June 14 – 16, 2015

Winning Papers
This year the ARF is pleased to present this digitally bound version of the white papers presented as both main stage and deep dive sessions at Audience Measurement 2015. Thank you to all the authors.

Enjoy!

*The ARF Team*
Table of Contents

Sunday, June 14
Design of a Smart TV ................................................................. 2

Monday, June 15
Applying Motivation to Digital Video Ads ....................................... 18
From Message Sent to Message Received ........................................ 27
How Much Time and Money Do People Really Spend on The Internet .. 39
Measuring Traditional Media Through Digital Data ......................... 41
Reach and ROI of Mobile and TV .................................................. 48
Why Aren’t You Looking at the Long Term ROI of Ads .................... 56

Tuesday, June 16
The Next Frontier ............................................................................ 74
Sunday, June 14
Design of a Smart TV Logging System considering Context of Audiences by Using Beacons and Smartphones

Jehwan Seo, Yeonhee Cho, Daesik Kim
Namjun Kang, Joongseek Lee, Bongwon Suh

May 5th, 2015
Abstract
In this paper, a smart TV logging system comprising a beacon system and smartphones is proposed. To
investigate the feasibility of our strategy, we designed and implemented a prototype system and
conducted a trial study. The study results show that the prototype can unobtrusively capture viewers’
various events embedded in TV viewing behavior. The results of the study also suggest that the
proposed method allows more robust and accurate data to be collected than do the TV viewing
behavior analysis approaches used in existing qualitative research studies, such as surveys and
interviews. As a future work, our system can be extended to the personalized recommendation service
through analyzing the collected data. Also, a Bluetooth-based system is useful for screening engaged
viewers. This system may have a tremendous impact in that it can match and classify the accurate TV
rating data in real-time. Lastly, it can implement the user context (local-based) N-screen service.

Author Keywords
TV viewing behavior, Smart TV, Beacon, TV rating, Audience measurement, Indoor Location-based
services

Introduction
TV rating is often measured by traditional research methods such as People Meter\(^1\), surveys, etc.
However, the capability of existing methods to observe and measure viewers’ behaviors is limited since
they are obtrusive. Recognizing a viewer’s behavior pattern could benefit both viewers and TV rating
firms: viewers can receive a personalized service based on sophisticated observed data while TV rating
firms can collect more precise data to build enhanced business models.

To address the issue, we design a novel system for observing the viewer’s behavior when located in
front of a TV by using a smartphone and Beacon. In this research, we are motivated to use common
digital devices to collect TV viewing behavior robustly and unobtrusively. We also collect app usage logs
to investigate a person’s multi-tasking activities on his/her smartphone while viewing TV. To verify the
feasibility of our strategy, we conducted a trial study in a single-occupant household. The results show
that a viewer’s accurate location can be measured unobtrusively using smartphones and Beacon
technology.

Related Work
Earlier approaches to research on TV viewing behavior
TV viewing behavior has been investigated by many researchers and organizations. Abreu et al.
examined the TV ecosystem through a survey in order to understand people’s TV viewing behavior
when simultaneously using smart devices [1]. Ericsson Consumer Lab conducted an annual TV and
Media study using quantitative and qualitative methods. In the study, the lab interviewed 23,000
people in 23 countries via online surveys and performed in-depth interviews with 22 people in San
Francisco, London, and Stockholm. They also interviewed 11 experts in the media industry [2]. However,
survey methods have limitations in that the collected information does not suffice for investigating
complicated behavior patterns.

---

\(^1\) People-Meter: http://en.wikipedia.org/wiki/People_meter
A number of research studies have, however, been conducted to investigate viewers’ behaviors by using digital devices. The analysis of log data generated by these digital devices allows researchers to measure audience activities beyond only the TV medium, broadening measurements to include the multi-screen experience and various other media systems [7, 8]. Among these devices, the people meter has been regarded as the de facto standard tool for measuring TV ratings.

**People meter**
The People Meter is an audience measurement tool invented by a British company called Audits of Great Britain Ltd. (AGB) to measure TV viewing behaviors. It is an electronic device that records when media are being used and who is viewing them. Members of participating households can transmit information about their TV viewing activities by pushing a button on the device. By collecting logs obtained through these devices, rating firms can observe and analyze behavior patterns in households. However, the People Meter can introduce bias and noise in the gathered logs. The device has a simple push-button interface for collecting the viewer’s status and information. Because of fatigue or ignorance of registering their behavior, viewers may avoid participating in the measurement system [5]. Thus, more unobtrusive systems that do not force viewers to participate have been developed, as follows.

**Portable People Meter and audio matching technology**
The Portable People Meter (PPM) is an electronic device developed by the Arbitron Company to track audience’s exposure to broadcasts, cable TV, and many types of digital media. Inaudible signals hidden in broadcast signals are detected by the PPM or software that can be downloaded to a mobile device. Unlike the People Meter, this portable device can be carried around all day and, therefore, can unobtrusively track viewer’s exposure to all media by analyzing the inaudible signals [3]. However, the auditory measurements contain potential biases due to environmental features. For instance, although a person is not viewing TV, the system can regard his/her behavior as viewing TV because of errors in the detection signals. In addition, auditory signals can be disrupted by noise in the person’s environment. Thus, more robust measurement systems to collect TV viewing behaviors need to be established.

**Design of a smart TV logging system**
In this paper, we propose a smart TV logging system that is capable of precisely detecting the TV viewing activities of each individual in a household. People Meter based on survey methods have limitations in the following cases.

- A new viewer does not register herself/himself as a viewer when another person is already watching the TV
- A viewer is outside the TV watching zone, but the TV is not turned off
- A viewer is not watching TV although she/he is located in front of it

In order to address these issues, we designed a system that comprises a beacon system and smartphones.
Figure 1. A prototype of the Smart TV logging system

Figure 1 illustrates the research prototype system we designed. The system procedures include:

1) Collecting the TV content information by recording the TV screen

2) Checking the location of the study participants through installed multiple beacon devices\(^2\), as shown in Figure 2

3) Gathering the smartphone usage logs by using the App Usage Tracker\(^3\)

The beacon system is a short-range communications technology for a next-generation smartphone that uses the Bluetooth Low Energy (BLE) technology. The technology can determine a user's location with an accuracy of a 1-m margin of error in an indoor environment. The use of beacon systems in indoor environments such as shopping malls is widespread. As shown in Figure 2, using beacon signals, researchers can accurately determine viewers' locations when they carry a smartphone. The fact that a smartphone can receive beacon signals without a pairing procedure allowed us to design the unobtrusive system that we propose in this paper.

\(^2\) Reco Co., Ltd: http://reco2.me/reco/?lang=en

The system assumes that people carry around their smartphones in the house. According to the Nielsen Cross-Platform Report of 2012 [6], 85% of tablet/smartphone owners use their device while watching TV at least once a month while 40% do so daily. The survey implies that the design of our system has significant potential.

An issue that was not previously addressed is that related to the third case in the list above where a person is not watching TV when s/he is located in front of it. In order to analyze this case, we collected smartphone usage logs to identify the relationships between using a smartphone and viewing TV. We implemented a prototype system in a real-world context.

**Trial Study**

To investigate the feasibility of our design, we conducted a trial study using one volunteer (male, 33 years old). The study was conducted using the prototype system in a single-occupant household for one hour (21:00–22:00) in the evening of a typical weekday. We selected a house that has one living room, one bedroom, and one bathroom. Three beacon devices were installed: in front of the TV, in the room, and in the bathroom (Figure 2). This setup allowed us to identify the viewer’s location accurately. We collected three types of data simultaneously for one hour as follows. 1) We recorded the TV screen with a video camera, 2) we collected beacon signal logs, and 3) we gathered smartphone usage logs by using the “App Usage Tracker” app on the participant’s smartphone. We also developed and installed a beacon collector app on the participant’s smartphone.
Figure 3. An example of collected beacon signal logs captured on the participant’s smartphone

Figure 3 shows an example of the collected beacon signal logs. Figure 4 shows the usage log of the “App Usage Tracker” app. We analyzed the collected data to identify patterns of TV viewing activities.

```
Mode: Low Energy Scan Wed Dec 17 02:01:15
GMT+09:00 2014 Lat/Long:
37.2943569/127.0464035RSSI:-61
Mode: Low Energy Scan Wed Dec 17 02:01:14
GMT+09:00 2014 Lat/Long:
37.2943569/127.0464035RSSI:-69
Mode: Low Energy Scan Wed Dec 17 02:01:12
GMT+09:00 2014 Lat/Long:
37.2943569/127.0464035RSSI:-61
Mode: Low Energy Scan Wed Dec 17 02:01:11
GMT+09:00 2014 Lat/Long:-1.0/-1.0 RSSI:-88
Mode: Low Energy Scan Wed Dec 17 02:01:09
GMT+09:00 2014 Lat/Long:
37.2943598/127.0464035RSSI:-61
Mode: Low Energy Scan Wed Dec 17 02:01:08
GMT+09:00 2014 Lat/Long:
37.2943598/127.0464035RSSI:-52
Mode: Low Energy Scan Wed Dec 17 02:01:05
GMT+09:00 2014 Lat/Long:
37.2943599/127.0464035RSSI:-52
Mode: Low Energy Scan Wed Dec 17 02:01:02
GMT+09:00 2014 Lat/Long:
37.2943599/127.0464035RSSI:-54
Mode: Low Energy Scan Wed Dec 17 02:00:58
```

Figure 4. An example of collected smartphone app usage logs captured on the participant’s smartphone

<table>
<thead>
<tr>
<th>App Name</th>
<th>Start Date/Time (YYYY-MM-dd HH:mm:ss)</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Messenger</td>
<td>2014-12-15 21:14</td>
<td>20s</td>
</tr>
<tr>
<td>Messenger</td>
<td>2014-12-15 21:14</td>
<td>7s</td>
</tr>
<tr>
<td>Messenger</td>
<td>2014-12-15 21:15</td>
<td>7s</td>
</tr>
<tr>
<td>Messenger</td>
<td>2014-12-15 21:15</td>
<td>15s</td>
</tr>
<tr>
<td>Portal Site</td>
<td>2014-12-15 21:15</td>
<td>2m 31s</td>
</tr>
<tr>
<td>Facebook</td>
<td>2014-12-15 21:18</td>
<td>1m 58s</td>
</tr>
<tr>
<td>Game</td>
<td>2014-12-15 21:33</td>
<td>8m 12s</td>
</tr>
<tr>
<td>Messenger</td>
<td>2014-12-15 21:56</td>
<td>17s</td>
</tr>
<tr>
<td>Messenger</td>
<td>2014-12-15 21:57</td>
<td>7s</td>
</tr>
<tr>
<td>Portal Site</td>
<td>2014-12-15 21:57</td>
<td>1m 10s</td>
</tr>
</tbody>
</table>

Today | Total Usage Duration | 14m 54s |

Results The results of the data analysis showed that the participant’s TV viewing activities were not simple. We built a timeline graph using the collected data, as shown in Figure 5. The figure shows various patterns related to the participant’s TV viewing behaviors. We observed that many other activities are embedded in the participant’s TV viewing behavior. The participant did not stay in one
location: after turning on the TV, he went to his bedroom first and stayed there for 10 minutes; while watching TV in the living room, he left to spend 8 minutes in the bathroom.

Figure 5. Timeline of TV contents along with participant’s locations and smartphone usage

The participant also used a smartphone several times while watching TV. He accessed several apps (e.g. the “Facebook” app, a portal site app, a messenger app, and a game app) while viewing TV (Figure 5). We observed that he played a baseball game for a few minutes. As soon as the sports news started, he stopped playing the game and resumed viewing the TV. These results imply that our design can more accurately measure TV viewing activities than existing People Meters.

We focused on various events of TV viewers and the data sources that can detect these events. As shown in Table 1, the “Moving the spot” event (A, B, G, and H in Table 1) can be easily detected by using beacon signal logs. Cases D, F, J and K, which denote the “Changing TV channel” event, can be recognized using TV content information. In addition, the “Immersive TV viewing” event and “Multi-tasking” event can be recognized by combining all the data sources.

<table>
<thead>
<tr>
<th>Events</th>
<th>Case</th>
<th>Data Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Viewer’s location change</td>
<td>A, B, G, H</td>
<td>Beacon</td>
</tr>
<tr>
<td>Changing TV channels</td>
<td>D, F, J, K</td>
<td>TV contents</td>
</tr>
<tr>
<td>Actively engaged in TV viewing</td>
<td>B-C, E-G, H-I, J-L</td>
<td>TV contents, Beacon, App usage</td>
</tr>
<tr>
<td>Multi-tasking with smartphones</td>
<td>C, I, L</td>
<td>TV contents, Beacon, App usage</td>
</tr>
</tbody>
</table>
Table 1. Events in Figure 5 and respective data sources

After the experiment, we interviewed the participant to gain further insights. He reported that he did not feel any differences caused by all the apps for logging data that operated in the background of his smartphone. He also reported that the camera used for recording TV screen made him feel uncomfortable for the first few minutes of the experiment, but that as time went by he was no longer conscious of it. These results show that the participant found the prototype system unobtrusive.

The study results imply that our system is able to unobtrusively recognize various additional patterns related to TV viewing activities. Table 2 shows the detailed TV viewing activities.

