# #IPA Social Works

# The Expert Guide to Measuring Not Counting

## How to Evaluate Social Media for Marketing Communications

About

# #IPA SocialWorks

- Authors
- Foreword
- Seven Key Messages

## About #IPASocialWorks

We are hugely fortunate in our industry to have the effectiveness 'case law' gathered from the IPA and The Marketing Society, amongst others. This vital work provides knowledge about how most 'traditional' communications work in delivering true business value and how they can be made to work even better.

However, this case law simply does not exist to the same extent when considering the various usages of social – not really a surprise given its relatively recent nascence. This is the gap that the #IPASocialWorks project was designed to fill.

So we launched a unique, industry-wide initiative, led by the IPA, in partnership with The Marketing Society and Market Research Society, supported by Twitter, Facebook and The London Business School. Our objectives are to find the elusive case law where we can see a proven ROI on social, and so draw conclusions for the industry about how to best measure and deploy social in its many guises as effectively as possible.

As well as rigorously tracking down and evaluating robust case studies, we're publishing this Guide to represent our learning to date, the lay of the land as we see it today, including emerging best practice.

We hope this will help the industry navigate its way through the labyrinth of data now available to us all and so deliver that elusive social media measurement that we know clients and agencies across the world are rightly demanding.

Stephen Maher
Chair of #IPASocialWorks
CEO MBA
Chairman of the marketing society

## Authors

### #IPASocialWorks cases

A team led by Fran Cassidy has been collating a large number of case examples from the UK and around the world, subjecting them to peer review, in order to identify those that show the effectiveness of social and which highlight the use of robust evaluation. The collection of case studies is growing all the time and can be accessed from the IPA website at:

#IPASocialWorks >

### The guide

Led by Ray Poynter, Fran Cassidy and Simeon Duckworth, a team of specialists have put months of research into this guide, also inviting contributions from academics and practitioners all over the world, to create a hub of best practice in social media measurement. This Guide is the first published iteration of this project. A short version of the Guide, for those looking for a more general guide to social media is also available *here.*



**Ray Poynter**
The future place

Author of The Handbook of Online and Social Media Research and The Handbook of Mobile Market Research, the founder of NewMR.org, and the Managing Director of The Future Place.



**Fran Cassidy**
Cassidy media partnership

Fran runs an independent marketing and research consultancy specialising in on line and off line media, the entertainment, and the marketing services industries. She works across Europe and the US and has experience as a client, in agencies and as a media owner. She is also a Board Director of The Marketing Society.



**Simeon Duckworth**
Mindshare

Simeon is Mindshare's Global Head of Business Planning and a member of its Global Exec leadership team. He has been at Mindshare for over 10 years. Previously he worked at Ogilvy Advertising and as a public health economist.

## Foreword

Social media marketing has come through the usual hype cycle and is now well established within the marketing armoury, although it's still developing and its role is still emerging.

Or, rather, roles: like other digital media, social can be used in many different ways under the broad headings of marketing communications, customer insights, customer experience and relationship management. It will take time for marketers to learn how best to exploit it under each of these headings.

**Patrick Barwise**
Emeritus Professor of Management and Marketing
London business school

This Guide focuses on the first category - the use of social media in marketing communications - although it also touches on the others. (Categories and distinctions in marketing are rarely clear-cut). The potential is obvious: social media has achieved huge global penetration and usage, and as marketers we naturally like to reach customers and prospects via all the media they use, provided we can find the right ways of doing so.

The potential is even greater today than five years ago, not only because of increased penetration and usage but also because of the parallel growth of the mobile internet: we can now reach consumers whenever we want and wherever they are and we also increasingly know where they are, when, as well as who else they talk to, when, and about what.

All this raises some big challenges. Clients need to know that money spent on social is delivering value. This means measuring impacts and outcomes, not just counting clicks, likes and interactions. Like the internet 10 years ago, social media marketing is now coming of age.

Measuring social is hard because:
• It needs to work in combination with other channels and activities.
• It is dynamic and interactive, which raises a number of measurement challenges.

• Social media campaigns can have many different aims.
• Campaign budgets tend to be smaller than those used for other types of media, such as TV, although larger ones are appearing and measurement approaches need to reflect this diversity.

On the plus side, like all digital media, social can generate a lot of data at little or no incremental cost and lends itself to the test and learn approach traditionally associated with direct marketing.

