Addressing Attribution Breakage
Through Post Purchase HDYHAU Solutions

Abstract

The ability to optimize campaigns relies heavily on knowing how your media is performing. This document describes how post-purchase How Did You Hear About Us (HDYHAU) surveys can be utilized to supplement direct measurement (vanity URL, promo code, pixel, etc.) to provide a holistic picture of what media campaigns are driving. This information allows for proper campaign optimization and ultimately makes profitable scaling possible.

Problem Statement

Despite direct response advertisement often having clear and unique call to actions (CTAs), it can be challenging to determine the impact of your ads accurately and comprehensively.

Adherence to the CTA (visiting a vanity URL, calling a specific phone, utilizing a specific promo code, etc.) tracking mechanism is considered a direct signal, and is usually easy to tie back to the originating media. However, there is usually a large subset of consumers where demand or interest was generated by the ad, but they engage with the brand through other methods, circumventing the CTA and direct tracking mechanism.

This activity that is not captured via the direct tracking is referred to attribution breakage. Without accounting for attribution breakage, which can be a significant proportion of activity, you do not have an accurate or complete picture of what the media is driving and how best to optimize (especially across different channels). Therefore, attribution breakage needs to be accounted for to effectively optimize campaigns.

Background

Before exploring ways to address attribution breakage, it's important to level set on terminology and understand the different components to attribution.

Direct Tracking – Measuring demand through adherence to call to action.

Within each ad or influencer integration should be a call to action (CTA). A CTA is the prompt for immediate response and encourages viewers to act in a specific way, such as visiting a website.

To help with sourcing of the response action to media exposure, vanity URLs should be communicated within the CTA. They should be specific to the media that it is run on – more granular tracking allows for better optimization.

Examples of Unique Direct

Ad Placement	Tracking URL	Promo Code
Podcast A	www.clientsite.com/podcastA	PromoA
Podcast B	www.clientsite.com/podcastB	PromoB
Podcast C	www.clientsite.com/podcastC	PromoC
Radio Station X	www.clientsite.com/radioX	PromoX
Radio Station Y	www.clientsite.com/radioY	PromoY
Radio Station Z	www.clientsite.com/radioZ	PromoZ

Although direct tracking does not capture all the activity that is generated by our ad campaigns, it is very helpful to see how individual placements or tactics are performing compared to one another. It allows for more granular performance breakouts that are often not available via indirect methods.

There are several factors that impact the strength of the direct signal, but biggest driver typically is a promotional offer or something to incentivize adherence to the CTA. In the absence of a compelling offer, advertisers must more heavily rely on indirect signals to gauge channel performance.

Indirect Tracking – Measuring overall demand generated from the ad campaign.

As mentioned before, the direct tracking will not capture all the activity generated by ads campaigns, so indirect methods can be utilized to supplement direct tracking and help gauge overall demand and activity.

Probably the most common indirect tracking method is lift over baseline. This is the exercise of observing the baseline activity levels that existed prior to the media campaign(s) launching, and then when the campaigns launch, attributing any incremental activity to the campaigns. Often, control or holdout markets (with no exposure to the media) are put in place to account for any organic fluctuations within the business (either increases or decreases in activity that would have occurred regardless of media presence). In these instances, changes over baseline in test markets would be compared to changes over baseline in the control markets, and the difference between test and control would be attributed to the media.

Lift over baseline is an effective way to determine holistic impact of media campaigns but can be challenging when there are multiple media channels being run. Also, when the existing baseline is high (or not stable), it can be challenging to get a read on performance. Thoughtful test design is required to help ensure that there is actionable insights and learnings from the campaign.

If both direct and indirect tracking mechanisms are setup for an ad campaign, direct and indirect signals can be compared to determine the level of attribution breakage (amount of activity that is not being captured via direct tracking).

Multipliers

This attribution breakage can be accounted for utilizing multipliers that are applied to the direct tracking signal. Doing so allows for a holistic view into each placement's contribution to overall performance.

