Practice papers

Using lift-testing to measure the true value of digital marketing in the cross-device world

Received (in revised form): 8th February, 2016

Dominic Williamson
is a marketing attribution specialist and currently a lead in the marketing science team at Facebook. With experience at analysis consultancy dunnhumby, media owner JCDecaux and major digital marketer eBay, he has faced the challenges of marketing measurement from all angles. Originally from Nottingham, his professional journey has taken him to roles in England, Switzerland and California and he now works in New York.

Facebook, 770 Broadway, New York, NY 10003, USA
E-mail: domw@fb.com

Abstract Digital marketing attribution faces a number of challenges as the online landscape evolves. Flaws in last-click models are being exacerbated by the growing move to cross-device usage. With this trend unlikely to change, it is clear that accurate digital marketing measurement requires an understanding of the behaviour of people, rather than devices. This paper presents lift-testing as a methodology that allows for the building and validation of a robust cross-channel digital measurement framework. Examples are provided to show how lift-testing can be used drive more accurate spend decisions than last-click models, as well as how it can be used for both between- and within-channel optimisation. This paper also calls for a shift in mindset within digital marketing analysis to ensure a focus on true incremental impact from both advertisers and publishers.

KEYWORDS: digital marketing, cross-device, lift test, measurement

'I DON'T KNOW WHICH HALF'

> Half the money I spend on advertising is wasted; the trouble is I don't know which half.

This quote, widely attributed to John Wanamaker, is comfortably the most famous adage in advertising measurement. It highlighted the marketer's plight. In a world of broadcast media, accurately identifying the effect of each component of a campaign was effectively impossible. Advertisers, however, still had reason to believe that, at the aggregate level, the impact justified the investment. As econometric modelling techniques developed, marketing mix models (MMMs) began to shine a light into the hitherto murky world of attribution. There was now a technique and, perhaps more importantly, a wider desire to calculate the impact of advertising spend. The difficulty and expense of MMMs placed them out of reach of many advertisers but the need for accountability remained. And then, around 100 years after Wanamaker's lament, the rise of digital marketing promised impact and accountability for all.

For the most part it delivered. The sheer amount of data allowed for a much more granular analysis of the influences a person had before conversion (in this context 'conversion' refers to the behaviour that an advertiser intends to drive, generally sales or sign-ups). The dominant technique that emerged was the last-click model. This model attributes a conversion to the last marketing channel a user clicked on prior to a conversion event (within a given latency window). For example, if a user makes a purchase from an online retailer and is known to have clicked on, say, a paid-search ad before the purchase, but within the latency window (often 24 hours), then the value of that transaction is attributed to the paid-search ad.

This felt like a huge leap forward from the broad assumptions made for broadcast channels. Since we know that this person clicked on the link and that they made a purchase, the level of insight would, on the surface, appear to be much greater. And to this day it remains the most common form of digital marketing attribution. Digital marketing is ever changing, however, and last-click faces some major challenges to its accuracy and relevance.

WHY THE LAST-CLICK?

Few marketing campaigns exist in a vacuum and thus we arrive at our first conundrum: why the *last* click? Most advertisers will run digital campaigns across multiple publishers simultaneously. If a person clicks on one or more ads before converting, then selecting the last one seems like an arbitrary way of attributing impact. Why not the first-click or a combination? If the order in which users click is random then this is a moot point as the results will be a wash, but the reality of marketing is that certain channels will fall consistently earlier in the purchase cycle. In the industry this problem is termed an 'arbitrary credit assignment … imposed upon the chain of advertising channel touches preceding conversion'.[1]

As a result, last-click models demonstrate a skew in the way in which they attribute value across different click types, but this also raises an even more fundamental question: why the last *click*?