<table>
<thead>
<tr>
<th>Detected feature</th>
<th>Traditional System</th>
<th>Smart TV Logging System</th>
</tr>
</thead>
<tbody>
<tr>
<td>Changing TV channels</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Viewers' identification</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Viewer's location change</td>
<td>X</td>
<td>O</td>
</tr>
<tr>
<td>Actively engaged in TV viewing</td>
<td>X</td>
<td>O</td>
</tr>
<tr>
<td>Multi-tasking with smartphones</td>
<td>X</td>
<td>O</td>
</tr>
</tbody>
</table>

Table 2. Traditional research methods vs. the smart TV logging system

Design Iteration

Bluetooth-based System

The lessons learned from the trial study allowed us to identify a number of ways to enhance the system. First, a smart TV can be used to identify the content of the TV screen. The initial prototype was not able to recognize TV content information simultaneously. However, a smart TV can take advantage of the Transport Stream (TS) packet of the MPEG-2 systems used by many broadcasting systems [4]. The header in the TS packet contains TV content information and a smart TV can parse the TS packet and acquire the content information directly from broadcasting signals. In this design, the smart TV also acts as a beacon. After the smart TV parses the content information acquired from broadcasting signals, TV content information can be easily embedded in the TV beacon signal and transferred to smartphones.

Second, detailed logging of smartphone usage can be collected for analysis. The engagement level of viewing a TV program can be measured based on logs. For instance, participants can share ideas and emotions about a TV program with friends through a social network service (SNS) and search information related to a TV program. Advanced logging could measure the viewer’s ignorance of and level of engagement in a TV program through the analysis of these logs.

Scenarios of Conventional System and Bluetooth-based System

Features of our system compared to the Conventional system are four points significantly. Based on those features, we will demonstrate the procedures of this scenario.
1. Integrating audience measurement devices into one device

**Figure 6. Configuration of whole devices  Conventional System (Left), Bluetooth-based System (Right)**

A conventional system consists of four different devices: Display Unit (DU), Base Unit (BU), Transmission Unit (TU) and Handset. Display Unit is the device on top of the TV set, which display the data, time and handset number with message. Base Unit is placed next to a TV set. It connects with TV or TV related device (video & audio signal output device), and has a function for storing the all audience information. Transmission Unit connected the telephone line plays a role to transmit the viewer data stored in the BU to the head office. Finally, the handset is a device that is used to register viewer to identify when they watch TV in a panel household. The panel grants the handset number to each individual member of the household and presses the handset number at the start or finish watching TV. In order to measure the TV ratings in conventional system as described above, it requires the additional peripheral devices. In contrast, Bluetooth-based system is consists of only two parts: smart TV and mobile devices (e.g. smartphone, smartwatch). The smart TV may include DU, BU and TU unit of the conventional system. This is possible in part with the development and connectivity features of smart TV OS\(^4\), and handset is replaced with the mobile devices.

\(^4\) OS: Operating System
2. Registering audiences automatically

![Figure 7. Viewers’ identification method Conventional System (Left), Bluetooth-based System (Right)](image)

In the conventional system, viewers had to press a button on the handset, in order to register the TV rating. In contrast, a Bluetooth-based system does not require additional actions, such as pressing a button. When the viewer enters the TV viewable area, the mobile devices receive beacon signals from smart TV automatically. TV viewable area entry confirms the viewer to determine the RSSI\(^5\) value of beacon signal. In addition, beacon signal is not necessary pairing procedure so that it can implement the registering audiences automatically.

3. Collecting audiences’ multitasking behaviors

\(^5\) RSSI: Received Strength Signal Indication
A conventional system does not reflect the user’s context since it only measures the individual TV ratings. A Bluetooth-based system, however, can analyze the individual TV ratings synthetically with use patterns and frequency on the audience of smartphone by utilizing App Usage Tracker. This implication proves that our system can measure the TV immersion level through the analysis of Multi-tasking behaviors.

4. Identifying audiences’ real locations

A conventional system does not know the physical location information of the viewer in front of the TV. However, a Bluetooth-based system can grasp the viewer’s actual location by analyzing the RSSI value of a received beacon signal to a mobile device carrying the viewer. This point demonstrates that we can obtain more robust and accurate TV ratings data.

**Limitations** In this paper, we proposed a smart TV logging method for investigating viewers’ watching behavior using a beacon system, smartphones and a smart TV. However, some limitations exist. First, the proposed method may not accurately capture viewers’ activities that do not involve digital devices (e.g. reading a newspaper). Viewers may fall asleep while watching TV, or conversations between viewers may distract them from watching TV.

The second limitation is caused by the beacon method we employed. TV viewers do not necessarily carry their smart phones with them all the time. However, using smaller wearable devices may resolve this issue. Thus, under the circumstance of increasing proximity devices, we believe that using Smartwatch which are more flexible to take in and out can help measuring TV rating more accurately in the future.
Lastly, we used a small number of beacons in the trial study. Depending on the structure and the size of a home, a greater number of beacons may be required. However, as more devices are equipped with Bluetooth technology [9], it may be possible to detect a sufficient number of signals to calculate the viewers' location accurately without installing additional beacons.

**Conclusion & Contribution**

In concluding it is worth reiterating the proposition presented earlier in this paper that design of a smart TV logging system as an innovative future TV rating system. The paper attempts to prove how much our system could develop the measuring TV rating comparing the previous systems such as People Meter. Through the research of related works, trial study, scenario, experiment, etc., we found out key findings to overcome the limitations of previous TV rating measuring system. Thus, our research satisfied the three main goals as follows: (1) non-invasively measure system using beacon and smartphones, (2) Collecting robust and accurate data and (3) efficient device maintenance. In conclusion, we demonstrate the contributions of each goal, and how it affects in academically and industrially.

1. **Non-invasively measure system**
   One of the merits using Bluetooth Technology (BT) was non-invasively measuring. Through the trial study, the research proved how much BT was unconscious. TV viewers do not need to install or register for participating TV rating system. Any multiple IT devices including BT can be received the beacon signal logs unobtrusively. The study results showed that the prototype can capture viewers’ various events embedded in TV viewing behavior. It means that the viewers can immerse TV programs without any interruption, and also they can be reduced tracking repulsion about all the following activities while watching TV.

2. **Collecting robust and accurate data**
   The system we proposed detecting TV viewing behavior more robustly and accurately than previous systems such as surveys and interviews. As shown the results of the trial study, the proposed method showed how much the beacon signals can be collected accurate data. Finally, it proved it as a respective data.

3. **Efficient device maintenance**
   According to James G. Webster from Northwestern University, conventional people-meters are expensive to manufacture, install, and maintain [10]. Thus, our proposed system was worth noticing in that all existing devices and technologies are united into one integrated system. It could save lots of expenses to utilize the peripheral devices. Thus, in the perspective of industry, the smart TV logging system can be an efficient method to maintain TV rating system.

As the contributions above, the system has plenty of strong points to develop future TV rating system. Moreover, we expect that its usage in the future work would be wildly utilized with other business services.

**Future Work**

The system we proposed can detect TV viewing behavior more robustly and accurately than previous systems. However, further researches are needed on the usage of the collected information. For example, real-time analysis of each audience’s TV viewing behavior information can be used to create
an important database for the development of a TV contents recommendation system. Moreover, studies can also be conducted on the situation where a smart TV functions as a hub system in a future home. We described scenarios of future work in detail below.

1. **Personalized recommendation service**
As shown Figure 10 below, our system can offer the personalized recommendation service based on analysis of individual TV rating data. This service can be possible because they know the audience information in front of TV and multitasking behavior of viewers while watching TV.

![Figure 10. The Configuration of a Bluetooth-based System](image)

2. **Screening engaged viewers**
Using social media can be a good tracking tool to identify the deeply-engaged viewers. Perhaps, an advertiser may want to find those engaged viewers. Our proposed Bluetooth-based System can be developed to track TV program viewing history, internet usage behavior on smartphones and mobile devices through the URL logs. Therefore, by analyzing the URL logs and the program, we can define whether the viewers are highly engaged or not. For instance, if the URL is related with the program, we can regard the viewer as a highly engaged viewer. In other case such as just surfing unrelated program, the viewer may regard as an un-engaged viewer. The system has a tremendous impact in that it can match and classify the two types (highly engaged viewer, un-engaged viewer) of viewers in real-time.

3. **Enhancing the immersion of contents (e.g. TV programs, advertisements)**
Our system can transmit to real-time viewers of the TV program contents and advertising contents by using beacon signal. This may give the opportunity to provide a variety of service on smartphone receiving beacon signal with content information. For instance, our system with beacon signal can provide the place in TV screen, the main character, BGM, as well as commercial information such as PPL in real-time.
4. Seamless TV (Local-based N-screen Service)

Integrating audience Seamless TV using beacons and smartphones: The main stream of our idea is to suggest seamless TV by using beacon. For better understanding about seamless TV, N-screen is one of the most popular technologies in these days. However, we found that N-screen has lots of limitations. Therefore, our idea desires to compensate the defect of N-screen. Our idea would show the flexibility while watching TV behavior from outdoor to indoor or on the opposite situation as well. A beacon senses the smartphone signal and defines its location. Let’s assume that a person went outside of the home after watching TV at a living room for a while. Even though the person was out of signal boundary, the seamless system will help to keep continuously playing through her/his smartphone. This principle is different than the existing mirroring system using Wi-Fi network. Smartphone and smart TV use each Internet network (e.g. Wi-Fi, 3G, 4G) and Application individually. It recalled the log data from smartphone to TV. The technology would activate in TV hardware and software business market. We carefully suggest the seamless TV system.

![Figure 11. The process of how beacon and smartphone are working as seamless devices](image)

1. Smart TV OS detects and parses the TV contents information from video streaming input signals. The signal format such as MPEG-2 (TS) has header information and payload information. Our idea is attempted to load TV contents information at the header. In this TV contents information, there are contents name and running time information etc.

2. Detected TV contents information transmits to Bluetooth Module on embedded smart TV. Bluetooth Module transmits the received TV contents information and transmits as beacon signals to show the real-time play. Users keep receiving the real-time signals from smart TV by transmission technology which is beacon. Beacon signals have RSSI parameter. By this principle, Users’ smartphones can sense people’s location and recommend the contents to watch continuously.
3. As shown in Figure 11 above, if they move to room or outside of house, the smartphone will recall the TV contents the user watching as pop-up application (e.g. Do you want to keep watching <Content name>?).

4. If the user pushes Yes button, then TV shows the contents watched by using smartphone service application (e.g. Netflix, Hulu etc.). The application database recalls the contents to display it.

It is our hope that these suggestions will serve as a platform for communication between doers and thinkers on a subject that is profoundly difficult and important.

Reference

[2] Ericsson Consumer Lab TV and Media study 2013


Monday, June 15
Applying Motivations to Digital Video Ads

Erika Lamoreaux
Associate Director, Digital Media, The Clorox Company

Heather Dougherty
Senior Director, Professional Services, Resonate

INTRODUCTION
Historically, media plans and investment decisions have been dominated by demographics that ultimately fall short in representing the true nature of consumer decisions. We have seen creative begin change and evolve, yet media planning has been slower to take advantage of new ways to reach the right audience. We still see examples of demographics leading the media planning process such as targeting men on websites like ESPN for grilling products, while millions of women grill, research recipes, and/or purchase the food. Advertising cleaning products is a great example where clichés in the messaging and targeting were the standard: providing housewives with images of pristine white sheets and clothes billowing in the wind without a stain in sight.

Now creative has become more realistic and reflective of real life cleaning situations. Take the father suiting himself up in plastic and armed with Clorox bleach to cleanup his toddler’s room in the creative titled “Poopocalypse – Language of the Domestic Jungle”.

18
Now every brand manager has a list of targets with differentiating characteristics in order to find relevant “moments” to message and influence them. For example, one Clorox persona is the “Joyful Guardian”, which starts as a detailed brand brief but, when applied to media targeting, ends up reduced to a definition of “women 25 to 54”. The challenge is then to find the additional micro-targets that live within that persona such as “new moms”, or others which may live outside of that persona such as “new dads”.

The following research identified and compared the motivations and behaviors of audiences that interacted with Clorox products across both owned and paid media. This research enabled Clorox to identify new micro-targets to engage more deeply through more relevant messaging and creative that speaks directly to specific, and sometimes previously unknown, consumer needs. The findings are not limited to just demographics, but also leverage motivations and attitudes to build out personas. This work is important because it combines multiple, historically disparate data sources with innovative data science techniques to add a new dimension of customer understanding for brand advertisers.

**BUSINESS QUESTIONS & OBJECTIVES**

- Who is the Clorox website visitor – what deeper dimensions describe their motivations and behaviors? What are the differences between those users who engage with the homepage, the pages for coupons and offers, and the specific product pages for disinfectant wipes?

- What defines the audiences that engage with Clorox’s paid digital media, and specifically, pre-roll video ads? These segments were defined as:
  - a) buyers of the Clorox product
b) buyers of a competitive product  
c) a control group

- What are the similarities and differences between the current Clorox website audiences and the target segments that viewed the online video ad?

- What differentiates responses between creative concepts? Are there significant differences amongst the audiences and how they engage with each message?