As the importance of social grows, along with the resources allocated to it, there is a growing need to evaluate social with the same rigour as that applied to more traditional channels, and increasingly with metrics that take an integrated, as opposed to channel-specific, approach.

This Guide is well timed to take stock as social media for marketing communications come of age. It summarises the current state of play and offers guidelines on how to make the most of this young but rapidly maturing medium, based on and illustrated by some great case studies. I'm not aware of any other publication anywhere that addresses these issues so comprehensively.
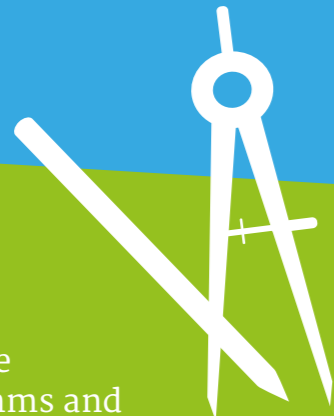
## Seven key messages

Use of social is still in its infancy and social tools, data and methods are fast-changing. Currently, there is no one best approach to measuring social activity. We are still learning. This Guide marks a moment in time, and a step on a journey. However, seven key messages have emerged from the project to date.

### 1.
**Social is more than marcomms and is challenging organisations**

Social is helping to bring the voice of the consumer to the heart of the organisation. Not only is it broadening the definition of media, but it is also blurring the traditional lines of responsibility for marketing and insight. It is operating as a communication channel, a service delivery platform and a source of insight. It is challenging the concept of a campaign with a clear start and end, as it is always on. For many organisations social data sets are now becoming part of their company-wide digital transformation. However, this can bring with it challenges for creating the sorts of reliable data sets suitable for accurate, predictive and attributional modelling. It is also shifting the balance of the organisation from collecting data to interpreting and analysing signals from multiple sources.

### 2.
**Social is changing the way we measure – its evaluation is more than a dashboard**

Social is a new and powerful source of insight for advertisers. For evaluation, it provides new ways to understand not only 'what' happened but 'how' and even potentially 'why'. Beyond generating new metrics, social is also changing the nature of measurement. Because its feedback is real-time, the evaluation process is being integrated with each stage of campaign management from strategy, targeting, content development, delivery and evaluation. Increasingly, faster learning will require a greater emphasis on predictive benchmarks and testing and not just metrics and dashboards. This 'always on' aspect should force organisations to adopt a much broader culture of test-and-learn than is currently evident, increasingly in collaboration with external data partners, agencies and platforms.

### 3.
Avoid a siloed approach to social measurement

Social tends to work in conjunction with other media. It cannot be measured in isolation. Social needs its planning and evaluation to be integrated with other channels in order to maximise its benefits, establish its value, and be more trusted as a mainstream option. Further, in the majority of cases, the success of owned and especially earned tends to be a product of paid and interaction with other media. The learning objective for social evaluation is to understand how it works with other marketing at all stages of the consumer journey.

### 4.
It is easy to overestimate the value of earned media and influencers

Accurately measuring causality for earned media is hard. Even with some of the most sophisticated statistical techniques, it is easy to see a causal link when in reality there is only correlation. Another reason to cultivate a broad 'test and learn' culture.

### 5.
**The commercial value of social will increasingly lie in the richness of its data**

Current methods of collection and analytics are not fully mature. Two areas in particular have further potential: sentiment analysis and Social CRM. Sentiment analysis will never be 100% accurate, but improvements in algorithms and data collection, will allow the signal to be stronger and more reliable. For Social CRM, given the potentially clearer value exchange for customers in offering personal social data, these data sets could be part of a gateway into much richer insight across an organisation.

### 6.
Social can learn from traditional planning

Social may be new, but the planning process is not, and the best way to make use of the new opportunities presented by social is to ground them in what is already known about campaigns and other communication activity, e.g. linking to objectives, based on clear assumptions, using comparable metrics. Social needs to adhere to the strategy and planning disciplines used across other marcomms activity and to be designed in from the start, not added retrospectively.

### 7.
Even short-term results need a long-term context

One of the benefits of social is that it provides measurements that allow campaigns and activities to be optimised in real-time. However, the management of campaigns should balance long-term success with short-term success, since they tend to depend on different elements and strengths. The IPA has shown that key factors such as profitability and loyalty result from long-term effects, not simply cumulatively from short-term successes.