Any model dependent on clicks, whether the last, first or otherwise, is underpinned by the important but unsubstantiated assumption that clicks cause conversions. Consider again the purchase on our theoretical online retailer. If this retailer expects to experience zero traffic without marketing, then the assumption of causality is well founded: the click, visit and corresponding purchase must be incremental. If this retailer has a high expected baseline of traffic, however, the assumption becomes weaker. If the converter was already an existing customer, visiting daily and purchasing regularly, then click and conversion may well be entirely coincidental. The assumption of causality therefore becomes increasingly weak as baseline traffic grows. In broader terms it is the unproven assumption that if a conversion occurs *through* a given ad then it also follows that the conversion occurs *because* of that ad.

View-through effect is an impact missed by last-click models. While it is clear that not all conversions are driven by the clicks that precede them, it is also true that conversions can be driven by exposure to advertising without a click. This is the 'view-through effect'. If any non-digital advertising has an impact, it must be view-through by definition. If television, press and outdoor advertising can influence behaviour without the need for a click then it is fair to assume that same phenomenon can exist in digital campaigns.

Last-click models can underestimate impact due to missed view-through, overestimate due to the assumption of causality, and either over- or underestimate due to click competition between channels. This problem is exacerbated by the overarching and growing impact of multiple device usage. More than 75 per cent of Americans who access the internet do so across both desktop and mobile devices.[2] Cookies, for so long the bedrock of digital attribution, work at a device rather than a person level. A user can therefore view, click or purchase from any or all of their devices without it being apparent that all the events are tied to a single person. Cross-device usage is growing and is unlikely to stop. In order to accurately attribute conversions to prior advertising events it is essential that measurement be done at a person level.

ALL ARE WRONG, BUT SOME ARE USEFUL

The shortcomings of last-click are well documented. Reports of 'the death of last-click wins' over five years ago were, as it turns out, greatly exaggerated.[3] It remains the most prevalent form of digital marketing attribution. The reason is straightforward enough, as Box and Draper (1987) succinctly noted: 'Essentially, all models are wrong, but some are useful.'[4] The last-click model provides undoubted utility as well as simplicity. In the absence of a better alternative, it became an entrenched part of the digital marketer's life. Better alternatives do now exist, however.

Multi-touch attribution models (MTAs) look at the full conversion paths for individual users through every 'touch' (view or click) and, using a range of modelling techniques, attribute the value of each conversion. MTAs allow for view-through impacts, do not assume causality, and take into account the impacts of other channels. There are benefits of switching to MTAs but again these come with challenges. MTAs are complex and the exact methodology used varies widely. Inevitably this means that the workings of the model remain opaque to the majority of people.

The results of the model are also very difficult to validate. A well-informed digital campaign will not target a random set of users but those users deemed most likely to convert. Disentangling users who are likely to convert because they are shown an ad, and users who were shown an ad because they are likely to convert is a massive challenge. It is not impossible, but it would be naïve to think that across the full range of MTAs this challenge is universally well solved.

Without a 'source of truth' to show the incremental impact of a campaign, the results of a model are benchmarked against intuition. Intuition has a habit of reinforcing our existing beliefs.

THE LAW OF PARSIMONY

Whatever the attribution approach, it is clear that people-based analysis is key.[5] As a general rule, the fewer assumptions needed the better. Lift-testing is a solution that is both people based and assumption free. In its basic form, a lift-test takes a randomly split target audience and assigns a test and a control. The test group is exposed to advertising and the control group is unexposed. Uplift is calculated as the difference between total conversions in the test and total conversions in the control. Conversions are independent of clicks and can be tied to individuals through pixel or offline data. The conversions in the control show the baseline (or 'ambient') activity, so the difference between the control group and the test group is the incremental impact (Figure 1).

This simple approach provides fixes for the full range of challenges listed previously. Measurement is performed at the person level. Conversion can be measured anywhere that identity is captured — desktop or mobile, online or offline. So if a person sees and is influenced by an ad on mobile but converts on desktop the impact is captured. In-store impact, impossible through last-click attribution models, simply becomes an additional conversion channel.

