Erasmus University Rotterdam

Thesis Project

James 't Hooft

Supervised by Kathrin Gruber

468288jh

Cooperative Game Theory Applications in Multi-touch Attribution Modelling

In order to dive traffic to a website, an online advertiser employs multiple online channels from social media to banner ads. The attribution problem describes the difficulty of knowing which of these channels is most effective in causing a favourable action, and it is integral in allocating an effective marketing budget. Large gains in accuracy as compared to industry standards can be made by utilising data-driven methods. This paper explores multiple modelling techniques, including a replication of the logistic regression of that of Shao and Li (2011). As an alternative the Shapley Value model offers a unique perspective, and with a novel adjustment of the weighting method, further modelling improvements are available. In combination with the findings on Markov Chains in Bovy (2021), this paper acts as a meta-analysis of several contemporary techniques that attempt to solve the attribution problem.

July 4, 2021

1 Introduction

Digital advertising spending worldwide is forecasted to reach a total of 646 billion U.S. dollars by 2024 (pub). The industry is growing in both its effectiveness at reaching online customers and its ease of implementation for companies. Digital ad spending covers everything from paying Google for raising the visibility of a search result to offering a portion of a sale to an affiliate website. There is therefore an expansive scope for ways in which marketers can influence the behaviour of online users. The subtle ways these ads interact with each other in combination with their scale (experts estimate most Americans are exposed to around 4,000 to 10,000 ads each day (Simpson, 2017)) makes it difficult for companies to accurately access the effectiveness of their online marketing strategies. This problem, defined as the attribution problem, aides in the bidding and budgetary decisions of advertisers and therefore has large consequences for the industry (Abhishek and Hosanagar, 2013).

One method, which has remained the industry standard for advertising companies, is last touch attribution (LTA). This method assigns all credit to the last advertising channel visited before purchase. In this way, the attribution for each channel simply describes its 'success rate' in causing a conversion after being clicked on. While, this metric is easily interpretable, it therefore over-estimates the attribution for whichever channel happened to occur last. It has been shown that applying simplified rules to increasingly complex customer journeys skews marketing strategies and distorts market prices (Berman, 2018). Consequently, there is a large amount of demand in the advertising industry for new forms of digital marketing attribution models (Corporation, 2018). The largest discrepancy within the LTA model is in its failure to consider the entire process a user goes through on the journey from exposure to conversion. This journey, called the conversion funnel, describes a process occurring in stages of awareness, information search, evaluation, and then conversion (Jansen, 2011). Each step, whether occurring online or not, is necessary for the next one to occur. This would suggest that every channel, however incrementally, that furthers a user along the funnel is responsible for the ultimate conversion. An intuitive response would therefore be to implement casual models such as those proposed in Dalessandro et al. (2012). However, these models encounter further reliability issues as online data often forms an incomplete picture of the entire conversion funnel.

For example, one user may first be exposed to a brand while reading an online news article. They may click on a banner ad out of interest and later receive a promotional email. Finally, after talking about the brand with friends, they search for it on Google, click on the top result and purchase a product. There are two complicating factors in relating this user path to a theoretical conversion funnel model. First, it is unclear which actions are user-initiated and which are firm-initiated, making it difficult to attribute a marketing channel with causing the user to convert. Even though the final touchpoint before conversion was a paid search channel, the user had already decided on converting. Second, there are multiple channels the user is being exposed to, though only a portion are clicked on. A website, through cookies, is only aware of the previous site a user was on before clicking through. As such, exposure to an ad, regardless of the change in interest or intention it induced, is only recorded if a user directly clicks on it through to the website.

An option that has been garnering popularity in the literature is to use data-driven methods. By aggregating over similar but distinct users that share a similar touchpoint history, an overall picture can be discerned of how channels affect conversions. Instead of knowing how an online marketing strategy is specifically effective, an approximation is made of the average effect channels have on conversion. This then becomes that channel's attribution.

This paper will compare three methods to arrive at this marginal effect estimation: logistic regression, Shapley Value and a simple probabilistic model. The logistic regression is a replication of that of Shao and Li (2011). Conversely, the Shapley Value is an implementation of Cooperative Game Theory. Each advertisement channel is modelled as an individual player cooperating towards a conversion. An average marginal effect is calculated by discounting the difference between two sets of advertisements by how likely the set is to occur combinatorially. This rests on the assumption that each channel is equally likely to appear in a user path. After initial data explorations, it is found this is not the case and therefore an alternative model, Frequency Adjusted Shapley Value, is proposed as an improvement. It is important to note that all three models have ultimately the same mechanism for calculating attributions, namely average marginal contribution to conversion rate, though the models all differ in the weighting given to each relevant data entry.

In order to maximise its usefulness, an effective model must not only be able to quantify the relative conversion credit of each channel but also provide an interpretable result. The ease with which a model can be implemented and then interpreted is an often cited characteristic that is desirable for advertisers. Therefore, a key goal of this paper is to improve Shapley Value to the level of reliability of alternative techniques. It should be noted that despite the moderate success and large inroads made into Deep Neural Networks, such as in Li et al. (2018), this paper will specifically address algorithmic data-driven models in order to focus on this interpretability characteristic.

This paper will start by giving a brief summary of the surrounding literature. It will go on to discuss characteristics of the dataset and the manipulations that were performed onto it. Lastly, results from multiple models will be given alongside an interpretation of what this implies for advertisers as well as for online journeys.

2 Literature Review

As the number of online users has exploded in the previous decade so too has online advertising, which in turn induced a similar rise in academic research. It is important to note that the attribution problem has yet to receive a widely applicable solution. Despite the financial incentives to do so, no method has been found that unequivocally performs better than all others. Often, the complexity of channel interactions demands that a choice between interpretability and accuracy be answered. Rather, the literature has an empirically-tested consensus that simple rule-based methods are consistently less reliable and accurate than statistical approaches. Nevertheless, heuristics such as LTA remain the industry standard even as academic models improve in effectiveness. This section will consider several frameworks that have shown promising results in modelling conversion attributions.

