Multi-Touch Attribution Calibration through Google and Facebook Conversion Lift Studies

Feasibility and lift calibration impact using Google Ads Data Hub

By Alexander Oude Elferink

This research was funded by Facebook

Introduction

A big question that many brands have is: what marketing channel gets credit for what part of a conversion? What was the impact of my display banner and that of my Facebook ad? To address these questions, there are many conversion attribution models out there. All these models analyse which touchpoints, or marketing channels, receive part of the credit for a conversion and all of them distribute the value of conversions across touchpoints differently. The majority of attribution models is simple and implies singular or rule based attribution. Some give all credit to the last touchpoint (last touch model) or some sort of position based system is used where 40% might be accredited to the first touchpoint, another 40% to the last touchpoint and the remaining 20% to the middle journey touchpoints also known as positional.

At Annalect we apply a Markov chain model as our attribution methodology. This model is fed data extracted from Google Ads Data Hub (ADH), which is Google's cleanroom solution to provide privacy-safe access to log level ad serving data. For this research project we used Google and Facebook conversion lift studies to adjust the attribution results from our Markov chain model. The goal was to calibrate the results for Google and Facebook advertising by using lift studies as a ground truth. We assessed the feasibility of these conversion lift calibrations and the impact of using conversion lift results in the calibration adjustment for Facebook. Our hypothesis is that Facebook conversions are underestimated in ADH, because we can only measure clicks and not impressions in our Markov model due to limitations on tracking Facebook ads in Google's Campaign Manager for attribution purposes.

Background

Rule-based attribution solutions are still popular as they do not rely on any complex data manipulation or modeling and are therefore quick and relatively easy to produce. Nevertheless, rule based methods can be considered subjective as they do not account for the full picture for several reasons. One reason is that they do not include non-converting pathways, so it will be difficult to truly assess if a certain marketing channel or partner is performing well. Another reason might be that if there is some 'cookie-bombing' going on, high frequency impressions will likely appear as the final touchpoint in a digital customer journey. A final reason is that if you use last click attribution, some channels are likely to appear before a conversion (e.g. search engine advertising) and will have inflated conversion rates due to their position in the journey. In Figure 1 we summarised frequently used rule based attribution methods versus our algorithmic data-driven attribution approach.

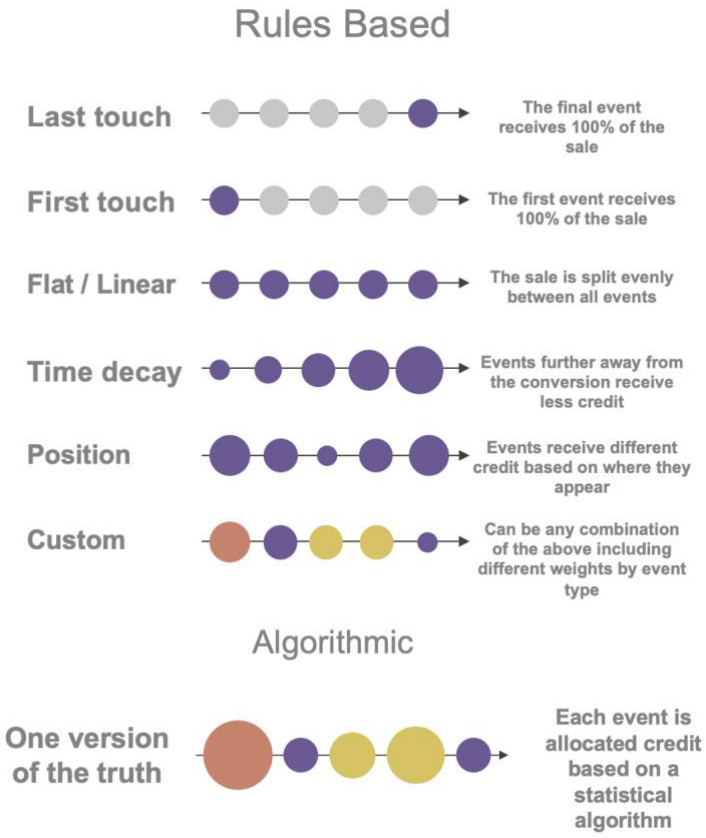

Figure 1. Rule based attribution approaches versus algorithmic data-driven attribution.