# What you need to know about this guide

Social is still in its infancy and rapid innovation has led to a proliferation of data, methods, and language. Nobody can claim to fully understand how social works in the marketplace and its nature and use are still evolving. Some key points about measuring social are becoming clearer, such as the need to measure the impact of social activities on relevant metrics and the limited value of simple concepts such as counting the numbers of posts, followers, and likes.

The lack of established practice on the evaluation and measurement of social is costly and is holding back innovation. It is costly because people are not sure of the best ways to optimise their use of social. It is holding back innovation and slowing down adoption because organisations are unwilling to invest more until they can evaluate the benefits more accurately. Therefore, a cross-industry initiative called #IPASocialWorks has been formed to help highlight good practice and advise on measuring and evaluating social.

The first time that such a wide collaboration across the marketing and advertising industry has been formed.

## Key assumptions

The project is shaped and guided by the following assumptions:

Social should be held to the same sort of standard as other forms of media and activity. If social is held to a higher standard, opportunities will be missed; if it is held to a lower standard then it will tend to be considered as a fringe option.

Social is not just a medium and its use is much wider than marketing. In order to make this project practicable it has focused on marketing related issues and specifically on the evaluation of marketing campaigns and activities.

Social does not always fit the traditional model of a campaign, with a clear start and finish. In social it is necessary to think of campaigns and activities comprising a variety of forms and timespans.

The measurement of campaigns and activities in social should build on what is already known about evaluating campaigns and activities in other media.

Organisations spending money need to be able to identify the value they are achieving, so social needs to be measured/evaluated in terms of metrics that relate to business objectives.

It is recognised that there are many interesting uses and applications of social that are not included within this project. Other uses include: using social to create marketing ideas; as a source of research about things like customer satisfaction, product usage, and brand awareness; and the use of qualitative research in the context of social. However, these wider issues are largely outside the scope of this Guide. It is envisaged that there might be future projects, for example, looking at social as a source of insight and social as a method of delivering services and interacting with customers/users.

## #IPASocialWorks objectives and scope

#IPASocialWorks is an industry-wide project to help advertisers, advertising agencies, researchers and social media owners develop a more robust approach to measuring the effectiveness of marketing through these new platforms. It is supported by the IPA in partnership with the MRS and the Marketing Society, and sponsored by Facebook and Twitter.

The project addresses the perceived mismatch between the measurability of social and limited amount of evidence into how and when it works. A large part of marketing's interest in deploying of social media is based on the ability of these platforms to create data and be measurable in the broadest senses. However, many businesses struggle to see the connection between social data (for example, a tweet) and outcomes (for example, sales or profit). The advertising industry has spent many years addressing similar perceptions about its effectiveness in the context of more traditional media and many of its learnings need to be leveraged in this new context.

The relevant learnings from evaluating marketing communications in the context of traditional channels include:

• Rooting effectiveness in business metrics, as opposed to marketing metrics.
• Distinguishing between correlation and causation.
• Baking in effectiveness thinking from the start.

One of the key challenges for utilising social is its lack of history and benchmarks. Traditional advertising media have developed reasonably slowly over the last couple of decades, allowing advertising and marketing systems to evolve with them.

The social media ecosystem is relatively new and is continuing to change and evolve. While the old media world suffered from a paucity of data, the new social world is awash with data. However, that data is of differing quality and standards. Whilst it took the UK advertising industry more than 30 years to create its current understanding, social media measurements need to be useful and reliable now.

#IPASocialWorks is creating a comprehensive knowledge bank of social media effectiveness – the equivalent of the IPA's 30 years of peer-reviewed advertising effectiveness data set or the Marketing Society's Awards for Excellence.

# Defining and using Social

Like many phenomena, social is hard to define, but easy to recognise. Key features that are at the heart of social are:

### Interactive and social

Users of social can share, like, link, create, amend, and connect with the material and with other users, i.e. there needs to be a peer-to-peer element.

### Digital

Whilst there has always been social, in the context of this project the focus is digital.

### Attributable media

Not all social media is attributable, but social is accepted as a highly measurable form of media. Attributable media means it is often possible to determine who saw what, who interacted with what, and what they did next.

## The evolution of social

When social media first burst onto the scene, with networks such as MySpace and Friendster, it was a distinct sector and one that created ambiguity about how it would be financed in the medium and long term. Since then there have been three major developments:

### Adoption

The adoption of social by a wide range of sites and services (for example news services, travel, and retail), seeking to get people to share, like, post, and co-create.