The area of research with some of the most effective outcomes is the Shapley Value approach. Google Analytics, the most widely used website statistics service, uses Shapley Values to provide their attribution estimates for its clients (w3t, 2021). Its unique Multi-Channel Funnels Data-Driven Attribution first forms a probabilistic model of the data such that each user has a probability of converting at every place on its path. From this, Shapley Values can be calculated, which results in the final attribution estimates (Google, 2021). Dalessandro et al. (2012) introduced the Shapley Value technique as a possible estimation of a causal model. The casual inference method found that overall, advertisers place greater weighting upon conversions that resulted from customer impulse, as opposed to conversions that resulted from the factual content of the ads. Shapley Value attribution frameworks often go beyond the fundamentals of cooperative game theory techniques. A popular modification is to utilise order effects within user paths in order to better model specific roles within the conversion funnel. This technique is utilised in Singal et al. (2019) and Zhao et al. (2018). The proposed attribution metric in Singal et al. (2019) further differs from previous iterations in adjusting for counterfactual alternatives. In short, this method looks at the difference in conversions when a channel is shown at a specific stage, as compared to no advertising channel being shown at that stage. Whereas, usual ordered Shapley Value techniques simply look at the resulting increase to conversions when a channel is introduced at a specific stage. The consideration of a counterfactual can offer a wide range of modeling improvements and was preliminarily explored by this paper. However, no results are discussed as the dataset proved inappropriate for this method.

Deep Learning has seen large contributions to the predictability of conversion models. Li et al. (2018) introduces a Deep Neural Net with Attention Multi-Touch attribution model (DNAMTA). The deep neural net component refers to the learning function of the model that maps channel interactions with nodes and edges in a manner similar to that of the neurological function of the brain. This is useful in areas where a large number amount of data is available and relationships between variables are often ambiguous. Therefore, for the subtle and specific ways channels interact in a user path, Deep Learning allows for the opportunity of unearthing interactions that are ignored in other structured and predefined models. DNAMTA uses its supervised learning function in order to approximate the conditional probability of conversion given the user path and user characteristics like demographics and behaviour. The "with attention" component describes the variation in attention the model gives certain touchpoints, allowing for consideration of larger and more complex data for the same computation time. The method exhibits meaningful performance in predicting conversions as compared to other areas of research. Li et al. (2018) also introduces an estimation of channel-specific time decay factors. The highly informational role these factors play is an additional contribution this paper makes to the literature. Deep Learning methods, for their part, are emblematic of the large trade-off between interpretability and performance. It is extremely difficult to inductively explain a channel's Deep Learning attribution by the sum of its variable effects. As a result, the industry has been apprehensive in implementing Deep Learning into its marketing analytics.

Markov models represent a more direct approach in considering the conversion funnel. From an econometric perspective, the sequential nature of a user transitioning from channel to channel is analogous to a Markov chain. If every channel and path consequence (conversion or NULL) is modelled as Markovian states, transition probabilities can be utilised in order to measure a channel's role within a user path. This attribution is taken to be the removal effect, which is the change to the conversion probability when the channel is taken from the overall state space. The method gained greater recognition with Abhishek et al. (2012). This paper formed a dynamic Hidden Markov Model in order to consider the order effects of certain channels. An approximation of conversions not attributable to online ads was a novel inclusion that helped to improve the validity of the results. One inference was that the removal effect of a given channel varied significantly depending on the order it was embedded in the chain. Anderl et al. (2016) expanded Markov implementations by applying them to four new data across three different industries. In a a comparison that used multiple forms of other models, the paper found direct type-in channels to be overvalued while firm-initiated channels to often be undervalued. Also found was that customer-initiated channels had large spillover effects, which describe the increased effectiveness with repeated showings. Firm-initiated channels however, exhibited no such behaviour.

Markovian methods nevertheless have distinct limitations. As Anderl et al. (2016) remark on their own Markov model, "the results are conditional upon a number of management decisions such as channels used or budget per channel". In other words, by using removal effects, what is partially being measured is how often a channel appears in a user path. Therefore if usage varies across channels, attributions could be affected. Bovy (2021) attempts to solve this endogeneity issue with a logistic regression that includes many user-specific variables. In out-of-sample forecasting the new model showed significant improvements to even higher-order Markov models.

Similarly, this paper will improve upon the severe limitations of the Shapley Value model by adjusting for channel usage. In order to compare results to a base model, a replication of Shao and Li (2011) will demonstrate the attributions of a logistic regression and a simple probabilistic model. Despite being written relatively early in attribution modelling research, Shao and Li (2011) makes significant steps in the direction of more reliable modelling. This paper will also contribute to the literature in proposing a novel Shapley Value model, an implementation of previous models to new datasets and a meta-comparison of four models with contrasting perspectives.

3 Data

The data is from a package entitled ChannelAttribution authored by Davide Altomare and David Loris and published on 2021-03-29. It was stochastically simulated on R and its goal was to create a realistic picture of the data online marketeers receive, which models could be tested against. Details of the dataset, such as the number of channels, the path length and the rate of conversion were all chosen specifically to relate to real life online marketing cases. While simulations always risk never truly approaching the randomness or idiosyncratic characteristics of recorded data, all variables are generated from a probability distribution that closely matches the stochastic nature of real life. One disadvantage, however, is that real world inferences cannot be made as to the different effectiveness of different channel types. Instead, each of the 12 channels are given an arbitrary title such as alpha, beta or theta, and therefore whether a channel is firm-initiated or user-initiated is unknown.

The dataset contains 10,000 rows of different user paths that are sometimes repeated. Of these paths, the total number of conversions and the total null or non-conversion value are known. In total, 19,785 conversions are generated from 88,387 users. However, as conversion metrics are given for large numbers of users, it is not known how many users converted at least once.