The unexposed control provides the counterfactual. It represents the 'what-if' scenario where advertising is not present, so the many complex assumptions required for attribution modelling become redundant. We are left with two very clear to interpret values: total conversions for people exposed to advertising and total conversions for

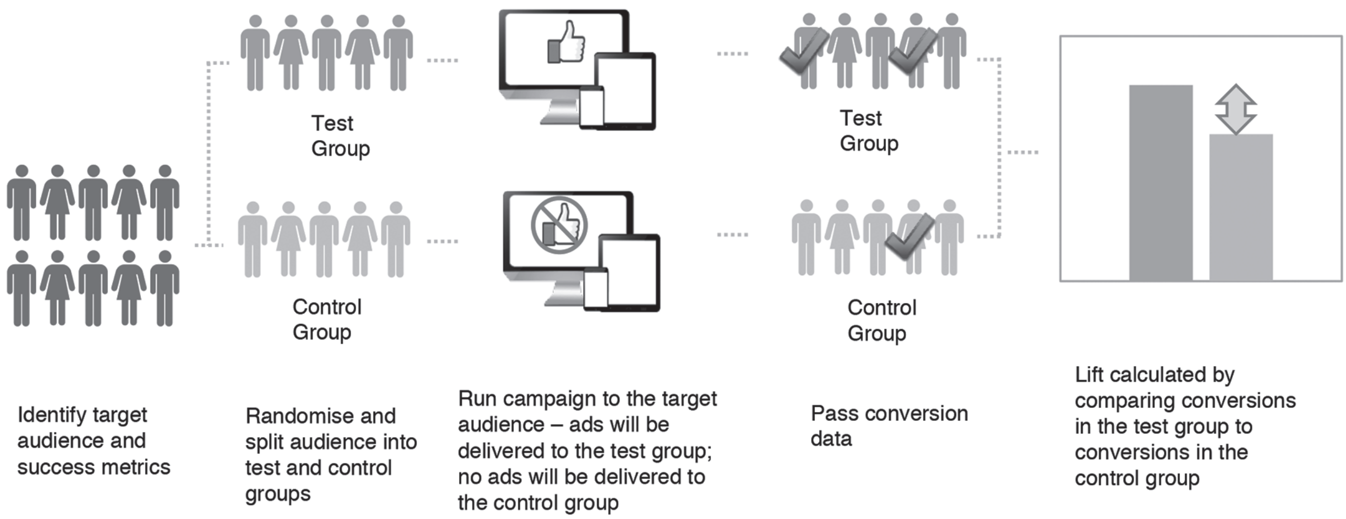

Figure 1: The lift-testing process

the corresponding people not exposed to advertising. Measurement is agnostic to clicks; therefore view-through impact is captured by default and the assumption of causality is nullified.

Lift-testing shares the same underlying methodology as the randomised control trials performed in medical testing. A target group is split, different groups receive different treatments, and responses are measured. RCT is the gold standard of measurement and a prerequisite for bringing new treatments to market.[6] When available, it behoves any client to apply that same analytical rigour to marketing campaigns.

Availability within a channel or publisher is dependent on the ability to target at the individual person level. E-mail lends itself well to lift-testing and direct mail has long relied on testing to drive optimisation. In digital channels, individual-level targeting generally requires a login to accurately tie identity across devices. In practice, the bulk of online advertising takes place outside logged-in platforms and this presents a challenge: once again measurement becomes reliant on cookies and their associated tracking shortfalls. Results can be diluted as gaps appear between exposure and conversion data.

One technique to mitigate this is the use of placebo ads (a nod back to medical testing). In a placebo test the unexposed hold-out is shown an ad unrelated to the advertiser (often for a charity). View tags can then be used to identify test and control groups. Placebo tests can also be inherently flawed, however. Algorithmic ad delivery targets people deemed the most likely to engage with a given ad. Delivery can therefore skew into different audience types. For example, charity placebo ads can skew towards charitable people. This skew violates the underlying principle that only one difference should exist between the two groups: exposure to ads.