### Movement

The movement towards mobile is shaping the way social is being used, particularly in the context of SoLoMo, the combination of social, location (e.g. GPS and beacons), and mobile.

### Creation

The creation of ad platforms that leverage their social members.

## Using social

There are a number of definitions and use cases for social. For example, Patrick Barwise, Emeritus Professor of Marketing at London Business School, likes to divide the marketing uses of social media into Marcomms (essentially, advertising), Customer Service, and Customer Insight (Barwise & Meehan, 2010). Similarly, McKinsey (Divol et al 2012) have defined four marketing related categories:

- **Monitor social channels for trends/ insights.** Monitoring implies researching what people are saying in social media (for example, posts) and what they are doing (for example, sharing, viewing, re-tweeting). This sort of activity is also referred to as social media mining, social media research, and listening research.
- **Respond to consumers' comments**. This includes crisis management and the provision of customer services.
- **Amplify** current positive activity/tone. Amplify is a broad term including: fostering communities, organising and promoting referrals and recommendations, and enhancing brand advocacy.
- **Lead** e.g. create changes in sentiment and behaviour. This category includes a broad range of marcomms and promotional activities, such as awareness, launches, and deals.

The scope of this Guide focuses on marcomms, and includes those types of activities covered in the remainder of this section. Issues such as service provision and crisis management are only tackled when they are closely linked marcomms related activities.

## Social media and the POEM framework

Social is often described in terms of the POEM model, which stands for paid, owned, and earned media (Corcoran, 2009). Each of the elements of the POEM model creates specific opportunities and challenges.

### Paid media

Paid media is the most traditional of the three types of media and refers to all paid media channels of course, not just paid social media. For example, paid media include:

ADVERTISING:

| TV | ONLINE DISPLAY | PAID SEARCH | AD WORDS |
| PRINT | RADIO | DIRECT MAIL | PAID ADVERTISING IN SOCIAL NETWORKS |

The Doritos campaign, highlighted in the case studies, centred on a paid media campaign. The paid media had as one of its aims driving people to owned media (i.e. its Facebook page). The Doritos campaign highlights the way that brands often seek to combine different elements of POEM.

### Owned media

Owned media means owned by the client, for example a corporate website, Facebook page, or Twitter account. Brands vary massively in terms of the scale of their owned media. Some brands have access to very large communities on platforms like Facebook, Twitter, and LinkedIn. For example, in January 2014, Coca-Cola had about 2.5 million Facebook fans from the UK and 93 million globally.

Owned media includes other media owned by the client, including campaign sites, competitions, apps, as well as older forms of owned media such magazines and newsletters. Although owned media is sometimes referred to as free, it still incurs the costs of time and effort, the materials shared via owned media cost money too, and there is the substantial capital cost of having owned media.

**2.5 Million**
UK Facebook fans

**93 Million**
Global Facebook fans

### Earned media

Earned media refers to material that is shared or distributed by people who are not being paid for what they are doing. In social media this term includes review sites, shares, uploads, replies, retweets, favouriting, discussions in forums and on bulletin boards. It is sometimes referred to as C2C – customer-to-customer, or peer-to-peer. The 2013 Nielsen Global Trust in Advertising and Brand Messages Report ranked earned media as the most trusted, followed by owned media, with paid media being the least trusted of the three. Interestingly, the trust levels reported for Europe are much lower for almost everything, compared with other regions.

One concern about earned media is whether it is genuine. There have been high profile cases where online reviews or blog posts that appeared to be from fans or neutral third-parties turned out to be paid placements. For example, author Stephen Leather has confessed to creating fake reviews on Amazon to help create buzz around his books (Charman, 2012).

### Converged media

Converged media is where two or more elements of the POEM model are combined. This tends to bring the quality of the content to the fore, rather than just the number of people it reaches. The idea underpinning a converged media strategy is that if material is sufficiently interesting people will want to share it, link to it, add comments, etc. The paid or owned media is used to provide the resources for the earned media to utilise, and to generate sufficient interest to cause people to create earned media.