	Direct Signal		Indirect Signal			Combined Signal	
Influencer	**A1** Vanity URL Orders	**A2** %	HDYHAU YouTube Response (Raw)	Survey Response Rate	HDYHAU YouTube Response (Gross)	**C1** Multiplier	**C2** Attributed Results (X)
Influencer A	30	30%					90
Influencer B	20	20%					60
Influencer C	10	10%	150	50%	300	3	30
Influencer D	40	40%					120
Totals	100	100%	**B1**	**B2**	**B3**		300

Multiplied results are commonly noted by an "(X)" **representing our best estimation of what the media is driving.**

A1 – Direct signal by placement

A2 – % of direct redemptions by placement

B1 – YT survey responses

B2 – Survey response rate: survey responses divided by total orders

B3 – Gross YT responses (simulating if everyone responded to survey): podcast response (B1) divided by response rate (B2)

C1 – Divide gross responses (B3) by direct signal (A1) to get multiplier

C2 – Apply multiplier (C1) to each of the placements (A1) to get attributed results

Multipliers can vary significantly by campaign and vertical, but it is common to see breakage multipliers between 2X – 10X (or even higher). When there is not a compelling offer associated with the direct signal, multipliers can be substantially higher. Multipliers should be calculated for each channel.

Offline Channels:

Relative Magnitude of Typical Breakage Multiplier

Podcast and satellite typically have the lowest multipliers (strongest direct signal) based on listeners being conditioned to follow CTA and strong affinity to the host (and desire to support).

Attribution breakage can occur across all channels, but typically is a larger challenge for offline channels (linear TV, terrestrial radio, podcasts, etc.). It is also common for digital channels where endorsements are often in-content and consumption is on-demand on the consumer's timeline.

Traditional Offline Channels

When someone refers to offline channels, they typically are referencing linear television or terrestrial radio. Direct tracking is straight forward for these channels and spike or spot time attribution is utilized to measure attribution breakage and supplement the direct signal. For spike attribution, ad spots within these channels run at a certain time and advertisers can make assumptions to which ads drove the activity based on response or decay curves occurring right after the ad running.

Spike or spot time type attribution works much better for television than for radio, primarily due to media consumption behavior:

- **Radio:** Radio is often consumed while commuting, so even if demand was generated from a radio ad, often the consumer cannot respond immediately because they are driving or moving from one spot to another. This makes spot time attribution less effective because response times are less correlated with when the ad ran, and response spikes are flatter. Also, since more time usually transpires between when the ad ran and the response, the likelihood of remembering the CTA is lower, so there will likely be more attribution breakage.

- **Television:** Spot time attribution is much more effective with television advertising because the consumer is more likely to be in a position where they can respond immediately. This results in larger activity spikes occurring shortly after the ad runs. This immediacy, as well as the fact that the CTA is visually displayed on screen, also makes it more likely for the consumer to remember the CTA, which strengthens the direct signal and reduces attribution breakage.

Podcast

Spike attribution is not effective to capture attribution breakage for podcast advertising due to the on-demand consumption of podcast media. Consumers download or stream podcasts on their own timelines and advertisers do not have insight into when the media is being consumed (or if it is consumed at all). Therefore, there are typically not

concentrated activity spikes that can be tied back to podcast ads. The direct tracking signal is still viable, but advertisers need alternative ways to measure the attribution breakage.

Originally, a pixel solution (fingerprinting consumers that were exposed to the media and later converted on the advertisers' site) like that commonly used in digital advertising, could not be leveraged for podcast attribution due to limitations in how the media was being delivered. However, over the last couple of years, pixel solutions have been developed to tie back the download of podcast to activity on the advertisers' website, which in theory can account for attribution breakage.