**METHODOLOGY**

The research design utilized two approaches to understand more deeply the Clorox customer across owned and paid media touchpoints.

- The first was to implement Resonate’s tags on the Clorox website, specific product pages and coupon pages to provide insights about these visitors. These tags enabled Resonate to associate site users with pre-existing models of those individual’s purchasing patterns and personal motivations.

- The second was to deploy tags across an online video campaign to compare the responses to four different creative messages across 3 segments – Clorox buyers, Lysol buyers and a control group. Over 1 million ad impressions were served to each of the segments. The purchaser segments were identified through Datalogix CPG POS data to provide a comparative view of behavioral differences in ad responders and video completion rates between creative messages and brands.

For each of the groups, approximately 300 descriptive attributes were analyzed for correlation between the segments. The attributes themselves were appended and modeled through a combination of observed behavior, online topic affinity (though natural language processing), trained against massive-scale survey response and purchase data.

Resonate conducts proprietary surveys of individuals online through research partners by selectively inviting users (through the presence of a cookie) that have amassed a significant threshold of online engagement to participate. Approximately 15,000 to 20,000 surveys are collected each month through a 25-minute online survey instrument. The survey responses are then fused with the observed content consumption data to establish recognizable and reproducible patterns. Resonate then models this data out to cookies representing approximately 90% of the total online population, with individual models for over 4,500 attributes. This addressable dataset may then be used to understand the values and motivations of online audiences to websites and digital campaigns through the Resonate Analytics interface or exported into other analytics tools. Defined audiences, consisting of any combination of attributes, may also be targeted via partnered DSPs and DMPs.

**INSIGHTS**

1. **Website visitors were older with a higher household income and environmentally aware**
   - The demographics of the website visitors confirmed that there is a good opportunity to target older consumers, especially women over 45 years old, earning $75k or more, some with kids in the household and others without. This finding served to help debunk previous theories of disinfectant wipes as a product preferred by Millennials.
Visitors to the pages for Clorox Disinfectant Wipes were 140% more likely to buy green products and 149% more likely to use reusable bags when compared to the online adult population as illustrated below.

**Environmental Actions of Website Visitors for Clorox Disinfecting Wipes**

![Environmental Actions Diagram]

Additionally, the website visitors have a much stronger propensity to use coupons (109% more likely) than those who completed the video ads (23% more likely on average among those that completed Clorox video ads). Coupon use is prevalent across many online users, with 61% of the online adult population citing a preference for coupons among their shopping behaviors. We can assume that these website visitors are most likely somewhat conditioned to visit to take advantage of the promotional offers found there and were probably going to purchase Clorox products anyway. The website visitors are already a unique group that are interacting with a CPG company and are probably not a core target for media spending. Alternatively, we also learned that this audience also under-indexed against the online population for loyalty programs and service, so price is more influential in this instance.

**Shopping Behaviors of Website Visitors for Clorox Disinfecting Wipes**
There was also greater differentiation between different segments of Clorox’s site visitors than with those exposed through the targeted video ads. Among all of the segments that were exposed to the video ad, they had a much stronger propensity to value Learning & Knowledge than the site visitors, which emphasizes an opportunity to provide product information and dynamic offers and promotions to those that have not yet taken any action. For Clorox and many other CPG companies, this can be successful since even if the media audiences never visit the website, the actual goal is ultimately purchasing the product.

Personal Values of Website Visitors for Clorox Disinfecting Wipes
2. Clorox purchasers are very similar to the Lysol purchasers
   - Because the products themselves are functionally very similar, the media audience comparisons showed a strong similarity among a large set of evaluated attributes, specifically among the audiences of Clorox and the competitive brand, Lysol. Personal values and motivations were very similar, but an important theme emerged around their Sense of Belonging and relationships, especially with family, across both Clorox site visitors (on average 136% more likely) and those exposed through targeted media (on average 53% more likely). This translates to their need for social acceptance, which is connected to having a clean and welcoming home. Other key motivations that extended across both website visitors and the media audiences were Longevity & Health and Financial Stability, suggesting a strong need to feel healthy and secure. **Focusing on these values and motivations can increase the relevance of the creative messaging to help differentiate the brands.**
Both audiences also consistently cite high Quality as the most important factor when making purchase decisions. These are the key messages to appeal to the motivations behind purchasing cleaning products.

3. Non-buyer response to informative creative

- Overall, the reactions to the creative were very similar among both Clorox and Lysol buyers since they are already familiar with the product, particularly noted for creative that was less informative. However, the creative that was the most informative (least funny) was the top performer with the highest completion rate, for the control group who would have the least experience with these products or brands. **Identifying the level of the relationship with the brand or product will be key for both audience targeting and understanding how to speak to consumers.**

**BUSINESS RESULTS & IMPACT**

As a result of the research and resulting optimizations, there was significant, measureable improvement in the following areas:

- **Identified Additional Micro-targets:** This research helped Clorox expand their targeting efforts to include smaller, more specific segments such as older women that are motivated by green behaviors and protecting the environment.

- **Improved Messaging Strategies & Content:** Based upon the interests of the Clorox audience, messaging that focuses on relationships with families and others, environmental friendliness and healthy living will be most impactful with this audience. These findings lead to additional opportunities to offer cleaning content and determine what will be most relevant. Are consumers more interested entertaining and sharable moments like cleaning with David Hasselhoff or functional tips that are further down the purchase funnel?

- **Improved Message Testing for Differentiation:** For Clorox and Lysol buyers that are already purchasers of the product type, there were more similarities observed between the
performances of the creative concepts. With such similar reactions, there is an opportunity to drive better differentiation in messaging and focus on health & positivity vs the typical Lysol approach, which focuses more on scare tactics like eliminating germs and bacteria. This is a particular challenge within a low interest category trying to make the connection with a brand and empower the customer to create a home environment conducive to social acceptance & belonging.

Example of Clorox advertising:

![Clorox Example Ad](image1)

Example of Lysol advertising:

![Lysol Example Ad](image2)

- **Validation to Increase Allocation to Social Media:** Clorox audience is highly engaged with social media and enjoys following companies and sharing content on social networks 182% more than the online adult population. The audience of Lysol buyers that completed the video are also 104% more likely to use social media and 24% more likely to follow companies on social media as well. This data points to social media as a strong platform to communicate and interact with current and potential customers and the importance of content they will actually want to share (informative versus entertaining).

These learnings have then led to the application of this research for new product launches as well as existing business lines to understand the impact of different creative messages across different types of media including social and display in addition to video.

**NEXT STEPS**

- **Unification of the workflow:** After completing this first phase of research, The Clorox Company is going to move into the next phase to take action upon the findings by launching a DMP to
activate and test creative against the micro-targets. This new marketing workflow will allow the Clorox Company to feed the results back to the DMP to continuously improve testing and targeting capabilities, not only across digital, but also influence messaging and creative across all channels, including TV.

- **Additional audience modeling & analysis** – The addition of the DMP will also allow additional flexibility to analyze the characteristics of the website and digital media audiences as well as the differences between them (e.g. people that register on the website vs. just visit). This knowledge can then be used to find and target similar prospective customers.
From Message Sent to Message Received

Fiona Blades - President & Chief Experience Officer, MESH, The Experience Agency
Catherine Willis – General Manager, Customer Research, Delta Air Lines

Introduction

Before a marketer asks “Where should I be investing my marketing budget?” they should be able to answer the question “How are people experiencing my brand?”

It’s easy to draw a pie chart of investment but how many marketers could draw one showing the ways people experience their brand? What would be the percentage of usage or consumption occasions in relation to TV advertising experiences, to seeing the product in the store experiences or coming up in conversation? What would be the percentage of Paid, Owned, Earned and Environmental experiences?

Figure 1: How do people experience your brand?

Unless we understand how people are experiencing our brands, how can we decide which experiences to invest in and which to eliminate?

We need to stop focusing on what messages we are pushing out and start measuring what people are picking up. We need to know what real people really experience.

To do this will require new Experience Metrics that allow us to compare Consumption experiences with TV advertising experiences. We need to make Experiences into a Currency used across all parts of the business that media agencies, creative agencies and finance directors all understand.

Delta Air Lines is one of the brands doing exactly this.

Challenges Facing Marketers

“It is no secret that the first ten years of the twenty-first century have shaken the media world.”

This is how the Reuters Institute report for the Study of Journalism on ‘Ten Years that Shook the Media World’ (2012) begins. In particular there is a continued expansion of media options, particularly online
and mobile, for marketers to contend with and this is leading to an erosion of audiences through more traditional media channels.

As a market research consultancy, these are some of the major challenges we hear marketers facing:

- The sheer number of places to invest marketing budget has grown phenomenally over the last decade. This is particularly true of the digital environment. Should we be transferring money from TV to Digital? Should we put this into Twitter, just standard banners or an app to aid shoppers? What metrics can help us to understand the answer?

- Each channel has different measures – GRPs, Click throughs etc. We can’t compare apples and pears making it incredibly difficult to know whether to transfer money from TV to digital in the first place. Furthermore, the measures don’t always get to what we are trying to impact, for example, if we are trying to change brand perceptions with an online ad and not generate click throughs, then this measure of ‘success’ is obsolete.

- Even with the media that gets the most attention, like TV, we have different companies measuring the impact of expenditure through this channel. The media agency (or media audit agency) gives us quantitative data on efficiency against the Reach and Frequency measures that were planned. The market research agency provides a data on recall and likeability. We get rational and emotional evaluation provided separately. With time-shifting and viewing TV programs online, people’s viewing behavior is changing and we need metrics that capture this. Andrew Lipsman and Joan FitzGerald from ComScore tackle this in their paper “Multi-platform takeover: From TV to total video - how integrated video planning can transition advertising from 'upfronts' to 'allfronts'”

- Current measures miss out on context. We know people liked the TV ad, we know we bought media at the right time to reach our audience, so why did the brand metrics line stay flat? Maybe it was the fantastic ad our competitor was running? Maybe that PR issue played its part? We can’t link our investment sufficiently well to impact.

- We know that some of our media spend doesn’t get picked up accurately in media measurement tools. But just because no-one in the company has confidence in the representation of digital media measurement, does it mean we shouldn’t engage here?

- More worryingly we know that there are major gaps in knowledge but we don’t know what we don’t know! Does Paid Media represent 50% of the experiences people are having with my brand or 5%?

A new Experience Driven Marketing Framework

Once we turn things on their head and stop thinking about all the different and diverse things we are doing and start thinking about what people are really experiencing, things become much simpler.

The currency of experiences embraces the notion of quantity (the number of experiences) and quality (how positive, persuasive and relevant are they). They can be long and nuanced (a flight) and fleeting
(driving past a poster). They happen in the context of other experiences – you saw 5 airline brands on a comparison website but only two really stood out.

Experiences are human centred.

As marketers, we try to put the customer first but we see stats and targets at our desks. Those GRPs get in the way of understanding real people. Experience Driven Marketing puts them back center stage.

How can we build a framework to measure experiences?

MESH has focused on capturing people’s experiences in real time since its launch in 2006. It was conceived to fill a need. The founder was working as a Planning Director for Claydon Heeley (Omnicom) on brands like Mercedes-Benz and knew that experiences such as seeing a neighbor’s new car, walking into a dealership and seeing the car on Top Gear were all going to be impacting on car buyers, yet there was no market research tool to measure these. Also, some experiences would easily be forgotten or mis-remembered so it would be important to capture these experiences ‘in the moment’.

What if you could get participants to let you know every time they had an experience with a premium car brand?

The mobile phone held the answer!

We developed Real-time Experience Tracking (RET) which 6 years later Harvard Business Review cited as “a new tool (that) radically improves marketing research”. But in 2006, capturing market research data via someone’s mobile was seen as a gimmick.

Typically we capture 4 metrics in real time (in addition to others captured passively, such as time and date):

- The Brand (e.g. Delta, American Airlines)
- The Occasion (e.g. TV, Online)
- Experience Engagement (how did the experience make you feel? Very Positive to Very Negative)
- Experience Persuasiveness (e.g. how likely did it make you to choose this brand next time? Much more to much less likely)

Beyond the real-time data capture, we capture data in near-time. We invite participants to expand the data about their experiences by asking questions that they are able to remember based on their immediate responses. These could include which website they were on, or did they see the poster at a bus stop or at a train station? We ask for photos of the experiences so we can see whether the poster’s tatty state was contributing to a negative experience and we ask participants to describe the experience in as much detail as they can.

Before the participant starts recording their experiences and after they finish (often a week), we capture data through a traditional survey so we can see whether perceptions and behaviors have changed. Through analytics we can then work out which experiences are driving which KPIs. Are Posters, Comparison Websites or Social Media doing the best job at driving consideration of my brand?
New Experience Metrics

This data provides us with new Experience Metrics that work across every experience from TV advertising to seeing a car on the road. It is important to capture the proliferation of touchpoints people come across most accurately, which real-time data collection addresses. The metrics build in engagement (captured in the moment which is vital because it is almost impossible to recall emotion. You might remember how you felt the first time you saw the poster but what about the 6th time?).

These experiences are the building blocks of brands. As Jeremy Bullmore famously said “People build brands like birds build nests, from scraps and straws they chance upon.” Whether someone says they trust the German engineering of a Mercedes or you see a neighbor driving one, each experience adds up to your impression of the brand.

