the optimal equilibrium bids and find that there is no α that solves the 8 equations.

To prove the third item, we calculate the first derivative of the expected profit when we set $p_L = pH(1 - p_H)$ and limit α such that $\frac{d+1}{d+2} \leq \alpha \leq \frac{1}{d+2}$:

$$\frac{\partial \pi^*_{FES}}{\partial \alpha} = \tag{9}$$

$$-\frac{(2\alpha - 1)\gamma(d+2)^2(p_H - 1)p_H^3 v^4 \left(p_H \left((p_H - 1)v \left(((\alpha - 1)\alpha + 1)d^2(p_H - 2)p_H v + d\left((1 - 2\alpha)^2(p_H - 2)p_H v - 4\right) + (1 - 2\alpha)^2(p_H - 2)p_H v\right) - 1\right) + 2\right)}{2\left((p_H - 1)p_H^2 v^2 \left(\alpha^2(d+2)^2 - \alpha(d+2)^2 + d + 1\right) - 1\right)^3} \tag{10}$$

Plugging in $\alpha = 1/2$ yields zero, while plugging in $\alpha = \frac{1}{2+d}$ is never zero.

Calculating the second derivative and plugging in $\alpha = 1/2$ is always negative (in this range of α), meaning that the Shapley value is a local maximum. $\alpha = 1/2$ is also the only value in this range that makes the first derivative vanish, and hence it's a global maximum, concluding the proof.

OA.4 Delayed Conversions

Our model thus far has assumed that conversions (or clicks) happen immediately after an exposure. In this section we drop this assumption. Delayed conversions happen when an ad influences a person in their decision to convert, but the conversion happens later after the campaign ends. In the setup of our campaign, a delayed conversion would mean that the revenue observed after a consumer is exposed to an ad from the first publisher will always be zero.

In analyzing this scenario, we note that the timing of the conversion does not impact the optimal allocation of ads with full information nor the allocation without attribution. The reason is that in both cases the optimization considers only a single revenue function which is not split between the publishers, and hence the timing of the conversion does not impact the result.

When the Shapley value is used for attribution, delayed conversions will also not impact the attributed revenue per publisher. The reason for that is that the Shapley value takes the average marginal contribution of the players in a cooperative game over all possible orders of player arrivals. Consequentially, in our model the externality in every state is split equally between the two publishers, and the timing of conversion does not impact the calculation.

The only analysis we need to consider is the impact on last-touch attribution profit. Examining equations 8 and 9, with delayed conversions, $A_1^{LT} = 0$, while A_2^{LT} gets allocated all of the revenue.

Performing the equilibrium analysis shows that in this case the global advertiser bids zero at the first publisher and γv_G (her expected value) in the second publisher, essentially acting as a local bidder that changes their bid locations to always bid at the second publisher. The expected profit of the global advertiser will be equal to the expected profit of a local advertiser, and the relevant results from the previous sections will hold.

From the point of view of the publishers, however, since the advertiser always bids at the second publisher, an advertiser which is later in the conversion funnel or one that can control the timing of their ad displays will highly benefit from delayed conversions if advertisers use last-touch attribution.

References

Brown, George W. 1951. Iterative solution of games by fictitious play. *Activity analysis of production and allocation* **13**(1) 374–376.

Conitzer, Vincent, Tüomas Sandholm. 2003. Complexity results about nash equilibria. *Proceedings of the 18th international joint conference on Artificial intelligence*. Morgan Kaufmann Publishers Inc., 765–771.

Conitzer, Vincent, Tuomas Sandholm. 2008. New complexity results about nash equilibria. *Games and Economic Behavior* **63**(2) 621–641.

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

Nandeibam, Shasikanta. 2002. Sharing rules in teams. *Journal of Economic Theory* **107**(2) 407–420.

Rabinovich, Zinovi, Victor Naroditskiy, Enrico H Gerding, Nicholas R Jennings. 2013. Computing pure bayesian-nash equilibria in games with finite actions and continuous types. *Artificial Intelligence* **195** 106–139.