Challenges with Pixel Attribution:

- **Does not track ad exposure for podcast:** It is important to point out that pixels can determine if a podcast is downloaded, not whether the consumer is exposed to the ad. This is important because many podcasts are auto downloaded and never get listened to. This can present a high rate of false positive matches and over-attribution.

- **Hard to determine appropriate lookback window:** Since podcast consumption is on demand, it is difficult to determine the appropriate lookback window. A lookback window is the amount of time from media exposure (or in this case podcast download) to activity on site. For many digital channels, such as online video, display, etc, the lookback window is set to 7-14 days from media exposure. Since for podcast, the lookback window begins at podcast download not media exposure, a longer lookback window of 30 days is typically used to account for latency between podcast download and podcast consumption. As you can imagine, results will vary dramatically based on what lookback window is used - often the volume attributed and therefore the perceived success of a campaign is dependent on the lookback window utilized. There needs to be a way to calibrate or configure the appropriate lookback window for an advertiser (and even for podcast genre).

Influencer Marketing

For influencer marketing, attribution breakage is often lower, and the direct signal is stronger given the consumption behavior (users are typically engaged and consuming through devices such as phone, where the action step is just a click away).

The on-demand consumption makes spike attribution challenging and social media platforms will not allow the pixeling of the in-content ads. Lift over baseline is a viable option, but the signal becomes muddied if there are multiple channels running.

Digital Advertising

Traditional Digital attribution also faces challenges. Pixel attribution is very common for digital advertising, but attribution breakage can still occur if media exposure occurs on one device and conversion occurs on another device – attribution partners try to account for this via device graphs, but this does not always capture everything. Also, evolving privacy regulation may hamper the effectiveness of this type of attribution in the future.

On the flip side, pixel attribution may over-attribute to digital channels. Within a 7-day period, a consumer could be exposed to multiple ads on different platforms, and all of those platforms will take credit for the activity. This duplication of conversions leads to inaccurate optimization, so there is the need for additional performance signals.

Solution

A post-purchase How Did You Hear About Us (HDYHAU) survey can be a simple but powerful tool that accounts for attribution breakage, duplicative data, and provides advertisers with the data needed to optimize campaigns. It can assist with offline attribution but also validate (provide guard rails) for digital advertising. HDYHAU surveys are also future-proof in the sense that they are not impacted by changing privacy legislation.

A HDYHAU survey is a way of determining how a consumer heard about the brand or product. It can be used to attribute activity back to an originating media and is a fair way to help evaluate the multiple and disparate performance signals across different channels and attribution methodologies.

HDYHAU surveys does not have to be the only performance or attribution signal. If possible, it should be used in conjunction and compared with other methodologies. Earlier we touched on the challenge of choosing the appropriate lookback window for pixel attribution – HDYHAU can be leveraged to calibrate lookback windows. Rather than simply utilizing industry standard lookbacks that may not be applicable to a particular advertisers' consumer path to conversion, HDYHAU results can be compared to pixel data to determine more appropriate lookbacks by channel.

HDYHAU surveys typically can be stood up relatively quickly given that there are several vendors and partners that provide these solutions.

There are several ways that this survey can be delivered, but it's typically served post purchase within the sales funnel. Some common alternate ways to serve the survey are upon site visit post purchase via an email. Here are some challenges with these alternate ways:

- **Site Visit:** Can have a detrimental impact on user flow and experience and it sometimes is challenging to track that through to an actual conversion.

- **Post purchase via email or Text:** Response rate are typically very low, especially if respondent is not incentivized to answer. Also, there is more time that passed from media exposure to conversion, so it may be more difficult for the respondent to accurately remember.

Post purchase, within the sales funnel (at checkout), is preferred because it's least disruptive to the consumer experience, you know that they actually made a transaction, and typically has higher response rates. Typical response rates for post purchase HDYHAU surveys at checkout is between 40-60% vs. single digit response rates for email surveys.