So what new Experience Metrics do we have?

- Experience Reach (the percentage of people who experience the brand or touchpoint – whether seeing Pepsi in the end shot of a KFC TV ad or seeing someone else drinking Pepsi).
- Experience Frequency (the average number of times they experience the brand or touchpoint – when frequency is too great, wearout occurs).
- Experience Share (the percentage of Mercedes experiences versus those for other premium cars).
- Experience Positivity (the percentage of positive experiences the brand or touchpoint has – some channels, like Cinema, are inherently more positive but others can be transformed through relevant messages. What people think are positive can be very different from what marketers think are positive. The most positive reaction we have seen – at double the positivity level to 3 award-winning TV ads – was for a brand featuring in retail advertising on promotion).
- Experience Persuasion (the percentage of persuasive experiences the brand has – this is closely correlated with Positivity but can indicate a more rational response).
- Experience Impact (the significance of the change that the experience has on key brand metrics, normally derived through Experience Maximizer statistical analysis).

For those marketers embracing work from the Ehrenberg-Bass Marketing Sciences Institute or Byron Sharp’s book “How Brands Grow”, these Experience Metrics help you to understand your Mental and Physical Availability.

For marketers embracing the McKinsey Customer Decision Journey, measuring experiences in this way will enable you to see exactly which experiences impact most at each stage along the CDJ.

For every marketer, the approach enables you to get to a level of detail that means you can answer the question, “How are people experiencing my brand?”

What have these Experience Metrics revealed?
Over nearly a decade, MESH has collected millions of experiences in real-time across categories and geographies. Some of these have been collected for clients on a continuous basis, as with Delta Air Lines.

We therefore have a rich database to mine.

These are some of the most valuable insights we have generated to date.

1. **Paid Media is often less than 50% of overall experiences in a category.** Across categories and countries we see that Paid Media normally represents 50% or less of overall experiences that people are having with brands in the category (see Figure 2).

   ![Figure 2: % Paid Media for a number of studies (% out of total experiences)](#)

   **Figure 2: % Paid Media for a number of studies (% out of total experiences)**

   - **Packaged goods**: 29%
   - **Electronics**: 28%
   - **Alcohol**: 47%
   - **Service**: 53%

   Where we have data over a number of years we have seen that in some categories, the percentage of Paid Media experiences has decreased with Earned increasing.

2. **Share of Experience (SOE) is more predictive of Consideration the following month than Share of Voice (SOV).** This stands to reason because SOV doesn’t include Owned, Earned and Environmental media. It is vital to monitor Share of Experience because we have seen this correlate with Share of Market (see Figure 3).

   ![Figure 3: Share of Experience correlates with Share of Market.](#)
We already know from the IPA’s paper in conjunction with Nielsen on “How Share of Voice wins Market Share” (2009) that Excess Share of Voice leads to growth in Share of Market. As SOE is a better metric than SOV at understanding the way people pick up experiences, monitoring SOE will help marketers to understand whether their Share of Market is likely to increase or decline.

3. **Sentiment matters.** By increasing your brand’s number of positive experiences, you will see a greater positive impact on brand metrics. We have seen that neutral experiences can actually have a negative impact on brand metrics. It is therefore vital that marketers monitor their overall Brand Experience Positivity as well as understanding how positive experiences are through each of their touchpoints, from Consumption to Online. Across categories we have seen that positive experiences impact on Brand Consideration at a greater level than simply having an experience.

4. **Context is king.** Traditional research often operates in a vacuum. A TV ad is evaluated whilst sitting doing an online survey. MESH has uncovered evidence that shows an ad can perform
well in traditional research but when it is experienced alongside an even stronger competitor, its result will be depressed. Negative PR can make a big difference to the way people respond to brand communication. Financial crises and adverse weather will all have an impact. Marketers need to know what to do in these situations. Should they cut all advertising and go dark? Should they change messaging? Through analyzing the impact of negative PR on people’s response to Paid advertising MESH has been able to advise clients on what actions they should take.

5. **It doesn’t always go to plan!** When we see the media laydown, we assume that this is the activity actually happening. MESH has picked up numerous occasions where this isn’t the case. We frequently pick up old outdoor activity that has been left up beyond the campaign end date. This can have both a positive and negative impression – great if people are still responding to the creative messaging, less good if it suggests the brand has nothing new to say. Even with relatively small media channels, like Cinema, we have gone back to clients to say that we would have expected to see more experiences, only to find out that instead of the campaign lasting for 4 weeks it has been spread over 12. Also, what you imagine is happening with your targeted communication might still not be targeted enough. One client was targeting commuters through specific underground stations in London. The number of experiences we picked up was lower than the client anticipated so we called participants who we knew should have been exposed to posters on their commute only to be asked whether the poster was on the right end of the platform because that was the only place they stood waiting for their train.

**How is Delta embracing an Experience Driven Marketing approach?**

Delta Air Lines faces similar challenges to other companies but there are some specific ones more relevant to the airline market:

- The purchase cycle is longer than for some categories making it harder to see the connection between activity and sales.

- A considerable amount of experiences are likely to be Owned, such as via the website, and we need to understand people’s experiences here in the context of the total path to purchase.

- We know that Word of Mouth is important in service industries and can pick up a certain amount of online buzz but what about the face to face conversations?

- Delta’s campaigns are truly 360 making it hard to pick up the impact of smaller elements. The brand also invests in sponsorship which needs evaluating.

- We have hubs and regional advertising as well as different audiences, adding to the overall complexity of measurement.
• The airline market is extremely competitive, so if we only understand Delta’s activity we miss out on the competitive context.

There is a lot we need to understand to see if we are genuinely getting return on our investment.

**How is Delta Air Lines using the insight from real-time experience data capture?**

With true 360 degree campaigns that are targeting different audiences and regional hubs, Delta Air Lines started working with MESH to understand which channels and messages are most influential.

In a competitive market with lots of ‘noise’ from many touchpoints that are not part of a campaign it was important to put Delta’s campaigns in context. Figure 4 shows the way that flyers are experiencing the airline market overall in the US.

**Figure 4: Pie chart for total market by touchpoint**

![Touchpoint Share for Total Market](image)

Within each broad touchpoint, our data enables us to get to sub touchpoints, of which there are many, for example, with Online. Clearly many of the experiences are from owned channels (like airline websites) as well as earned (through conversations and social media). The challenge was how to maximize the Paid media budget to make investment work harder.

One way we did this was to understand the impact of different media channels amongst a newly defined target audience. What immediately became apparent was that Print was working particularly effectively with this target audience (Target Audience A), see Figure 5.

**Figure 5: Touchpoint quality for Target Audience A**
Print was only reaching a small percentage of Target Audience A, indicated by the size of the bubble, and following discussion with our media agency, we realized that there was an opportunity to increase reach and frequency through this channel for this audience.

Whilst marketers are aware that reach through digital is increasing substantially as is marketing investment, it is difficult to know which digital activations and methods to employ with what return. Through the Real-time Experience data, Delta was able to invest in more highly targeted digital channels which we knew would reach our particular audiences. In Figure 6 we can see the importance of YouTube and airline websites.

Figure 6: Online quality by sub touchpoints

Understanding experiences didn’t simply help us with channel reallocation but also helped us to understand the noise from other competitors that we needed to combat. The humorous messaging from a competitor was being well received by one of the Delta audiences. Comments and photos enabled us to see exactly why this audience was responding well to this tone of voice and how it was focusing attention on the key messages.
Clearly the Delta response was not to copy this competitive tone of voice, rather the insights about what this target audience liked were used to brief our creative agencies and bring our flyers to life in a very human way – where are they coming across airline brands and what messages and tones of voice are getting the best reactions? How are these experiences impacting on brand perception for Delta and competitors? Therefore what types of experiences should Delta be creating to attract flyers? Whereas with traditional communication tracking we would normally only have been evaluating the Delta brand ads and a few competitor ads (or survey length and engagement become an issue) with the real-time experiences, we were capturing every type of competitive communication people were coming into contact with so could develop our creative strategy with a much richer understanding of the competitor context. Figure 7 shows how well the competitor brand experiences were resonating with one of our audiences.

**Figure 7: Quality of Brand Experiences among Target Audience A**

The level of detail we have been able to reach in understanding has helped to encourage constructive conversations with our agencies (creative and media) on how to develop strategies. It has also given the Delta team greater confidence in our decisions, for example to change investment from one channel to another or to introduce a new creative strategy. The range of data from the study – quantitative metrics, statistical analysis, qualitative comments and photos – means that we have been able to look at different perspectives to draw conclusions. If the initial quantitative real-time spontaneous reactions are indicating a particular insight and the statistical analysis confirms this, with comments explaining why, this helps to take an initial insight through to the business taking action.

Now that the MESH data has become understood within the business we have created an Experience Score metric that we use on our executive scorecard. This incorporates both the emotional and rational quality of experiences. In general it reflects our expenditure in the market and helps to act as a barometer of how well our investment is impacting our audiences. It also has the benefit of showing movement over short time periods (whilst many of our other KPIs remain relatively constant).

When we need to dig into why we are seeing a particular Experience Score, we have layers of data to excavate, from quantitative numbers right through to individual comments. If we see that at a time of increased expenditure that our Experience Score is not improving sufficiently, we can immediately see what we should do to optimize this.
Over time we are discovering more and more ways to use real-time experience data within the business to help us with our decisions. Some of the next areas we are planning on include:

- Understanding what is most differentiating from a messaging perspective in relation to supporting our brand strategy.
- Identifying the best channels to deliver these messages.
- Incorporating our data into econometric modelling to provide greater explanatory value.

As the Delta Air Lines business utilizes more channels with a regional strategy across different target audiences, we need both a sophisticated approach to research to deep dive into the detail and a simplicity of measurement that illuminates what opportunities to target. Using Real-time Experience Tracking with our new Experience Metrics is guiding the way on this journey.

**What implications are there for the industry?**

Real-time Experience data has challenged the way we see the world. Once we have looked at things from a customer perspective in a competitive and market context we have discovered new ways to engage with people. There have been some surprises! We weren’t expecting retailer advertising to be the best way to change brand perceptions for some of our clients, like PepsiCo where we found that On Premise (bars, restaurants and pubs) experiences and TV advertising (not brand, but KFC featuring Pepsi) were the two channels impacting on making people feel closer to the brand. This seems like heresy. Yet we have the evidence and when we analyze it, we can see why. Pairing Pepsi with food at moments with friends and family, whilst having a treat, gave the brand a role in people’s lives. It changed the On Premise from somewhere to push volume sales to a place to build brands.

Once we have come to this realization we can’t understand how any marketer can justify making investment decisions without knowing how people are experiencing their brand in the first place. How do they know which experiences are most influential? What are they doing about trying to eliminate the highly damaging negative and even neutral experiences that erode perception and sales?

Taking a “message received” rather than a “message sent” perspective actually simplifies the complex world we are living in. All of a sudden an experience currency allows us to talk apples and apples as we compare across channels. We can speak to media and creative agencies using the same metrics as they include both quantitative channel and qualitative messaging metrics.

It is perhaps not surprising that articles in the industry have been calling for a change in the title of the Chief Marketing Officer to reflect the changing nature of the role. One of these titles is Chief Experience Officer. If Marketing signifies what we are pushing out, Experience more appropriately conveys that it is what people are picking up that is important. Listening to these experiences and working to improve them should be fundamental to any marketer, whatever their title.
Sources:

“Ten Years that Shook the Media World: Big Questions and Big Trends in International Media Development.” Rasmus Kleis Nielsen, October 2012, University of Oxford, Reuters Institute for the Study of Journalism.

“How share of voice wins market share: new findings from Nielsen and the IPA Databank”, IPA, July 2009

“Multi-platform takeover: From TV to total video - how integrated video planning can transition advertising from 'upfronts' to 'allfronts.’” Andrew Lipsman and Joan FitzGerald, ComScore Audience Measurement 9.0, June 2014.

How Much Time And Money Do People Really Spend On The Internet?

If You Want The Correct Answer, Don’t Ask Them

Gian Fulgoni

Co-Founder & Chairman Emeritus

comScore, Inc.

(Excerpted from June 2014 Journal of Advertising Research)

A number of years ago while at Morgan Stanley, Mary Meeker popularized the concept of the “digital advertising gap” by comparing the percentage of media time that consumers claimed they spent on the Internet (obtained through online surveys) with the percentage of ad spending that was allocated to digital. Since the former appeared to be larger than the latter, then marketers were under-spending on digital … or so the argument went. Of course, one can take issue with this rationale, with some critics pointing out that time spent is too simplistic a metric upon which to base ad spending decisions and that one has to also take into account the context in which an ad is delivered and its relative effectiveness. More recently, reports have surfaced (again predominately based on online surveys) from eMarketer claiming that in 2013 consumers spent 20% more of their time online than watching TV. As a result, the call for increased digital ad spending has gotten louder.

