FRONTIERS OF MARKETING DATA SCIENCE JOURNAL

At the Forefront of Smart Data Marketing

Improving B2B Marketing Efficiency with Multi-Source Intent Data Targeting

Multi-Touch Attribution for Consumer Journey: Snapchat's Learnings

Attribution and the Marketing Mix Model

Can AI Reduce Noise in Decision Making?

ISSUE No.4 | 2023

Sponsored by marketscience

I-COM GLOBAL FORUM FOR SMART DATA MARKETING

Marketing Attribution and the AI Illusion

Marketing attribution aims to quantify the *incremental* contribution of each element of the marketing mix on the consumer path to purchase, typically in the form of a sales conversion. Outputs form the basis of ROI calculations and optimal budget allocations, helping to inform marketing strategy and the planning cycle.

Attribution measurement approaches generally fall into three camps: disaggregated Multi-Touch Attribution (MTA), aggregated Marketing Mix Modelling (MMM) and (unified) combinations of the two. However, given concerns surrounding user-identifiers, walled gardens, GDPR, third-party cookies and overestimation of short-term performance media, MTA is now rapidly falling out of favour. This has seen a resurgence in the popularity of MMM, notably amongst key players such as Google and Facebook.

Whereas it is certainly true that aggregated MMM solutions can overcome many of the privacy concerns of MTA, bias towards short-term performance marketing remains a core problem. Yes, the time series dimension can help redress the balance between short-term activation and long-term brand-building. However, like MTA, the *selection bias* inherent in much online media confounds correlation and causation, where (part of) the 'treatment' outcome (sales) is caused by a factor that predicts the likelihood of selection into treatment rather than the treatment itself. For example, consumers with a greater propensity to buy predicts the level of search traffic, which in turn predicts the sales outcome. As such, a large proportion of site visits are simply an artefact of the sales process. This creates an *endogeneity* or *identification* problem, leading to biased estimates of the search-sales impact and all marketing effects that work through it.

The growing popularity of AI and automated machine learning (ML) techniques has only exacerbated this problem, where rigorous causal analysis is sacrificed for the promise of rapid 'real-time' MMM delivery at scale. This is fine if the goal is simply a set of best-fitting predictive models: after all, stepwise regression has been around for a long time. However, if the aim is to uncover the 'structural' cause and effect relationships necessary for accurate budget allocation, standard AI/ML approaches simply don't cut it. This has led to the introduction of 'causal' AI techniques, which attempt to discover the causal pathways present in an observational data set, based on Directed Acyclic Graph (DAG) structures (*inter alia*, Pearl, 2000). However, since human context is always required, such techniques 'do not yet work as stand-alone methods for causal learning' (Peters *et al*, 2017). Furthermore, there is rarely one unique chain: endogeneity bias stems from ignoring the simultaneous likelihood of all plausible DAGs in the data. Consequently, AI techniques are useful for uncovering sets of competing (causal) chains but cannot automatically solve the selection bias problem *per se*.

So what is the way forward? One popular solution lies in experimental design research such as A/B testing. Advocates contend that this is the only way to assess true incrementality. Results can then act as Bayesian priors to provide some form of 'ground truth' in both conventional and AI-driven MMM solutions, thereby solving the selection bias problem. However, even setting aside objections to Randomised Controlled Trials as the gold standard of causal inference (*inter alia*, Deaton and Cartright, 2018), experimentation results are rarely available as part of the MMM data collection process. Even if they were, experiments would need to be run across all endogenous variables at considerable expense: an impractical solution in the vast majority of cases.

In practice, therefore, we generally have to do the best we can with observational data and seek recourse in more traditional parameter identification techniques: namely, statistical methods designed to separate correlation from causation such that the estimated parameter reflects a true incremental effect. Candidates range from Instrumental Variables (IV), Difference in Differences, regression discontinuity designs and Heckman correction factors, through to Latent IV, Gaussian Copulas and weighted DAG analysis. Whichever route is taken, the message is clear: for meaningful attribution, the chosen identification scheme needs to be clearly specified as part of any MMM engagement. This requires careful human judgement. No amount of automated AI bluster can escape this central fact.

References

1. Deaton, A. and Cartwright, N. (2018). "Understanding and misunderstanding randomized controlled trials", *Social Science & Medicine* (210), Pages 2-21.

2. Pearl, J. (2000), "Causality", New York: *Cambridge University Press*.

3. Peters, J., Janzing, D., & Scholkopf, B. (2017). "Elements of Causal Inference: Foundations and Learning Algorithms" (Adaptive Computation and Machine Learning series). *The MIT Press*.

Author

Peter Cain
Executive partner & Co-founder
Marketscience,
UK

Table of Contents

Improving B2B Marketing Efficiency with Multi-Source Intent Data Targeting

Olivia Kenney
Foundry, Canada

Sreejata Chatterjee
Foundry, Canada

Kundan Kumar
Foundry, Canada

Classifications, Key Words:

- Intent data
- Marketing experiments
- Ad targeting
- Marketing performance
- Data-driven marketing

Abstract

Intent data has emerged as a critical asset for B2B marketing teams, enabling faster growth and more effective customer acquisition and retention. According to a 2020 Gartner study, by the end of 2022, more than 70% of B2B marketers will be using intent data (Antin & Berkowitz, 2020). With various B2B intent data vendors providing different, singular silos of first-party, second-party, or third-party data, marketers face a challenge in choosing which vendors and data to rely on. 71% of marketers at US companies with +$10M in revenue use three or more intent data sources, and discerning them falls on the marketer, making the process even trickier (Crane, 2021). Some intent signals exist in immeasurable or inaccessible places, such as in-person conversations. However, first, second, and third-party data can be combined for a more comprehensive picture of the buyers' journey. Layering diverse data on buying signals results in more effective and efficient B2B targeting.

Our research compares the results from two types of datasets: a subset of Ideal Customer Profile (ICP) fit prospects which acted as a control audience, and ICP fit prospects displaying buying intent across a combination of first, proprietary second, and third-party platforms. We found that using intent-based targeting from multiple sources performs more effectively and efficiently than traditional ICP fit targeting in B2B advertising.

1. Introduction

The premise of the research and solutions focuses on B2B software companies.

B2B companies typically determine ad targeting based on their Ideal Customer Profile (ICP). Crunchbase defines ICP as "the type of company that would benefit the most from your product or service. Companies that fit your ICP are most likely to buy and continue to use your product, making them extremely important for business growth." (Robinson, 2022). Organisations typically determine ICP by identifying commonalities in existing customers with the highest Customer Lifetime Values (CLV) and where the product meets customer needs. Companies may have more than one segment of their ICP and typically base it on the industry, company size, job title or decision maker, and processes in place. For example, Directors of Marketing at B2B Financial Technology companies with 51-200 employees based in the United States with an established outbound sales motion.

Purchase intent data is a category of behavioural data that identifies a strong likelihood that a buyer is in-market for a product or service. The term has been shortened to buying intent or intent data.

B2B intent data has become an essential resource in marketing strategies, with 59% of B2B marketers using intent data (Crane, 2021) and 71% using three or more different sources. Companies leverage this data to make informed decisions for prioritisation, messaging, and targeting with more efficient resource usage.

The three major types of intent data are:

• First-Party Intent: Data is only accessible and controlled by the owner. For example, website page visits, whitepaper download lists, or first-party CRM data. First-party intent users typically install a tag on their websites to analyse website visitors at the organisation level (for example, "IBM" or "SAP"). Contact-level website activity is only available following opt-in permission from the contact or contacts logging in and consenting to data being shared. Contact-level first party data was not used in this experiment.

• Second-Party Intent: Data arises when a first-party provider can access or host first-party data and acquires the ability to sell it. For example, review sites, such as G2 or TrustRadius, that analyse visitor activity on their websites and then provide the relevant engagements, such as a visitor viewing a competitor's profile, to their customers. Another example is publisher sites monitoring visitors viewing relevant content, then providing data on that activity to their customers.

• Third-Party Intent: Data generated and hosted by a non-connected entity, other than first-party from a consumer's point of view, is considered third-party. It is often collected from various sources and provided to businesses by vendors who do a combination of collecting data and working with other data providers to aggregate and enrich the data points. For example, all public web data can be considered third-party data. A vendor may collect that data, then contract another data vendor to enrich those public web engagements with firmographic data. This data is typically collected by mining buying engagements and discerning them to insights relevant to a firm's ICP and product category (for example, CRM software or ERP system).

Data delivery and usage of intent data vary based on the data type, vendor, and use case. Intent data vendors, regardless of the party type, often integrate and push data directly into the users own CRM or Marketing Automation tools at the account or contact level. Second- and third-party vendors may also deliver data via CSV, server, or lead generation service.

This research aims to understand if intent-based ad targeting performs more effectively than traditional methods that rely solely on public sources of static firmographic and demographic dimensions such as company size or job title. This experiment used two audiences: an exposed audience and a control audience. The exposed audience is based on buying intent signals from first-party, proprietary second-party, and third-party sources. The control audience is based on publicly sourced firmographic and demographic targeting dimensions. We ran identical ads to each audience for three separate brands to evaluate ad performance, with targeting as the variable. We found that the exposed, intent-based audiences outperformed the control audience across all tests, improving efficiency by 2.5x.

2. Methodology

For this experiment, we used a blend of first-party, proprietary second-party, and third-party data, combining and prioritising them using a smart scoring algorithm described later. The database comprises 52 million intent signals per week, including social activity, website activity, and proprietary second-party data from owned publishing networks. Intent signals were mapped to a database of 4 million companies and over 40 million B2B contacts. Within the database, intent signals are grouped into 7,560 category taxonomies such as Accounting Software, Competitive Intelligence Tools, and Data Entry Services Providers. On average, using a proprietary fuzzy match algorithm, we can match 60% of users to their corporate profiles and up to 90% back to their account level.

2.1 Intent Data

First-party intent data was identified using tracking tags on each brand's website. Website visitors were mapped back to the companies over the previous 30 days and enriched with third-party signals and contacts. Essentially, we recorded accounts visiting each company's website for 30 days and enriched that data with third-party signals and ICP-fit contacts.

Third-party intent was collected by crawling the public web across social networks such as Twitter and LinkedIn, job boards, technographic data, blogs, forums, company websites, SEC filings, content consumption information from partners, meta-search engines, and industry publications. We crawl millions of data points every day and find approximately 1.2M intent signals daily. We use a fuzzy matching algorithm that looks at the username, location, bio, profile photo, and other available metadata to map these signals to the individual who took the actions online. When that is impossible due to sparse data or the person not being in our database, the intent signal is matched to the associated company and then enriched with contacts based on ICP criteria for targeting.

Second-party data was captured and included based on engagements with owned and operated publishing networks and editorial brands. Engagement data occurred at the contact level and was sorted into relevant category taxonomies based on the content associated with the engagement. For example, a Director of DevOps downloading a whitepaper on how often PC and other devices should be refreshed may signify they are evaluating a new hardware purchase.

An advantage to having Foundry's proprietary and first-party data is that the accuracy is typically high, especially when we filter out ISPs from the list. IP address mapping is a technique that has been validated over the ages, so confidence is high. In the case of our first-party data, for a company to be identified, we must be at least 75% confident in the fidelity of the data. This confidence score is based on multiple signals of identification, human and machine learning sourced, and includes score adjustments based on frequency, time decay, and non-conflicting signals.

The third-party data that Foundry collects is fuzzy matched to accounts with 70-90% accuracy, depending on the data source and industry. For example, the accuracy is higher from LinkedIn vs from Twitter, due to it being a professional network. The accuracy also differs by the industry as internet conversations about schools and hospitals could be very different from conversations about buying million-dollar software contracts. Since we're only collecting B2B activity, the accuracies are higher. We merge all this data based on the domain or website URL, which is commonly used as a unique and reliable identifier.

2.2 Scoring

All intent signals are scored at the account level using trend-based scoring. This model is closely derived from the statistical Z-Scoring model often used in financial industries ("*How to Calculate Z-Score and Its Meaning*", 2022). The model ranks accounts out of 100 using a 90-day baseline. Using a baseline ensures large organisations do not falsely score higher due to relative volume. The score considers various factors, each with a different intent priority or weight: number of intent signals normalised by company employee size, size of the company, number of unique individuals showing intent within a company, titles, and seniority of individuals showing intent to buy and the type of signal(s) they have shown. Some of the signal types include “competitive engagement,” “awareness of the industry,” “company growth,” “company funding,” “event attendance,” etc. For example, the algorithm would score a company discussing an industry term lower than a company receiving funding. The different signal types vary in strength and thus have different weights attached, which are used for the rule-based part of the scoring model. In general, a company having “n” daily intent signals can have scored 0 since they have “n” because “n” is their baseline. If their signals increase, the score begins to climb. The companies with the most intent have a score of 100. We classify any score above 70 as “hot.”

Signals across all three data types (first party, proprietary second party, and third party) were aggregated and scored. The account scores were sorted in descending order, and the top-scored accounts were used, targeting the most relevant companies and contacts.

For the score, we use Z-Score, which is famously used for predicting and measuring stock market growth and surges. It can be calculated as:

$$Z = \frac{\text{standard deviation}}{\text{data point} - \text{mean}}$$

We use a similar approach for scoring the accounts where we combine a few Z-scores.

We used a rolling 7-day average to find the mean number of intent signals. The signals are broken up into five categories, each with its own Z-Score. Competitive signals, event visits, signals for an influx of resources such as funding, acquisition, large contract signing, hiring signals, and everything else (reading blogs, commenting, following someone on Twitter, etc.). They are all weighted differently, with some being priority signals which do not follow the rule of medians. The default weights and signals are as per the formula below:

{'priority_group': {'triggers': [{'name': 'expansion', 'priority_days': 90}, {'name': 'investment', 'priority_days': 90}, {'name': 'leadership', 'priority_days': 90}, {'name': 'technology', 'priority_days': 90}, {'name': 'partnership', 'priority_days': 90}, {'name': 'company:funding', 'priority_days': 90}, {'name': 'company:newlocation', 'priority_days': 90}]}, 'weighted_groups': [{'name': 'behavioral-intents', 'weight': 0.8, 'triggers': ['engaged', 'followed', 'related', 'competitor', 'following', 'company:content', 'event', 'company:event', 'company:award', 'recognition']}, {'name': 'growth-intents', 'weight': 0.2, 'triggers': ['company:hiring', 'jobposting']}], 'default_weighted_group': 'behavioral-intents'}

2.3 Audiences

Each of the three companies ran separate tests with identical processes and settings. Each test had two audience groups: a control audience and an exposed audience. The firmographic and job title criteria remained consistent for control and exposed audiences to reduce the risk of selection errors, but the control group did not factor in behavioural signals.

The control audience was based on a randomised group based on firmographic dimensions, including company size, location, job title, and industry. This imitates typical B2B targeting dimensions used in advertising platforms like LinkedIn (“Best Practices for Ad Targeting”, n.d). The control audience was not generated with behavioural intent signals to allow a comparison of ad performance between ads shown to the randomised control audience, and those shown to the exposed audience that was generated based on behavioural data. The exposed audience was based on behavioural data of individuals who had shown intent to make a relevant purchase and used firmographic data as a filter to ensure ICP fit.

- **Control audience:** a randomised group with B2B marketing job titles at companies ranging from 50-5,000 employees globally, based on static, public data.

- **Exposed (intent-based) audience**: based on the likelihood of being in-market, filtered by the ICP and the company's product or service, based on behavioural signals of buying intent across first, second, and third-party data.

Tests ran during overlapping periods +/- three days. We have defined them as Test A (KickFire), Test B (LeadSift), and Test C (Triblio). The three tests were run based on these companies (KickFire, LeadSift, and Triblio) based on access to Google Ad Accounts, as well as budget and time resources. These companies were not used as the ad platform in this experiment but rather as the test subjects. Each one is a technology company in the B2B space.

Test A was established based on the company's relevant intent signals and ICP. The two audiences, or groups, used as variables were:

1. Control: A randomised group of 2,178 individuals with B2B marketing titles employed by a company with 50-5000 employees globally.

2. Exposed (intent-based) audience: A group of 2,168 individuals made up of first-party website visitors and individuals who have shown intent toward Buyer Intent Data Tools, Visitor Intelligence Software, Visitor Identification Software, named competitors, and custom relevant keywords. Individuals must have senior marketing job titles and be employed at a company with 50 to 5,000 employees.

Test B was established identically to Test A, other than updating the intent signal criteria to the specific company. The two audiences, or groups, used as variables were:

1. Control: A randomised group of 3,456 individuals with B2B marketing titles employed by a company with 50-5000 employees globally.

2. Exposed (intent-based) audience: A group of 3,446 individuals made up of first-party website visitors and individuals who have shown intent towards Buyer Intent Data Tools, Lead Generation Services, Other Lead Generation Software, named competitors, and custom relevant keywords. Individuals must have senior marketing job titles and be employed at a company with 50 to 5,000 employees.

Test C was established identically to Test A and B, other than updating the intent signal criteria to the specific company. The two audiences, or groups, used as variables were:

1. Control: A randomised group of 2,542 individuals with B2B marketing titles employed by a company with 50-5000 employees globally.