HDYHAU data is self-reported, so there is some skepticism regarding the accuracy of the data, but based on extensive historic analysis, Havas Edge is confident that if setup correctly, HDYHAU surveys are reflective of the impact that media is having on the business. Some reasons contributing to this confidence are:

1. We consistently see response patterns from surveys that mimic direct response patterns:
2. When a campaign launches, we see the gradual increase in survey responses for the new channel as frequency builds.
3. Conversely, after a campaign concludes, we see the gradual decrease in survey responses for the channel as ad stock wears out.

4. We have tested out "dummy" options within the survey and have seen consistent response. We have introduced media channels that clients are not advertising in to get a readout of survey noise. While we do get some responses for these dummy options in the survey, the volume is very minimal (usually well less than 2%) and is consistent week over week.

5. Once mature (typically a couple weeks into the campaign), we see consistency in the week over week attribution breakage and multipliers. If we see dramatic fluctuations, we are usually able to identify the cause and account for it. This volatility is usually indicative of one of the following:

    a. Promotional offer change – if the discount goes up, the direct signal will improve, and the multiplier will go down.

    b. Leaked promo code – if a promo code leaks and is not scrubbed from results, we will see an artificially strong direct signal and we will see a decrease in the multiplier.

    c. Sitewide promo or flash sale – sitewide promos will weaken the direct signal. Consumers will utilize the sitewide discount promo code rather than promo code within the ad if it's prominently displayed on screen but will still indicate the demand-generating media on the HDYHAU survey. This will increase the multiplier.

    d. Media coverage (outside of advertising campaign) – if advertiser is featured on a podcast, or there are several review videos on YouTube, the multiplier will increase giving false performance signals.

Certain steps that can be taken to help ensure accuracy and consistency of the data received:

- **The survey should not be mandatory:** Brands should allow respondents to skip the survey if they do not want to participate. Not allowing for this may result in random or inaccurate selections. Non-responses are preferred to forced and inaccurate responses. Responses should be grossed up to full response meaning if 50% of the people take the survey, responses should be multiplied by 2 (or divided by 50%) to simulate full response.

- **Calculate and exclude false positive baseline:** Surveys should be live prior to campaign launch with tested channels. This is to account for noise and to establish a baseline of respondents who may randomly select media channels even though advertising is not present. This baseline is considered false positives and should taken out of survey responses once the campaign begins.

- **Streamline survey choices:** Survey should not be too lengthy, otherwise we risk the respondent not reading through and just randomly clicking – depending on the scope of the campaign, you typically want to list channels, not individual placements on the survey.

- **Clarity in survey choices:** Help respondents respond accurately by making it very clear what the selection represents. For example, we recently have seen 'Influencer' become a more common survey option – this can be confusing to respondents because if they heard about it from an influencer on YouTube, should they respond "YouTube" or should they respond "influencer"? Avoid industry jargon and provide examples to clarify if needed.

- **Options should be rotated:** Randomize survey choices to mitigate recency bias. A static alphabetized list of choices may result in overrepresentation of the first or second choice in your results. To prevent bias in survey response, options should be rotated so that the same option is not always first.

- **Collect and analyze Open-Ended "Other" Responses:** Open-ended responses can often yield insights about new opportunities for marketing mix diversification. There may be a channel or method that consumers are finding the brand that the advertiser is not currently capitalizing on.

Conclusion

HDYHAU surveys are not the perfect attribution solution, but they are an often-overlooked tool that if used properly, can be utilized in conjunction with other measurement methodologies to fill in the gaps and corroborate performance signals. HDYHAU's are helpful in cross-channel measurement and in essence, represent the voice of the consumer.

About the Author

Brian Kim is the Director of Data Science and Modeling at Havas Edge. With over 14 years of working in the advertising industry, Brian has developed a mastery of data analytics and attribution. His personal mantra (or as he likes to say "battle cry") is to "manage by the numbers." Brian's analytic, numbers driven approach to partnerships has branded him as a source of truth for many of the businesses he works with.