Rule based attribution is way too simple when you try to model people's online buying behaviour, given that people got more online than ever, especially during COVID-19. Accordingly, the amount of multi-channel interactions increased and individuals are increasingly online targeted from many different angles. This makes rule based attribution not a fruitful and accurate path to take. On the contrary, our algorithmic data-driven attribution approach, i.e. a Markov chain model, is able to analyse each touchpoint in both converting and non-converting online media journeys. This allows us to quantify each channel's actual contribution in driving conversions without making rigorous assumptions.

Our approach is impacted by the data available though. We use Campaign manager log level data files in Google's Ads Data Hub (ADH) to do our analysis. Although we are able to measure Facebook ads in Campaign Manager, this only includes cases in which a user clicked on a Facebook ad and not the impressions. Furthermore, with Facebook campaigns, it also depends on the buying objective on what is optimised for. So if one optimises for example, add to cart or purchases, this will be maximised but not necessarily clicks. We therefore expect that the number of conversions attributed to Facebook may be underestimated. To test this hypothesis, we carried out two conversion lift experiments; one for Google DV360 and one for Facebook. We used these experimental results to adjust the results of our attribution analysis.

This research was executed for the French automotive brand Renault, that manufactures passenger cars, company cars, trucks, tractors, airplane engines and even formula 1 engines. Our analysis focussed on three passenger cars: Captur, Clio and Zoe. For these cars Renault runs a great variety of ads via for example Google display and search, Facebook and also various local publishers. To measure the incrementality of advertising, Renault differentiates between 'soft' and 'hard' leads. These relate to touch points higher in the sales funnel of buying a car versus touch points that are closer to buying a car. Respectively you can think of: page landings, e-brochure downloads, the start of a car configurator (soft leads), a test drive, dealership appointment, or finishing the process of care trade-in (hard leads). Another feature that distinguishes soft and hard leads is the visibility of customer information, widely known as Personal Information Identifier (PII), which hard leads will include when filling in a form for example.

Methodology

Conversion lift studies

For this research two conversion lift studies are conducted for Google and Facebook in the Dutch market for Renault passenger cars. Conversion lift[1] refers to the measurement of causal differences in conversion behavior. It divides your advertising reach into two groups; test (people that see your ads) and control (people that do not see ads). Afterwards conversions among both groups are compared to determine the incremental lift caused by the ads. When it comes to timing, both conversion lift studies were carried out between 12-02-2021 and 05-04-2021, to subsequently calibrate the ADH attribution analysis.

From a conversion point of view, we both considered soft leads and hard leads. Commercially the goal was to focus on hard leads whenever possible. In the end selected conversion points concern landing pages for cars in the research scope (Captur, Clio and Zoe), e-brochure download, the start of a car configurator and the trade-in finish.

A side note on the conversion point selection is that Facebook did not show the car configurator start in its lift results. Though, it showed the car configurator finished. Seeing that lift confidence levels for both conversion points were above 90%, we decided to upweight the first Facebook conversion lift result. We did this according to the 49.2% difference in total conversions between the start and finish configurator, which was derived from Google's Campaign Manager 360. Although this extrapolation can be considered a caveat on deciding

[1] "Conversion Lift: Helping Marketers Better Understand the Impact of Facebook Ads" https://www.facebook.com/business/news/conversion-lift-measurement Accessed 20 May 2021.

incrementality for car configurator start on Facebook, we assumed a comparable drop off rate despite channel and hence regard this as sound. Trade-in finish had to be excluded after attribution analysis, since it couldn't be calibrated given incremental lift conversions were higher than the total attribution conversions. For this hard lead conversion point, we further elaborated on implications for advertisers in the discussion and conclusion section.

Multi-Touch Attribution

Instead of using a singular or rule based attribution solution, we used a Multi-Touch Attribution (MTA) solution through Google Ads Data Hub (ADH)[2]. ADH is Google's tech stack where you can join various sources of Google event-level ad campaign data (e.g. YouTube, Display & Video 360), while using Google's high standards of data security. Since ADH aggregates results from a group of users, it respects the latest user privacy regulations. ADH stores all results within a owned cloud project in BigQuery[3], which is Google's scalable cloud data warehouse. So, how does the ADH ecosystem look like (see Figure 2) and how does it work in a nutshell[4]? ADH reads all Google campaign ad data from various buying platforms such as Display & Video 360, Campaign Manager 360, Google Ads and YouTube, that reside in our Google-owned cloud project. Based on the queries you run in ADH, these results are aggregated and exported into BigQuery within your Google cloud project. ADH allows linking own data to the Google ads data, this wasn't done for this research project.