This got me thinking about the accuracy of an online survey approach to measurement because when I looked at media engagement based on nationally representative TV and digital behavioral data, which are obtained from passively-measured electronic panels, a far different picture emerged. Importantly, this type of behavioral measurement is not subject to consumer recall error (a well-known source of error in surveys). The behavioral data showed digital engagement lagging TV by a wide margin. Specifically, it showed that in Q3 2013 consumers’ time spent online via desktop, smartphone and tablet represented only about 50% of the time spent watching TV, a far different result from how people respond when asked in a survey. Such a disparity begs the question of the representativeness of people who join online survey panels, and whether their behavior might be biased toward the internet?
To test this, we conducted research using comScore’s 1-million person U.S. panel by identifying those panelists who completed an online survey on behalf of any market research company (comScore’s measurement technology captures this information). Comparing the passively-observed online behavior of the survey-takers to the rest of the panelists, we found that members of online survey panels are 2X to 3X heavier users of the internet than average across virtually all behavioral dimensions. In particular, online survey panel members spend more time and money online. This bias is likely the result of the recruitment of survey panelists via online means. Heavier internet users have a greater likelihood of encountering a recruitment ad and joining a survey panel. So, bottom line, if one tries to measure online engagement using surveys, one will inevitably come up with numbers that significantly overstate reality.

The same overstatement issue occurs when trying to measure e-commerce spending using online surveys. For example, the results of an October 2013 survey conducted by Accenture, which sought to understand online vs. offline holiday shopping, concluded that “two out of five consumers would spend the majority of their holiday purchases online.” In fact, the survey results also indicated that consumers were saying that between 30% to 40% of their 2013 holiday spending would occur online. These data, however, stand in marked contrast to numbers released by the U.S. Department of Commerce (which obtains actual sales numbers directly from retailers) stating that Q4 2013 e-commerce spending represented only 7% of all retail sales. And even if one excludes autos, food and beverage (which are predominately bought offline) DOC data show that e-commerce represents no more than about 14% of consumers’ retail spending. Either way, it’s clear that responses from online survey takers overstate actual online behavior by multiples of the actual behavior.

The rate at which consumers are shifting to digital is an issue of great interest to marketers, but using online surveys to measure the rate of media and channel shifting is clearly problematic because of respondent bias and recall error. Consequently, the use of online surveys to measure cross-platform and cross-channel behavior will likely lead to sub-optimal decisions in allocating media spend between TV and digital or focusing marketing efforts between e-commerce and in-store. The far better approach is to use the emerging electronically-measured cross-platform behavioral databases in conjunction with in-market testing or market-mix models to determine optimal allocation of marketing investments. Only then will marketers be able to properly align media engagement and ad spending.
Measuring Traditional Media Audiences Through Digital Data

Ori Stitelman, Blake Aber, Brian Dalessandro, Lauren Moores
Dstillery, New York, NY

Introduction

In this age of digital advertising, traditional media channels such as outdoor home (OOH), radio and TV tend to be ignored in the digital measurement discussion. Historically, traditional media has relied on sources such as diaries, demographics, surveys and ethnographies for intelligence. For out-of-home (OOH), similar to the original measurement of in-store visits,

OOH media providers have relied on human counters or gross approximations of audience based on geographical areas.

Location signals derived from our mobile devices are the key to bringing traditional media measurement into the digital age. We have started to achieve that with in-store measurement, where the power of location has already proven to be an effective barometer for both desktop and mobile campaigns. Now, a similar application of location signals has the power to provide digital insights and audience optimization to offline media channels such as out-of-home (OOH) and terrestrial radio.

In this whitepaper we show how using mobile location signal and desktop digital data can provide a robust and efficient methodology for analyzing, targeting and measuring audiences exposed to OOH, with the intent of determining how to help marketers use digital data to create a full media package for their needs. By identifying OOH locations, we can use mobile device location data to identify the exposed mobile audience, match the mobile device data to desktop data and score attributes of the local OOH audience compared to the national online user population.

Measuring Local Audiences

Picture the advertising you see every day on the train platform or on your route to work. Most of the time, the decision to place ads in these locations is based on history, local brand experts or ZIP-based demographics. By capturing digital data on people in proximity to the ad location, we can start to paint a much more nuanced and dynamic view of the relevant audiences.
The location signals can come from apps and advertising with opt-in location tracking. The relevant behavioral data, which can come from a marketer’s own CRM, a publisher, an ad network or a programmatic supplier, can help characterize and define the audience near an OOH asset during a particular period of time. Using predictive analytics to match consumer devices, it’s possible to marry the behavioral and location data sets to give an unprecedented perspective on OOH audiences.

Specifically, the data associated with various devices allows us to map app, site and place behavior to an otherwise offline local channel. From these behaviors, it is possible to derive brand signals and optimize brand placements. Compare the characteristic local behaviors to a broader population, either regional or national, and you can map the propensities of each audience to certain brands or brand categories, finally freeing the industry of coarse demographic proxies.

**Methodology**

Our method starts with determining the fixed location asset that we would like to measure which could be an out-of-home billboard, a terrestrial radio station region or a local TV broadcast region. We then match qualified location signals from an area around the asset to other behavioral content cross-device signals by using a probabilistic device matching solution. The behavioral data used can be brand specific, desktop specific, mobile specific or any combination thereof.

**Data:**

Given a set of fixed locations that represented OOH inventory for a particular provider, we chose a representative set in various DMAs to enable a comparison of insights for different locations.

Dataset $D_1$: Two weeks of anonymized mobile in-app bid-requests observed in online advertising exchanges, detailing the lat/lon from which the bid request originated.

Dataset $D_2$: Two weeks worth of anonymized, cookie based brand engagements. The brand engagements are observed via standard retargeting pixels that marketers implement as part of a standard Dstillery advertising campaign.

It is important to note that any location data in $D_1$ used for the method has been thoroughly tested for accuracy as over 60% of location data as provided through the mobile bid-request stream is bad. For $D_2$, non-brand, internet behavioral data or pre-defined marketing segments can also be used.

**CrossWalk:**

To connect these cross-channel data sources, we create a device-to-device mapping using a ‘crosswalk’. The crosswalk method is a bridge that can connect mobile data, such as location to unique consumer transactions or segments on desktop or other channels. This bridge consists of probabilistic device matching algorithms which link users across devices and platforms based on a weighted random walk algorithm to determine link similarity [1].

The output of this process is a match between mobile devices and desktops, and for each matched pair we have the locations observed from the mobile device as well as the brand interactions associated with the cookie.
**Action Scoring:**

Let $Y$ be a binary variable representing some **action of interest**. $Y$ can represent a visit to a brand web page, a visit to any web page of interest, or the use of an app. Thus, $Y$ equals 1 when the action of interest occurred and equals 0 otherwise.

The goal of our approach is to understand for a **location of interest**, $c$, ($c$ can be a billboard, a train station, etc.) and the **geofence around** $c$, $G$, the probability that $Y$ occurs within $G$ as compared to the probability of $Y$ in the general population:

$$P(Y|G)/P(Y)$$

This ratio, let’s call it $LS_{YG}$ represents how much people in the vicinity of location $c$ or $G$, over index in the action of interest $Y$. So if $LS_{YG}=3$, that means people within $G$ are 3 times more likely to do the action of interest, $Y$, than the general population.

The end result of the method is a ranked list of behaviors proxied through action. See Table 1 for an example where the action of interest is website visitation.

<table>
<thead>
<tr>
<th>Top URLs For People in Vicinity Of This Shelter</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>major LA radio station</td>
<td>12.65</td>
</tr>
<tr>
<td>local LA news portal</td>
<td>10.73</td>
</tr>
<tr>
<td>state lottery</td>
<td>8.92</td>
</tr>
<tr>
<td>LA site for news network</td>
<td>7.52</td>
</tr>
<tr>
<td>regional zoo</td>
<td>7.26</td>
</tr>
<tr>
<td>LA entertainment venue</td>
<td>7.18</td>
</tr>
<tr>
<td>SoCal restaurant chain</td>
<td>6.46</td>
</tr>
<tr>
<td>SoCal university</td>
<td>6.15</td>
</tr>
<tr>
<td>online surf shop</td>
<td>5.81</td>
</tr>
<tr>
<td>national magazine</td>
<td>5.50</td>
</tr>
<tr>
<td>local news portal</td>
<td>5.46</td>
</tr>
<tr>
<td>national surf retailer</td>
<td>5.41</td>
</tr>
<tr>
<td>design portal</td>
<td>5.37</td>
</tr>
<tr>
<td>national surf magazine</td>
<td>5.10</td>
</tr>
<tr>
<td>national restaurant chain</td>
<td>5.18</td>
</tr>
<tr>
<td>regional news site</td>
<td>5.09</td>
</tr>
<tr>
<td>regional news</td>
<td>5.02</td>
</tr>
<tr>
<td>national education site</td>
<td>4.97</td>
</tr>
<tr>
<td>LA sports team</td>
<td>4.93</td>
</tr>
<tr>
<td>regional news portal</td>
<td>4.72</td>
</tr>
<tr>
<td>game developer site</td>
<td>4.70</td>
</tr>
<tr>
<td>regional sports store</td>
<td>4.66</td>
</tr>
<tr>
<td>national clothing store</td>
<td>4.59</td>
</tr>
<tr>
<td>national talk show</td>
<td>4.55</td>
</tr>
<tr>
<td>national furniture store</td>
<td>4.53</td>
</tr>
</tbody>
</table>

Table 1. Scores for top 20 scoring URL (host names) for bus shelter.
Media Optimization

Once actions are ranked for a particular location, we can choose a brand category or brand or behavior to measure the audience propensity for the brand or behavior across a DMA or region to allow for optimal media placement in a broad area.

Essentially one can score and rank each fixed location asset according to the characteristics from the original audience and map these propensities visually.

Methodology

We can generalize the method above to understand the propensity of the action of interest across a particular area. More specifically, \( c \), can be represented by a set of coordinates. We also observe millions of coordinates a day for mobile devices in \( D_1 \). Thus, for each mobile device we can determine it's distance from, \( c \). Let, \( m \), equal a mobile device. Given that we have the coordinates for a mobile device, \( m \), and the coordinates for a location of interest, \( c \), we can determine for each mobile device within \( G \) all coordinates observed.

Let's call \( d(c,m_j) \) the distance of mobile device, \( m \), from \( c \), where \( j \) is an index on the coordinates we observed from \( m \) over a fixed period of time. Given that many devices provide us with 100s of coordinates a day, we can calculate \( \min(d(c,m_j)) \) over all \( j \) to determine the closest mobile device, \( m \), ever was to the location of interest within \( G \) over any set of coordinates provided by the device.

For each mobile device, in addition to location, we can observe the app behavior associated with \( m \). Using the crosswalk, we can also link many of the devices we observe, as discussed above, to a desktop. Through these linkages we can create a rich set of \( Y_n \). Finally, we can define \( L_{YG} \) as:

Table 2. Scores for marketer segments for city billboard.
\[ P(Y|I(\min(d(c,m_j)) < G)) / P(Y) \]

where \( I(\min(d(c,m_j)) < G) \) is an indicator function that equals 1 when the minimum distance between a mobile device and the location is less than \( G \), and 0 otherwise. Note: \( G \) can be chosen based on the measurement use case.

Using large data sets, the value \( L_{SG} \) can now be calculated empirically. In practice we run the above calculation for a large set of locations of interest. This way we can score locations relative to each other for an outcome of interest \( Y \). See Table [3] and [4] for examples of brand propensity maps.

Table 3. Scores for all OOH locations in Irvine, CA to a major surfing retailer
Conclusions

The research proves the ability to find a dynamic methodology for analyzing, targeting and measuring audiences exposed to OOH.

Since OOH locations permeate across the country, this method can be used to score all of these locations with respect to any of the URLs, combination of URLs, or any other observed actions on the desktops or the mobile devices seen near the location.

The rich insights that come from combining digital behaviors with location allows marketers to much more powerfully define audiences in traditional media channels, such as OOH and radio. Further, the scale of the data allows for time of day and week analyses that could benefit brand messaging or content, particularly with digital OOH. This methodology can be used for optimizing OOH content placement and brand sales because it allows one to identify the audience within meters of any fixed OOH location using more precise data for decision making over the legacy use of zip code and DMA level information only.

We can also help identify opportunities on cheaper, or less desired billboards that are similar in audience to the more highly desired billboards. This methodology can also be used for other purposes than advertising such as urban economics, identification of gentrification in process or store expansion economics, where to build your next franchise, etcetera.
References

The Reach and ROI of Mobile and TV

NikShah, Facebook

Andreas Bisping, Florian Renz, Stephan Knäble: GfK Advanced Business Solutions

Digitization has complicated the business of media planning enormously, by opening up the possibility of reaching people on multiple platforms and devices, both online and offline. The fact that these varied touchpoints often interact and feed back on each other makes it even more challenging for marketing and media leadership to evaluate the impact and efficiency of each campaign and elements within a campaign.

Cross-media channel planning is becoming the industry norm but media planners need better information to understand how to best optimize the mix between online and offline. Currently there is no standard currency available which is able to indicate which parts of offline and online advertising are most efficient. The market needs cross-media KPIs like the real gross/net return on investment for each media channel individually. With the increasing role of mobile advertising the need for a cross-channel evaluation is increasingly important.