Shares

Replies

Review sites

Retweets

Favourites

Earned Media

Uploads
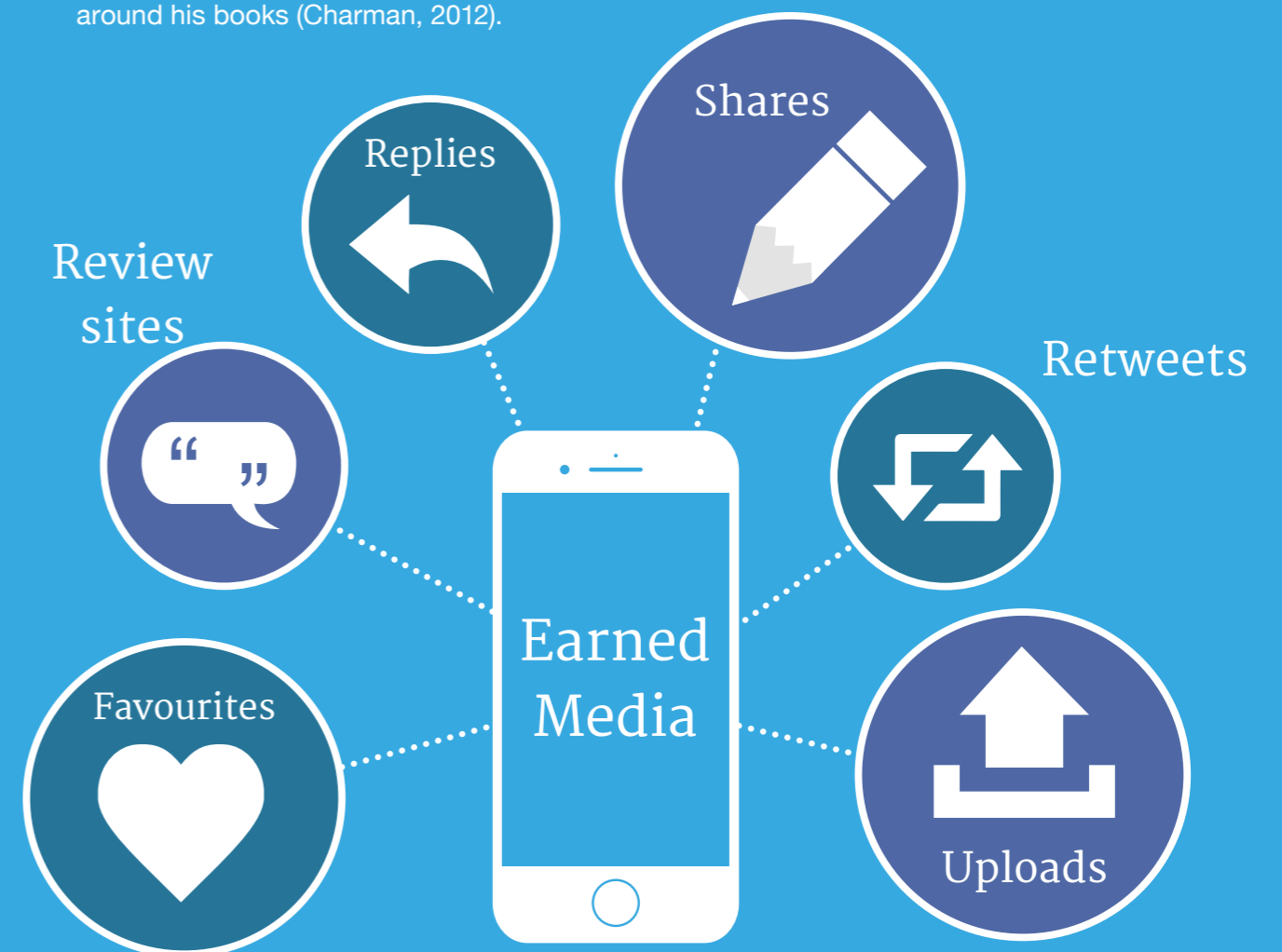
Whilst most traditional forms of marketing are amenable to being utilised via social, eWOM (e-Word of Mouth) and COBRA (Consumers Online Brand Related Activities) are more specifically associated with social.

Social works, in marketing campaigns and activities, in a variety of ways (please see points 1-6). These six are explored more fully below. Naturally there were a number of other uses that could have been added to these, for example making product and service delivery social. These may be explored in subsequent editions of the Guide. However as the focus for this Guide is more communications we have concentrated on this list at present.

### eWOM

Word of mouth (WOM) is as old as marketing and has long been considered the most powerful element in marketing, and the digital equivalent is eWOM. A simple example of eWOM might be a customer review section on a hotel website site, where the hotel seeks to create positive flows of information and advocacy from visitors to the site.