In short, the limitation of lift-testing is tied to a platform's ability to execute. When available, however, lift-testing is the most powerful measurement option.

This power is reinforced by the unyielding march of big data. The increasing availability of data leads to sample sizes much larger than were previously available. Properly used, these data mean higher powered tests and the ability to detect smaller (but perhaps still business-relevant) uplifts.

Lift-testing is therefore more valuable than ever and the shift of users to mobile is fast making it indispensable. People are increasingly consuming ads and converting online, and it is essential to use an attribution technique that is able to tie those events together. This means using people-level analysis in order to neatly sidestep the problems that cross-device usage creates for tracking. Lift-testing does not just provide academic benefits; it is a very pragmatic way of dealing with the evolving consumer landscape.

MEASURING IMPACT, NOT CLICKS

The question for advertisers is whether business decisions based off lift-tests are more accurate than decisions based off alternatives. As of January 2015, Facebook offers managed clients the opportunity to measure campaigns through person-level lift-tests. Swiss online retailer DeinDeal ran a Facebook advertising campaign and measured the impact both through lift-test analysis and last-click modelling.[7] The results showed last-click underestimated the incremental new buyers by around 30 per cent or, to put another way, for every new buyer identified through last-click the true incremental value was closer to 1.5. UK health-food retailer Graze found similar results with last-click underestimating incremental online subscriptions by around 28 per cent.[8]

Consistent misestimates of publishers or channels leads to suboptimal cross-channel budget allocation. Within-channel optimisation, however, represents an even larger opportunity. Last-click inevitably draws spend into segments of people with a high likelihood to click, rather than people with the highest propensity to respond favourably to advertising. Consider the chart in Figure 2 showing responses from an online fundraising campaign for a senate race. The campaign was targeted at previous donors, so it is unsurprising that people in the unexposed hold-out group show a high baseline level of donation. The assumption of click causality is particularly weak here. This weakness manifests itself in a disconnect between segments that demonstrate high levels of uplift and segments that demonstrate high click conversions.

Figure 2 shows that click models imply most efficiency in 55–64 and 65+ age groups, whereas true incremental value can be seen in the 35–44 age group. In short, the older demographic clicks and donates more but would also donate more without