The GfK Crossmedia Link

In order to answer the question of how mobile social media and TV interact as media channels, it was necessary to build some new tools. Before describing our methodology it is important to describe the tools used. In 2008 GfK introduced a single-source approach called “GfK Crossmedia Link”, formerly known as “GfK Media Efficiency Panel” (19,000 panelists). These cross-media single source panels allow advertisers, media owners and media planners to measure the effectiveness and the efficiency of specific crossmedia advertising campaigns on short term sales using data which link actual exposure to online advertising, along with electronic measures of offline advertising exposure such as TV, to real sales data. The core of this project is the ability to create a single source file containing actual measured ad contacts for TV and Online and actual measured purchase data.
The Crossmedia Link Panel is a Single Source Subsample of the Consumer Panel, encompassing 19,000 Panelists

Crossmedia Link Panel Sample Size and overlapping structure

**Figure 1: GXL Germany with Panelsizes and Subsamples. The Facebook cross-device data are available for all panelists within the Crossmedia Link population.**

**Data Link**

Until recently panels such as the GfK Crossmedia Link have relied on pixels, cookies and URL detection to track exposure of panelists to online advertising. This approach suits a world where an individual person uses a single device to access the internet and all digital ad contacts can be observed upon that device. However, the explosion in smartphone and tablet usage has made this approach untenable as panelists are exposed to advertising on various devices. A recent study by Omnicom Media Group’s BrandScience unit found that 91% of German internet users access the internet on at least two separate devices. This challenge affected Facebook long before other digital publishers as the majority of usage on Facebook’s service swung towards mobile devices very early on.

After GfK won the Facebook/ARF Innovation challenge in 2012 for their single-source panel approach, Facebook and GfK partnered to innovate a Data Link which enables GfK to receive all ad exposures for their panelists, regardless of the device they took place on, in a privacy-safe process that does not involve the transfer of personally identifying information. By using a trusted third party as well as salt-hashed user information, GfK are able to receive the data for all Crossmedia Link panelists without transferring panelist information to Facebook, and without learning anything about Facebook users who are not on their panel. This enables more comprehensive, privacy-safe measurement of mobile ad effectiveness than has previously been possible for any single-source panel, and also a direct comparison of TV and Facebook mobile exposures, which is the focus for this meta-analysis.
Modeling

To quantify advertising effectiveness a binominal logit regression model is used. The base is single purchasing acts measured in “GfK ConsumerScan” household panel (30,000 households), of which the media panels are subsamples. For each household and each individual in the households covered in the analysis, the information concerning product purchases as well as the contact dose of TV and Facebook campaigns is available. Furthermore, additional control variables like promotion pressure, loyalty and household demographics are taken into account to isolate the influence of media on purchasing decisions.

The logit regression model is one of the most commonly used tools for applied statistics and discrete data analysis, most useful for understanding the influence of several independent variables on a single dichotomous outcome variable (in this case, whether someone bought the product or not). As in other regression models the logit approach faces the statistical phenomenon called multicollinearity where independent variables are highly correlated. To assess the level of multicollinearity in the final aggregated model we run two different diagnostic analysis:

1. Examination of bivariate correlations between independent variables to detect the level of association.
2. VIF (Variance Inflation Factor) -> This factor measures the impact of collinearity among the independent variables in a linear regression model.

Both diagnostics showed that all independent variables were only weakly correlated (<0.15) and therefore there was no cause for concern over multicollinearity.

Using this logistic regression approach, GfK are able to estimate the short-term sales uplift contributed by each advertising medium in the campaign. From this, GfK can derive the incremental short-term revenue by medium, and the short-term ROI for each medium. The ‘short-term’ approach examines sales that took place within the campaign period and up to two weeks after campaign end. The performance of these media over the mid- and long-term is not analysed here (many of the campaigns finished only briefly before the analysis period) and may differ from the results presented here.

GfK were commissioned by Facebook to conduct a meta-analysis of seven German FMCG campaigns run between October 2014 and March 2015 in which both TV and Facebook were part of the media mix. The campaigns were chosen to be large enough to allow sufficient sample on the panel to study and compare the sales impact resulting from TV and from Facebook individually. Around 80% of the impressions delivered as part of these campaigns were delivered on mobile devices. Thus while the rest of this study looks at the overall impact of Facebook, this is largely a story about mobile. They included campaigns that had been studied individually on behalf of advertisers as well as others that were commissioned specifically by Facebook as part of this study. Here GfK present results from the individual campaigns as well as results from a dataset formed by aggregating the seven campaigns into a single “meta-campaign” to allow us to look at the effects of flighting, and impacts on different sub-populations with increased statistical power.

The total efficiency of a campaign can be summarized in three terms:
While reach and cost are relatively easy metrics to compare across media, a consistent way to directly measure impact across TV and mobile has been more elusive, and it is on this that our analysis concentrates.

**Findings**

**OVERALL LIFT BY MEDIUM**

Perhaps the most surprising findings from our meta-analysis revolve around the relative individual-level impact between Facebook and TV (defined as the increased likelihood that someone reached by either medium will go on to purchase a product, compared with if that person had not been reached). It has long been supposed that the more engaging creative format of TV (sight, sound and motion) mean that the impact is necessarily higher than that for other media, especially digital, and the advantages of other media lie mainly in the realm of cost. However in this set of seven studies where GfK can directly compare the impacts of TV and Facebook on the same set of people, it is seen that in fact Facebook is driving comparable person-level impacts to TV. In three of the campaigns, Facebook performed significantly better than TV in this respect. In three further studies the lifts from Facebook and TV were comparable, and in one campaign TV had a stronger impact than Facebook.

![Figure 2: Sales impact by campaign. Results based on analysis of seven FMCG campaigns run between October 2014 and March 2015 in Germany on both TV and Facebook.](image-url)
(Note: a lift of 1.00 implies the medium had no impact on sales; a lift of 1.50 implies that a person reached by the medium was 50% more likely to purchase the product after controlling for all other relevant variables.)

Looking at the aggregate results from the “meta-campaign”, Facebook has a significantly stronger impact on the reached audience than TV (p<0.05). Panelists exposed to Facebook are 26% more likely to purchase, compared with 12% for TV. One reason for the disparity in performance may be attention – the ideal setting of full attention being paid to a TV ad is rarely actualized – 74% of the German population claim to stop paying attention to the TV when the ads are on, with 60% of Facebook’s mobile users saying they use the internet while watching TV (figures from Dentsu Aegis Network’s Consumer Connection System study). In contrast, on the Facebook News Feed, each piece of content must be viewed, even if only briefly, before deciding whether to pause and read more or to swipe to the next item. A second reason may also be the accuracy of Facebook’s targeting. While the TV campaigns typically delivered many of their GRPs outside the stated target audience, Facebook by its nature is more precise. While TV may be having the same impact on people in the target audience, its overall impact is spread across a wide range of people who may not be the most ripe customers.

**LIFT BY FREQUENCY**

A direct consequence of the higher lift overall is that the impact for Facebook at a higher frequency is greater than the impact for TV at a higher frequency. The results of the model by frequency class are shown in Figure 3 below. Thus if an advertiser were to increase the frequency with which people are reached by a campaign, they may be advised to increase that frequency on Facebook, particularly as the current frequency on Facebook campaigns tends to be lower than that on TV campaigns. The exact optimal frequency depends not only on sales lift but also on cost, so the cost of increasing frequency needs to be taken into account alongside the increased impact on revenue.

![Sales impact by frequency](image-url)
LIFT BY EXCLUSIVE REACH GROUP

GfK also looked at the impact of each medium among people who were reached on that medium only, versus those who were reached on both TV and Facebook. The results show that Facebook has a significantly higher impact among those people who were not reached by TV. Further research would be required to know if this is because those people are differently disposed to the brands being advertised, or because TV is somehow pre-empting Facebook’s impact among the overlap population.

Figure 4: Lift by medium, broken out by overall lift vs. lift among people exclusively reached by this medium. This result is derived from the combined dataset of all seven campaigns viewed as a single ‘meta-campaign’.

Among people exclusively reached on TV, the sales impact was in fact lower than among the total TV audience. People reached exclusively on TV tend to be slightly older than the group reached by both channels, and they also tend to be more loyal to brands, with less propensity to switch. This means that media in general have a smaller influence over this group, as their buying habits are more ingrained.

FLIGHTING

Our final area of investigation revolved around the flighting of cross-platform campaigns. It is often thought that the order in which media are used affects their impact. In all but one of the campaigns studied, TV was on air before the Facebook campaign began. Anecdotally this appears to be standard practice. Our hypothesis was that flighting does not in fact matter, and media should be run simultaneously.
Figure 4: Change in impact when a medium is viewed first. Derived from the combined ‘metacampaign’ dataset. Only one campaign started with Facebook, while the rest started with TV first.

If flighting were to have an effect it would be enacted through the fact that the media that started first would be seen first by people. Therefore we were able to simulate the effects of flighting by identifying individuals in the “meta-campaign” who had been first exposed to a campaign on Facebook and others who had been first exposed on TV. GfK then modelled sales impact among these two populations. If one medium should be flighted before the other, we would expect to see a significant difference between the total impact caused by that medium and the impact among people reached first by that medium.

The results contradicted our hypothesis. Although the impact of TV is indeed almost unchanged whether it is seen before or after Facebook (TV lift is 2% higher if it is seen first), Facebook is significantly more impactful if it is the first touchpoint in a campaign (lift is improved by 10%).

Recommendations

MAXIMISE REACH ACROSS THE MEDIA PLAN

While the results presented above have focused upon the relative individual impact of Facebook and TV, there are other elements to the total efficiency of a campaign as noted above: reach and cost. In all of these campaigns, both the reach of TV and the amount spent on TV were significantly above the equivalent metrics for Facebook. Therefore, while Facebook was driving significant impact among those it reached, TV drove more sales overall due to its higher reach. It should be remembered that this study focuses on short-term sales impact (i.e. campaign period plus two-weeks). Longer-term sales impacts may be different for both media.
This analysis shows not only that Facebook is comparable to TV in terms of its sales lift, but also that a significant part of its impact comes from the populations it can reach but that TV does not. Given the comparable impact of Facebook impressions to TV impressions, it seems clear that maximising Facebook’s reach among the target population will improve the effectiveness of campaigns, just as is true for TV. While the physical and technical characteristics of each medium are different, the effect is comparable, and so should the planning be.

CONSIDER DIFFERENT FREQUENCY TARGETS FOR DIFFERENT MEDIA

Our analysis shows that Facebook can continue to increase its impact at higher frequency levels than are usual for classical media. The exact optimal frequency will vary according to the revenue generated by the incremental impact and the cost of the additional impressions, but given that the average frequency on Facebook for brands in this campaign was lower than that for TV, for many brands there may be room to consider higher levels of frequency on Facebook than they are currently planning.

UNDERSTAND THE VALUE OF FLIGHTING MEDIA

Currently most campaigns begin with TV. This research shows that TV impact is not significantly changed whether it is seen first or second. However, Facebook impact is significantly improved if it is the first touchpoint in a person’s experience of the campaign. Given that it makes no difference to the effectiveness of TV whether it is seen first or not, but it could raise the impact of Facebook, we would recommend that future campaigns begin simultaneously on both media.
Why Aren't You Looking at the Long Term ROI of Ads?

A Study Of The Current State Of The Art
And The Opportunities For Improvement

Jim Spaeth, Alice Sylvester

Conducted by Sequent Partners

For The Council For Research Excellence
The Question

Complete the thought:

Tony the Tiger says “They’re ...”

What do you deserve today?

Got what?

What stays in Vegas?

What kind of beauty does Dove promise?

15 minutes will save you what?

Who do P&G spokespeople say Thank You to?

When it absolutely, positively has...

You don’t have to be a Quintile 5 media consumer to know the answers to most of those questions. They’re some of the most famous advertising taglines in history. Do you know the sponsor? Can you see the ad in your head? Do you remember the industry buzz when people recognized the simplicity and staying power of those big advertising ideas?

That’s what we’re talking about here. Staying power. The ability of an ad to do its job long after a flight wraps up. Long after the agency and the brand manager get bored. Long after its contributions to sales have been tallied in marketing mix models.

How do you quantify the long-term effects of advertising? How do you structure analytics to give advertising credit for staying in the minds and hearts of consumers – for driving sales, while creating tight bonds, preferences and associations between consumers and brands?
In today’s ROI evaluations, for the most part, the focus is on short-term effects. The long term is not routinely considered – no one has time for it and few brand managers are incented to protect the long-term health of the brand. It’s a short-term world.

But does advertising get all the credit due in marketing mix models and other methods of evaluating return on investment? Not if the focus is only on the short term.

Why the disconnect? Why isn’t advertising’s tremendous power to stick around in consumers’ heads part of the routine evaluation of its contribution?

Read on!

**The Long History of Long-Term Effects**

Our world certainly knows about advertising’s long-term effects. There have been hundreds of academic and commercial studies/papers on the topic in the past 50 years. Please see the appendix of this paper for the bibliography and enjoy this historical overview.

In the ‘70s, as scanner data drove the popularity of marketing mix models, advertising data scientists (though they weren’t called that back then) and academics developed long-lasting concepts of adstock, half-life and decays ... and the Koyck lag co-efficient factor ... that are broadly used today. Some of you will know what that means.

Long-term was defined and validated ... papers cited solid evidence of an advertising effect occurring 3-9 months after the campaign ran -- and the impact was supported with actual market results. In the ‘70s!!