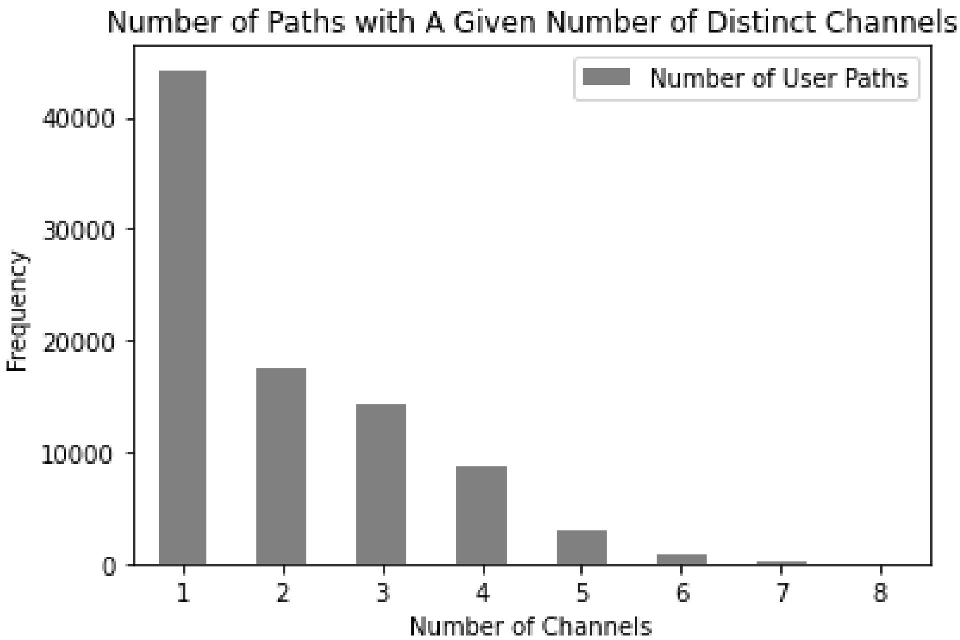


Figure 1: Number of Paths per Channel

The majority of user paths consist of very few channels. Figure 1 indicates the overwhelming lack of user paths exhibiting 8 channels or more. This may appear surprising as there are 12 possible channels a user can click through, however it can be theoretically described with the conversion funnel. A single-channel path most likely describes users who convert after a single touchpoint, not requiring marketing 'persuasion' so much as the awareness the intended website exists. Overall, 44,194 of user paths consist of only a single channel (50% of the entire dataset). This is common for online user path data, appearing also in the Google Merchandise store dataset analysed by Bovy (2021). On the other hand, the more times a user clicks through to a website, the more opportunities they have to convert and end the path.

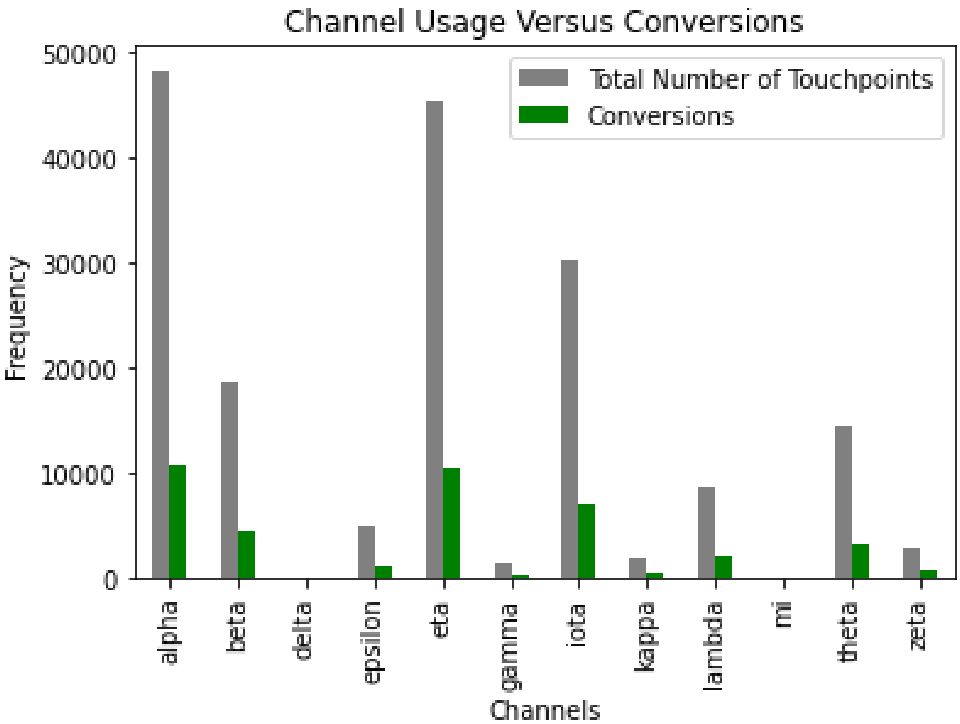


Figure 2: Number of Paths per Channel

Figure 2 demonstrates the large range in usage and conversions between channels. This is realistic and unsurprising when compared to real life data sets such as the Google Analytics sample dataset. This is due to a combination of network effects, where popular channels are able to gain greater popularity, and the specialised uses for certain channels. Email promotions are an example of a channel that serves a distinct role within the conversion funnel. Its effectiveness is largely determined by its permission status and the channel therefore necessarily requires a minimum amount of user engagement (Grubor et al. (2018)). This factor suggests email promotions are only likely to be effective at a specific place on the conversion funnel, namely towards the end. Banner advertisements, on the other hand, are more likely to improve with visibility and therefore frequency. The targeting that does occur is often to do with behaviour types and not where a given user lies on the conversion funnel (Sherman et al. (2018)). The empirical consequence of these heterogeneous channels is that channel usage is highly variable. Conversions per channel, which are given in the same figure, are expectantly shown to be highly correlated with usage. The result of this is that conversions are similarly variable between channels.

In itself, the distribution of conversions across channels is not an obstacle to attribution modelling. However it does break one key assumption of Shapley Value analysis in its fundamental form. As later described in Section 4, basic Shapley Value weights the marginal increase in conversions by assuming each channel is equally likely to occur. Just by looking at the differences in path frequency or conversions between the channels *alpha* and *delta*, this can be shown to be an inaccurate assumption. The channel *alpha* appears in 48,285 user paths, 10,736 of which result in a conversion. However, *delta* is only featured in 42 paths, resulting in only 9 conversions.