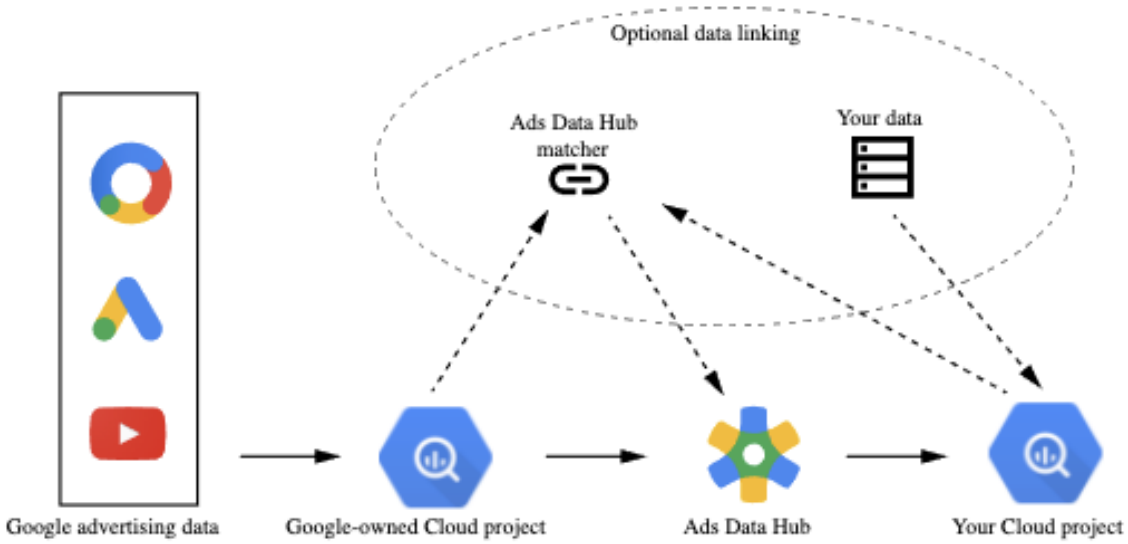

Figure 2. The Ads Data Hub (ADH) ecosystem in a nutshell.

The algorithmic attribution solution that we used with ADH data, concerns a data driven approach where each touchpoint - click or impression in a certain digital channel - is allocated credit based on a statistical model; a Markov chain model. This probabilistic model computes

[2] "Ads Data Hub" https://developers.google.com/ads-data-hub Accessed 26 May 2021.
[3] "BigQuery" https://cloud.google.com/bigquery Accessed 26 May 2021.
[4] "How Ads Data Hub works" https://developers.google.com/ads-data-hub/guides/intro Accessed 26 May 2021.

the probability of transitioning from one ‘state’ to the next. A state in this context refers to a touchpoint with a digital channel or an end state on the path of the digital customer journey. In the end this customer journey may end in a conversion or no conversion over a specific period of time. By ingesting aggregated ADH cookie-level log data as input, the Markov chain model assigns the appropriate weight to each touchpoint in a user’s pathway, based on their ability to positively impact conversions. More specifically, the Markov chain model uses thousands of converting and non-converting online pathways to decide the share of credit each channel should take for each conversion. A simplified example of a Markov chain is visualised in Figure 3.