In the eighties, the long-term impact of advertising on brand equity was studied, and another long-lasting concept, advertising’s impact on company stock prices was developed and advanced. Despite growing evidence of the long-term impact, the c-suite (they weren’t called that back then, either) and brand managers remained steadfastly committed to short-term effects – next quarter – or by year-end. Driven to perform well, all of marketing was geared to read and react to short-term results.
The 1990s were probably the hey-days of long-term measurement. The concepts of “Purchase Reinforcement” and “Memory Effects” were introduced and the world-famous IRI “How Advertising Works” study showed long term effects were twice short term – 2x – and advertising still worked two years later. Penetration and repeat buying effects from consumer purchase panels were studied. More and more evidence supported advertising’s ability to reduce price sensitivity and modelers reveled in discussions of how to “decompose base sales” – that behemoth chunk of sales that happen without any marketing stimulation – in order to account for advertising’s long term effects.

By the 2000’s, work continued on base sales evolution, price elasticity, penetration and repeat purchasing, and single source data showed promise for assessing long-term effects. And finally, the effect of advertising on long-term brand building was advanced again with the integration of equity measurement and purchase dynamics.

Interestingly, in looking at the academic and industry papers over the course of the last 30 years, the hand-off from academics to practitioners is evident. These days, there are fewer theoretical explorations and more evidence showing how long-term effects of advertising work. The mechanisms.

To summarize, the big ideas in the study of long-term effects of advertising are: impact on brands or buyers -- brand equity enhancements over time … or purchase dynamics and customer lifetime value over time.

But so far, there is no unified field theory here -- no fully integrated approach to advertising’s long-term effects, spanning short term, intermediate term on buying behavior and brand attitudes.

**Just What Is The Long Term?**

Note the different ways modelers define and operationalize the long term:

“Our definition of long term depends on industry, but it’s usually three to five years.”

“For us, short term is one year. Long term is Year 2 and part of Year 3. If you go dark, you lose some of the long-term effect. Five years is a long time -- things change.”
We look at results over 6-12 months ... at most!"

We see that the long term ranges from 6 months to three years – and everything in between. That’s quite a range. It makes sense that category dynamics define the boundaries between short and long term, but it certainly hampers the industry’s ability to develop and evolve standard measurement and analytic practices. Several modelers asked for a standard definition of long term.

**Current Practices**

Frankly, we were a little surprised to learn that modelers can, and do, look at long-term effects of advertising despite the overwhelming demand for short-term or intermediate-term analyses. For the most part, it doesn’t happen routinely, but on request. Clients reach for long-term analyses when they need additional support for an element of the marketing mix that may have smaller short-term results than they need. Here’s how modelers described the prevalence of the practice:

“*It’s standard practice to provide a perspective on long term, but there are costs to going deeper. We do it more than half the time -- it’s bigger outside of CPG, but we also do it in CPG.*”

“We do it regularly when clients have continuous tracking – in more than half of our studies.”

“*Measuring the long term will be the current practice. But it costs extra and advertisers don’t believe it changes that often.*”

“We do it sometimes, not always. Not every client is asking for it. Different client think differently about the long term. Sometimes they look at long term more from an ROI perspective, other times from a brand health perspective.”

The data inputs to long-term models can be a real challenge – three to five years of data are required, and marketing variables can change a great deal in that timeframe. Many marketers keep only 52 weeks of data and Nielsen/IRI reportedly discard data after five years. And studying the long term can be an expensive addition to other modeling efforts.
How Well Are Long-Term Effects Received?

We wondered how well marketers responded to modeler’s reports on the long-term effects of advertising. After all, we are putting hard numbers to estimates of how long a campaign continues generating sales after it’s done or how long it will stay in the minds of consumers. Are marketers skeptical? Do they believe the numbers? Here’s what modelers told us.

“[They react] generally pretty well. The most push-back comes when the results don’t align with the brand’s pre-conception.”

“Some clients get it and some don’t. CMOs are comfortable with risk; some clients seem interested. A lot of heads nod when you talk about the longer term outcomes, but not everyone is willing to act on them.”

“There is a human factor. Some individuals just don’t believe in long-term effects, and then there are those who have made long term investments and need to justify them.”

On Average, How Big Is The Long Term Effect?

The “How Advertising Works” study taught the industry that the long-term sales effect of CPG advertising was equal to the short-term effect. That means over the course of the next two years, advertising campaigns continued to produce sales that eventually, in total, doubled the short-term sales. So ... how close or how far off this “norm” are long term effects the modelers see these days?

Their answers surprised us...

- “2x is not a bad estimate.” (Whoa, what? Wouldn’t that be surprising if it were really that simple?)

- “1.6x is so prevalent – but I don’t believe it – it has to differ by brand and campaign.” (Surprising how close 1.6x is to 2x!)
• “We don’t really have a quotable benchmark – long-term effects range from 0.5 to 2x or 3x the short-term effect.” (There’s that 2x again! But the range is good, and what we’d expect to see across brands and categories.)

• “It varies – the long term multiplier can range from -2 to 9x! Low equity, price-driven brands have low multipliers, while other brands have higher.” (Ah ... a big range that doesn’t average out to 2x. And an observed effect of brand equity on sales!)

We learned that in general, the long-term effects of advertising either vary considerably by brand and category or hover around 2x. While we expected to hear reports of a lot of variation by campaign and category, and a lot of precision, perhaps we need to be open to the idea that the long-term effect of some ad campaigns is about 2x – even if it seems too easy. It makes us think, though. Is the 2x a product of some central tendency in data or analytics approach, or the inexorable pattern of repeat-purchase behavior? We don’t know yet.

But it seems, as an industry, we’re a long way from knowing the answer to the size of long-term effects of advertising in aggregate. And maybe the aggregate shorthand answer isn’t really important. So many variables – brand history, the media and marketing mix, message and creative quality – contribute to short-term sales, and we would expect the same over the long term, too. We occasionally heard the question, “how should management use these results?” Currently they are most often used to provide a more positive ROI for financial justification of a strategic marketing decision. We rarely heard them used to make strategic marketing decisions toward building, or maintaining, profitable brands.

**How Do Long Term Effects Build?**

We were particularly interested in the mechanism of long-term effects in order to understand how marketers can manage the long term. Our investigation confirmed the How Advertising Works conclusion that there must be a short-term sales effect in order for there to be a long term effect. That’s useful to marketers. And we learned the effect builds and decays, but beyond that, there are few rules of thumb we can cite.
• “We only measure some of the long term effect in the short term – especially if there is a sustaining value such as repeat purchase. But there is only long term if there is short term, so we see the beginnings of the effect right away.”

• “There is a build up and a decay, but it’s hard to generalize.”

• “We’ve seen advertising half-lives of 0.5 for one or two years. Effects build quickly, but it’s a cumulative effect. If there’s no short term effect, there is no long term effect.”

How Big A Factor Is The Creative?

It is reasonable to assume certain types of creative appeals last longer than others. But surprisingly, we learned that this is not uniformly true. Perhaps this finding underscores the incredible complexity of marketing and advertising and the difficulty modelers have isolating specific variables. Or perhaps there’s another reason. In study after study, we have found that creative approach and strength are not routinely factored into marketing mix models. And yet, creative explains something like 70% of advertising effect. What will it take for creative power and message strength to become standard inputs into marketing mix modelers? Who will lead this charge?

Anyway ... some modelers have seen a difference between different types of creative approaches.

• “Branding ads have longer term effects than promotional ads. Celebrity and sponsorship benefits can be longer term – they create resonance in the marketplace.”

• “Call to action is more short term. Branded equity building ads have a more sustained flow from year to year.”

• “Creative is a big factor. Think of the Dove “Real Beauty” campaign. Emotional appeals do better, but we don’t see any other consistent patterns.”

Other modelers have seen no differences in creative approach.
“There is no difference in long-term effects by nature of ad. It both drives penetration and enjoys the long-term consequence of repeat purchases or it doesn’t.”

Others simply aren’t sure.

“We don’t usually look at this. Advertising is a multiple exposure effect, so it’s hard to know which ad is responsible for the effect.”

“Adstock depends on creative (and flighting). We see holiday ads in short bursts – and the half-lives are terrible. But they have a long-term effect because of the holiday tie-in. Sometimes emotional ads don’t break through. It’s hard to generalize.”

“I don’t know about creative effect. Usually a change of creative strategy is accompanied by many other changes, so it’s hard to parse out the creative effect.”

**Approaches For Modeling The Long Term**

Many modelers tend to be technique agnostic. The approach they take to evaluating long-term effects is driven by the interests of the client, the marketing strategy and the quality and availability of longer-term data.

In the literature search, we learned there are a number of modeling approaches for looking at the long-term effects of advertising. And we saw these approaches in action today.

- Repeat purchase model (Ehrenberg, vonGonten, Hess/Ambach, Wood)
- Price elasticity effects (Mela)
- Baseline decomposition/drivers model (Mela, Cain)
- Brand tracking (Millward- Brown/Vermeer)
- Brand value (Accenture, Bera)
Repeat Purchase Model

One of the most common approaches to measuring long-term effects, especially in CPG, is modeling repeat purchase patterns. It’s associated with a classic benefit of advertising -- advertising helps keep consumers in a brand franchise – and keeps them buying over time. Modelers tend to fix a point in time, say at the beginning of a campaign, and monitor the effect of advertising on subsequent purchases.

Baseline Drivers Approach

In marketing mix modeling, incremental sales are separated from the brand’s baseline sales. Sales associated with all advertising and promotion efforts drive incremental on top of a massive amount of baseline sales. For the most part, the baseline is largely unexplained. It just is. Because it’s unexplained, it’s also unmanaged by marketers.

Another popular approach to measuring long-term deconstructs the baseline – it looks at what drove the baseline, what created it, and what’s in it. With this understanding, marketers can manage and improve the baseline. Baseline sales are undiscounted, un-promoted sales – high value sales that a brand receives in the absence of advertising or promotional discounting efforts. It would stand to reason that the long term effects of advertising would be found in the baseline – presumably advertising creates awareness, differentiation, likeability and preference for a product that persist over time.

Price Elasticity Approach

Closely linked to the evolving baseline approach, another analytical approach looks at analyses of advertising’s impact on long-term price elasticity. This approach is founded on one of the most pervasive beliefs in marketing -- that over time, advertising creates differentiation, brand preference and a willingness on the part of the consumer to pay more for a product. Advertising reduces price elasticity. Several modelers look at the long-term effects of advertising through this lens.

Brand Tracking Approach
Though it is a challenge, modelers also look for long-term effects of advertising in brand tracking data. They work to determine how advertising impacts brand health and consumer attitudes towards the brand over time. Presumably advertising messages are motivating, interesting and entertaining, and enhance consumers’ thoughts and feelings towards the brand. We should see a long-term effect of advertising played out in brand tracking data. Stronger brand attitudes should lead to stronger brand sales. However, a direct measure of advertising, branding and short-short-term and long-term sales remains elusive.

**Brand Value Approach**

Brand value – assessing advertising’s contribution to a brand’s financial value – is an approach several modelers discussed. The operational model is that advertising entrances consumers to stay in the brand franchise. Their purchasing power over time has a value, in financial terms, the net present value of which is attributable to advertising.

**Are the Approaches Validated?**

We were interested to learn whether the results of long-term effects analyses hold up -- whether they are routinely validated, and how they are validated. If we want marketers to consider the long-term effects of advertising, it’s important that the forecasts hold up. Here’s what we learned:

Some modelers definitely have an approach for validating their long-term forecasts.

- “You can compare forecast results with actual results for any long term projection. We routinely do this. Take data from out-years and use them to predict Years 1 and 2. Then use the actuals to see where the model is over/under representing.”

However, validation is a challenge.

- “We’ve done validations, but they’re hard to do. We look across different approaches for convergence – and see how our analyses play out across the marketplace – a lot of times, new competitors enter or changes in the economic environment intervene.”

Generally, estimates from the models are self-validated over time.
• “We see if results match the forecast. One year, three to five years typically, forecast well.”

• “We look at face validity. When we have upper funnel models and forecast businesses, the forecasts tend to be better.”

• “It’s self-validating. If you get the model right, what you forecast is almost ordained in the aggregate.”

Final Thoughts

We learned a lot throughout this exploration and see a number of ways to advance the industry’s thinking on how to measure long-term effects.

First, let’s share more learning. Many modelers have specific brand case studies and would be willing to share their learning (pending client permission, of course.)

We think there’s an opportunity to take all the case studies modelers share and see if there are any normative effects on the size of the long-term effects and the processes with which they are modeled.

In addition, we’d like to see renewed interest in brands, brand equity, brand value and the role of advertising. There’s an opportunity to develop a management learning event and promote long-term vision. The ANA could also participate and engage marketers directly.