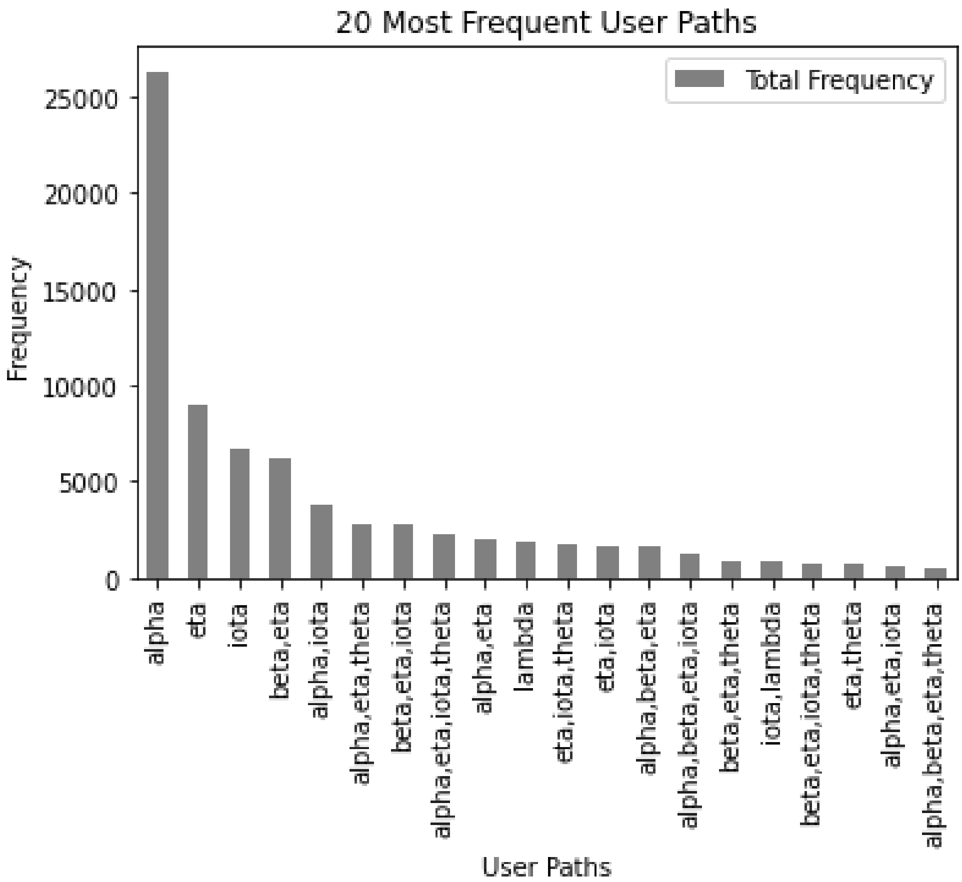


Figure 3: Number of Paths per Channel

The heterogeneity extends to individual user paths themselves. Figure 3 shows the irregularity across the 20 most frequent channel combinations. Clearly, *alpha* at 26,325 user paths dominates as the most common single-channel user path. A surprising observation is the highly skewed distribution of frequencies across subsets. The trend of a disproportionate number of users exhibiting a handful of paths continues further past the 20 channels presented. In fact, the number of users per path quickly trends towards zero. Put statistically, the 20 most frequent channel combinations out of a total of 341 represent 84.2% of all user paths. This may be partially explained by certain channels grouping together. For example, it is common to observe search channels (organic search, paid search and direct) together in the same path as this describes a user who is researching the website on their own (Bovy, 2021). In comparison, it is far less likely to observe a user path consisting of display, social media and referral channels; a user only needs to pass the 'awareness' stage once. The large disparity between certain subsets creates further difficulties for the validity of the basic Shapley Value model. In Section 5 the results of attempts to better model the distribution of channel combinations is shown.

4 Methodology

This paper addresses simple probabilistic, logistic and Game Theory models. A replication will be made of the logisitic and probabilistic model considered in Shao and Li (2011). The Shapley Value estimation from Dalessandro et al. (2012) will be used for the Cooperative Game Theory model. A novel frequency adjusted Shapley Value model is proposed as an improvement on several large flaws Shapley Value has in implementation. Two distinct methods for adjusting for frequencies will be considered, one that uses independent probabilities and another that uses conditional probabilities.

4.1 Mean Absolute Error

In order to compare the estimation accuracy of different frequency adjustment methods, the mean absolute error will be utilised. First, an estimation for every possible channel combination is generated. Then, the absolute difference between the estimate and the true number of paths for that combination is taken and a summation is made. Finally, the number is divided by the total number of possible channel combinations, which for 12 channels is $2^{12} = 4096$. This metric gives implications exclusively for the accuracy of estimating combination frequencies. Only a fraction of these errors are incorporated into the Shapley Value attributions as the method only considers converted combinations.

4.2 Simple Probabilistic Model

Shao and Li (2011) introduces the Simple Probabilistic model as a benchmark method with which gains in accuracy can be compared against. It is a rudimentary aggregation of first and second order conditional probabilities of conversion. The method therefore represents a basic but intuitive calculation of average conversion rates per channel. The attribution is given as follows:

$$A(x_i) = p(y|x_i) + \frac{1}{2(N-1)} \sum_{j \neq i} \{p(y|x_i, x_j) - p(y|x_i) - p(y|x_j)\} \tag{1}$$

For their estimation of conditional probabilities, Shao and Li (2011) count the number of converted paths that feature the channel subset over the total number of paths that feature the channel subset. For example:

$$p(y|x_i) = \frac{C_{converted}(x_i)}{C_{converted}(x_i) + C_{unconverted}(x_i)} \tag{2}$$

Where C describes the enumeration of paths with that feature. Additionally, N describes the number of channels overall, thereby averaging each second order effect.

4.3 Logit Model

The logit model is a regression technique used when the dependent variable is binary while the independent variables are continuous. For the attribution problem, this outcome is either a conversion (1) or no conversion (0). This decision of whether to convert is a gradual transition as a result of incremental changes to multiple variables. It is what allows the gradual curve of the logisitic function to be appropriate for this data. In this way, the statistic can also be seen as a probability model. Accordingly, the input variables X_i are defined as the channels included in user path i and Y_i as the binary variable for the conversion of user path i. As such, β_i describes the average marginal effect of a channel in improving the logarithm of the probability of conversion across all possible user paths. The coefficients are then taken to be approximations of each channel's attribution. Therefore, the following formula for the probability of converting customer i is used:

$$P(Y_i = 1|X_i) = \frac{e^{\beta_0 + \sum_{n=1}^{k} \beta_n x_{n,i}}}{1 + e^{\beta_0 + \sum_{n=1}^{k} \beta_n x_{n,i}}} \tag{3}$$

The model in its regression form demonstrates that it is simply a linear regression of the logarithm of the probability of conversion on its explanatory variables:

$$logit(P(Conversion)) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + ... + \beta_p x_p \tag{4}$$

Here, alpha is a constant and beta are the coefficients. From the above equations, basic assumptions, and therefore restrictions, can already be identified. The most significant is that the channel effects are linear. This assumption does not allow for certain combinations of channels to have different marginal effects on $logit(P(Conversion))$ than the sum of its individual channel effects. Naturally, this restriction contradicts the theory of the conversion funnel discussed in section 1 that suggests a conversion process as being greater than the sum of its stages.