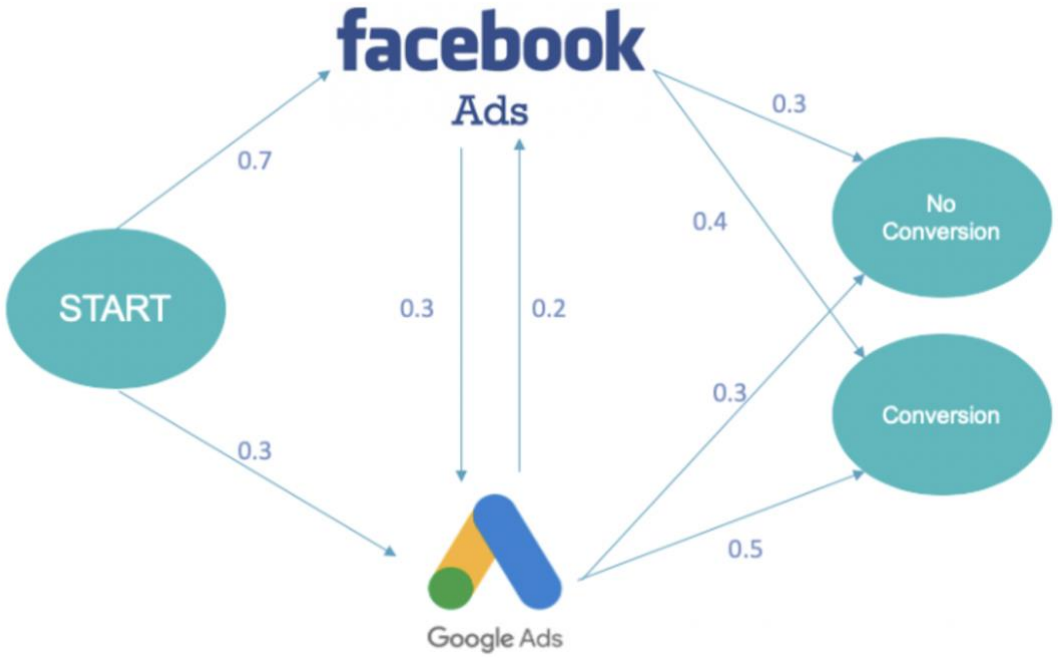

Figure 3. A simplified Markov chain example including Facebook and Google ads.

In order to determine the contribution of either the Facebook or Google ads in Figure 3, so-called removal effects need to be calculated. Removal effects are calculated in three steps. First you look at the total probabilities of conversion. Second you consider the probabilities of conversion when a certain channel is left out (removed). Lastly, you compare the relative differences: what percentage of the conversion will still be reached, when a specific channel was left out; i.e. not in the customer journey. This is your removal effect for that channel. Using Figure 3, we'll demonstrate how the Markov chain model calculates the removal effect of Facebook by taking the following steps:

Example: calculating the removal effect of Facebook using Figure 2.

1. Calculate the total probability of conversion
   a. START -> Facebook ads -> Conversion = 0.7 * 0.4 = 0.28
   b. START -> Google ads -> Conversion = 0.3 * 0.5 = 0.15
   c. START -> Facebook ads -> Google ads -> Conversion = 0.7 * 0.3 * 0.5 = 0.105
   d. START -> Google ads -> Facebook ads -> Conversion = 0.3 * 0.2 * 0.4 = 0.024
   e. Total probablity of conversion: 0.28 + 0.15 + 0.105 + 0.024 = 0.559

2. Calculate the probability of conversion removing the channel Facebook
   f. START -> Google ads -> Conversion = 0.3 * 0.5 = 0.15

3. Calculate the removal effect
   g. 1 – (0.15 / 0.559) = 0.732

Figure 4. Removal effect calculation for Facebook using state probabilities of Figure 3.

So considering Figure 4, the corresponding conclusion is that by removing Facebook as a channel, you'll have nearly 75% less conversions in total.

Calibration

We have taken a four step approach to calibrating the conversion lift results into our MTA results (see Figure 5). First, we compare the absolute lift result in terms of conversions to the MTA results. For the second step we subtract the attributed value for the conversion lift media channel from the MTA results. That leaves us a lower total of MTA conversions. Then we take the amount of conversions per media channel and divide it by the remaining amount of total conversions. This results in the split percentage of conversions each media channel represents, if our conversion lift channel (i.e. Facebook or Google) is removed.