2. Exposed (intent-based) audience: A group of 2,671 individuals made up of first-party website visitors and individuals who have shown intent toward Account-Based Advertising Software, Account-Based Analytics Software, Account Based Orchestration Platforms, Account-Based Web and Content Experiences Software, Display Advertising Software, Marketing Account Intelligence Software, named competitors, and custom relevant keywords. Individuals must have senior marketing job titles and be employed at a company with 50 to 5,000 employees.

The exposed audience was sorted by intent score and refined to include individuals whose account scores were over 70. A score above 70 is an organisation-wide standard users have focus-grouped to identify the most relevant in-market accounts. This addresses any potential data quality issues and focuses on the most relevant audience members based on their buying propensity. Prioritising by score is not relevant to the control audience as they were not identified based on intent signals, and therefore do not have an associated score. In practice, building audience targeting on standard B2B ad platforms (LinkedIn or Google, for example) does not provide scores on individuals or accounts sourced by firmographic details.

Audiences were made the same size based on the unique count of last names per list. Simple randomization was used during this process for the control audience by randomising the entire list of individuals on the control list and extracting the same number of individuals as the exposed audience. The use of a unique last name count accounts for having up to 5 emails per contact to aid in match rates.

2.4 Advertising

All three tests were run on the Google Display Network with identical setups. Each test aimed to drive individuals to a piece of marketing content. A campaign was created with two ad groups in it. Ad groups shared a campaign-wide budget and were enabled simultaneously. The only difference between each ad group was the targeting settings. The control ad group exclusively targeted the control audience, whereas the intent ad group targeted the exposed audience.

Both audiences used Google's "Targeting" setting rather than "Observation" so as to not reach individuals outside of each ad group's respective audience. Google audience optimization was off for both ad groups, and no other targeting settings were in place. This allowed us to measure the effectiveness of each group based on intent versus control audience rather than Google's optimizations or differences in ad creative or landing page.

Ad assets matched associated brand colours and styles with three image assets per test. Assets were 250x250, 120x600, and 300x250 pixel image ads (see appendix 1). Campaigns used a "Maximise Conversion" bid strategy and set the goal to drive website traffic. Campaigns were budgeted at approximately $66 per day and ran for nine days.

2.5 Compliance

The data used in this experiment was collected and used compliantly in accordance with data protection regulations. Audiences were strictly business/company-based and B2B in nature. They do not include contact-level location, phone number, or any sensitive information. Data has been obtained via licensing from partners or from public data processing. When licensing data from third-party partners, partners have either obtained the information from public sources or collected consent from their trusted partners. In the case of public data processing, we collect data across multiple publicly available sources that can be manually accessed without any authorization or login required. We follow all necessary terms and conditions of third-party sites to process and share the data. We only process B2B and professional information, and the user would reasonably expect to use it for that purpose (e.g., Head of Research at a company announcing a strategic initiative in a press release, the VP of Marketing asking a question about a product on a public social network, a Director of Sales hiring for new sales folks, etc.). We also leverage APIs whenever available and adhere to strict crawling guidelines (rate limits, robots.txt) to collect and process the data.

The personal data processed and used is standard, publicly available business card contact information. None of the information processed poses significant risks to individuals' rights and freedoms or perpetuates any social stereotypes or segregation. All the information is stored securely and is only accessible via password-protected web applications. We maintain and check against an updated list of individuals who have requested their information to be deleted. Any individual can request confirmation, access, or erasure of their personal data by emailing datarequest@idg.com. We also work with downstream data partners to maintain a blacklist of individuals who have opted out or have explicitly mentioned not to be processed.

All audiences were matched and reached via Google Ads Customer Match, which uses hashing and restricted data processing to serve ads in a way that ensures compliance based on certain unique identifiers and their use in ad delivery, reporting, and measurement. Any reporting on campaign activity or performance is anonymized and does not include the actions of any specific individual. For example, it is possible to see how many clicks an ad received but not which individuals clicked the ad.

3. Results

3.1 Impressions

Google defines an impression as any time an ad appears on the search page or other Google Network pages. Our research finds that the ad groups reaching the exposed (intent-based) audience resulted in 83.5% more impressions on average than the control audience (see **Figure 1**). This effect remained true and repeatable across all three independent tests.

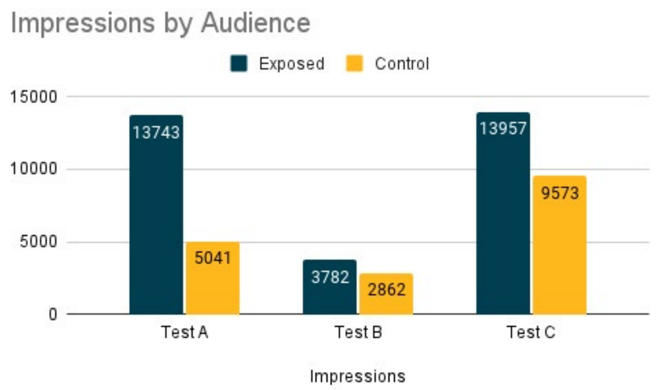


Figure 1: Comparison of impressions by audience for each test.

3.2 Click-Through-Rate

Google defines Click-Through Rate (CTR) as a ratio measuring how many viewers clicked an ad out of the total viewers. A high CTR suggests that the ads shown are relevant to their audience. This experiment showed an average CTR increase of 220% on ads shown to intent-based audiences in all three tests (see **Figure 2**).

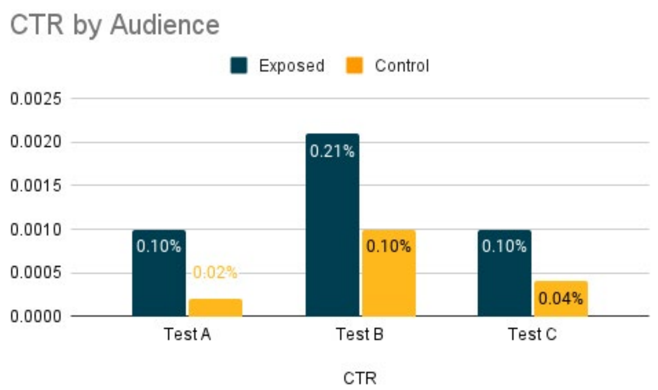


Figure 2: Comparison of click-through rate by audience for each test

Increased CTR suggests that individuals showing intent across the sources captured have a higher likelihood of clicking an ad compared to those not actively showing the same types of intent signals.

3.3 Average Cost-Per-Click

Cost-Per-Click (CPC) is the amount paid for each click on an ad, as defined by Google. Average CPC is calculated based on the total cost of clicks by the number of clicks. We assessed CPC based on the Average CPC associated with each separate ad group. Targeting, ad assets, landing page, and budget impact CPC. It is important to note that all factors other than targeting were identical for these tests. Based on our research, intent-based audiences saw a 59.6% lower CPC than the control audience based on a 95% confidence score (p=0.02) (see **Figure 3**). Even though the total cost of both ad groups in each test was similar, CPC is significantly lower due to the higher likelihood of that audience clicking an ad. This improved efficiency is reflected across all three tests and coincides with the elevated CTR in the exposed audiences.

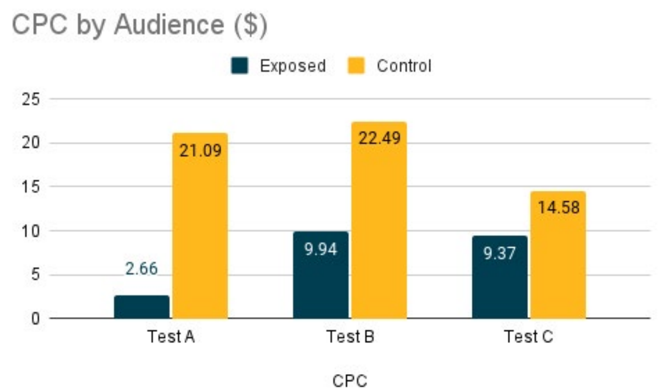


Figure 3: Comparison of cost-per-click by audience for each test

This research suggests that moving away from traditional targeting methods based only on ICP definitions and shifting to multi-source intent targeting may result in the ability to drastically improve a given marketing campaign's effectiveness while reducing cost.

3.4. Impact of Intent-Based Targeting

With Gartner reporting a drop in marketing budget as a percentage of total revenue from 11% in 2020 to 6.4% in 2021, the lowest ratio in recent years, marketers must find novel ways to do more with less (Blum, 2021). The improved efficiency seen in this research enables marketers to strategically spend each dollar so that no amount of budget is wasted to reach individuals who are not showing buying propensity. For example, based on these results, with the same budget using an intent-based audience, B2B marketers could receive 2.5x the number of clicks than if they were using standard firmographic and demographic targeting.

To validate that data is statistically sound, means and standard deviations for each result metric by audience group as well as totals for both (see **Figure 4**). This method was chosen over a more typical T-Test due to data being on three groups, limiting the relevance of a test of that nature.

	Impressions	CTR	CPC
Control Mean	5825	0.05%	$19.39
Exposed Mean	10494	0.14%	$7.32
Control Standard Deviation	3423.56	0.04%	$4.22
Exposed Standard Deviation	5813.75	0.06%	$4.05
T otal Mean	8159.67	0.10%	$13.36
Total Standard Deviation	4974.65	0.07%	$7.57

Figure 4: Means and standard deviations by ad metric and audience

4. Limitations

For evaluation purposes, we used Google Display Ads engagements as a proxy for showing that individuals are in an active buying stage based on the relative efficiency of intent-based targeting. Tests were performed on three B2B technology companies that target a similar ICP but operate in different verticals. These results may vary if we were testing marketing for dental offices or commercial radiation machine manufacturers, for example.

Not all intent can be viewed, captured, or included. We did not have access to search engine data or specific non-digital engagements over the phone, email, or in-person conversations. Depending on other non-tech industries, these engagements may affect purchasing decisions.

A third limitation is the audience size. This test ran with relatively conservative audience sizes resulting in relatively short campaign run times. Since audiences in this test were static, ad placement frequency will become too high over time, and the test will see diminishing returns. In practice, this could be mitigated with regular audience updates. In this test, it meant that the scale of each campaign was limited.

References

1. Antin, A. and Berkowitz, T. (2020, February 11). *Emerging Technology Analysis: Leveraging Intent Data for Marketing and Demand Generation.* Gartner Research.
2. Crane, D. (2021, May 5). *New Survey Report: The B2B Marketer's State of Intent Data.*
3. Robinson, J. (2022, July 19). What Is An Ideal Customer Profile (ICP) And How Do You Create One? *Crunchbase*
4. *Best practices for AD targeting: Linkedin Ad tips.* Best Practices for Ad Targeting. (n.d.). Retrieved December 8, 2022
5. *How to Calculate Z-Score and Its Meaning.* (2022b, December 21). Investopedia.
6. Blum, K. (2021, July 20). Marketing Budgets Drop to Lowest Levels in Recent History. *Gartner.*

Authors

Olivia Kenney is a Director of Marketing at Foundry, focused on Product Strategy and Innovation. Excited by the connection between data and creativity, she graduated from Saint Mary's University with a Bachelor of Commerce in Marketing and Computing & Information Systems. Olivia led the marketing team at LeadSift, an intent data company in Halifax, NS, up until its 2021 acquisition by IDG Communications (now Foundry). Today, she works to drive growth by innovating and optimizing across channels, connecting people with technology through experiential marketing strategies.

okenney@foundryco.com

Sreejata Chatterjee was Co-founder and CPO at LeadSift prior to being acquired by Foundry in 2021. Now she leads Product at Foundry, focused on intent data. She graduated from Dalhousie University with a degree in Computer Science and has over 15 years of experience building products loved by customer success, engineering, sales, and marketing teams, but most importantly, customers.

Outside of work, she's a guest lecturer at Saint Mary's University, co-hosts a Data Science meetup group, and sits on the board of a non-profit, but you're most likely to find her chilling over wine and food after a stroll that she'll insist was a hike!

sreejata.chatterjee@foundryco.com

Kundan Kumar is currently a Tech Lead at Foundry, helping to build an innovative intent solution for the MarTech industry. He completed his Master's in Computer Science from Dalhousie University and has over 12 years of experience in the software research and development industry. In the past, he has been Director of Data Science at LeadSift and Senior-Engineer at Samsung R&D, where he worked extensively on developing scalable data pipelines, text analytics, machine learning, and algorithm design and development.

Apart from work, he loves exploring and learning AI and machine learning advancements. He loves to be part of hackathons, play table tennis, and travel.

kundan.kumar@foundryco.com

Appendix

Example of ad assets used in testing.

Multi-Touch Attribution for Consumer Journey: Snapchat's Learnings

Dinesh Krithivasan
Kantar Analytics Practice, India

Satya Sarvani Malladi
Kantar Analytics Practice, India

Seema Dhamija
Kantar Analytics Practice, USA

Jeremy Jones
Kantar Analytics Practice, USA

Takeshi Tawarada
Snap Inc., USA

Classifications, Key Words:

- Machine learning
- Recurrent neural networks
- Campaign optimization
- Brand outcomes

Abstract

The consumer path-to-purchase is a complex topic to track and analyze, as a growing number of advertising touchpoints can be accessed in an infinite number of combinations. This is why performance evaluation of digital media has often defaulted to conversion attribution on a last-touch basis. In other words, the last advertising touchpoints receive all of the credit for conversion and no credit for any of the other touchpoints preceding it. Additionally, with the rise of iOS 14 and ATT, reduced signals have led to an even diminished view into conversion attribution. But this doesn't stop the desire for marketers to evaluate and optimize their campaign spend. In this paper, we present a novel utilization of recurrent neural networks to evaluate contributions to brand lift performance of campaigns run across the Snapchat ad stack that can help marketers understand specific formats and tactics that drive performance and to improve outcomes.

1. Introduction

Digital advertising is a massive online industry, with $500 billion spent on online advertising in 2021 (Statista, 2022). With the ability to serve advertising tailored to individual needs and track them precisely, measuring ad effectiveness becomes of paramount importance. Advertisers want to know how effective their current ad campaigns are, how best to tailor them to any given customer segment, and how they can increase campaign effectiveness on-the-fly. These questions can be answered by building an attribution model that can tease out the relationship between the exposures a user received and their eventual conversion behaviour. Depending on the context of the study, conversion behaviours can range from a positive brand awareness/perception in the upper funnel of a consumer journey to a web visit/online purchase at the lower funnel of the consumer journey, with a range of possibilities in between. Attribution is, in general, a very difficult problem to solve because any sample of users will be very heterogeneous with respect to their pre-existing brand perceptions, purchase behaviours, timing, cadence of exposures being served to them, etc. A common simplifying assumption that is often used in this space is to attribute all credit to the first or last exposure that a user saw prior to their conversion event (Berman, 2018). Such simplifications fail to capture the inherently temporal nature of digital advertising, where a series of exposures served at the right cadence can incrementally accumulate in impact until their combined effect eventually tips the user over into conversion.

This article presents a solution to the multi-touch attribution problem that uses recurrent neural networks (RNN), a deep learning architecture specifically developed to understand temporal sequences.

Our approach allows us to establish a quantitative relationship between the exposure sequence seen by a given user (including the number of exposures, attributes of each of the exposures, and the time elapsed between successive exposures) and the conversion behaviour of that user while controlling for other factors such as demographics and prior brand engagement. The insights gleaned from our model can then be used to pinpoint whom to serve (audience selection), how much to serve them (frequency capping) and what to serve them (ad attribute optimization). We also present a formal optimization framework that can prescribe budget reallocations that move advertising dollars towards optimal combinations of users and ad attributes subject to constraints dictated by business considerations.

While the techniques described here can be applied equally well on any conversion metric, we focus on attribution for consumer journey brand metrics in this paper. To make things concrete, we will assume that the measured brand metrics are Ad Awareness and Intent, i.e., how likely the respondent is to be aware of the ad in question and how intent they are on purchasing it. The results in this paper are for campaigns run on Snapchat.

2. The Data

For a given advertising campaign, we build our attribution models at the user level with two pieces of information for every user:

1. Granular information about all the digital exposures from the campaign under consideration seen by any given user. This will include precise information about what exposures the user saw (exposure attributes such as video vs. image ad etc.) and when they saw it (timestamp).

2. Demographics and attitudinal information, including the user's perception of the brand under consideration and other consumer journey metrics. This information is obtained from the demographic data captured by Snapchat.

Snapchat offers advertisers a native brand lift measurement solution powered by Snapchat's in-app polling tool. Brand lift studies compare the brand perceptions of these groups of survey respondents. Snapchatters in targeted audiences are randomly split into one of these two groups before a campaign goes live. Brand survey responses are gathered from the exposed group based on their exposure to the advertising campaign and from the control group based on their opportunity to see these advertisements. The differences between the two groups can be attributed to Snapchat media exposure, as all other factors are held constant. For the study, 52 total campaigns were ingested, including over 50,000 respondents and 630,000 Snapchat advertising exposures across the US, KSA, France, and Australia. We ingested event-level media delivery data and user-level brand lift survey responses, to understand how variables like ad format, frequency, and creative features contributed to lift.