4.4 Shapley Value

A game theory interpretation of attribution modelling considers all advertisement channels to be players cooperating towards creating a conversion. In this way, the attribution given per channel must be equivalent to its average marginal influence on the conversion rate. Under only a small set of assumptions it can be proven that the Shapley Value achieves the optimum spread of attributes. The formula for the Shapley Value for channel r is as follows:

$$\pi_r^{Shap} = \sum_{\chi \subseteq P \backslash \{r\}} w_{|X|} \cdot \{v(X \cup \{r\}) - v(X)\} \tag{5}$$

The characteristic function $v(\cdot)$ describes the number of conversions that contain this set of channels as its user path. Therefore, the identity $v(X \cup \{r\}) - v(X)$ describes the increase to conversions that the inclusion of channel r brings to the set X. This calculation is then made for every possible subset of channels that channel r does not belong to, written as $\chi \subseteq P \backslash \{r\}$. Each subset χ is given a certain weighting $w_{|\cdot|}$, which adjusts for the number of user paths that result in subset χ as well as the number of players. In a conventional Shapley Value method, this weighting is calculated by addressing the number of distinct ways a combination of size $|\chi|$ can be made from the set of all players:

$$w_{|\chi|} = \frac{|\chi|! \cdot (n - \chi - 1)!}{n!} \tag{6}$$

In this definition, n is the number of players. Naturally, in using this combinational sum as the proxy calculation for the number of ways this channel theoretically occurs necessarily assumes each channel has an identical likelihood of appearing in a subset. Or, as Dalessandro et al. (2012) explains, the distribution of channels across the dataset must be uniform.

4.5 Frequency Adjusted Shapley Value

The weighting laid out in Equation 6, while appropriate for systems of similar players, is too constricting to reliably be implemented to user paths. As discussed in Section 3, the rate each channel occurs within the dataset varies greatly between channels. A necessary improvement to reliability is therefore to make use of the characteristics of online user paths in order to give greater weighting to subsets that occur more frequently. Two statistics will therefore be utilised, the frequency of each channel and the frequency of each path length. The reason the frequency of each channel subset is not taken into consideration is that modelling with this information is likely to create overfitting. Channel and path length frequencies however, are relatively homogeneous across online user path datasets, as Section 3 demonstrates. This information can therefore be considered exogenous to the model and appropriate to consider within attribution estimates.

Importantly, some knowledge of the distributions of channels across user paths must be known. For example, consider a situation where channel subset {A, B, C} contains a far greater number of conversions than set {B, C} simply because channels B and C occur significantly less frequently than A. The conventional Shapley Value will over-estimate the contribution of A to this set as the weighting given to {B, C} exclusively depends on the number of combinations a set of 2 elements can be made from 3 players. On the other hand, an accurate estimation of the frequency of {A, B, C} versus {B, C} would allow one to consider the increase in conversions differently. The difference in conversion rates could be adjusted for the subsets' frequency thereby revealing the true effectiveness of A in causing conversions. Done for every channel and subset, this would result in more accurate attribution estimates.

4.5.1 Conditional Probability Weightings

The most theoretically accurate calculation for the likelihood a subset occurs is to dissect the estimate into the probability a subset of that particular size occurs and the probability it would consist of those particular channels. As a demonstration, take subset {*alpha*, *beta*, *gamma*}. A path with 3 channels occurs with probability 0.161. Then, to calculate the probability a subset of size 3 would consist of these three channels, the probability of each permutation is considered. For each permutation the conditional probability the channels occur in the particular order is calculated. This is done with: channel frequency/(total paths - frequencies of channels already of the chain). The summation of each permutation probability is made. For {*alpha*, *beta*, *gamma*} this is 0.017. Finally, the product of both probabilities is taken for the likelihood that this subset represents a given user path, 0.003.

4.5.2 Independent Probability Weightings

A far more intuitive approach is to take the product of each channel probability. For {*alpha*, *beta*, *gamma*} this becomes:

$$\frac{\text{alpha frequency}}{\text{total number of paths}} \cdot \frac{\text{beta frequency}}{\text{total number of paths}} \cdot \frac{\text{gamma frequency}}{\text{total number of paths}} \quad (7)$$

This calculation is given as $0.55 \cdot 0.21 \cdot 0.016$ to equal 0.002.

5 Results

5.1 Frequency Adjustment Comparisons

Frequency Adjustment Method	Mean Absolute Error
Independent Probability Estimation	18.23
Conditional Probability Estimation	21.54

Table 1: Aggregate Estimation Errors between Independent and Conditional Frequency Estimations

Comparing estimation accuracy between independent and conditional probability methods yields an unintuitive result. Despite conditional probabilities taking into account more information, such as the distribution of path lengths and single channel path frequencies, Table 1 demonstrates that the method has greater errors in estimating the frequency of a path than taking the product of independent probabilities. Aggregately, conditional probability estimation does only marginally better than predicting the frequency of each path to be zero, which would produce a Mean Absolute Error of 21.58. The fact that additional knowledge of the channel subset leads to little to no gains in modelling accuracy would suggest that the touchpoints on a user path are being generated iterativeley and independently of previous touhpoints on the path. The alternative would be for the user paths themselves to be stochastic, thereby warranting the use of combinatorial effects. To illustrate this effect, the 30 most frequent user path channel combinations are plotted against their estimates.