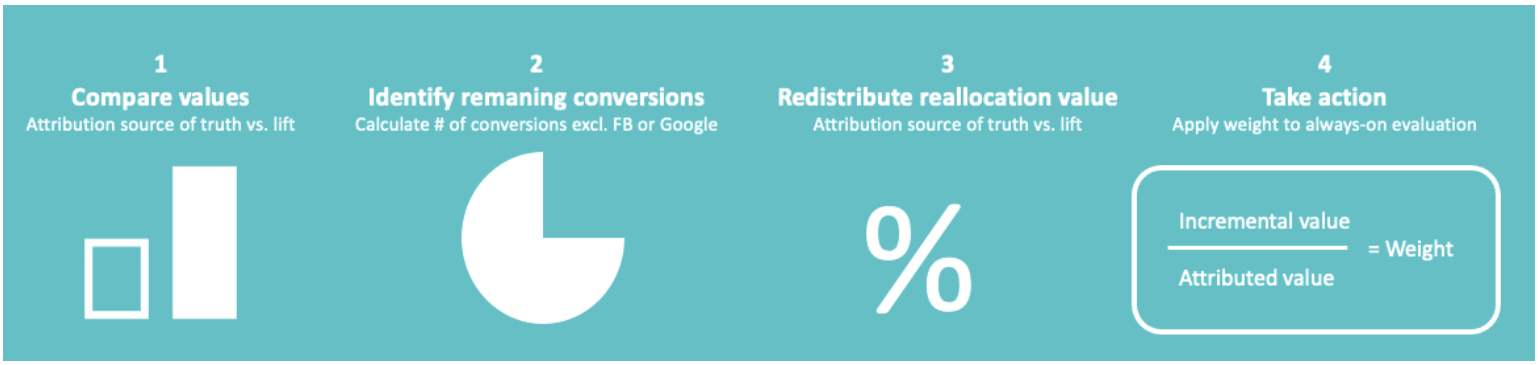

Figure 5. Step-by-step MTA calibration with conversion lift.

Third, we take the total of attributed conversions and subtract the lift conversions; that forms the reallocation value. Afterwards we get the new reallocated attributed values per media channel by multiplying the split percentage with the relocation value. In our last step, step 4, we calculate the weight that can be applied to always-on budget allocation and campaign

evaluation. It is a division of the increment of the conversion lift study through the attributed MTA value. So, what does that mean in practice? When your Facebook conversion lift study shows Facebook generated 150 conversions, while your MTA analysis shows 100 conversions, it means that for every future conversion, (150 / 100 =) 1.5 conversions will be attributed with the use of the weight.

By means of this four step approach, the weighted channel represents the true value while the other channels have the same representation, only redistributed when accounting for the lift result(s). Nevertheless, these four steps were only a sub step in the whole project process. Figure 6 briefly summarises the full project process on the methodology side: from data collection to MTA analysis, running the Facebook and Google conversion lift experiments and finally the calibration of the MTA outcomes with lift results.