3. Methodology

We assume that every user starts with some predisposition towards the brand being advertised. In other words, each user has a certain baseline probability of answering the survey question about Ad Awareness or Intent in the affirmative. This baseline probability $p_0[i]$ for user i will naturally depend on the demographics $\vec{Z_i}$ and behavioral attitudes $\vec{Y_i}$ of that user. This probability will evolve with each digital exposure $\overrightarrow{X_{ik}}, k = 1,2, \ldots, N_i$ that the respondent sees until eventually, after N_i exposures, the probability changes to $p_{N_i}[i]$.

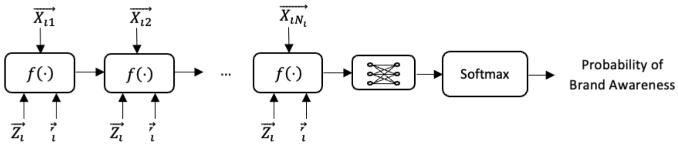
Note that each exposure $\overrightarrow{X_{\iota k}}$ is itself parameterized by a series of ad attributes (format of the ad, site in which it was served, time since previous exposure etc.).

The attribution problem can be cast as learning a function $f(\cdot)$ that takes as inputs the demographics $\overrightarrow{Z_\iota}$, behavioral attitudes $\overrightarrow{Y_\iota}$, and the exposure sequence $\overrightarrow{X_\iota} = \left(\overrightarrow{X_{\iota 1}}, \overrightarrow{X_{\iota 2}}, \dots, \overrightarrow{X_{\iota N_\iota}}\right)$ of respondent i and produces $p_k[i]$, the probability of that respondent answering the Ad Awareness question in the affirmative after k exposures for $k = 0,1,2,\dots,N_i$.

Once we have learned this function $f(\cdot)$, the baseline probability of any user can be written as $p_0[i] = f(\overrightarrow{Z_\iota}, \overrightarrow{Y_\iota}, \vec{0})$ and the final probability as $p_{N_i}[i] = f(\overrightarrow{Z_\iota}, \overrightarrow{Y_\iota}, \overrightarrow{X_\iota})$. This formulation makes clear the role of control users in our approach – the learning algorithm will rely on them to learn the baseline probability $f(\overrightarrow{Z_\iota}, \overrightarrow{Y_\iota}, \vec{0})$ for all users. It would then rely on the test users to learn the impact of the exposure sequence $\overrightarrow{X_\iota}$ in producing a lift $l_i = p_N[i] - p_0[i] = f\left(\overrightarrow{Z_\iota}, \overrightarrow{Y_\iota}, \overrightarrow{X_\iota}\right) - f\left(\overrightarrow{Z_\iota}, \overrightarrow{Y_\iota}, \vec{0}\right)$ over this baseline probability. Note that this is just a conceptual way of understanding the role of test and control users – in reality, we train the model on all users simultaneously. This training methodology is described in the following section.

3.1 Recurrent Neural Networks

We use a deep learning architecture called Recurrent Neural Networks (RNN) to learn the function $f(\cdot)$. RNNs are a powerful class of algorithms that are especially adept at handling sequential inputs. The specific architecture we use, LSTMs (Long Short-Term Memory), have proved to be adept at learning both short-term and long-term temporal patterns in input sequences (Hochreiter & Schmidhuber, 1997). Our attribution solution leverages this unique ability of LSTMs to move past simplifying assumptions like first/last touch attribution and instead assign some importance to all the exposures seen by the user. The model arrives at this attribution in a purely data-driven way by learning any patterns present in the exposure sequences.



3.2 Insights from Model

Once we have trained the LSTM model on our dataset of control and test users, we can use the model to predict a given user's response probability at the end of each of their exposures. This gives us the ability to interpolate the evolution of that user's probability of Ad Awareness with each exposure that they see.

For respondent i, we can create the sequence of predictions $p_k[i] = f\left(\overrightarrow{Z_\iota}, \overrightarrow{Y_\iota}, \overrightarrow{X_\iota}^{(k)}\right), k = 0,1,..,N_i$ where $\overrightarrow{X_\iota}^{(k)} = \left(\overrightarrow{X_{\iota 1}}, \overrightarrow{X_{\iota 2}}, \dots, \overrightarrow{X_{\iota k}}\right)$ are the first k exposures seen by the respondent. We call this sequence of predictions $\overrightarrow{p_\iota} = (p_0[i], p_1[i], \dots, p_{N_i}[i])$ the "probability curve" for respondent i.

These probability curves encode all the information the model has learned about how demographics, user attitudes, and ad attributes affect brand awareness. Appropriate manipulation of these probability curves enables us to quantitatively attribute a share to each of these variables in driving a user's overall Ad Awareness behaviour. For instance, to measure how users of different age groups respond to the campaign, we can create average probability curves.

$$\vec{p}_{<18} = \frac{1}{N_{<18}} \sum_{i \in <18} \overrightarrow{p_\iota} \quad , \quad \vec{p}_{18-20} = \frac{1}{N_{18-20}} \sum_{i \in 18-20} \overrightarrow{p_\iota} \quad \&$$

$$\vec{p}_{>21} = \frac{1}{N_{>21}} \sum_{i \in >21} \overrightarrow{p_\iota}$$

Level differences between these probability curves are indicative of a baseline difference in brand perception across different age groups while, slope differences indicate varying degrees of response to the campaign by different age groups. **Figure 1** shows the probability curves for different age groups for campaigns run on Snapchat and the overall probability curve across all users.

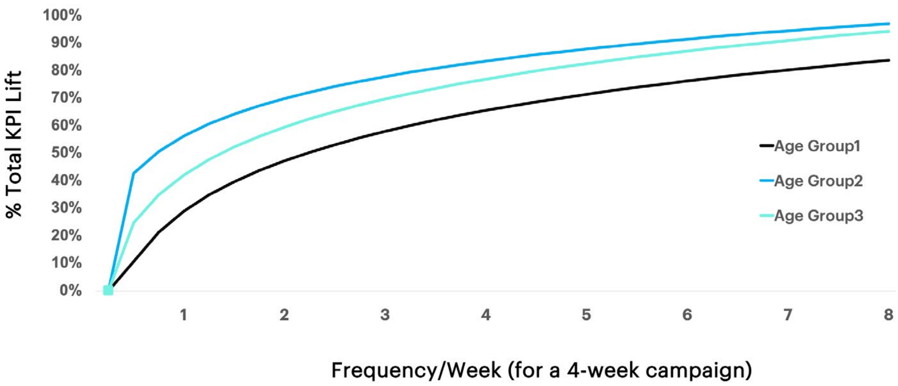


Figure 1. Effectiveness of different Age Groups at lifting Brand Awareness by Frequency Level

While we do see an increase in KPI lift with every exposure we also see that age group 3 starts saturating around 3-4 exposures per week. The use of our models helps in identifying over-exposed or under-exposed groups and optimising accordingly and frequency caps are built into our optimization systems.

These probability curves can be used to derive the overall lift in probability of Ad Awareness for a typical exposure for a typical user in the following way:

$$l_{<18} = \frac{\sum_{i\in<18} \quad (p_{N_i}[i] - p_0[i])}{\sum_{i\,\in<18} \quad N_i}, \; l_{18-20} = \frac{\sum_{i\in18-20} \quad (p_{N_i}[i] - p_0[i])}{\sum_{i\,\in18-20} \quad N_i}, l_{>21} = \frac{\sum_{i\in>21} \quad (p_{N_i}[i] - p_0[i])}{\sum_{i\,\in>21} \quad N_i}$$

Of course, age group is just an example here and the same methodology can be used to derive lifts and probability curves for any demographic/attitudinal variables.

Deriving lift numbers for ad attributes is slightly more involved since they are not constant across a given probability curve $\vec{p_i}$ (user i might see ads of different formats, say format1 and format2, for example).

However, since the probability curves are available at the exposure level across users, lifts can still be computed for a given ad attribute (for example, the creative format that takes two values – format 1 and format 2) by measuring the lifts induced by that ad attribute across all incidences of that ad attribute in the data. Using the notation we have introduced above, these lifts can be written as

$$l_{format1} = \frac{\sum_{X_{ik}\in format1} \quad (p_k[i] - p_{k-1}[i])}{\sum_{X_{ik}\in format1} \quad 1} \quad \& \quad l_{format2} = \frac{\sum_{X_{ik}\in format2} \quad (p_k[i] - p_{k-1}[i])}{\sum_{X_{ik}\in format} \quad 1}$$

Of course, creative format is just an example here, and the same methodology can be used to derive lifts for any ad attribute variable. These per-exposure lift numbers can be contrasted with per-exposure cost numbers (CPM) to understand their cost-effectiveness.

The training time for the model scales linearly with the number of observations, i.e., with the number of test and control users used in modelling. Even for problem sizes as large as several tens of thousands of users, the algorithm can still be efficiently implemented using modest computing resources and reasonable run times.

It must be mentioned here that there are other flavours of the attribution problem where the scale of the data can be considerably larger. For instance, a "web visit attribution problem" can track digital advertising and associated website visits and aim to attribute the impact of digital exposures in leading to a website visit. The same concepts we have discussed thus far (using RNNs to model digital exposures as a time sequence) can still be applied to such large problems but with some additional modifications to allow for the massive data size. In such cases, we leverage specialised cloud-based computing infrastructure for both data handling and model training.

4.Optimization in MTA

After building an LSTM model to apportion the credit of conversion to all the observed touch points in every customer's journey, the marketer will be able to draw broad insights on the relative performance of ads or levels of an ad attribute (e.g., publisher). To extract actionable insights (e.g., budget outlay to a publisher or frequency capping on a publisher's site), the marketer must be equipped with an optimizer created to leverage the capabilities of the LSTM prediction engine and determine the best course of action for the marketer. The granularity of the optimised decisions can vary depending on the scope of the marketer's actionability.

In a campaign, selected users are served exposures based on current business wisdom. Every exposure is characterised by attributes that capture:

1) user demographic information,

2) ad creative and delivery information, and

3) user information that is exposure-dependent.

Ad creative features and ad delivery choices are controllable levers that let the decision-maker leverage learnings from a given campaign to make optimised decisions in similar settings later.

We formulate the budget allocation problem by various dimensions as a constrained knapsack problem. We present the mathematical formulation below:

$$\max \frac{\left(\sum_{i \in \cup_{k \in \{1,\ldots,K\}} I_k} \quad l_i z_i\right)}{K} \tag{1}$$

subject to

$$\sum_{i \in I_k} z_i \leq N_{current} \ \ \forall\, k \in \{1, \ldots, K\}, \tag{2}$$

$$\sum_{i \in I_k} c_i z_i = B_{current} \ \ \forall\, k \in \{1, \ldots, K\}, \tag{3}$$

$$m_i - \delta_i \leq z_i \leq m_i + \Delta_i \quad \forall\, i \in I_k \ \forall\, k \in K, \tag{4}$$

$$z_i \ \in Z_0^+ \ \ \forall\, i \in I_k \ \forall\, k \in K. \tag{5}$$

The notation used in the formulation above is described in **Table 1**. The objective function (1) is the average total lift summed over all levels of all attributes of ads served to the users of interest. Constraint (2) ensures that the total number of exposures does not exceed the current total number of exposures while constraint (3) fixes the total media spend to the current budget. You will note that these summations work because the parameters c_i and l_i are additive for a particular ad – for example, the total lift of a format1 Ad of type Creative Theme 25, served on Platform A would have its total lift computed as $l_{format1} + l_{Theme25} + l_{Platform\ A}$. We impose guardrails on the optimised allocations of exposures to various levels of attributes in constraint (4). Constraint (5) requires the exposure allocations to be integer quantities; however, this constraint is often relaxed to allow for faster optimization without significantly compromising on objective quality.

The optimizer described above returns the exposure allocations by levels of attributes as guidance for the marketer to action on, when making future outlays of budget across attributes.

Notation	Description
K	Set of actionable ad attributes
I_k	Set of levels with attribute k
c_i	Cost of an exposure containing level i of attribute k
l_i	Aggregated lift per exposure containing level i of attribute k
$N_{current}$	Current number of exposures
m_i	Current number of exposures for level i of attribute k

Notation	Description
δ_i, Δ_i	Lower and upper guardrail offsets from m, at level i of attribute k
$B_{current}$	Total current spend
z_i	**Decision variable** capturing the number of exposures with level i of attribute k

Table 1. Notation of the knapsack formulation

4.1 Insights from Snapchat

While we evaluated vertical and region-specific insights to develop a foundation for learning agenda frameworks, we identified four patterns across the outputs that may help guide all marketers in activating their first branding campaign on Snapchat.

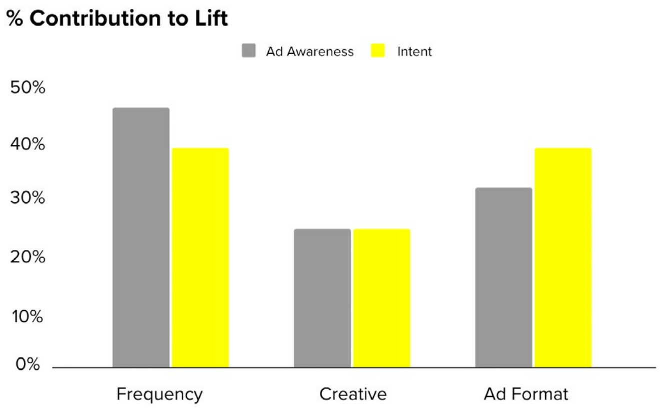


Figure 2. Contributors to success*

Firstly, the performance of a campaign is not solely dependent on a single factor, like the creative or frequency alone. In fact, they work together in varying ways, and we found that contributions can differ depending on the metrics we looked at.

For example, we found that frequency was a more significant contributor to Ad Awareness lift, while ad format was found to be more important to Intent lift. This would mean that frequency of exposure can help raise awareness metrics, while the ways in which Snapchatters can engage with your brand across the app is more meaningful for driving action.

The best way to plan a Snapchat campaign starts with leveraging multiple ad formats on Snapchat, which not only ensures that you are reaching all types of Snapchatters but provides increased performance over a single format alone. Furthermore, leveraging both content and camera ads provides the most benefit.

To take this further, our second finding showcased how higher budget levels benefited from increased allocation of resources into camera-based formats, including Lenses and Filters, to maximise lifts in Intent. As target audiences become saturated with Content Ads, which include Commercials, Snap Ads, and Story Ads, the reallocation of budget towards Camera Ads helps with incremental reach and building differentiated frequency.

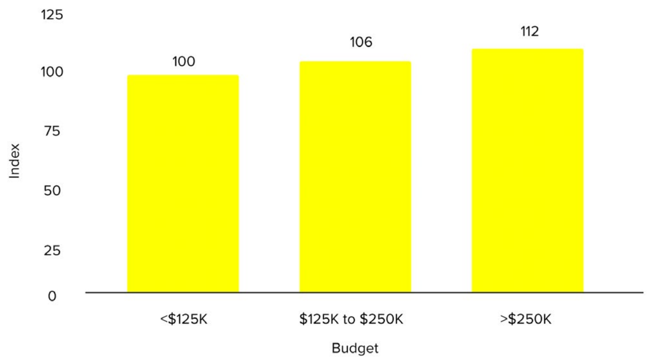


Figure 3. Camera ads budget allocation index*

When customers see a brand from different perspectives and get to interact with it in those contexts, not just on Snapchat, but across the brand's entire marketing plan, a relationship is built that ultimately drives them towards purchase. With the utilisation of multiple ad formats providing incremental reach between each other, leveraging a wider library of creatives to power those formats creates synergies in building differentiated frequency. In other words, an exposure that delivers the same message as other exposures can still be effective if it does so in a different way compared to other creatives in the mix. This allows us to emulate how an audience gets exposed to a brand in the wild, while still staying within the Snapchat platform.

# of Exposed Creatives	Purchase Intent Index
1 - 5	100
6 - 10	172
11+	263

Table 2. Creative diversity index*

*Source: Kantar Balance Attribution study commissioned by Snap inc Q4 2021

And finally, in knowing that additional formats and creatives can help drive action amongst Snapchatters, it's important to know that thinking about the platform as a single marketing channel can leave performance on the table. On average, across the studies included in this campaign, we found that a frequency of 4x/week drove about three-quarters of total possible Intent lift, with the model predicting total saturation at 8x/week.