WOMMA (the Word of Mouth Marketing Association) draws a distinction between organic and amplified WOM. Organic refers to naturally occurring patterns of posts, links, shares etc. Amplified refers to marketing campaigns created to utilise the mechanics of WOM. One key feature of eWOM is that it can be positive or negative, in contrast to most other routes which are only occasionally negative.
McKinsey suggest dividing WOM into three categories (Bughin et al, 2010).

• Experiential, which they claim accounts for 50%-80% of WOM and results from people's direct experience with a brand or service. This usually refers to where the experience does not match expectations. If something is as good or as bad as expected then there is little to say. Something like lost airline baggage (or a broken Taylor guitar) is so powerful because it is not what is expected.
• Consequential. This describes what happens when people are exposed to marketing.
• Intentional. This describes paid WOM, for example, celebrity endorsements or paid influencer campaigns.

### COBRAs

COBRA stands for Consumers Online Brand Related Activities. A COBRA can be as simple as persuading customers to view a video, through to uploading a picture of their new shoes/hat/meal to Facebook or engaging in an ongoing activity such as Sharpie's series of social media campaigns to get people to create and upload doodles in a variety of forms to a variety of sites, using Sharpie pens.

COBRA can also be thought of as comprising three elements: consuming, contributing, and creating (Mutinga et al, 2011).

1. Consuming, for example, viewing videos, sending branded gifts, and playing branded games.

2. Contributing, for example, engaging in brand related conversation, including:
   • Rating or scoring products, e.g. TripAdvisor.
   • Engaging in branded conversation, for example, on Facebook page.
   • Adding comments to a brand site.

3. Creating, including uploading brand related videos, photos, drawings etc, writing brand related articles, creating product suggestions and ideas, for example, Lego or MyStarbucksIdea.

Another example was seen when Mattessons brand 'Fridge Raiders' partnered with an online games celebrity 'Syndicate Project'.

View the case study *here*

### 1. Telling deeper and richer stories

Social facilitates telling deeper and richer stories in a number of ways, including:
• Providing more space/time for a story to be told in its entirety, rather than fitting it to a conventional format.
• Allowing the story to expand over time, in response to reactions, questions, suggestions.
• Utilising co-creation, where the story is made deeper and richer through the contributions of the crowd.

Case study: Visit Iceland case
View >

### 2. Leveraging social behaviour

One of the strengths of social is that it can utilise social behaviour, for example, to spread a message or promote an activity.

Case study: New Zealand Bank ASB
View >

### 3. Enhancing/amplifying other activities

Generally, the most effective way to use social is in conjunction with other activities, in particular to amplify the impact of other channels.

Case study: Department of Tourism for the Philippines

View >

### 4. Monitoring and responding to discourses

The ability to listen to conversations taking place in social media creates a number of opportunities, including:
• Gathering a 360 degree picture of the brand, its campaigns, and its activities.
• Sourcing of information about the activities of other brands.
• Gathering inspiration and advice.
• Checking for problems, underperformance, and alerts.
• Monitoring ....discourses, "Making brands culturally relevant"

Case study: TfL (Transport for London)
View >

### 5. Allowing real-time management of campaigns

Historically, the measurement of campaigns and activities was considered separately to the implementation and management of the campaigns/activities. Typically, measurement was used to evaluate campaigns after they had finished. With social the measurement and evaluation can be integrated and the measurement process used to guide and influence the implementation.

Case study: Cadbury's Crème Egg
View >

### 6. Targeting customers

Social platforms can enable varied and better targeting of customers to increase efficiency. Examples of this include using geotargeting for mobile promotion, or demographic targeting.

Case Study : IKEA / Onken
View >

## 2.6

### The interaction of social with other media

In most cases, social is used in conjunction with other media, and in most cases its impact will be partly direct and partly through its interaction with other activities. This level of interaction needs to be designed into the campaign/activity at the outset and should be addressed by the evaluation, see Figure 1.