And finally, the industry needs to keep going – advance technically. Store level data, the fuel of most marketing mix models, will never reveal the importance of advertising long term, so we need a true consumer-level understanding. We need to understand repeat purchase behavior and
other behaviors – as well as consumer attitudes and perceptions on the value of brands. We need to unify definitions – especially on something as straightforward as what is the long term? And we believe there is room to study more thorough validation approaches.
Appendix - Bibliography


Appendix - Contributors

- Many thanks to Leslie Wood and MASB, WARC and all the experts who pointed us in the right direction for the literature review and discussion guide

<table>
<thead>
<tr>
<th>Company</th>
<th>Person</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nielsen</td>
<td>Carl Mela</td>
</tr>
<tr>
<td>Analytic Partners</td>
<td>IRI</td>
</tr>
<tr>
<td>MPhasize</td>
<td>Marketshare</td>
</tr>
<tr>
<td>Brand Science</td>
<td>NPD</td>
</tr>
<tr>
<td>Georgia Pacific</td>
<td>Kellogg</td>
</tr>
<tr>
<td>Accenture</td>
<td>Nielsen MBI</td>
</tr>
<tr>
<td>NCS</td>
<td>ThinkVine</td>
</tr>
<tr>
<td>Concentric</td>
<td>P.M. Cain</td>
</tr>
<tr>
<td>Bera Brand Management</td>
<td>Mike vonGonten</td>
</tr>
<tr>
<td>Millward Brown</td>
<td>Marketing Evolution</td>
</tr>
</tbody>
</table>
Tuesday, June 16
The Next Frontier: Content Analytics

*Discovery and Application of the Video Genome*

Bill Harvey, Chairman, ScreenSavants and Bryan Mu, Vice President, Research & Analytics, NBCU Cable Entertainment, UCP

Précis

A systematic process found 273 metatags that substantially shape viewing behavior to television and motion pictures. This paper reports the story of that discovery, its validation, and how the data are being applied today.

Background

In the beginning, storytellers sat in the firelight and entertained their audiences face to face. The feedback was instantaneous and visceral to the storyteller, he or she got most of the signal from the audience, with very little noise. Fast forward to the 21st Century, where people who create screenplays for features or for video, receive feedback in the form of box office and Nielsen statistics, dial tests and other data tables which replace the audience’s direct reactions.

The information collected and reported, and the much smaller quantity of that data which actually gets attention, is currently focused on the need to add some form of objectivity to the decision to go or not go. There is not a great emphasis on giving feedback that might help the creatives dream up even more of what works on a given theme or in a given character. The latter is “termed” prescriptive with extreme negative connotation.

As a result, of the estimated $50 billion invested in the creation of screen content annually, the percentage of properties reaching profitability as a percentage of properties started is extremely low. For example, in the case of series television, only about ~20% of concepts become pilots, only about ~20% of pilots become series, only ~33% of first season shows make it to a second season. The percentage of pilots that reach profitability (five years on network to create enough episodes for syndication) is under 2%. As a benchmark, 40% of the tests of heavying up a television advertising campaign behind a specific TV commercial (also content) more than pay back the incremental media costs with sales.

We might say that the higher predictability of advertising results as compared with entertainment results is a reflection of the historically greater use of data in advertising than in entertainment. Many would question this view out of a correct understanding that creativity transcends data. However the writers can envision a plausible scenario in which creatives, instead of being given report cards, are given audience reactions in a form highly digestible to creatives, such as in the language they use when they think about storytelling. With the right execution of measurement and reporting for the particular circumstances, why could not creativity and science harmonize effectively?

Objectives

1. Understand what drives viewing behavior. Transcend simple reporting of who watched what, and learn why they watched it. Codify this as science.
2. Be able to delight the average human being by increasing his or her viewing pleasure through the recommendation of fare personalized to the user’s deepest viewing motivations.

3. Have creatives hungry for the stimulation they receive from audience feedback digested into thought starters, i.e. clumsy examples of efforts to do the writer’s job.

4. Optimize advertising by having the metatags of the ad and the program environment in harmony. TiVo Research has reported that the creative is 65% of the ROI lift effect of using direct-match singlesource at scale. In order to optimize creative, one must have levers to pull. DTags are the first levers for creative the industry has ever had. DTags are analogous to the Human Genome and the Mendeleev Table of the Elements in exposing hidden structure. These are the hidden structures that motivate what people watch.

5. Help media executives make decisions by adding the dimension of motivations. Lead-in/lead-out scheduling and tune-in placement decisions, etc.

6. Recalibrate the metatags for each culture.

7. Validate at each stage of the objectives.

**History**

In the 1970s, Harvey and Dr. Timothy Joyce worked on several experimental projects around Harvey’s concepts of “psychographics that matter”. Over 10,000 words were systematically drawn from the Oxford dictionary and used in national self-scaling surveys. Factor analysis reduced the mass of data to 20 psychographic questions still used to this day in one of the leading syndicated services in the world, MRI.

Harvey used about 1000 of these words along with others developed in the research connected with the launch of many new cable networks in the 1980s to compile a list of ~1500 words describing people, content (e.g. programs/movies), human values, emotions, moods, and situations. These words had inductively risen to attention in the course of research including open end questions.

In the 1990s Harvey’s company Next Century Media (NCM) did three things for the first time: (a) addressable commercials (b) set top box data to research grade (c) data driven program recommender. These software solutions were partially deployed and tested with four leading MVPDs and then used by a fifth MVPD to analyze the Aurora addressable trial.

In the preferred implementation, the viewer would hit the remote’s “A” button to receive a recommendation for the next time period. For example, if at 8:55PM a viewer were to hit the button, he or she would receive one most recommended show starting at 9PM. The system would look back at all the set top box data, at the shows this set top box had watched, and the metatags that NCM’s experts had attributed to those shows. It would then eliminate all shows starting at 9PM that the set top box had viewed for at least 10 minutes in the past, and make a best match of weighted metatags between the subject’s profile and the choices at 9PM. The viewer could then hit the “B” button to applaud the
recommendation or the “C” button to boo the recommendation. In tests, applause ran more than 90% over boo.

In the course of the 1990s work it was seen that of the ~1500 metatags, only 273 of them had high predictivity that the viewer would become loyal to the recommended show. This was corroborated independently by set top box data and by an online interactive questionnaire which used the concept of “accelerated time”. The viewer was given a roster of programs/movies and asked to rate on probability of viewing. Those assigned higher probabilities were inspected for their metatags. The next roster shown was loaded with programs having those metatags. In test after test, the scores were seen to go up on every pass. Eventually a viewer would exhaust all the shows in our test that could possibly satisfy him or her and then the scores would hit a ceiling and begin to go down.

Figure 1

The Experimental Procedure

Exhausting Available Positives

Validation
More recently the 273 DriverTags™ were re-validated as predictive of Nielsen ratings, and conversely of series cancellations.

In the first study, the evidence was arrayed heuristically. In Figure 2 you see a number of DTags™ arrayed down the vertical, while across the horizontal the failed and cancelled shows are to the left and the successful, high rated and not cancelled shows are to the right. The eye can see that there is more density of the Orange DTags in the upper left and conversely higher density of green tags in the lower right. This means the orange tags are negative drivers to the success of a sitcom, and the green tags are positive drivers to the success of a sitcom. (The actual DTags are not shown, instead showing anonymizing handles on the DTags themselves.)

**Figure 2**

One can see in this graphic that the upper left hand quadrant is most densely covered in orange (negative DTags), correctly predicting that these shows would be cancelled. The green is most dense in the lower right hand quadrant, the happy land of the successful shows, where the positive DTags again correctly predicted and explained the actual results.

In the next study we moved to the use of the metric “Adjusted R Square”. This is the percentage of variance accounted for in some dependent variable, in this case being the Nielsen Live+7 day household rating. The household rating accounts for 85% of the variance in the people ratings and is passively collected, hence its choice as the most robust “payoff measure” against which to scientifically measure other measures.⁴

“R” is the coefficient of correlation. For example, a top agency found that TiVo Research had a .9533 correlation with Nielsen ratings.⁵ “R Squared” is “R times itself” so in the latter example the R Squared is .9533 times .9533 which equals .8663. R Squared is obviously a tougher test, a higher bar, than R. A still higher bar is the metric “Adjusted R Squared” which is adjusted to discount R Squared based on how many variables it uses. Adjusted R Squared is the most rigorous test and so is the one we chose.

**Metrics of the Context**

In evaluating any new metric, it is appropriate to use the same yardsticks to also appraise the other metrics which practitioners have already been using. Research Measurement Technologies (RMT), the company acquiring NCM’s metatags and renaming the key 273 DriverTags, carried out two surveys with...
Vision Critical’s national U.S. panel. The first, conducted November 4-7, 2014 measured 40 new series using the standard intent to view question and the standard quality rating question used by most practitioners. Trailers were the stimuli in that wave and in a second wave February 18-20, 2015 which measured 22 additional new series. Each of the two waves had intab samples of 300 respondents. In the second wave the Vision Critical second by second tool called Media Impact was used as a sort of a textured “dial” to get second by second reactions to the stimuli. Ten different buttons allowed the collection of reactions along ten different dimensions. In normative practice today, one is using a slider or dial that measures a single dimension, representing enjoyment, which was one of the buttons (“Enjoying It”) used in the second wave.

It therefore became possible to compare directly the Adjusted R Squared of DriverTags with the dial, intent to view, and quality rating scales used today. Those results are as follows for all shows used in the second wave study just described.

Note the extremely low predictivity of the conventional metrics. However, this could be accounted for by a number of factors, including:

- Because the stimuli were trailers (typically 2 minutes or less) and cherrypicked to include the best scenes and cuts, the range in scores in the survey metrics were relatively tightly clustered, unlike the range in the currency ratings which fell across a far wider range. Specifically, the highest value on the survey metrics was from 2.8X to 4.7X the lowest value depending on which of the four metrics one chooses, indicating a truncated range of mostly high values due to the cherrypicking. By stark contrast, in the Nielsen ratings being predicted, the highest rating was 64.4X the lowest rating for these same 22 shows.
- These 22 programs have in most cases just recently launched and the currency ratings used in the analysis may yet to have stabilized.

We expect that as more data are collected, we will see that the traditional metrics of dial, intent and quality will emerge as having their own significant value as predictors, when at least 8 minute samples are tested, that are more actual stories and less sizzle reel of top moments. We predict that .80 Adjusted
R Squared will be achieved on samples of over 100 series once this combination of traditional with ScreenSavants metrics is fully implemented.

The next analysis combined the sample sizes of all series for which DTags, Historic Factors, and published currency ratings were available, for a reappraisal with more stable sample, but now lacking the dial, intent, and respondent perceived quality metrics.

Practitioners use variations on the theme of “recent historic rating averages enjoyed by the network that later carries the specific program, in the specific time period and day where that program will appear”, as a predictor. ScreenSavants has its own variation on this metric, called the ScreenSavants Proprietary Historic Factor (Historic Factor for short), with proprietary aspects felt to increase predictive power. Here, then, for the largest sample of shows possible to date, is a comparison of DriverTags with the Historic Factor.

The DTags on their own, without any survey measures, attain .64 Adjusted R Squared (a correlation of 0.80). This is on a very large sample of 152 shows, approximately the top 150 series. When the Historic Factor is added to the DTags, this brings the 0.64 up to 0.76 (a correlation of 0.872). These high degrees of predictivity of currency are unprecedented.

This suggests that as the industry continues its evolution, the use of all of the traditional and new metrics including the DriverTags will cause an increase in the predictivity of ratings. When 8 minute or longer actual samples of what a series will be to experience are used with the traditional metrics, we expect to see that the dial, intent, and quality rating scores have significant predictivity they will add to the DTags and Historic Factor. An expectation of 0.80 Adjusted R Squared is not unreasonable. For samples of up to 78 shows each we have seen 0.86 without survey measures.

Tests with Real World Creatives

In September-December, 2014, NBCU conducted a series of three tests with ScreenSavants, and a different network conducted a fourth test. In all four cases, feedback was given to the creatives, with the intention to:

1. Break the mold of the old “research versus creatives” box
2. Walk dangerously close to being prescriptive without falling into that category
3. Provide insights that the creatives would take to heart
4. See if this had a noticeable effect on the direction of the property

It was realized that this was too early a test to yield actual ROI proof. In the future some of these properties might be producing data that could be aggregated into an analysis of test vs. control groups.

Anecdotally the reception given was extraordinarily positive. The number of cases was small and yet enough studios and networks became interested in DriverTags as a way of penetrating to the next level of the game, such that the news went viral within the community. Not only did creatives like the input, it agreed with some of the things they had been thinking but hesitating to bring forth. It was found that the feedback could also be used by studios in pitching to networks, “a new vocabulary to upgrade the stories” was how one wise person put it.

Future

The ScreenSavants Vision is to create a machine learning system inspired by Watson. Humans will continue to apply tags to shows, however, so will robots. Seven different robotic methods are described in the RMT patent. The human and robot inputs will be separately and continuously measured and given a weight by Adjusted R Squared against ratings, box office, and in the case of the recommender application, applause:boo ratio and loyal conversion rate to recommended programs. The weights will be optimized to maximize the prediction.

Creatives will be able to start to get feedback from the artificial intelligence (AI) as soon as they can start to key in their first notes.

RMT is in discussions with technology and media companies about partnering to bring this about in the most expeditious way. DriverTags promise to add a level of understanding of process that can help increase success rates in the creation of new screen content. With screen content consumption time exceeding everything but sleep in some cases, this is an important area in which to increase effectiveness and efficiency, but cannot be allowed to slow down creatives, the process at all times must be alert to creatives as the prime users.

References

1 These statistics derived from Variety compilations, one author’s own 10-year track record compilation, inputs from Melva Benoit, and conversations with leading program researchers.

2 ARF Adworks I study, compilation of IRI BScan tests. Please contact authors for details.