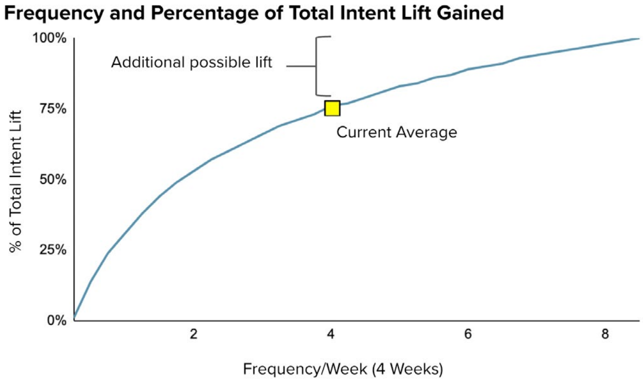


Figure 4. Lift by Frequency*

7. Conclusion

In this paper, we have presented a powerful framework that leverages a sophisticated deep-learning architecture to arrive at an attribution solution. Our technique does not make any simplifying assumptions such as first/last touch attribution and is completely data-driven. We have also demonstrated how our model's insights can be used to optimise budget allocations to maximise overall campaign effectiveness. Snap Inc. has leveraged these methods in the real world by deriving value from the model insights and optimizer prescriptions to create a starting point for clients in approaching campaign planning. While this paper focuses on the crux of our methodology, our modelling and optimization approaches are flexible enough to admit many variations and extensions such as accounting for other channels beyond digital exposures (TV, print media, etc.), modelling other outcome metrics of interest (sales, web visits, etc.), and attribution modelling at scale (billions of digital exposures across millions of devices). While the current solution focuses on advertising formats and channels within the Snapchat platform and it focuses only on digital exposures, the approach is readily extendable to other venues such as video platforms, programmatic, web apps, and social media.

References

1. Statista. (2023, January 6). *Digital advertising spending worldwide 2021-2026.*

2. Berman, R. (2018). Beyond the Last Touch: Attribution in Online Advertising. *SSRN Electronic Journal.*

3. Hochreiter, S., & Schmidhuber, J. (1997). *Long Short-Term Memory. Neural Computation, 9*(8), 1735–1780.

*Source: Kantar Balance Attribution study commissioned by Snap inc Q4 2021

Authors

Dinesh Krithivasan is a Partner at the Data Science & Innovation team at Kantar Analytics where he has been leading several R&D initiatives on developing state-of-the-art solutions powered by advanced data science techniques. He has extensive experience in using deep learning and machine learning algorithms across a variety of problems such as computer vision, time-series modeling, text analysis and ROI modeling. He has directed studies for multiple clients across different geographies and industries, addressing issues such as segmentation and targeting, customer satisfaction and loyalty, advertising effectiveness, brand performance, and brand strategy. Dinesh received his Ph.D. from the University of Michigan, Ann Arbor.

dinesh.krithivasan@kantar.com

Satya Sarvani Malladi is a Senior Director of the Data Science & Innovation team at Kantar Analytics. During her time at Kantar, Sarvani has played a key role in developing several efforts into full-fledged solutions that rely on advanced optimization techniques and data science engines. Equipped with expertise in Operations Research and Analytics, she has innovated solutions geared towards actionability for marketing effectiveness via multi-touch attribution, marketing return on investment, and global investment allocation. Sarvani holds a PhD from Georgia Institute of Technology, USA and a bachelor's degree from IIT Madras, India. She worked at the Technical University of Denmark as a postdoc before taking up the current role.

satya.malladi@kantar.com

Seema Dhamija holds a Ph.D. in Physics. She is an effective hands-on leader and innovator with extensive experience at the intersection of data and analytic modeling. She has also worked in multidisciplinary teams in the field of High Energy Physics and organizations like the CERN Facility (Europe), Thomas Jefferson National Accelerator Facility (Newport News, VA), Brookhaven National Lab (Upton, NY), to name a few. She left academia's world for a career in industry, where she brought AI-driven innovation in healthcare and retail. She uses Artificial Intelligence/Machine Learning/Deep Learning and advanced statistical techniques to solve a wide variety of real-life problems. Seema joined Kantar in 2021 and works on Multi-Touch Attribution projects. Her most recent experience is on the Walmart Data Science team, with a prior stint at Nielsen.

seema.dhamija@kantar.com

Jeremy Jones brings vast experience in leveraging AI to modernize Marketing Mix Models, Multi-touch Attribution and Unified Media Measurement. He is focused on the intersection between the problems that marketers face every day and cutting-edge technology. Currently he is a VP, NA Head of Media Performance Analytics at Kantar and the solution lead for Kantar's Balanced Attribution. Balanced Attribution is an AI enabled Full-Funnel Multi-Touch Attribution platform that provides marketers tools to understand the why behind the ROI.

Jeremy.Jones@kantar.com

Takeshi Tawarada currently leads Augmented Reality ads effectiveness research at Snap Inc. and was previously a Marketing Science partner for Entertainment, Tech, Social Media, Health, Energy, and Retail brands. He has a wide breadth of analytics, data science, and story-telling experiences within the agency and social media worlds, which gives him a unique perspective on the types of problems that come from both the buy- and sell-side of digital marketing. He received a B.S. degree in biopsychology, cognition, and neuroscience from the University of Michigan, which he utilizes to approach these problems in a human-centric way and seeks novel and innovative ways to solve them.

ttawarada@snapchat.com

Attribution and the Marketing Mix Model

Peter Cain
Marketscience, UK

Nitesh Sahay
Marketscience, UK

Classifications, Key Words:

- Endogeneity
- MMM
- MTA
- Consumer Journey
- Unobserved Component Models
- VAR
- VECM
- Cointegration

Abstract

Inconsistent user-identifiers and the walled garden policies of dominant social media players, together with the (imminent) abolition of third-party cookies has led to renewed interest in the marketing mix model as an attribution tool. However, to be useful in a post-MTA world any 'next generation' MMM framework needs to deliver on three fundamental business issues. Firstly, to serve as a true attribution solution, MMM needs to focus on causal estimation methods. Too often we see reliance on consumer journey solutions to address the problems of last-touch-attribution. However, these ignore the critical issues of selection bias endemic in much online media – leading to endogeneity bias and misallocation of the marketing mix. The growing popularity of automated machine learning approaches to the mix model only serve to exacerbate this problem, where the focus is on prediction not causation.

Secondly, MMM needs to quantify the long-term (base-building) effects of marketing and so inform brand-building strategy. Standard approaches are simply not set up to measure these effects, with fixed baselines and a focus on short to medium-term lag structures or Adstocks. Alternative time series structures are required that can quantify both short and long-term (base) variation – coupled with dynamic network models that can explain the causes of base variation and the economics of brand-building.

Finally, next-generation MMM needs to fill the gap left in a cookie-less world to deliver granular and swift insights on marketing ROI and optimal budget allocation. Suitably identified high-dimension mix models – across consumer cohorts by day or hour – can fit the bill. This can provide many of the claimed benefits of MTA such as granular online media effectiveness ranking by publisher and placement, together with the ability to quantify the impact of pricing, offline media, economic factors and longer-term brand-building.

1. Introduction

Marketing attribution attempts to quantify the incremental impact of each element of the marketing mix on consumer demand, typically in the form of a purchase conversion. With the advent of multi-channel marketing and the growing proliferation of off and online channels, accurate attribution represents a significant challenge. In response, modern analytical approaches have evolved into three broad strands.

Firstly, we have Multi-touch attribution (MTA). This focuses on the contribution of online touchpoints to a binary online conversion outcome such as 'buy or not buy', across a network of observed consumer journeys. Measurement is carried out at individual (cookie) level using parametric discrete choice modelling approaches such as Logit regression, or non-parametric machine learning algorithms such as random forests or neural networks. Outputs allow the marketer to assign credit to each element of the digital mix, addressing tactical questions such as how, where and when to spend the allocated budget and which publishers.

Secondly, we have nested (system) approaches to the marketing mix model (MMM). Here, consumer journey theories attempt to provide a complete explanation of the final purchase decision with pricing, paid, owned and earned media working together to drive demand. Measurement is typically carried out at an aggregated level, using least squares econometric methods applied to groups of consumers at store, chain, regional or market level for example. Outputs are used to quantify ROI, advise on optimal budget allocation across off and online channels and produce sales forecasts.

Finally, there are solutions that attempt to unify MTA and MMM into one framework. Treated separately, each approach is often seen as a competing attribution solution with conflicting outputs and recommendations: MTA is too narrow a representation of consumer demand, with no control for pricing, offline media and economic factors, whereas MMM is too broad to address the granular aspects of online marketing with little capacity for 'in-flight' measurement and 'real-time' optimisation. This dichotomy has led to attempts to unify the two approaches (Nail, 2015, MMA, 2021), where MMM deals with the wider macro view of consumer demand across off and online touchpoints, leaving MTA to focus on the narrower micro online view.

Notwithstanding the often-contentious issue of how best to combine the two, the very nature of customer-level MTA analysis is increasingly problematic. Firstly, the impact of traditional market-level media is notoriously difficult to measure at an individual consumer level. Secondly, the growing use of multiple online platforms has led to an increasing inability to obtain consistent user-identifiers. Thirdly, the walled-garden policies of dominant social media players, such as Facebook, have now made such identifiers unavailable altogether. Finally, and perhaps most crucially, GDPR and the ending of support for identifiers stored in third-party cookies will further impede the ability to connect on-site transactions with third-party ad placements.

In light of these issues, the focus is shifting in favour of the marketing mix model, where the aggregated nature of the data inputs can facilitate the estimation of market-level factors and capture the (de-identified) sum of individual actions across platforms. However, to constitute a valid attribution framework, any next-generation MMM solution needs to address three fundamental criteria: causal inference, short and long-term measurement, and granular 'real-time' insights. In this article, we explore these topics in detail, paving the way for more credible and actionable MMM approaches.

2. MMM and causal attribution

Marketing attribution and budget allocation rely on accurate causal attribution to each element of the marketing mix. Modern MMM attempts to inform this process via path-to-purchase theories of demand, where paid, owned, and earned media work together to drive sales. For example, a common hypothesis contends that marketing investments stimulate a journey that starts with natural online search, continues

[1] It is sometimes argued that experimental designs can handle walled garden data gaps. Even if this were possible, the problems of multiple cross-platform usage still remain.

[2] Although, note the potential for Clean Rooms (Forbes, 2021), Self-Attributing Networks (Apple) and Attribution API (Google), where some (narrowly-focused) attribution data would be available.

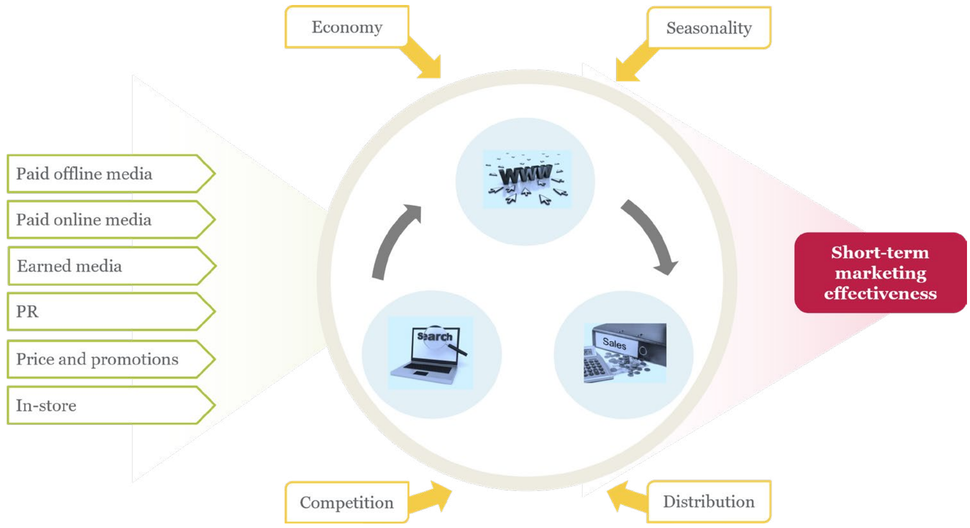


Figure 1. Network model of short-term sales

through to website research, and finally onto online and offline product purchase. This can be depicted as a Directed Acyclic Graph (DAG), illustrated in **Figure 1** and set out in equations **(1)-(3)**.

$$LnS_{it} = \dot{\alpha}_{it} + \dot{T}_{it} + \dot{\sigma}_{it} + \sum_{j=1}^{n} \sum_{k=1}^{n} \dot{\beta}_{ijk} \ln X_{ijkt} + \sum_{k=1}^{K} \dot{\delta}_{ik} D_{kt} + \dot{\gamma}_{is} lnW_{st} + \dot{\varepsilon}_{it} \tag{1}$$

$$LnW_{st} = \ddot{\alpha}_{st} + \ddot{T}_{st} + \ddot{\sigma}_{st} + \sum_{j=1}^{n} \sum_{k=1}^{n} \ddot{\beta}_{jks} \ln X_{jkt} + \sum_{k=1}^{K} \ddot{\delta}_{ks} D_{kt} + \ddot{\theta}_{s} lnNS_{t} + \ddot{\varepsilon}_{st} \tag{2}$$

$$LnNS_{t} = \dddot{\alpha}_{t} + \dddot{T}_{t} + \dddot{\sigma}_{t} + \sum_{j=1}^{n} \sum_{k=1}^{n} \dddot{\beta}_{jk} \ln X_{kt} + \sum_{k=1}^{K} \dddot{\delta}_{k} D_{kt} + \dddot{\varepsilon}_{t} \tag{3}$$

S_{it} = Off and online product sales by cross-section i (e.g., household, store, region)

W_{st} = Unique web traffic visits by source s (e.g., paid search, SEO, direct to site, display, social)

NS_t = Natural keyword search (branded and generic)

α_t = Model intercepts

T_t = Linear trends or drift

σ_t = Seasonal terms

X_{jkt} = k market level/cross-sectional own and competitor marketing variables (j=1-n) covering pricing and off and online media

D_{kt} = k interventions and external controls

ε_t = Equation error terms

β, δ, γ and θ denote estimated parameters

Equation (1) analyses the sales behaviour of groups of consumers over time and cross-section i, in terms of a range of marketing and economic covariates. **Equations (2)** and **(3)** describe the relationships between, and the drivers of, web traffic sources and natural search.[3] This type of 'nested' structure is designed predominantly to address the problems of *last-touch attribution*, helping to reattribute a part of media, such as paid search, back to sources earlier in the chain, such as TV advertising. In this way, it can be seen as an aggregate form of MTA, where appropriate credit is allocated to each touchpoint leading to improved budget allocation. However, much like MTA, this approach ignores the fundamental problem of *selection bias* leading to serious consequences for the estimated relationship between sales and web traffic sources in **Equation (1)**.[4]

[3] All equations as expressed in natural logarithms, capturing non-linear response and synergies between the driver variables. Other types of transforms such as Generalised Additive Models are also used.

[4] Some practitioners bypass **Equation (2)** and use online impressions directly in **Equation (1)**. This is based on the notion that consumers do not have to click through to the site for impressions to impact demand. While this is certainly possible, the selection bias problem remains.

2.1. The endogeneity problem

Selection bias arises when (part of) the difference in the 'treatment' outcome (sales) is caused by a factor that predicts the likelihood of selection into treatment (paid search) rather than due to the treatment itself. That is, consumers with a greater propensity to buy predict the level of search traffic, which in turn predicts the sales outcome. In this way, a large proportion of site visits are simply an artefact of the sales process. This creates an *endogeneity* or *identification* problem, leading to biased estimates of the traffic-sales impact of **Equation (1)** and all marketing effects that work through it. Consequently, even if offline TV advertising does lead to more paid search activity, it does not necessarily mean it drives *incremental* sales in this way.

The textbook solution is Instrumental Variable (IV) estimation, where the causal effect of an independent variable (paid search) on an outcome variable (sales) is estimated using an instrumental variable z which affects sales only through its impact on search (the *exclusion* principle). If successful, z provides the necessary *exogenous* variation in search, such that the outputs are more akin to those of experimental trials.[5] However, valid instruments are notoriously difficult to find. Consequently, many alternative solutions have been proposed ranging from the difference in differences, regression discontinuity designs, and Heckman correction, through to Latent IV (Ebbes *et al*., 2009), Gaussian Copulas (Park & Gupta, 2012), DAG analysis (Chen *et al*., 2018, Pearl, 2000) and incorporating experimental results as Bayesian priors (Ugena *et al*., 2021).[6] Whichever route is taken, the message is clear: for meaningful attribution, the chosen identification scheme needs to be clearly specified as part of any MMM engagement.[7]

3. Short and long-term marketing effectiveness

For a complete view of marketing ROI and optimal allocation, marketing mix models need to reflect both short and long-term marketing effects. Short-term effects explain mean-reverting or transitory sales variation. Long-term effects explain persistent changes in underlying base sales, reflecting permanent additions to the loyal customer base. Measuring the true long-run impact of marketing investments, therefore, requires a focus on the base sales component of the mix model.

3.1. The standard approach

Standard mix models use ordinary or generalised least squares regression techniques, with fixed or deterministic baselines, and focus solely on short to medium-term sales effects with stationary Adstock transforms. Consequently, all such models fail to reflect any persistent changes in core brand preferences by construction. A popular remedy simply adds attitudinal brand metrics to the short-term model together with sub-models in terms of advertising variables. The indirect effects of advertising on sales are then interpreted as long-term effects.[8] However, this approach is flawed in several respects.

[5] On face value, the nested mix model structure of **Equations (1)-(3)** *appears* to fit the bill. Provided (at least one of) the variables driving web traffic satisfies the exclusion principle, the web traffic fitted values could be substituted into the sales equation to give a two-stage least squares estimate. However, X_t and D_t generally affect both web traffic and sales and cannot serve as valid instruments.

[6] Experimental priors in MMM rely on valid A/B testing or 'lift' studies for all endogenous variables. However, these are rarely available as part of the routine data collection process.