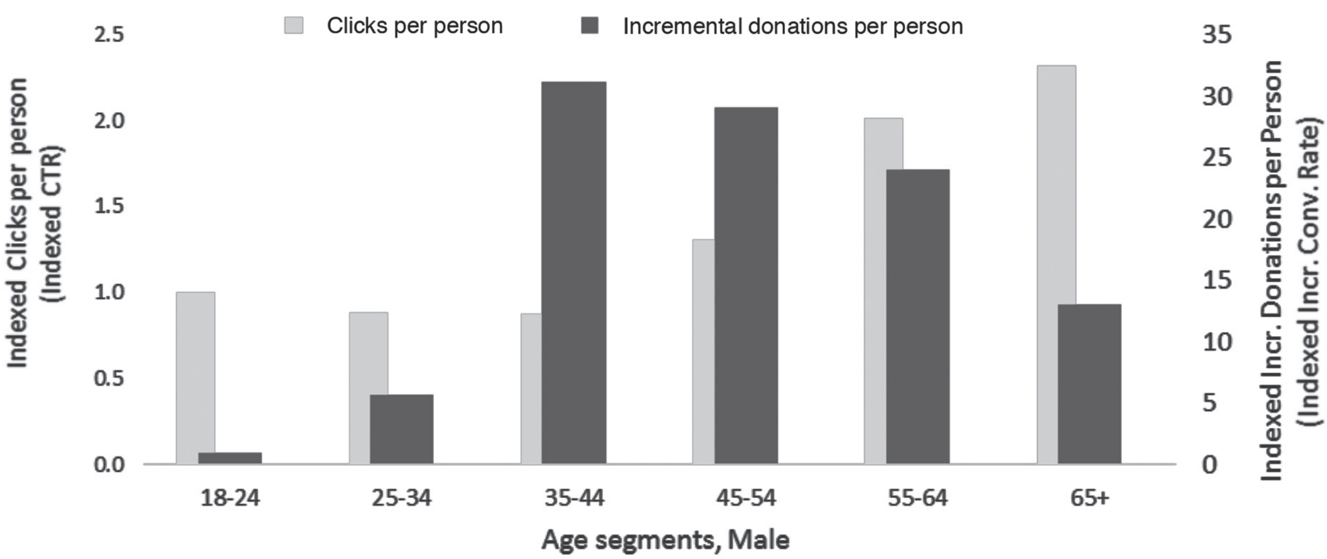

Figure 2: Responses from an online fundraising campaign for a senate race

Source: Facebook Marketing Science research with Trilogy Interactive, Conversion Lift analysis of a US fundraising campaign, Q3 2014.

advertising. The 35–44 group shows the greatest increase in donation as a result of exposure to the ads. Effective optimisation would target the 35–44 group against the recommendations of a last-click model.

FROM 'WHY' TO 'BECAUSE'

The biggest hurdle to lift-testing is the relative lack of availability. Last-click and MTA models provide a framework to provide some degree of measurement across any digital channel with tracking capabilities. In order for measurement to be fit for purpose, the mindset needs to shift from 'what happened through a *given* ad' to 'what happened *because* of a given ad'. In the medium term, MTA models provide the most viable way to execute this shift across all channels. Where available, however, it would be remiss not to use lift-testing as a way of building and validating these models. It is therefore essential, as an industry, to push for the greater usage and availability of lift-testing in order to evolve the measurement and effectiveness of digital advertising. Advertisers need to demand lift-testing from their media channels, and media channels need to build out the infrastructure needed to meet these demands.

In conclusion, while attribution has always faced challenges and will continue to do so, these challenges are not insurmountable. Indeed, the development of ad tech methodologies allows for both better measurement and more efficient execution. And, as much as we find comfort in the status quo, measurement evolution is inevitable, so it is incumbent on all of us to drive it forward.

References

1. Marr, J. A., Olmedo, O. O. and Borne, K. D. (2015) 'Advanced attribution analysis: A data science approach', 20th May, Syntasa Data Science Services, Herndon, VA, available at: http://cdn2.hubspot.net/hubfs/407782/SYNTASA_Attribution_White_Paper.pdf (accessed 5th February, 2016).
2. Dreyer, K. (2015) 'Mobile internet usage skyrockets in past 4 years to overtake desktop as most used digital platform', 13th April, *comScore*, available at http://www.comscore.com/Insights/Blog/Mobile-Internet-Usage-Skyrockets-in-Past-4-Years-to-Overtake-Desktop-as-Most-Used-Digital-Platform (accessed 5th February, 2016).
3. Lee, G. (2010) 'Death of "last click wins": Media attribution and the expanding use of media data', *Journal of Direct, Data and Digital Marketing Practice*, Vol. 12, pp. 16–26.
4. Box, G. E. P. and Draper, N. R. (1987) 'Empirical Model Building and Response Surfaces', John Wiley & Sons, New York.
5. Hogan, M. (2014) 'The challenge of digital marketing attribution across internet devices', *Applied Marketing Analytics*, Vol. 1, pp. 6–12.
6. Meldrum, M. L. (2000) 'A brief history of the randomized controlled trial: From oranges and lemons to the gold standard', *Hematology/Oncology Clinics of North America*, Vol. 14, No. 4, pp. 745–760.
7. 'DeinDeal', *Facebook for Business*, available at https://www.facebook.com/business/success/deindeal (accessed 5th February, 2016).
8. 'Graze', *Facebook for Business*, available at https://www.facebook.com/business/success/graze (accessed 5th February, 2016).