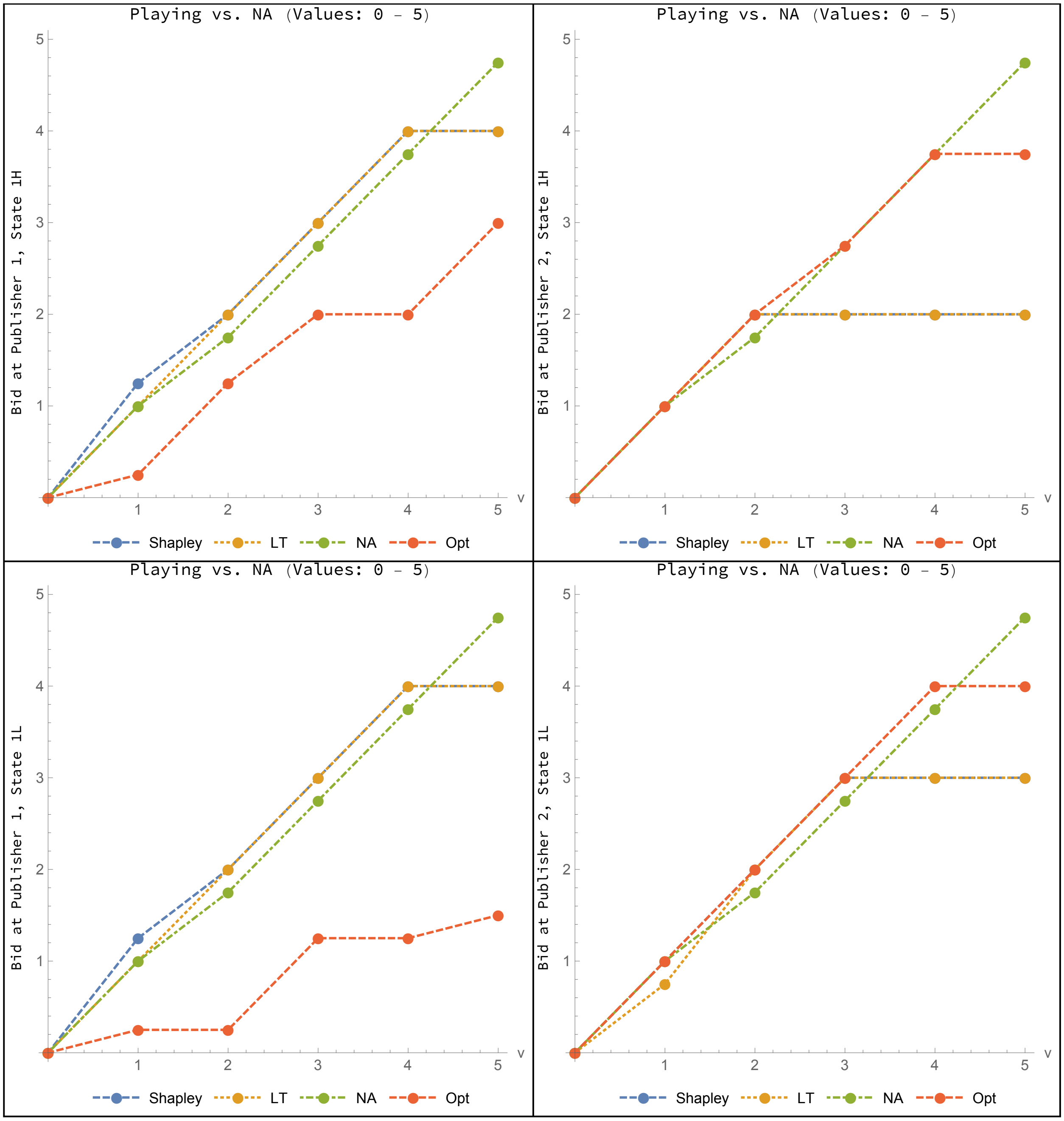

Figure 2: Bids of the attributing global player when competing with an NA advertiser. $\gamma = 1$, $p_H = 1$, $p_L = 1/2$ and $d = -1/2$.