[7] Note that 'causal AI' methods seek to automatically identify DAGs such as **Figure 1**. However, since human context is always required, such techniques 'do not yet work as stand-alone methods for causal learning' (Peters *et al.*, 2017). Furthermore, there is no one unique chain. Endogeneity bias stems from ignoring the simultaneous likelihood of all other plausible DAGs, leading to the correlation between search and the error term in **Equation (1)**. We need to control for all paths to help identify causal effects.

[8] Alternative 'long-term' approaches simply extend the short-term structure, either by adding Adstocks with very high retention rates or multiplying the short-term effects by an *ad hoc* scaling factor.

1. Ignores the fundamentals of time series econometrics

If (observable) brand-building effects exist, sales should exhibit evolutionary behaviour.[9] If not, then the impact of brand metrics on sales can only be a short-term relationship by definition. If, on the other hand, sales are evolving, we cannot just run simple regressions of sales on marketing and brand metrics. Firstly, if marketing and brand metrics are stationary, then the mix equation is unbalanced, and we must first-difference sales. Alternatively, if brand metrics are also evolving, then there is potential for spurious regression problems. Consequently, brand equity metrics must also be first-differenced, and valid cointegrating relationships between sales and brand metrics need to be incorporated.

2. Mindset metrics are regressed directly on short-term sales

A plausible theory of brand-building needs to link the long-term brand preferences embodied in mindset metrics directly to the long-term purchase demand revealed through base sales. This follows naturally from the fact that base sales and attitudinal data both represent brand health (*inter alia*, Kamakura & Russell, 1993, Hanssens *et al.*, 2014). As such, they are essentially two sides of the same coin. Therefore, the use of actual sales is inconsistent and obscures long-term movements risking contamination with short-term transactional effects.

3. Does not reflect the brand-building process

Simply adding brand metrics as additional regressor(s) precludes feedback between (base) sales, earned media, and other long-term drivers. Feedback effects mimic word-of-mouth as consumers talk about brand experiences leading to new trialists and growth of the loyal customer base, which wears in over time. Only by identifying these *endogenous* relationships in a suitable network structure can we estimate the true incremental long-term impact of marketing on base sales via brand perceptions.

3.2. An alternative approach

To resolve these issues, brand metrics need to be linked directly to variation in base sales in a long-term network model of brand-building. To achieve this, the marketing mix model needs to be re-cast in a form that allows measurement of both short-term sales and long-term base variation. One candidate is the Unobserved Component Model (Harvey, 1989), illustrated in Cain (2005, 2008), where sales behaviour is decomposed into a trend, seasonal, regression effects and measurement error. This re-writes the marketing mix model **Equations (1)-(3)** as:

$$LnS_{it} = \dot{\mu}_{it} + \dot{\sigma}_{it} + \sum_{j=1}^{n}\sum_{k=1}^{n}\dot{\beta}_{ijk}\ln X_{ijkt} + \sum_{k=1}^{K}\dot{\delta}_{ik}D_{kt} + \dot{\gamma}_{is}lnW_{st} + \dot{\varepsilon}_{t} \qquad 1(a)$$

$$LnW_{st} = \ddot{\mu}_{st} + \ddot{\sigma}_{st} + \sum_{j=1}^{n}\sum_{k=1}^{n}\ddot{\beta}_{jks}\ln X_{jkt} + \sum_{k=1}^{K}\ddot{\delta}_{ks}D_{kt} + \ddot{\theta}_{s}lnNS_{t} + \ddot{\varepsilon}_{st} \qquad 2(a)$$

$$LnNS_{t} = \dddot{\mu}_{t} + \dddot{\sigma}_{t} + \sum_{j=1}^{n}\sum_{k=1}^{n}\dddot{\beta}_{jk}\ln X_{kt} + \sum_{k=1}^{K}\dddot{\delta}_{k}D_{kt} + \dddot{\varepsilon}_{t} \qquad 3(a)$$

$$\mu_t = \mu_{t-1} + \lambda_{t-1} + \eta_t \qquad 4(a)$$

$$\lambda_t = \lambda_{t-1} + \xi_t \qquad 4(b)$$

$$\sigma_t = -\sum_{j=1}^{p-1}\sigma_{t-j} + \kappa_t \qquad 4(c)$$

The intercepts α in each equation are replaced with a time-varying (stochastic) trend μ_t comprising two components. **Equation 4(a)** allows the underlying level of each time series to follow a random walk with a growth factor λ_t analogous to the conventional trend term T. **Equation 4(b)** allows λ_t to also follow a random walk. Depending on the estimated values of the

[9] The absence of evolution does not imply the absence of brand-building *per se*: merely that it is *unobservable*. Evolution could be offset by customer churn rendering observed sales stationary.

covariance parameters η_t and ξ_t, the system can accommodate both stationary and non-stationary product demand allowing the data to decide between them. **Equation 4(c)** specifies seasonal effects, which are constrained to sum to zero over any one year. If κ_t is zero, then seasonality is deterministic.

Equations 1(a)-4(c) provide a direct separation of sales behaviour into short and long-term components. The estimated regression parameters capture short-term (transitory) marketing effects, informing short-term ROI and budget allocation decisions. Long-term effects can then be analysed through a network model of the permanent sales component μ_t in terms of consumer brand perceptions and external long-term controls (Cain, 2010, 2022). A representative example is illustrated in **Figure 2**, where marketing investments stimulate brand awareness, drive brand consideration, and increase social media interest leading to underlying base sales growth. If we can show that marketing significantly impacts the permanent (baseline) component, then we can state that marketing campaigns have persistent long-term effects, as existing purchase incidence increases and/or new buyers are converted into permanent loyal consumers. These effects are then combined with short-term effects to provide total ROI and budget allocation recommendations.

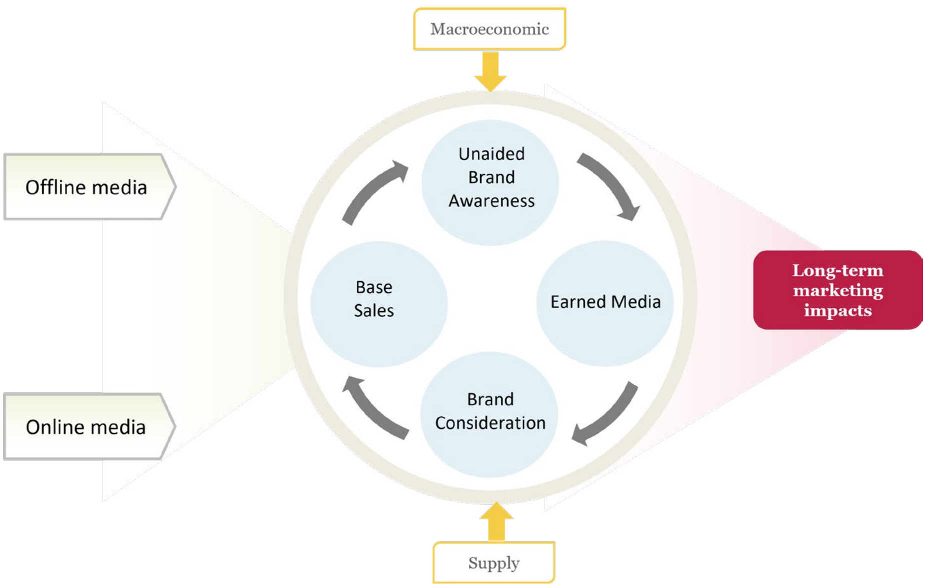


Figure 2. Network model of long-term sales

Estimation of the long-term network model requires a suitable systems approach to capture the long-term relationships between the nodes of Figure 2 and the persistent brand-building role of media. Popular systems approaches are Path Models or Structural Equation Models. However, these frameworks are typically static and ignore the dynamic relationships between the network variables. As such, they are unsuitable for long-term trend and cointegration analysis and cannot measure feedback between the nodes and the dynamics of how brand-effects *wear in* over time. To overcome these issues, a dynamic systems approach such as a Vector Autoregression (VAR) is required (*inter alia*, Hendry, 1995), written as a cointegrated Vector Error Correction Model (VECM):

$$\Delta lny_t = \Psi_1 \Delta lny_{t-1} .. \Psi_{l-1} \Delta lny_{t-l+1} + \alpha\beta^{'} lny_{t-1} + \Omega_k(L) lnx_{kt} + \Upsilon_k D_{kt} + \varepsilon_t \tag{5}$$

where y_t denotes a vector of n endogenous variables capturing base sales and path-to-purchase or brand-building 'steps', x_t denotes a set of k marketing variables with lags L and D_k denotes a set of dummy variable events. The $\alpha\beta' lny_{t-1}$ term represents the error correction component, comprising r cointegrating (equilibrium) relationships β between the nodes and associated error-correction parameters α. With n endogenous variables, there may be up to n-1 such relationships with a minimum of one common trend driving the non-stationary (brand-building) properties of the system.

Equation (5) is then estimated using the Johansen technique (1988) and identified using either a Cholesky decomposition, restrictions based on economic theory or instrumental variable techniques (Juselius, 2006). Once identified, impulse response analysis traces out the dynamic long-term base sales impact of changes in brand metrics and earned media. The long-term impact of marketing activity x_t then cumulates indirectly and permanently into the level of base sales.

3.3. Worked example

The complete short and long-term modelling approach is formally demonstrated in Cain (2022). Here we present a simple example to illustrate the principles involved. We first take daily data for sales, web traffic, and natural branded search for a seasonal brand, together with a range of off and online marketing factors, pricing, monthly unaided awareness data, and an index of monthly business economic activity. Monthly business activity and awareness data were then disaggregated to daily level and introduced directly into the sales equation (1). Standard OLS (fixed base) estimation gives an awareness coefficient of 0.045 but is insignificant with a t-ratio of 1.1. Furthermore, the base price coefficient is positive and a Durbin Watson (DW) statistic of 1.08 indicates significant model error autocorrelation. The implication is that neither awareness, price, nor economic growth manages to adequately capture long-term sales movements.[10]

We then applied the UCM framework of 1(a)-4(c) and aggregated the extracted baseline to the weekly frequency - illustrated in Figure 3 alongside unaided awareness, base price, and business economic activity (BEA). Standard ADF tests indicate that all are non-stationary

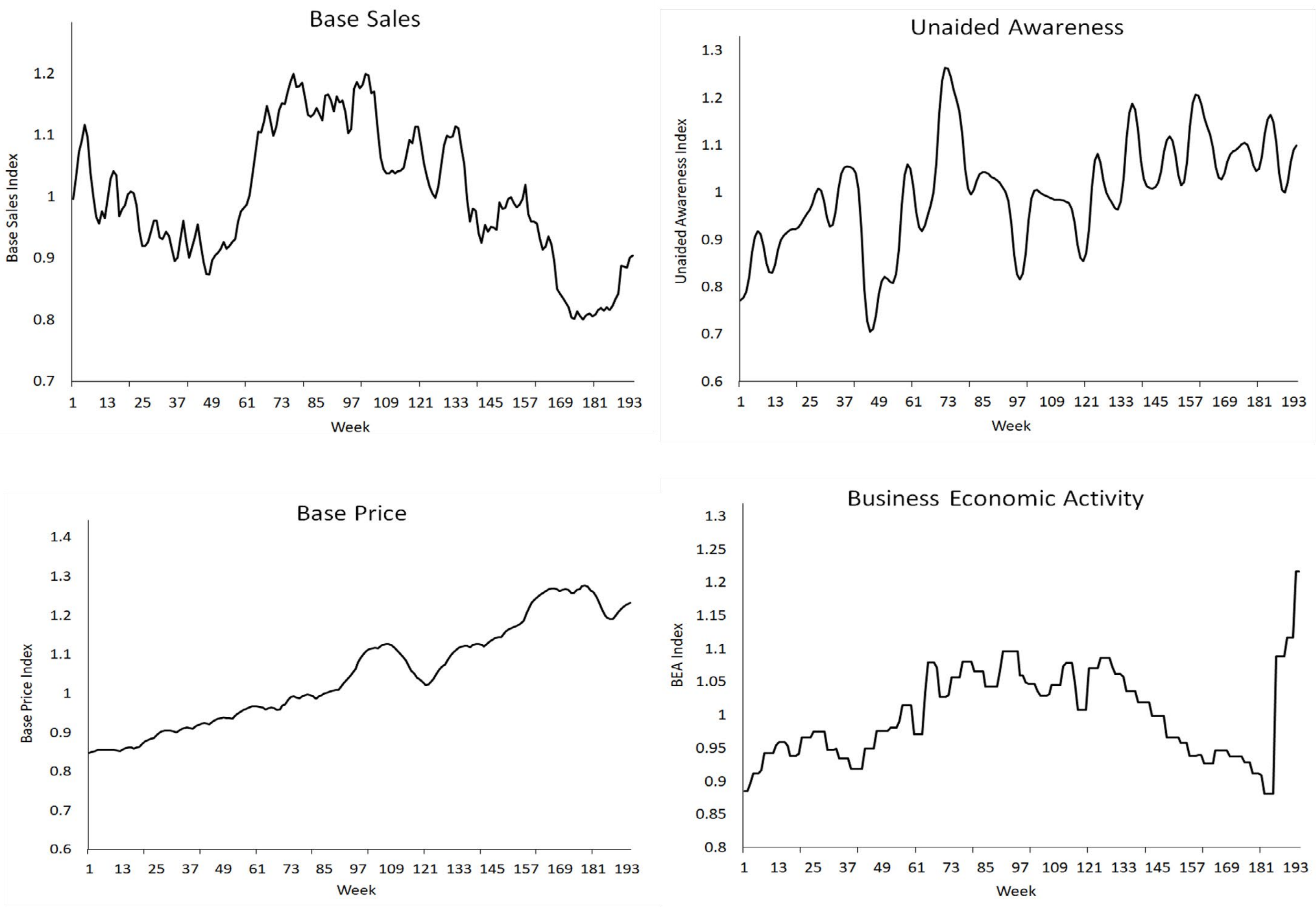


Figure 3. Extracted base sales, brand awareness and controls

[10] An AR(1) error structure improves autocorrelation with a DW stat of 2.01. However, the awareness coefficient is -0.03 and insignificant. Weekly frequency models made little difference to the results.

I(1) series. Equation (5) was then estimated with two lags of the endogenous variables to ensure well-behaved residuals. Marketing regressors x_{kt} comprise paid TV GRPs and social media commentary (earned media).

The cointegrating relationship between the variables is illustrated in the left-hand panel of **Table 1**. This captures the underlying equilibrium (attractor) relationship between base sales, price, unaided awareness, and business activity, with feedback reflected in the alpha (error correction) parameters. The full long-term (impulse response) coefficients are illustrated in the right-hand panel, where the first row shows the final (permanent) elasticity of a 1% impulse in unaided awareness on base sales of 0.13 with a significant t-ratio of 2.7. Note too, that the final long-term effects of base price and economic activity are also correctly signed and significant.

Figure 4 then illustrates the corresponding pattern of dynamic adjustment of base sales to unaided awareness, where the full impact wears in over approximately 16 weeks.[11]

The corresponding VECM is given in **Table 2**, which shows how marketing investments impact the dynamic adjustment of each of the network variables. Here TV and earned media impact unaided awareness with elasticities of 0.002 and 0.022, respectively. Weighted by the long-term impact of unaided awareness on base sales gives final base elasticities of 0.0003 and 0.003. These are used to quantify long-term base contributions over the sample and extrapolated over a 3-5 year forecast horizon. Combined with the short-term effects from the UCM, this provides total ROI and budget allocation recommendations.