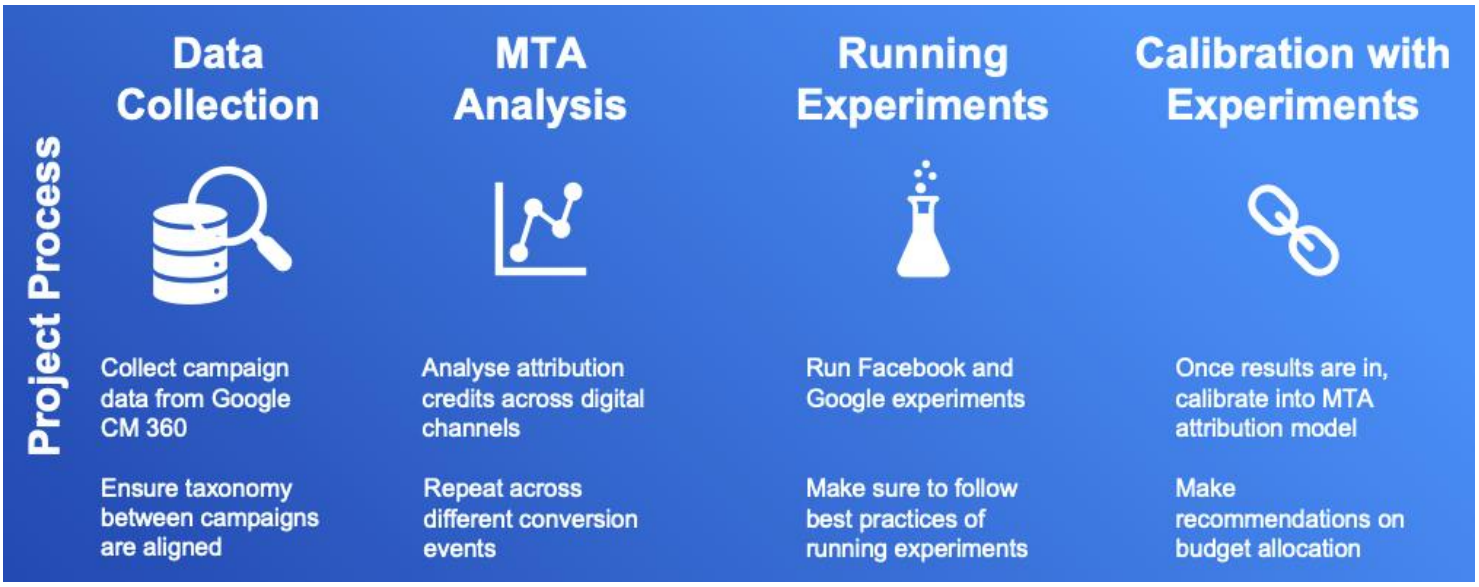

Figure 6. End-to-end project process from data collection to experimental calibration.

Results

In this section we'll discuss in order the results of our three selected and calibrated conversion points: Renault research landing pages (Captur, Clio and Zoe), e-brochure download and car configurator start, by taking the four step approach from Figure 5. We'll also shed light on our conversion lift findings for trade-in finish.

The Renault research landing pages were only found significant in the Facebook lift study with a confidence level above 95% and the largest lift result of 12.300 conversions. Hence this conversion point was calibrated for Facebook only. The step-by-step results can be found in Table 1.

Step 1: ***compare values***			**Step 2:** ***identifying remaining conversions***	**Step 3:** ***redistribute reallaction value***	**Step 4:** ***take action***
Media channel	MTA conversions	Lift conversions	Split percentage	New attributed value	Weight (incremental value / attribution value)
Social test	3.269,95	12.300	-	12.300	3.761
Social other	675,20	-	0.42%	636,52	
Display	23.388,64	-	14.84%	22.048,57	
Display Other	1.244,32	-	0.78%	1.173,03	
Other*[5]	83.903,84	-	53.23%	79.096,51	
PPC Google	46.710,74	-	29.63%	44.034,41	
PPC MSN	1.681,27	-	1.06%	1.584,94	
Total conversions	**160.874**		**Remaining value (total - Facebook test): 157.604,05**	**Total newly attributed conversions: 160.874**	

Table 1. Renault research landing pages (Captur, Clio and Zoe) - MTA calibration with Facebook conversion lift study.

Our first calibration results (Table 1) that relate to Renault research landing pages as conversion point, show the first evidence that Facebook is underestimated in ADH; specifically with a factor of 3.761. This weight of 3.761 translates to that one MTA conversion for Facebook actually represents nearly four true conversions.

For e-brochure download, Facebook lift results showed 636 lift conversions with a confidence level of 95%. Google DV360 lift results however, showed a nearly significant result (confidence level < 80%) with a confidence level of 75.2% and twice as many conversions: 1.509. Since only Facebook showed a statistically significant result with a confidence level above 95%, we calibrated our attribution results with Facebook conversion lift in Table 2.

[5] * See Appendix 1 Explanation of channel “Other” in MTA results.

Step 1: ***compare values***			**Step 2:** ***identifying remaining conversions***	**Step 3:** ***redistribute reallaction value***	**Step 4:** ***take action***
Media channel	MTA conversions	Lift conversions	Split percentage	New attributed value	Weight (incremental value / attribution value)
Social test	79,65	636	-	636	Social: 7.984
Social other	83,27	-	0.27%	81,72	
Display	4.646,26	-	15.52%	4.559.90	
Display Other	368,40	-	1.23%	361,55	
Other*[6]	9.251,79	-	30.90%	9.079,83	
PPC Google	15.071,17	-	50.35%	14.791,04	
PPC MSN	511,42	-	1.70%	501,91	
Total conversions	**30.011,99**		**Remaining value (total - Facebook test): 29.932,34**	**Total newly attributed conversions: 30.011,99**	

Table 2. E-brochure download - MTA calibration with Facebook conversion lift study.

Calibration for the e-brochure download conversion point shows that Facebook conversions were underestimated with a factor of 7.984.