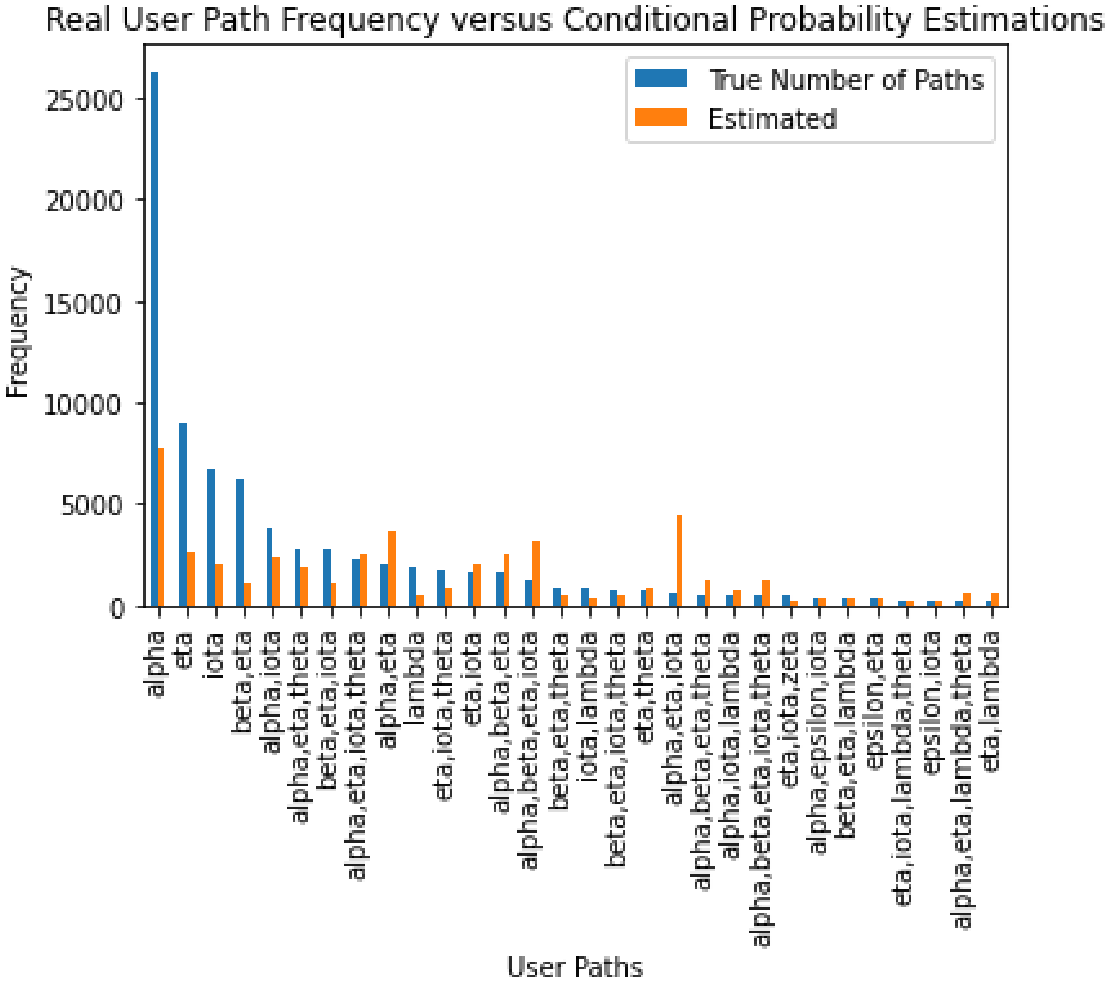


Figure 4: Estimated Frequencies of Key Combinations by Conditional Probability

For conditional probability estimation, the frequency of single-channel paths are taken into consideration when estimating combinations of multiple channels. The channel subsets $\{alpha\}$, $\{eta\}$ and $\{iota\}$ all occur far more frequently than other combinations. The result of this is that the permutation probabilities of $\{alpha, beta, eta, iota\}$ are over-estimated as the sample size for each iteration is assumed to be significantly smaller once either $\{alpha\}$, $\{eta\}$ or $\{iota\}$ are accounted for. Figure 4 demonstrates this over-estimation as well as the over-estimation of many combinations that contain the channels $alpha$, eta and $iota$.

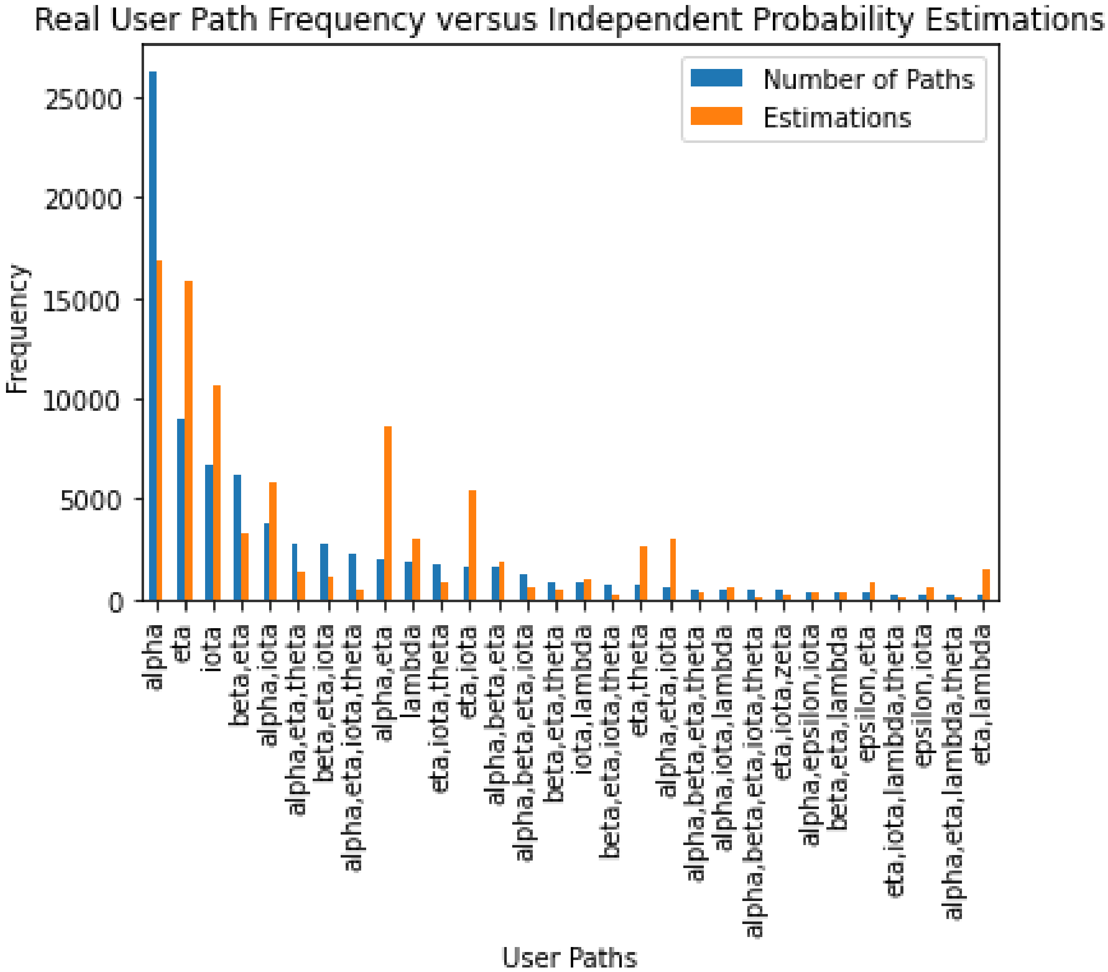
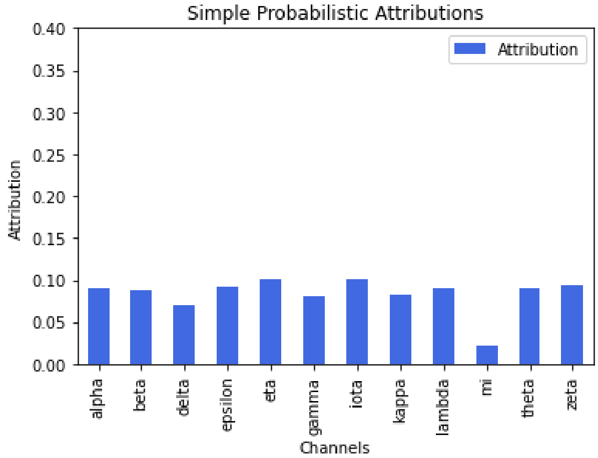