Regressor	CV	Alpha
Base sales	1	-0,151 (-3.8)
Base Price	-0,588 (5.1)	0
Awareness	**0.280 (2.1)**	0.082 (2.10)
BEA	1.81 (4.9)	0.060 (3.70)

Equation	$\sum \hat{\varepsilon}_{Base}$	$\sum \hat{\varepsilon}_{Price}$	$\sum \hat{\varepsilon}_{Aware}$	$\sum \hat{\varepsilon}_{BEA}$
Base sales	0.455 (6.40)	-0.688 (-3.6)	**0.130 (2.7)**	0.960 (4.0)
Base Price	-0,005 (-0.1)	1.37 (10.0)	0.006 (0.20)	-0.020 (-0.10)
Awareness	0.196 (1.80)	0.080 (0.30)	0.889 (12.5)	-0.721 (-1.21)
BEA	0.221 (5.70)	0.054 (0.50)	-0.063 (-1.1)	0.639 (4.9)

Table 1. Cointegrating economic structure and impulse response matrix

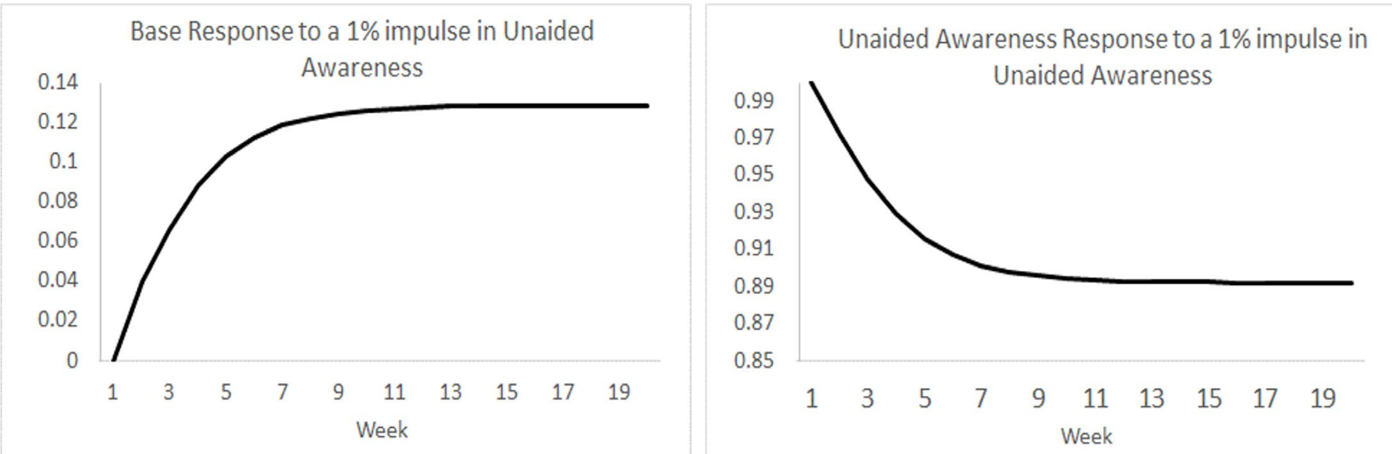


Figure 4. long-term base sales adjustment

[11] Note that these results imply that long-term effects are under-estimated using the traditional approach. However, the relationship(s) between brand metrics and sales can often be over-estimated if the long-term network dynamics are not accounted for. It depends on the data and model structures, requiring careful modelling on a case-by-case basis.

Equation	$\Delta Base_t$	$\Delta Price_t$	Aware	ΔBEA_t
Intercept	-3.14 (-3.76)	0.002 (1.31)	1.69 (2.09)	1.25 (3.68)
$\Delta Base_{t-1}$	-0.188 (2.63)	-	-0.130 (-1.86)	0.019 (0.94)
$\Delta Price_{t-1}$	-0.334 (-1.82)	0.28 (3.96)	-0.15 (-1.01)	-0.054 (0.98)
$\Delta Aware_{t-1}$	-	-	-0.014 (1.01)	-
ΔBEA_{t-1}	-0.365 (-2.06)	-	0.221 (1.26)	0.273 (3.78)
ECM_{t-1}	--0.151 (-3.76)	-	0.082 (2.10)	0.060 (3.68)
TV	-	-	0.002 (2.10)	-
ΔEM_t	-	-	0.022 (2.80)	-

Table 2. dynamic network adjustment

4. MMM and tactical planning

It is often argued that MMM is too slow and lacks the necessary granularity to handle the tactical and 'real-time' attribution problems that solutions such as MTA purport to solve. Given the detailed and rapid solutions marketers have now come to expect, any viable MMM framework needs to be able to rise to the challenge.

4.1. Tactical decision making

Whereas MTA focuses on cookie-level data over very short time windows, MMM can provide similar learnings across both off and online through higher frequency time series data. The process is illustrated in **Figure 5**, where the dynamic aggregated framework set out in Sections 2 and 3 is first estimated at daily level, providing trend and seasonal factors and incremental contributions for off and online media investments. We then take the hourly data for sales and remove the (proportions of) trend and the contributions for all variables not available at an hourly level. We then model the remaining portion of hourly sales in terms of the detailed 'sub-tactic' elements of all off and online media variables – subject to the estimated 'upper-level' contributions. In tandem with offline media effectiveness, media synergies, and long-term brand-building of the main daily MMM model, this type of approach can then provide granular online media effectiveness by day-part, ranking by publisher, placement, and web page.[12]

4.2. Real-time attribution

Daily network UCM marketing mix models, as set out in Sections 2 and 3 and summarised in the left-hand panel of **Figure 5**, typically take approximately eight weeks to build – depending on the number of models and cross-sections. Updates in response to business needs, or potential structural/parameter changes typically take place every three to six months. Hourly level models – summarised in the right-hand-

[12] Since consumer cohorts are time-based rather than geography-based, this approach is not subject to the matching problems typically faced with cookie or household-level models, where a mapping between individual data and more aggregated (mass-market) offline data is required.

side panel of **Figure 5** – are constructed on the last three months of data used to build the main daily models, with contribution/parameter constraints set from the daily models to ensure consistency. The hourly models are then updated weekly to deliver the types of rapid in-campaign attribution illustrated in **Figure 6**.

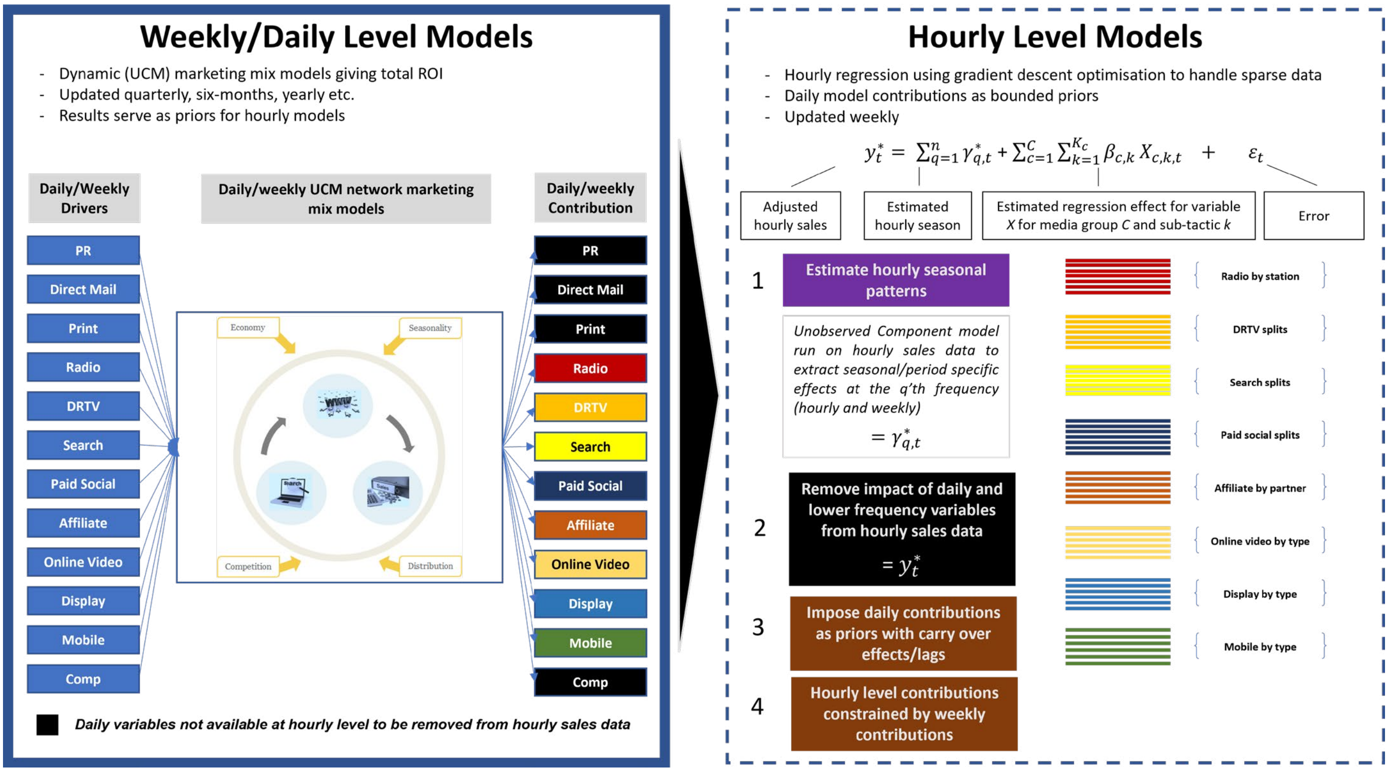


Figure 5. High frequency MMM

Figure 6. Campaign response attribution

5. Conclusions

With the potential demise of MTA, the focus is once more back on the marketing mix model as an attribution framework. However, to be useful and live up to the exacting standard that marketers have come to expect, any next-generation MMM approach needs to satisfy three fundamental criteria.

Firstly, to serve as a true attribution solution, MMM needs to focus on causal estimation methods. Too often, we see reliance on consumer journey solutions to address the problems of last-touch-attribution. However, much like micro-level MTA methods, these ignore the endemic selection bias in many online media – leading to endogeneity bias and misallocation of the marketing mix. The growing popularity of automated machine learning (ML) approaches to the mix model only serves to exacerbate this problem. To address this issue, all MMM work - whether based on regression, neural nets, or other ML methods - needs a transparent identification scheme to isolate true incrementality.

Secondly, MMM needs to quantify the long-term (base-building) effects of marketing and so inform brand-building strategy. Standard approaches are simply not set up to measure these effects, with fixed baselines and a focus on short to medium-term lag structures or Adstocks. Alternative time series structures are required that can quantify both short and long-term (base) variation – coupled with dynamic network models that can explain the causes of base variation and the economics of brand-building.

Finally, next-generation MMM needs to deliver near 'real-time' granular insights on marketing ROI and optimal budget allocation. Suitably identified high-dimension mix models – by day and hour – can fit the bill. This can provide many of the claimed benefits of MTA, such as online media effectiveness ranking by publisher and placement, with the added benefit of controlling for the wider economic environment and quantifying the contribution of pricing and offline media.

6. Acknowledgements

The authors are grateful to David Dixon for helpful comments and to Arnab Pal and Kunal Rambhia for input into section 4 and model outputs of **Figure 6**.

References

1. Cain, P.M. (2005). "Modelling and forecasting brand share: A dynamic demand system approach". *International Journal of Research in Marketing*, 22(2), 203–220.
2. Cain, P.M. (2008). "Limitations of conventional market mix modelling". *Admap*, 48-51.
3. Cain, P.M. (2010). "Marketing Mix Modelling and Return on Investment". In: P. J. Kitchen (ed.) *Integrated Brand Marketing and Measuring returns*, Palgrave Macmillan, 94-130.
4. Cain, P.M. (2022). "Modelling short and long-term marketing effects in the consumer purchase journey". *International Journal of Research in Marketing*, 39(1), 96–116.
5. Chen, A., Chan, D., Koehler, J., Perry, M. L., Wang, Y., Sun, Y., & Jin, Y. (2018c). "Bias Correction For Paid Search In Media Mix Modeling". *ArXiv (Cornell University)*.
6. Ebbes, P., Wedel, M., & Böckenholt, U. (2009). "Frugal IV alternatives to identify the parameter for an endogenous regressor". *Journal of Applied Econometrics, 24*(3), 446–468.
7. Hanssens, D. M., Pauwels, K., Srinivasan, S., Vanhuele, M., & Yildirim, G. (2014). "Consumer Attitude Metrics for Guiding Marketing Mix Decisions". *Marketing Science, 33*(4), 534–550.
8. Harvey, A. C. (1989). "Forecasting, Structural Time Series Models and the Kalman Filter". *Cambridge University Press*.

9. Hendry, D. F. (1995). "Dynamic econometrics". *Oxford University Press*.

10. Johansen, S. (1988). "Statistical analysis of cointegration vectors". *Journal of Economic Dynamics and Control, 12*(2–3), 231–254.

11. Juselius, K. (2007). "The Cointegrated VAR Model: Methodology and Applications" (Advanced Texts in Econometrics). *Oxford University Press.*

12. Kamakura, W. A., & Russell, G. L. (1993). "Measuring brand value with scanner data". *International Journal of Research in Marketing, 10*(1), 9–22.

13. Mobile Marketing Association (MMA) (2021). "MTA is dead, long live MTA".

14. Nail, J. (2015). "Introducing Unified Marketing Impact Analytics". *Forrester*.

15. Park, S, and Sachin G. "Handling Endogenous Regressors by Joint Estimation Using Copulas." *Marketing Science, 31*, no. 4 (2012): 567–86.

16. Pearl, J. (2000), "Causality", New York: *Cambridge University Press*.

17. Peters, J., Janzing, D., & Scholkopf, B. (2017). "Elements of Causal Inference: Foundations and Learning Algorithms" (Adaptive Computation and Machine Learning series). *The MIT Press*.

18. Schoen, M. (2021). "Can Data Clean Rooms Be The Answer For Privacy-Safe Marketing?" *Forbes. (CLE) blending. Deloitte working paper*.

19. Ugena, C.R, Conde de Simón, M, Answer, M. and Fabrizio, D., (2021), "Exploring Marketing Mix Modeling (MMM) and Conversion Lift Experiment (CLE) blending", *Deloitte working paper.*

Author

Peter Cain is executive partner and co-founder at Marketscience, an award-winning independent consultancy specialising in advanced econometric modelling, statistical analysis and bespoke advice on marketing investments. Prior to this, Peter was the founder and CEO of Marketscience Consulting, established in 2012 with the goal of blending academic, commercially relevant analytics and strategic advice for business. He has more than 20 years of commercial and academic experience in economics and marketing science designing econometric business solutions for blue-chip companies and organizations across a wide range of industries. He writes extensively on economics and econometrics in marketing and published in top peer-reviewed journals. Before marketing research, Peter was in academia, specializing in monetary economics and econometrics. He holds BSc and MSc degrees in Economics from the University of Warwick, and PhD in Monetary Economics from the University of Nottingham.

peter.cain@marketscience.co

Nitesh Sahay is a Partner at Marketscience. Prior to this, Nitesh worked at different roles in analytics at GE Capital, Symphony Marketing Solutions, Datamonitor and Ninah (The Publicis Groupe). He has more than 20 years of commercial experience across multiple industry verticals and clients in Fortune 500 companies working on techniques such as Marketing Mix Models, Market Structure, Customer Segmentation and other predictive methods. He has written 2 papers on economics published as Working Paper series and another in an edited book. Before marketing research, Nitesh was in academia, specializing in macroeconomics and econometrics (computable general equilibrium models). He holds a BA in Economics from Delhi University and MA and MPhil degrees in Economics from the Jawaharlal Nehru University, New Delhi.

nitesh.sahay@marketscience.co

Can AI Reduce Noise in Decision Making?

Fahad Daniyal
Publicis Groupe, UK

Jason Smith
Publicis Groupe, UK

Classifications, Key Words:

- Decision making
- Artificial intelligence
- Marketing effectiveness

Abstract

Largely decision making is made by people but we're unreliable decision makers for a host of reasons, but fundamentally we are strongly influenced by irrelevant factors, such as mood, beliefs, values etc. This means a lot of our decision making, whether individually or collectively contains noise – which is unwanted variability. This noise is present in the decisions and the process of making decisions. What if we could help reduce system noise by using data and technology to help us with our decision making? Could we use algorithms or artificial intelligence, to replace/augment human decision making?

1. Introduction

In the marketing/advertising world, it Is largely us humans that make decisions, because how could a machine possibly be any better? However, it turns out we are unreliable decision-makers for a host of reasons, because, fundamentally, we are strongly influenced by irrelevant factors, such as mood, beliefs, values, etc. This means a lot of our decision-making, whether individually or collectively, contains noise – which is unwanted variability. Noise is present both in the decisions we make and in the process of making decisions. What if we could help reduce system noise by using data and technology to help us with our decision-making? Could we use algorithms or artificial intelligence to augment or replace human decision-making?

2. What do we mean by decision making?

In the context of marketing/advertising, there are a myriad of decisions that need to be made across the production media process. Some of these are more creative in nature – such as what is the best content to run; across what channels; to what audience; over what timeframes – and others are more commercial – such as assigning budgets; how long to run campaigns, etc. Many of these decisions also involve an element of forecasting and prediction.

3. How do we make decisions?

We often start the process of decision-making with an inclination towards reaching a particular conclusion, whether we are conscious of that or not, because we tend to employ our fast, intuitive System 1 thinking first. Then, we either jump to that conclusion and simply bypass the process of gathering and integrating information, or we mobilise System 2 thinking, which is engaging in deliberate thought, to come up with arguments that we believe support our decision. We use this simplifying operation to make fast decisions – called heuristics, which are quite useful in certain circumstances (think 'fight-or-flight'). Our minds 'construct' a point of view very quickly. We check whether that 'feels' right and if it does, we go with it. It allows us to act very quickly but it also produces variability.

When we do adopt System 2 thinking by considering information, we collect and interpret evidence selectively to favour a decision that we already believe or wish to be right. We believe that these post-rationalizations were the cause of our decision, when in fact it is really a trick of our minds.

In general, we jump to conclusions, then stick to them because we determine what we think by consulting our feelings. That's why marketeers work so hard to attach a positive emotion to a brand. We like to think that we base our opinions on evidence, but the evidence we consider and our interpretation of it is likely to be distorted, at least to some extent, for instance by our feelings, and to fit our initial snap decision. As a result, our minds maintain the coherence of an overall story that has emerged, but the truth is somewhat different. All of this is fine when the decisions are correct but often when they are not, we tend to stick to them even in the face of contradictory evidence.