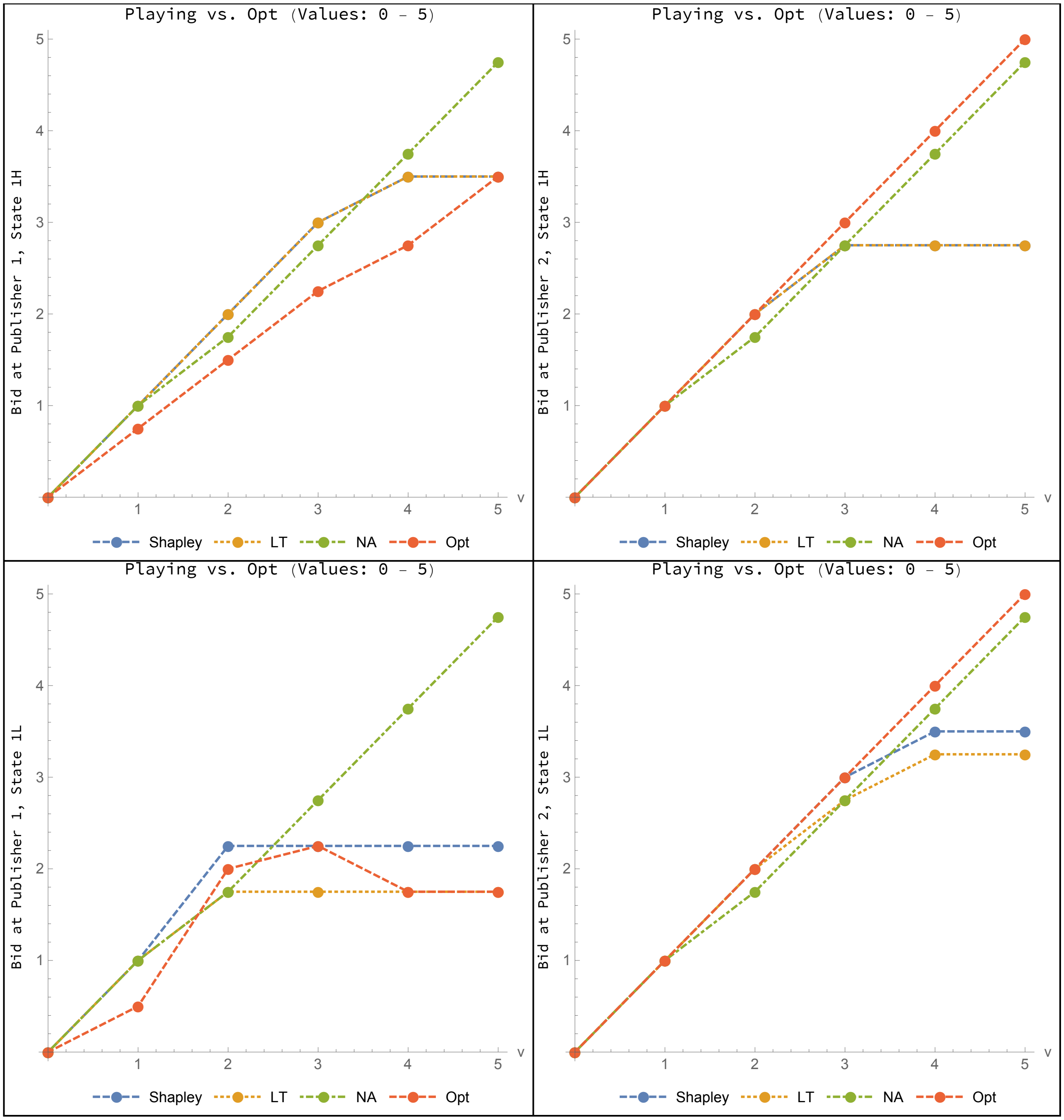

Figure 3: Bids of the attributing global player when competing with an *Opt* advertiser. $\gamma = 1$, $p_H = 1$, $p_L = 1/2$ and $d = -1/2$.