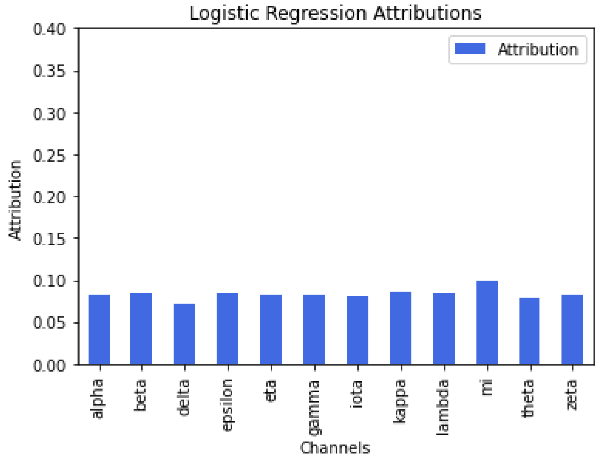
Figure 5: Estimated Frequencies of Key Combinations by Conditional Probability

Independent probabilities show clear improvement in modelling the true skewness of the data towards key channels. Despite over-estimating combinations of very frequent channels, as in $\{alpha, eta\}$, the method nonetheless manages to estimate the large discrepancy in frequency between small combinations of frequent channels and larger combinations that include rare channels. This further supports that channels in a user path are being simulated independent of channels already in the path.

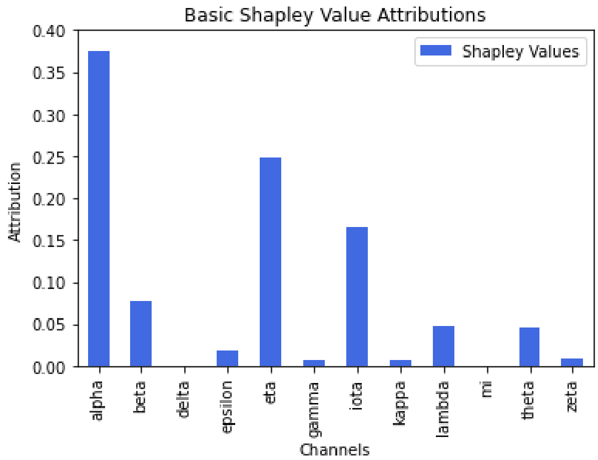
5.2 Attribution Comparisons



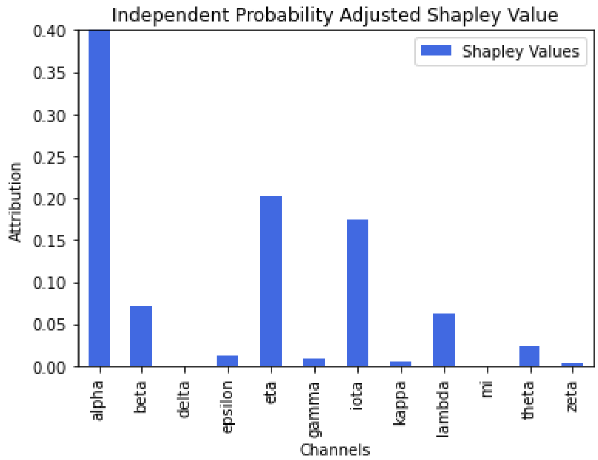
(a) Simple Probabilistic



(b) Logistic Regression



(c) Basic Shapley Value



(d) Frequency Adjusted Shapley Value

Figure 6: Channel Attributions Given by Each Model

The striking differences in given attributions across different models is that Shapley Value appears highly frequency dependent in comparison to a simple probabilistic model and a logistic regression, even after adjusting with independent probabilities. When compared against channel usage in Figure 2, the Shapley Values show many similarities with the number of touchpoints per channel. This was expected when considering the Shapley Value assumptions in Section 4, however even adjusting for this frequency using independent probabilities (which was chosen over conditional for its improvements in estimation accuracy) shows only incremental changes when compared to the differences with simple probabilistic and logistic. Even with some frequency adjustments in the attribution of eta (24.8% to 20.1%), larger corrections must be made in the marginal conversion differences between combinations if Shapley Value attributions are to be independent of the underlying differences in channel usage. Using conditional probabilities to adjust for frequency offers a picture of what less endogenous results may look like. The attributions, which are given in the appendix, show far less variability between channels while differences in effectiveness are still visible. Nonetheless, the inaccuracy of the method's corrections yields its attributions unusable.

It is likely the data was generated with relative homogeneity regarding the effectiveness of different channels in promoting conversions. This is the take-away from both the simple probabilistic model and the more accurate logistic regression. In this situation, the logistic regression proposed by Shao and Li (2011) would prove to be a reliable measurement of channel attributions. It must also be noted that while the relative differences between attributions vary between models, each model arrives at a similar conclusion of which channels perform best. A notable exception is mi, which the logistic regression awards 9.9% , far above any other model. For a full list of attributions, see Table 2 in the appendix.