We are not the same people all the time. As our mood varies, some features of our cognitive function also vary (something we are generally not consciously aware of). Our moods in the moment can influence our decisions. Other factors such as stress and fatigue also have an impact. These psychological biases are universal, and they often produce errors. When there are large individual differences in biases, or when the effect of biases depends on context (different triggers), there will be what is called 'noise'.

What is 'noise'?

Noise is caused by the variability of our minds. Our opinions change without apparent reason, and often we do not produce identical decisions when faced with the same set of facts on two, or more, occasions. Noise is the unwanted variability in decisions caused by how our minds operate.

The prevalence of noise has been demonstrated in several studies. Academic researchers have repeatedly confirmed that professionals often contradict their own prior judgements when given the same data on different occasions. For instance, when software developers were asked on two separate days to estimate the completion time for a given task, the hours they projected differed by 71% on average (Grimstad & Jørgensen, 2007). The same was true for pathologists (Einhorn, 1974); stockbrokers valuing stock[1]; auditors (Colbert, 1988), and even judges sentencing criminals (Anderson *et al.*, 1999).

Noise can be present in individual decisions but also in the process of how decisions are made. Very often businesses do not have processes for decision-making. Decisions are left to chance. How many times have we heard the comment, "nobody knows who makes the decision"? Unsurprisingly, there is a lack of consistency in how (and sometimes whether) decisions are made, which leads to unwanted variability.

5. Can AI help?

AI can now perform many tasks formerly regarded as quintessentially human – recognise faces, translate languages, read radiology images, and even make scientific breakthroughs (see Deepmind Alphafold work on protein structures). Very large datasets are essential for sophisticated analysis and the increasing availability of data is one of the main reasons for the rapid progress in AI in recent years. More data means more relationship patterns that no human can detect. These patterns can then be modelled to make better predictive decisions.

[1] https://psycnet.apa.org/record/1969-15267-001

AI can find significant signals in combinations of variables that might otherwise be missed. Given the advantage of exploiting much more information, AI algorithms can now not only outperform many human decisions, but the algorithms are noise free.

An example of such a system is inherent in creative and media execution and targeting. A creative campaign is initiated by identifying the audience tribes and segments that constitute the target audience for the campaign objective. Not only is this heavily biased and based on the pre-conceived notions that the analyst might carry but also tends to be noisy with significant variation in process and analysts' individual decisions. Even when we perceive such decisions to be data driven, they are in fact centred around a null or statistical hypothesis, i.e., people with such attributes can be pushed down the consideration and/or conversion funnel by such stimulus. For instance, people with kids tend to look for cars with more space, or younger audiences need to be fed content that resonates with their social footprint. While, this sort of targeting, might consider what we can refer to as attitudes does not consider the situational signals and is as such flawed.

This impact is magnified when we start thinking about media execution where another layer of pre-conceived notions of an ideal candidate for creative messaging is laid over. Complex decision trees are created, that try to capture a user's journey through an experience loop, but they tend to be outdated as soon as the campaign is launched.

6. Background & Related Work

As businesses try to stretch the media budget, they want to minimise ad wastage so that the ads are relevant to the user and can stimulate the user to move further down the conversion funnel, or in fact complete the purchase. To this end, we see optimization efforts in three main areas: 1) real-time bidding; 2) targeting; and 3) activation.

In the first category, Wang *et al.* (2016) present a detailed survey of bidding adjustment frameworks. Most of the methods in this branch of work try to compensate for the so-called black-box bid adjustment mechanisms of the publishers (Chapelle, 2015), (Balseiro *et al.*, 2015). The bid adjustment considers three metrics (*Facebook - About Ad Auctions*, n.d.): the bid placed by an advertiser for that ad, estimated action rates: an estimate of user engagement propensity with the ad, and finally, ad quality which is a measure of the quality of an ad as being relevant. These three metrics[2] then come together to create a total score that determines an auction win. However, it was shown by Ali *et al.* (2019) that there is bias in the bidding frameworks based on how ad relevance is calculated. To this effect, Tunuguntla and Hoban present an online learning algorithm to simultaneously estimate impression values and learn the bidding policy, whereas Waisman *et al.* (2019a) use an adapted Thompson Sampling (TS) algorithm to solve a multi-armed bandit problem that succeeds in recovering such bids. As in this work, we are more focused on Type-2 bias, hence, such algorithmic bias is considered out of the scope of this, and experimentation is set up so that the test and control branches are equally affected by such bias.

The second area of research focuses on audience selection or user ranking mechanisms. Most publishers provide very similar audience targeting mechanisms based on demographics (age, gender, location) and on-site activity and linked attributes (likes, dislikes, interests, etc.). In more popular publishers (such as Facebook or Google) these attributes can be in the range of 1000s. There are two other audience targeting mechanisms: 1) based on a predefined list from a CRM or a 3rd-party audience provider, and 2) based on retrieval modelling (Tyler *et al.*, 2011). In both cases, the advertisers provide a seed list to the publisher and then the publisher identifies the updated or refined user groups. Various success metrics (retention, value, etc.) can be used to identify the optimal audience segments for a campaign. Grunt *et al.* (2018) use Markov

[2] Similar metrics exist in all majors DSPs, i.e. Google, TikTok.

Decision Process method modelling customers with assigned rewards corresponding to the expected return value. Provost *et al.* (2009) utilise affinity networks to group users with similar exposure and interests. Therefore, once a user in the network is interested in a product, the ad is targeted to the other components of her affinity network. Tree expansion of the target space to learn the partition that efficiently maximises the campaign revenue is proposed by Gasparini *et al.* (2018) and combined with tree search to drive the tree expansion. They formulate the problem of target optimization as a Learning from Logged Bandit Feedback (LLBF) problem.

Creative activation constitutes the third leg of the optimization frameworks, where a creative is tied to an audience based on the perceived relevance of the creative strategy. Generally, a set of creatives is activated against an audience segment, and most advertisers let the publishers cycle through the creative set to identify and match the optimal creative to the selected audience segments. A stream of literature (Geng *et al.* (2019), Ju *et al.* (2019), Ba *et al.* (2022)) is aimed at applications of bandit models to web content optimization. Huang *et al.* (2019) propose an RL-based framework that identifies the optimal ad within a rec list.

The selected framework stands out because it does not focus on optimising ads' revenue and instead aims to gauge the positive and negative user experience to decide the next best. We demonstrate with the breadth of choices of decisions and the fact that we treat audience and creative as malleable, we avoid any human bias and noise towards identifying the optimal combination and hence improve decision-making around creative, media, and targeting.

7. Methodology

To cater for biases and noise, we start by proposing/establishing an optimization framework that connects the creative and media strategies, and captures the campaign objectives, performance, audience, creative signals, and the relationships between them. Observable and quantifiable features from creatives and audiences are used to represent these dimensions, and the selection of such represents the actions that you can take to maximise the campaign performance. The audience or audience segments can be represented as a group of attributes (demographic, situational, and behavioural) and individual users claim memberships to these audience groups based on these attributes. The creative can be represented by passing it through a Computer Vision AI engine (G. Huang *et al.*, 2017a) or manually labelling the content with the visual brand language (VBL).

Each campaign is initiated with a large set of creative options that are pre-selected loosely based on the canonical definition of the ideal target group (strategic audience) and represent creative strategies for a brand. For the sake of consistency of strategy and audience targeting, we create a mapping between the strategic audience and the addressable audience.

Our objective is to identify an optimal combinatorial mapping between the creative and audience, where both are malleable, and this malleability constitutes our action set. The creatives can be updated by changing the various segments within the ads, while maintaining the creative narrative, and the audience can be adjusted by adding or removing certain attributes from the audience segment[3].

Then the quality score ρ_t^i of the i^{th} combination of content $(\mathrm{T_t^i})$ and audience $(\mathrm{A_t^i})$ at time t can be represented as

$$\rho_t^i = \mathcal{M}(\mathrm{T_t^i}, \mathrm{A_t^i})$$

However, given the range of audience attributes and creative elements, it is infeasible to do heuristic-based optimization and avoid bias creep. Moreover, we want this optimal combination to be flexible enough to achieve desirable performance across various engagement and business KPIs. To this end, we

[3] Please note that in the case of audiences we operate at segment level and thus are not affected by the privacy rules

have developed an optimization framework that $\forall\, t$, tries to identify the optimal combination of audience and creative realised as a Markovian Decision Process (Daniyal & Cavallaro, 2011) which we outline below.

7.1 Framework

Within the optimization framework, we first map the observations, selected creatives, and audience to a reward within the system; then the optimal policy is defined as the one that maximises the gain over the action space, i.e. the combination of creative elements and the audience attributes that generate the maximum 'resonance' and maximise the targeted KPI. Let us refer to this solution as our policy π. An important point to note here is that there are other latent factors which may or may not be observable, and thus cannot be affected by us. Hence the solution is only true for a slice of time. To this end, we model the system as a Markovian Decision Process, with partially observable states where each decision takes our system to the next state (s_{t+1}). This can be shown as an influence diagram as shown in **Figure 1**.

Let the state of the creative-audience mapping set C^i be represented at any time t as $s_t^i \in \mathbb{R}+$ where the state space is $S \in [0,1]$. Thus, the state space for the system at time t can be expressed as

$$s_t = (\rho_t^1, \ldots, \rho_t^N) \in (\mathbb{R}+)^N$$

Let the action space be represented as $\boldsymbol{C}$ and the action at any time be represented by the transition $c_t^j \rightarrow c_{t+1}^j$. Please note that within the scope of this article we assume that only one mapping can be activated at any given instance such that action at any time t is represented as a N dimensional vector Ω_t^i which has only 1 only in the i^{th} location and 0 elsewhere. Then the reward $u(s_t, c_t^i)$ of selecting a content-audience mapping $c_t^i \in \boldsymbol{C}$ given state s_t , can be represented by one step reward function as

$$u(s_t, c_t^i) = \alpha\rho_t^i + (1-\alpha)\upsilon_t^i$$

Where α is a scaling factor and υ is a Boolean based on the previously selected set.

$$\upsilon_t^i = \begin{cases} 1 & if\ c_{t-1}^i = 1 \\ 0 & \text{otherwise} \end{cases}$$

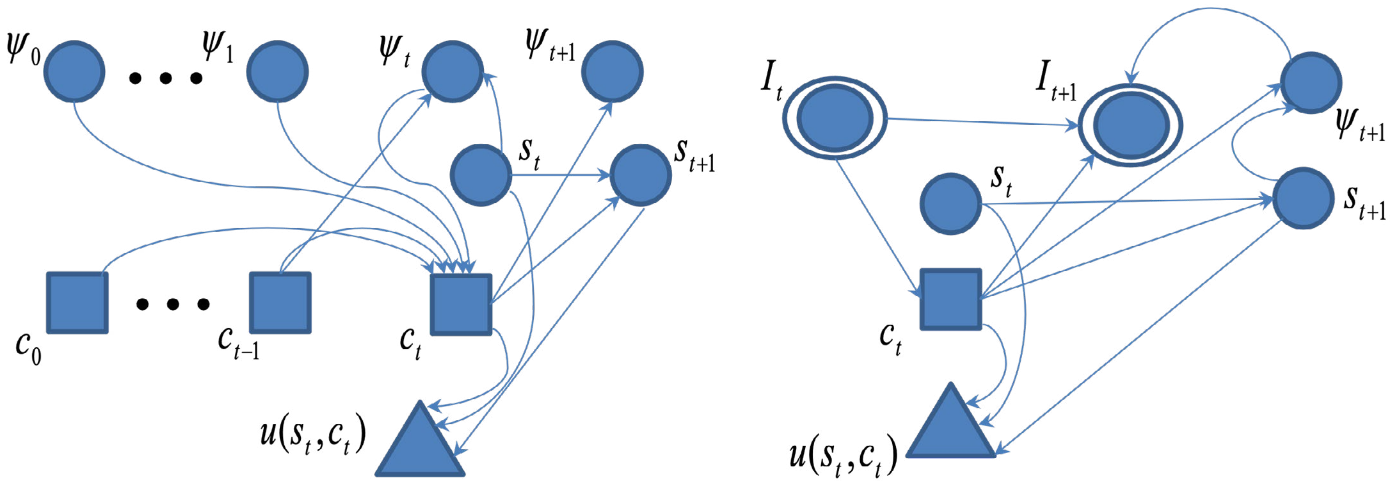


Figure 1. Influence diagram describing model. Rectangles correspond to decision nodes (actions), circles to random variables (states) and triangles to reward nodes. Links represent the dependencies among the components. s_t, c_t, ψ_t , and $u(.)$ denote the state, action, observation, and reward at time t. Information states (I_t and I_{t+1}) are represented by double-circled nodes.

(a) Note that an action at time t depends only on past observations and actions, not on the states. (b) An action choice (rectangle) depends only on the current information state.

The one step cost function described above is an integrated metric that accounts for both content relevancy and audience size given by the accumulated quality score at each time. The state space of such a solution is prohibitively large and modelling such state is intractable. Thus, approximate solutions to this state space are formulated [REF]. These solutions assume that the state space S is quantized with a factor g such that the quantized state space S_d is represented as $s_k^d = g.s_k$, where $g = (g1, g2, \dots gS)$: $g1 > g2 > \cdots > g_S$. For brevity we will drop the superscript d and refer to this discrete space as S.

Then we can represent our solution, the policy $\boldsymbol{\pi}$ as

$$\pi = \{\mu(p(s_t| \ I_t \ ))\}$$

Such that for each t, $\mu(p(s_t| \ I_t \ ))$ is a state feedback map that specifies an action $c_t^j \to c_{t+1}^j$ on $\boldsymbol{C}$ depending on the belief state $p(s_t|I_t)$. A graphical representation is shown below where the posterior probability distribution of state s_t is conditioned on the observable history I_k such that

$$I_t = \begin{cases} p_0 & if \ t = 0 \\ (p_0, \varpi, \dots, \varpi_t) & \text{otherwise} \end{cases}$$

Here $\varpi_t = (\Omega_t^i, (\psi_t^1, \dots, \psi_t^N)), p_0$, p_0 is the initial probability distribution and $\psi_t^i \in \Psi$ are the observations drawn from the observation space Ψ given by the observation equation as

$$\psi_t^i = h(s_t^i, w_t)$$

$$\psi_t = (\psi_t^1, \dots, \psi_t^N)$$

Where h represents the observation map and w_t represents the randomness in the observation at time t. We assume here that w_t is and independent and identically distributed random variable with zero=mean Gaussian distribution.

Then the sequence of states is generated such that at time $t = 0$, the system starts at an initial unobservable state s_0, and creatives are shown to a set of pseudo random audiences given the initial distribution p_0. Then at any time t, the system, which is now in state $s_t \in S$, and taking an action $c_t^j \to c_{t+1}^j$ (selecting the audience-creative set i, given that the j^{th} set was selected at the previous time instance $t - 1$) takes the system to the next state s_{t+1} and an immediate uplift in campaign performance (reward) $u(s_{t+1}, \Omega_t^i)$ is achieved. This state transition is governed by the state transition equation

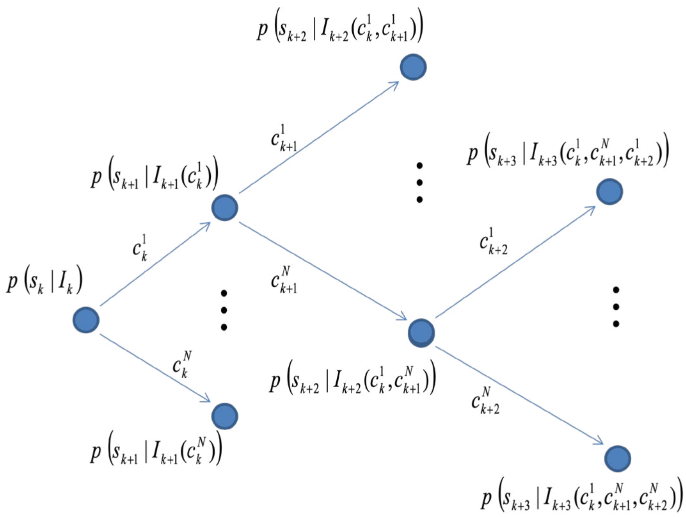


Figure 2. Belief state distribution for three consecutive time stamps. Please note that $I_k(c_k^n, c_{k+1}^m)$ signifies the observable history I_k given $c_k^n = 1$ and $c_{k+1}^m = 1$.

$$s_{t+1} = f(s_t, \Omega_t^i, v_t)$$

Where f and vt represent the state dynamics and randomness in the state transitions, respectively. Because the state equation is composed of two segments, the state dynamics can be decomposed as $f(s_{t+1}, \Omega_t^i, v_t) = [f^s(s_t, v_t), f^C(\Omega_t^i)]$. Now all the components in $f^C(\Omega_t^i)$ are 0, except for the i^{th} component that corresponds to the selected mapping set C^i where it is 1. The specific form of f^s represent the model for the quality-score evolution which we have approximated with a multivariate Gaussian distribution and can be represented for each audience-creative mapping as

$$\rho_t^i = N(\mu^i, \Sigma^i, \psi_t^i)$$

Please note that the belief state probability $p(s_t|I_t)$, i.e., the probability of being in state s_t is the posterior probability distribution of state s_t conditioned on the observable history I_t. Then the estimated belief state probability $\overline{s_{t+1}}$, given s_t after selecting the audience-creative set C^i and observing ψ_t is given by the Bayes' rule as

$$\overline{s_{t+1}} = \eta\, p(\psi_t|s_t, c_t^i) \sum_{s_t \in S} p(s_t|\psi_t, c_t^i) p(s_t|I_t)$$

Where $\eta^{-1} = p(\psi_t|p(s_t|I-T), c_t^i)$ is the normalization constant.