As mentioned in the methodology section, we upweighted the statistically significant Facebook lift result for car configurator finish (1.216 conversions) with 49.2% to car configurator finish (2.471,54 conversions), seeing that the Google DV360 lift conversion study directly showed the car configurator start as statistically significant (90% confidence level). After upweighting lift conversions for Facebook, we calibrated both lift results for this conversion point with our MTA results (see Table 3).

[6] * See Appendix 1 Explanation of channel “Other” in MTA results.

Step 1: ***compare values***			**Step 2:** ***identifying remaining conversions***	**Step 3:** ***redistribute reallaction value***	**Step 4:** ***take action***
Media channel	MTA conversions	Lift conversions	Split percentage	New attributed value	Weight (incremental value / attribution value)
Social test	207,96	2.471,54	-	2.471,54	Social: 11.884 Display: 0.047
Display test	2.126,68	101	-	101	
Display non-test	5.956,88	-	8.43%	5.936.82	
Display Other	620,28	-	0.87%	618,19	
Other*[7]	38.195,3	-	54.05%	38.066,69	
PPC Google	24.843,85	-	35.16%	24.760,19	
PPC MSN	976,53	-	1.38%	973,24	
Social other	61,49	-	0.08%	61,28	
Total conversions	**72.988,97**		**Remaining value (total - Facebook and Google test): 70.654,33**	**Total newly attributed conversions: 72.988,97**	

Table 3. Car configurator start - MTA calibration with Facebook and Google DV360 conversion lift study.

Looking at the results in Table 3, MTA conversions for the experimental Google DV360 display banners are overestimated when comparing it to the true value of 101 lift conversions. Consequently the calculated weight for Google DV360 display banners is 0.047. This weight corresponds to that 1 MTA conversion truly represents 1 conversion for the Google DV360 display banners when having 22 MTA conversions for that channel. More importantly, it can be seen that the incremental value for Facebook (Social test) is nearly 12 times as high compared to MTA conversions. So also for this conversion point there is clear evidence of Facebook conversions being underestimated in our ADH MTA solution.

For trade-in finish we found a statistically significant Facebook conversion lift result (confidence level > 95%) with 520 conversions. Unfortunately, this conversion lift result couldn't be calibrated with MTA results, as the incremental conversion lift was higher than the

[7] * See Appendix 1 Explanation of channel "Other" in MTA results.

total amount of conversions (463,99) found with MTA. Step two in our calibration method requires the other way around; the total amount of MTA conversions needs to be larger than the conversions from the lift study result.

Overall, we can deduce that by calibrating our Google and Facebook conversion lift results for different conversion points, Facebook is indeed underestimated in ADH. Specifically for each step going deeper into the funnel, i.e. from landing pages to e-brochure download to car configurator start, we see that for our ADH attribution solution it is twice as hard to quantify true Facebook conversions. Being explicit, we see that Facebook conversions are underestimated for landing pages with a factor of nearly 4 (3.761), for e-brochure download nearly 8 (7.984) and for car configuration start with a factor of nearly 12 (11.884). From a feasibility point of view, it's evident that using this four step calibration approach is an intuitive, simple and overall quick approach for MTA results to be calibrated by conversion lift. Three important things that can be deduced from Table 1, 2 and 3 is that 1) the total number of conversions (e.g. Table 2: 30.011,99) did not change after calibration, 2) the remaining value (e.g. Table 2: 29.932,34) was reallocated according to the split percentage and 3) the weighted channel represents the true lift value, while others have the same representation.

Discussion and Conclusions

This research used a Markov chain model as a Multi-Touch Attribution (MTA) solution through using Google Ads Data Hub (ADH). The goal of this research was two-fold. Firstly we assessed the feasibility of using conversion lift studies via Google and Facebook to adjust our MTA results accordingly. Secondly we assessed the impact of using the conversion lift study results for MTA calibration, through comparing the MTA results pre and post calibration. Our hypothesis was that Facebook conversions are underestimated in our ADH MTA solution since Facebook is only represented in ADH by impressions and no clicks. Using lift conversion studies to adjust the results was regarded as sound, given lift conversion studies are widely considered as the purest form of testing for evidence. Looking into the conversion lift results, we selected: Renault research landing pages of car models Captur, Clio and Zoe, e-brochure download, the start of a car configurator and the finished process of a car trade-in as conversion points for the MTA analysis. For the start car configurator, Facebook did not show the start car configurator in its lift results, but the car configurator finished. Given lift confidence levels above 90%, the Facebook car configurator finished lift result of 1.216 conversions was upweighted by 49.2% to 2.471,54 conversions. Seeing that this percentage was the difference in total conversions between the start and finish configurator in Google's Campaign Manager 360.