6 Conclusion and Discussion

This research demonstrated how a novel Shapley Value method can be formed by adjusting for frequency and how this compares to a simple probabilistic model and a logistic regression. Moderate success was shown in estimating the frequency of different channel combinations by taking the product of each channel probability. As mentioned previously, channel frequencies are relatively constant across datasets that utilise the same channels. Therefore, the same frequency adjustment mechanism can be utilised for many different datasets. This would allow the easily interpretable Shapley Value to be used in areas of heterogeneous channel usage. The complicating factor is whether the dataset this paper used is true to real life user paths. It is suggested that user path channels are generated at every new touchpoint and that the distribution with which a new channel is simulated is independent of the channels already featured in the path. This is a possible reason for the success of independent probability estimation. This is perhaps unrealistic as the research of Markov Models into user paths in Bovy (2021) implies a clustering behaviour of sequential touchpoints. This is where certain channels are more likely to appear after each other.

More research is needed across varying databases into the different probability distributions of channel combinations. Shapley Values must then be fully corrected for this distribution in order to be most likely to show similar results to reliable methods such as logistic regressions. To fully prove the effectiveness of a frequency adjusted Shapley Value technique, attributions must be compared to models that incorporate order effects or user characteristics. Another feature that may be possible to investigate is the time-delay between touchpoints. Survival theory has already shown success in probabilistic models (Zhang et al. (2014)), it would therefore be interesting to also introduce as a method of conversion adjusting for Shapley Values.

References

Usage statistics of traffic analysis tools for websites, 2021. URL `https://w3techs.com/technologies/overview/traffic`$_a$*nalysis*.

Vibhanshu Abhishek and Kartik Hosanagar. Optimal bidding in multi-item multislot sponsored search auctions. *Operations Research*, 61(4):855–873, 2013.

Vibhanshu Abhishek, Peter Fader, and Kartik Hosanagar. Media exposure through the funnel: A model of multi-stage attribution. *Available at SSRN 2158421*, 2012.

Eva Anderl, Ingo Becker, Florian Von Wangenheim, and Jan Hendrik Schumann. Mapping the customer journey: Lessons learned from graph-based online attribution modeling. *International Journal of Research in Marketing*, 33(3):457–474, 2016.

Ron Berman. Beyond the last touch: Attribution in online advertising. *Marketing Science*, 37(5):771–792, 2018.

’t Hooft Bovy, de Vreede. How company controlled channels influence the online customer journey. *Erasmus Universiteit Rotterdam*, 33(3):457–474, 2021.

Teradata Corporation, Jun 2018. URL `https://www.prnewswire.com/news-releases/teradata-lunexa-partnership-introduces-digital-marketing-attribution-solution-for-teradata- aster-analytics-platforms-157198315.html`.

Brian Dalessandro, Claudia Perlich, Ori Stitelman, and Foster Provost. Causally motivated attribution for online advertising. In *Proceedings of the sixth international workshop on data mining for online advertising and internet economy*, pages 1–9, 2012.

Evert De Haan, Thorsten Wiesel, and Koen Pauwels. The effectiveness of different forms of online advertising for purchase conversion in a multiple-channel attribution framework. *International Journal of Research in Marketing*, 33(3):491–507, 2016.

Ivana S Domazet, Ines Đokić, and Olja Milovanov. The influence of advertising media on brand awareness. *Management: Journal of Sustainable Business and Management Solutions in Emerging Economies*, 23(1):13–22, 2017.

Anindya Ghose and Vilma Todri. Towards a digital attribution model: Measuring the impact of display advertising on online consumer behavior. *Available at SSRN 2672090*, 2015.

Google. Bounce rate. URL https://support.google.com/analytics/answer/1009409?hl=en.

Google. Mcf data-driven attribution methodology - analytics help, 2021. URL

7 Appendix

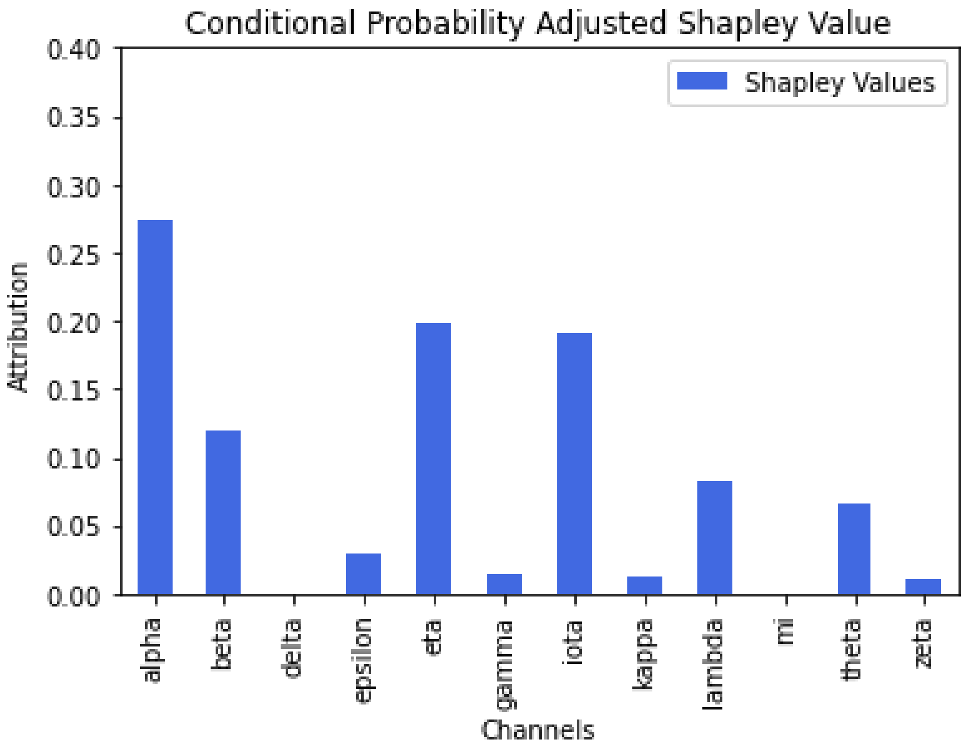


Figure 7: Estimated Frequencies of Key Combinations by Conditional Probability

Channel	Independent Probability	Basic Shapley Value	Logistic	Probabilistic
alpha	43.4	37.5	8.2	9.0
beta	7.2	7.7	8.4	8.9
delta	0	0	7.2	7.0
epsilon	1.3	1.8	8.4	9.2
eta	20.1	24.8	8.3	10
gamma	0.9	0.7	8.2	8.1
iota	17.5	16.5	8.2	10
kappa	0.5	0.8	8.6	8.3
lambda	6.3	4.8	8.5	9.0
mi	0	0	9.9	2.2
theta	2.4	4.7	7.9	9.0
zeta	0.3	0.8	8.2	9.3

Table 2: Attributions for Each Channel in %