To calculate the optimal policy π^* and the optimal value $\mu^*(p(s_t|I_t))$ we construct the value to action mapping

$$\pi^*: \mu^*(p(s_t|I_t)) \to \boldsymbol{C}$$

This can be estimated using the Bellmans equation [22]

$$\mu^*(p(s_t|I_t)) = \max_{c_t^i \in \boldsymbol{C}} \left[\sum\nolimits_{s_t \in S} u(s_t, \Omega_t^i) p(s_t|I_t) + \lambda \sum\nolimits_{s_t \in S} (p(\psi_t|p(s_t|I_t)., c_t^i) \mu^*((p(s_{t+1}|I_{t+1})) \right]$$

Where $\lambda \in [0,1]$ is a discount factor. Then the corresponding optimal policy select the value maximizing action as

$$\pi^*(p(s_t|I_t)) = \underset{c_t^i \in \boldsymbol{C}}{\operatorname{argmax}} \left[\sum\nolimits_{s_t \in S} u(s_t, \Omega_t^i) p(s_t|I_t) + \lambda \sum\nolimits_{s_t \in S} (p(\psi_t|p(s_t|I_t)., c_t^i) \mu^*((p(s_{t+1}|I_{t+1})) \right]$$

The optimal value function μ above or its approximations can be computed using the value iteration algorithm [23] and can be determined [24] within a finite horizon by performing a sequence of value-iteration steps if the sequence of estimates converges to the unique fixed-point solution. To this end we need to rewrite the Bayes equation above in the value-function mapping form. Let the real-valued bounded functions μ^* be such that value function mapping H for all information states can be written as $\pi^* = H\mu^*$ and the value mapping function can be written as

$$(H\mu)(p(s_t|I_t)) = \max_{c_t^i \in \boldsymbol{C}} h(p(s_t|I_t), \Omega_t^i, \mu_t)$$

Where H is an isotone mapping and such that value-functions are estimated per each iteration as

$$\begin{aligned} h(p(s_t|I_t), \Omega_t^i, \mu_t) \\ &= \sum\nolimits_{s_t \in S} u(s_t, \Omega_t^i) p(s_t|I_t) \\ &+ \lambda \sum\nolimits_{\psi_t \in \Psi} \sum\nolimits_{s_t \in S} (p(\psi_t|p(s_t|I_t)., c_t^i) \mu^*((p(s_{t+1}|I_{t+1})) \end{aligned}$$

The error in the belief state is estimated using the error in the estimated and observed belief state

$$g(s_t, \Omega_t^i) = E\|s_t - \overline{s_t}\|^2 + (1 - u(s_t, \Omega_t^i)$$

Ideally the estimations should continue until $g(s_t, \Omega_t^i) = 0$. However, in practice we stop the iteration well before it reaches the limit solution (10^{-5}). Finally, the optimal audience-content selection is performed $\forall\, t$ using the belief to action mapping.

7.2 Experimental Setup

Marketing and advertising provide stimulus to the user, who may react in a particular way (positively or negatively) based on their situation and/or brand attitude, which are too broad to codify. To test this framework, we set up a multi-market campaign with an international travel client on Meta (previously Facebook), to achieve cultural relevance and audience nuance while activating centrally across multiple markets. The objective was to demonstrate higher engagement rates by deviating from the practice of using a single, or small, number of video creatives, in order to drive relevance for targeted audience segments who were identified as being in market for travel, with no preconceived notion of good or bad creatives or strategy.

The campaign messaging was set up under the narrative of destination of travel, reason to travel, and a set of activities at that destination. These creatives along with copy texts constituted our ads. Initial guard rails were put in place to make sure that content mapped to each of the destinations was culturally and contextually relevant. The campaigns were set up in 7 markets for a period of two months and with over 7.2 million possible valid versions of the ads. From the audience point of view, the variability was in the audience dimensions, such as age, household compositions, travel affinity, lifestyle statements, etc.

Once we combined these audience attributes with the creative attributes, we ended up with over 83 million potential audience-creative combinations. Thus, trying to find the optimal combination of audience and creative manually, without noise and bias is intractable. To gauge the efficiency of the creative and audience selection, we set up an in-platform A/B experimentation setup[4] on Meta which ensures that a user exposed to the control is not included in the tests branch and hence not exposed to any test creative. The test branch was set to receive regular creative updates when our optimization framework deemed appropriate, i.e. there was a candidate with higher predicted performance as compared to the current/observed. Meanwhile, the control branch was business as usual with manual optimization. The selection of the control branch was based on an expert (manual) decision of the SME. This included both the audience groups and the creative ads. The experts identified three audience groups, young affluent professionals travelling for business or leisure, families travelling for holiday or relaxation, and the final catch-all category (people in market for travel). This targeting was done based on in-platform targeting on Meta. The experts then identified creative ads, which based on their belief, would resonate well with certain audiences in each market. To ensure comparability between the two branches the test was also restricted to three ad-sets[5]. Thus, when a new belief point is identified within the belief set (Gordon *et al.*, 2021), we overwrite the targeting specifications for the ad-set and attach the recommended creative against it. The in-platform optimization target was set as “Consideration”, with View Completion Rate (VCR: 100% of the video ad viewed) as the metric, and pacing was set such that the daily budget was spent as soon as possible.

8. Results

During the campaign, we served 47.11 million impressions at a frequency of 6.91 ads per person and recorded view completion rates (VCR) and the Cost Per Dated Search (CPDS). This measures how many of the users exposed to lower funnel (usually) brand activity then go on to search for a brand/product keyword. As for the lower funnel activity it might be hard to measure intent (that drives sales/revenue), CPDS helps us quantify this by measuring the cost of impressions [CPM] that led the user to demonstrate their intent via search. Please note that CPDS is recorded in averages post-campaign. On the test branch (optimised via

[4] https://www.facebook.com/business/help/1738164643098669?id=445653312788501

[5] An ad-set is the placeholder for which the auction takes place in Meta, i.e., where audience strategy & creative come together in the platform.

AI) of our campaign we served a total of 824 different ad-versions, while on the control branch (manually maintained and optimised) the setup consisted of 9 manually selected ads chosen by the SMEs.

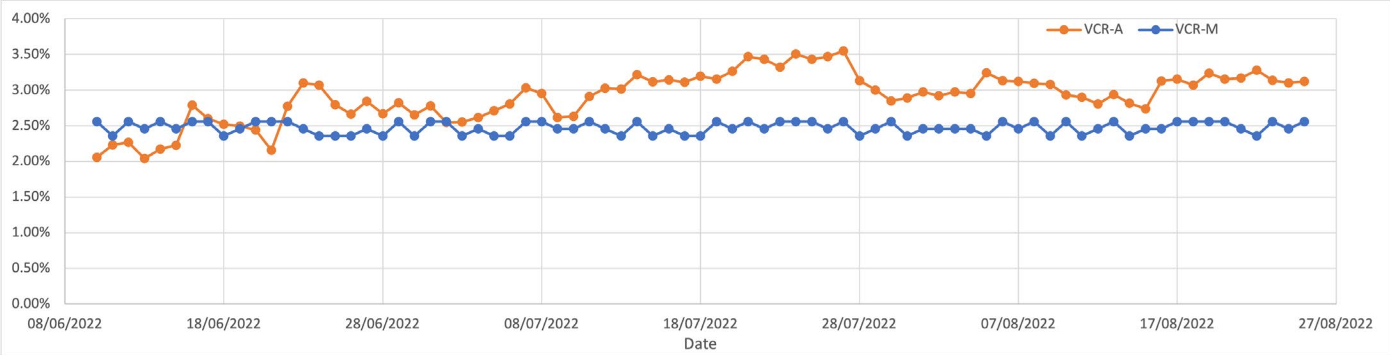


Figure 3. Comparison of Blended VCR between the AI-optimised Ad-Sets (VCR-A) and the manually optimised Ad-Sets (VCR-M).

Initially, we see that control ad-sets (VCR-M) were performing better than the test ad-sets (VCR-A) with the test still in its calibration phase. However, around the 11-day mark, we see that VCR-A started to stabilise. By the end of the experiment, we saw that engagement on control had risen by 31.4% and VCR-A outperformed VCR-M by 17.8%. The corresponding gap in CPDS between the test and control measured at 18.6%. Meaning that the test was directed more toward search than control.

The two major outcomes from this experiment can be summarised as:

1) The massive scale of candidate sets prohibits the delivery of the optimal ad by human and manual decision making, which in turn limits the stimulus that the user receives. Even more so, the manual selections are heavily biased and limit the campaign performance.

2) Any decision-making on identifying the next best creatives not only looks at the immediate reward/uplift but also the long term (within the horizon H) to ensure that campaign fatigue and frequent creative switches are minimised.

8.1 But don’t algorithms have flaws?

A common argument against using such algorithms is that they can perpetuate discrimination and bias and are black boxes. It is true that whilst it is possible to produce algorithms that eliminate noise, they may contain bias. The bias can be caused either by design (using predicators that may correlate with bias/discrimination) or could come from the source training data. If an algorithm is trained on a data set that is biased, then the algorithm will be biased. However, we can test whether an algorithm displays bias, and we are starting to understand how we can reduce and/or remove these biases. The fact that we can do this becomes a distinct advantage of algorithms over human decision-making. It is much harder to subject humans, where decision-making is opaque, to the same scrutiny as we currently subject algorithms. On this basis, there is a strong argument that algorithms are more transparent than humans.

9. Conclusion

People are willing to give an algorithm a chance but stop trusting it as soon as they see it makes mistakes. We all make mistakes, but it seems we're not prepared to share this privilege with AI. We expect machines to be perfect. Unfortunately, due to our intuitive expectations in marketing & advertising, we are unlikely to suddenly start trusting these algorithms and are more likely to keep using human decision-making, even when it produces inferior results. Our attitudes are only likely to change when AI starts producing near-perfect predictive accuracy.

Our inescapable conclusion is that, although a predictive algorithm in an uncertain world is unlikely to be perfect, it can be far less imperfect than noisy and often biased human decision-making. The challenge for those of us involved in improving automated decision-making will be to design and build algorithms that can not only do better than human-based decision-making, but are also accurate, fair, and free of bias. Only then may our flawed intuitions be overcome.

References

1. Grimstad, S., & Jørgensen, M. (2007). Inconsistency of expert judgement-based estimates of software development effort. *Journal of Systems and Software*, *80*(11), 1770–1777.

2. Einhorn, H. J. (1974). Expert judgement: Some necessary conditions and an example. *Journal of Applied Psychology*, 59(5), 562–571.

3. Colbert, J. L. (1988). Inherent risk: An investigation of auditors' judgments. *Accounting Organizations and Society.*

4. Anderson, J. M., Kling, J. R., & Stith, K. (1999). Measuring Interjudge Sentencing Disparity: Before and After the Federal Sentencing Guidelines. *The Journal of Law and Economics*, *42*(S1), 271–308.

5. Wang, J., Zhang, W., and Yuan, S. (2016b, October 7). *Display Advertising with Real-Time Bidding (RTB) and Behavioural Targeting. arXiv.org.*

6. Chapelle, O. (2015). Offline Evaluation of Response Prediction in Online Advertising Auctions. *The Web Conference.*

7. Balseiro, S., Besbes, O., & Weintraub, G. Y. (2015). Repeated Auctions with Budgets in Ad Exchanges: Approximations and Design. *Management Science*, *61*(4), 864–884.

8. Facebook - About Ad Auctions. (n.d.). Facebook.

9. Ali, M., Sapiezynski, P., Bogen, M., Korolova, A., Mislove, A., & Rieke, A. (2019). Discrimination through Optimization: How Facebook's Ad Delivery Can Lead to Biased Outcomes. *Proceedings of the ACM on Human-Computer Interaction, 3*, 199.

10. Tunuguntla, S., & Hoban, P. R. (2021). A Near-Optimal Bidding Strategy for Real-Time Display Advertising Auctions. *Journal of Marketing Research, 58*(1), 1–21.

11. Waisman, C., Nair, H. S., Carrion, C., & Xu, N. (2019). Online inference for advertising auctions. *Research Papers in Economics*.

12. Tyler, S. C., Pandey, S., Gabrilovich, E., & Josifovski, V. (2011). Retrieval models for audience selection in display advertising. *Conference on Information and Knowledge Management*.

13. Grunt, O., Plucar, J., Šťáková, M., Janečko, T., and Zelinka, I.. Modeling of marketing processes using markov decision process approach. In: *Advances in Intelligent Systems and Computing. Volume 679*. Cham: Springer, 2018. p. 468.

14. Provost, F., Dalessandro, B., Hook, R., Zhang, X., & Murray, A. (2009). Audience selection for on-line brand advertising. *Knowledge Discovery and Data Mining*.

15. Gasparini, M., Nuara, A., Trovò, F., Gatti, N., & Restelli, M. (2018). Targeting Optimization for Internet Advertising by Learning from Logged Bandit Feedback. *International Joint Conference on Neural Network.*

16. Geng, T., Lin, X., & Nair, H. S. (2019). Online Evaluation of Audiences for Targeted Advertising via Bandit Experiments. *ArXiv (Cornell University)*.

17. Ju, N., Hu, D., Henderson, A., & Hong, L. (2019). A Sequential Test for Selecting the Better Variant. *Web Search and Data Mining*.

18. Ba, W., Harrison, J. M., & Nair, H. S. (2022). Advertising Media and Target Audience Optimization via High-dimensional Bandits. *ArXiv (Cornell University)*.

19. Huang, X., Gu, C., Zhang, H., Yang, X., Liu, X., Tang, J., & Liu, H. (2019). DEAR: Deep Reinforcement Learning for Online Advertising Impression in Recommender Systems. *Proceedings of the AAAI Conference on Artificial Intelligence, 35*(1), 750–758.

20. Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017a). Densely Connected Convolutional Networks. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*

21. Daniyal, F., & Cavallaro, A. (2011). Multi-camera Scheduling for Video Production. *Conference on Visual Media Production.*

22. Bellman, R. (1957). Dynamic Programming. *Princeton University Press.*

23. Cassandra, A.R. (1998). Exact and approximate algorithms for partially observable markov decision processes.

24. Denzler, J., & Brown, C. (2002). Information theoretic sensor data selection for active object recognition and state estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 24*(2), 145–157.

25. Gordon, B. A., Jerath, K., Katona, Z., Narayanan, S., Shin, J., & Wilbur, K. C. (2021). Inefficiencies in Digital Advertising Markets. *Journal of Marketing, 85*(1), 7–25.

Author

Fahad Daniyal is Lead Data Scientist, for the Data & Commerce Practices, at Publicis Groupe and heads up the Data Sciences, Engineering, and Development within the practice. He received his BE in Computer Systems Engineering, an MSc in Digital Signal Processing, and PhD in computer vision delving in the Markovian and Bayesian paradigms. Within the group his work is focused on inferring causality in both content and media space. He is working towards the adaptation of techniques focused on agent-based, generative and reinforcement learning models. In the past he has worked on developing systems for biometric and document verification, risk aversions in supply chain, and multi-modal content analysis.

fahad.daniyal@publicismedia.com

Jason Smith's focus at Publicis Groupe is on decision intelligence. He is a strategic, technical and commercially astute executive who has over 20 years of experience of starting, growing and managing data & tech businesses. He has worked at corporate level, leading a global data innovation team focusing on privacy, data sharing, consent & personalisation. He has also co-founded data technology start-ups and data research labs, the former being a social media data intelligence software using AI natural language processing for sentiment and emotional analysis and topic clustering, and the latter, a research data lab as part of ScaleUpNation, in Amsterdam, using machine learning & network science to research 'what lies behind a successful scale-up'. Prior to joining Publicis Jason was working with two start-ups, one developing verifiable credential & token technology infrastructure and another developing an AI governance platform.

Jason has written, presented and produced 3 radio documentaries for BBC on technology ('Being Social' on social media & data; 'Becoming Artificial' on AI & what it means for humans and 'Going Viral' on the use of technology during the first Covid lockdown). Jason has also written presented on data & AI and he is a member of the European AI Alliance.

jason.smith@publicisgroupe.com

marketscience

OUR DIFFERENTIATORS

+ Academic Validation and Backing
+ Total Demand Measurement and Optimization
+ Software, Enablement and Scaling
+ Independent and Transparent

OUR SERVICES

Consulting · Studio · Academy

CONTACT US AT
INQUIRIES@MARKETSCIENCE.CO

VISIT US AT
WWW.MARKET.SCIENCE

FRONTIERS OF MARKETING
DATA SCIENCE JOURNAL

ISSUE No. 4 | 2023
www.i-com.org