When assessing feasibility of our calibration solution, it's crystal clear this solution is highly scalable as it is easy to use, given an intuitive and simple four step approach consisting of only a few calculations. Even if you have multiple lift results, you can use this method to calibrate your MTA results. Adjacent to it, this solution doesn't require any programming skills or knowledge from econometric modeling. Nevertheless, its simplicity may come at the cost of accurately modeling the conversion contributions of your media channels. Hence a more advanced calibration method can be a Bayesian one, like in our latest study where we calibrated lift results into time series[8]. In this context MTA could be used as fundamental prior

[8] Bayesian time series for Marketing Mix Modeling" https://github.com/annalectnl/bayesian-mmm/blob/main/Bayesian_MMM.pdf Accessed 22 June 2021.

knowledge in the Bayesian model that subsequently gets merged with the conversion lift priors. As such, multiple conversion lift studies and even MTA results can be statistically combined to calibrate and estimate media impact.

In terms of pre and post calibration impact, the Facebook conversion lift result for the Renault car landing pages (models: Captur, Clio and Zoe) showed that Facebook was underestimated nearly four times in ADH. Going deeper into the car sales funnel, e-brochure download showed even a larger discrepancy. For this conversion point, Facebook was underestimated with a factor of nearly eight (7.98). Even more remarkable is the joint Facebook and Google DV360 conversion lift calibration for car configurator start. Facebook was underestimated nearly 12 times (11.88). On the contrary, Google DV360 display banners were overestimated with a factor of 22.

We also found that car trade-in finish, which is a hard lead event, was statistically significant in Facebook lift (520 conversions) but when comparing with MTA results, was difficult to calibrate. So what should advertisers do when they find lift results higher than MTA total results? First we have to comprehend on measuring incremental value, second to understand how to choose the right approach knowing the constraints to answering a measurement question. Conversion lift as mentioned earlier, is a ground source of truth of incremental impact, whereas attribution is assigning credit to touchpoints driving conversions; conversion lift out of the two will output the incremental impact of ads. Now onto choosing the right approach to answering the measurement question, we have to consider the data and measurement constraints for each solution. For conversion lift, it can only answer the impact of Facebook ads and not cross channel (other lift solutions may be available), for MTA it is dependent on the quality of tracking media and non-media via corresponding DCM pixels for each campaign. If the MTA wasn't set up for a particular event, it might not be wise to analyse that event using MTA. If conversion lift for a specific campaign ran during other Facebook campaigns, this might skew results and may not be reliable. This is to highlight the constraints and why choosing the right approach is important. We found car trade-in finish to be statistically significant from Facebook conversion lift so we cannot ignore its impact. Therefore we would recommend to use lift over MTA results in this instance for hard leads. More evidence will need to be collated on measuring incremental impact of hard leads and how to calibrate with MTA results or how to ensure MTA is representative for hard leads.

In conclusion we have shown that our hypothesis was true: Facebook conversions are underestimated in ADH. It appears that the higher in the car sales funnel, Google DV360 display conversions are to an increasing extent overestimated. In contrast, Facebook conversions are increasingly underestimated in the car sales funnel. This verifies that Facebook plays a larger role in generating conversions when being closer to a car sale, compared to Google DV360. Furthermore, this study illustrates that one should differentiate media optimisations per conversion point, or put differently; per touchpoint in the customer journey. By a continuous process of experimentation, one can increasingly better calibrate for online media impact. This can either go through averaging in multiple conversion lift results, per channel, over time or by approaching it in a Bayesian way, as it allows you to merge multiple experimental results due to the unique property of viewing each result as a distribution.

Appendix 1. Explanation of channel “Other” in MTA results

Our MTA solution requires a predefined hierarchy of placements or campaign ids that map channels and subchannels for analysis. If the time window of analysis has placements or campaign ids that weren’t included in the hierarchy or a conversion happened and we didn’t record a media touchpoint before it - touchpoints different than “Other”, then the conversion will be attributed to “Other”. Generally these conversions are a result of SEO and direct to site traffic and may cover up to 30-50% of the total conversions.