(Fig 1)

Display
Outdoor
Press
Print
Paid content
Paid search
Radio/TV

WOM natural search

Owned media

Earned media

Sales
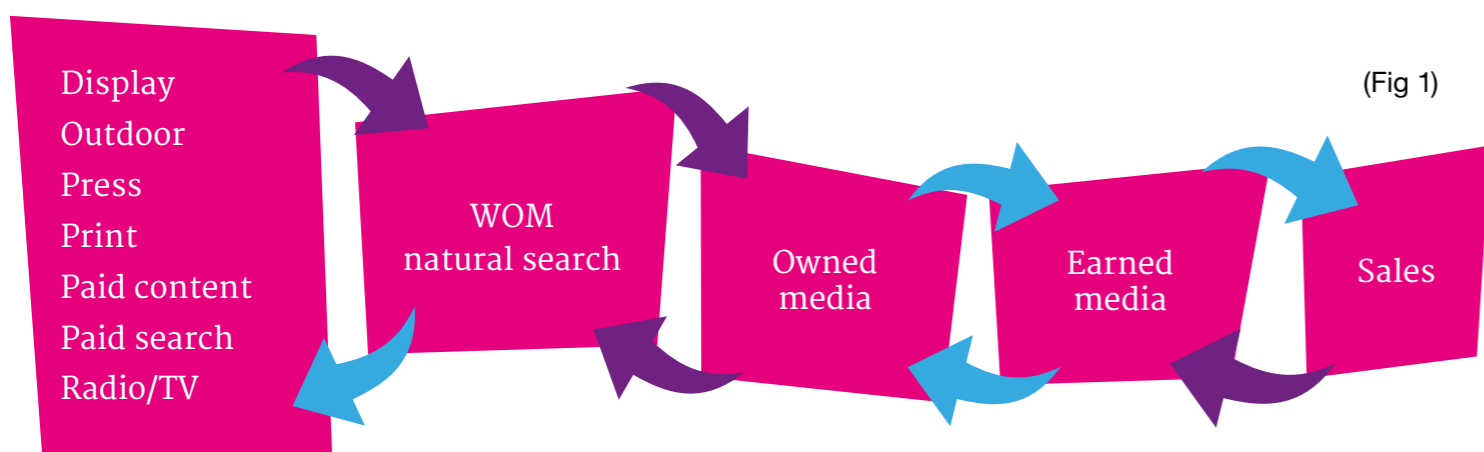
The measurement challenge is not only to evaluate the direct effectiveness of the social element – but also its role in amplifying other media.

## 2.7

### The importance of social data for measurement

Social data is a new and powerful source of insight to help evaluate all marketing activity, whether it is explicitly social or not. It provides new ways to understand not only 'what' happened but 'how' and even potentially 'why'. And its immediacy and availability can help fuel the shift toward real-time, adaptive marketing. The explosion of new evaluation data sources, metrics and techniques is principally in four areas:

**1.**

**Sentiment and text analytics.**

The ability to be able to use social as an indicator - or even predictor - of brand health. For example, using topic analysis to see whether a brand is more associated with a desired positioning post-campaign. Or even to evaluate specific pieces of content.

**2.**

**Context.**

Using social data to better understand what predictable influences drives consumer response and campaign ROI (eg timing, cultural factors etc).

**3.**

**Network.**

Using social networks to measure and understand how ideas spread and hence improve targeting and the role of influencers.

**4.**

**Relationship data.**

Using social ID's to link to other data sources and provide more robust, large scale and efficient testing solutions.

## The value of social as a data source for marcomms

The scope of this Guide is focused on the value of social as a communications medium rather than as a data source, but the latter will play an increasingly important role for brands going forward and should therefore be factored into their marketing and social media strategy.

Data is becoming central to the marketing function, driven by the rise of digital, the availability of customer-level data and the addressability of media. Brands are seeking to harness this customer-level data to better understand their audience and increase their relevance to the individual through personalised experiences.

Social networks can be a rich source of information on demographics, networks, interests/affinities and behaviour, across a broad spectrum of data shared by users from standardised data points such as gender through to more complex text analytics of commentary. It is this richness of persistent data at a customer level that will provide incremental value to existing customer profiles.

Use of customer data from social networks is still relatively nascent, however. Within marketing it is it is often used by brands at an aggregate level, on an ad hoc basis, or in silo from other marketing channels and the rest of the business rather than integrated into a single customer view.

**For example:**

• Mining publically available data at an aggregate level for social media research, e.g. mining data for brand mentions online, including applying sentiment and text analytics, to understand brand perception.

• Using publically available and privacy compliant data at an individual level, e.g. using influence scores or follower count data to prioritise engagement with or response to individuals.

• Accessing permissions based data through apps, social logins and connected accounts, often campaign-based and ad hoc and/or in silo, e.g. accessing Facebook Like data obtained through app registration or social login to personalise communications based on interests.
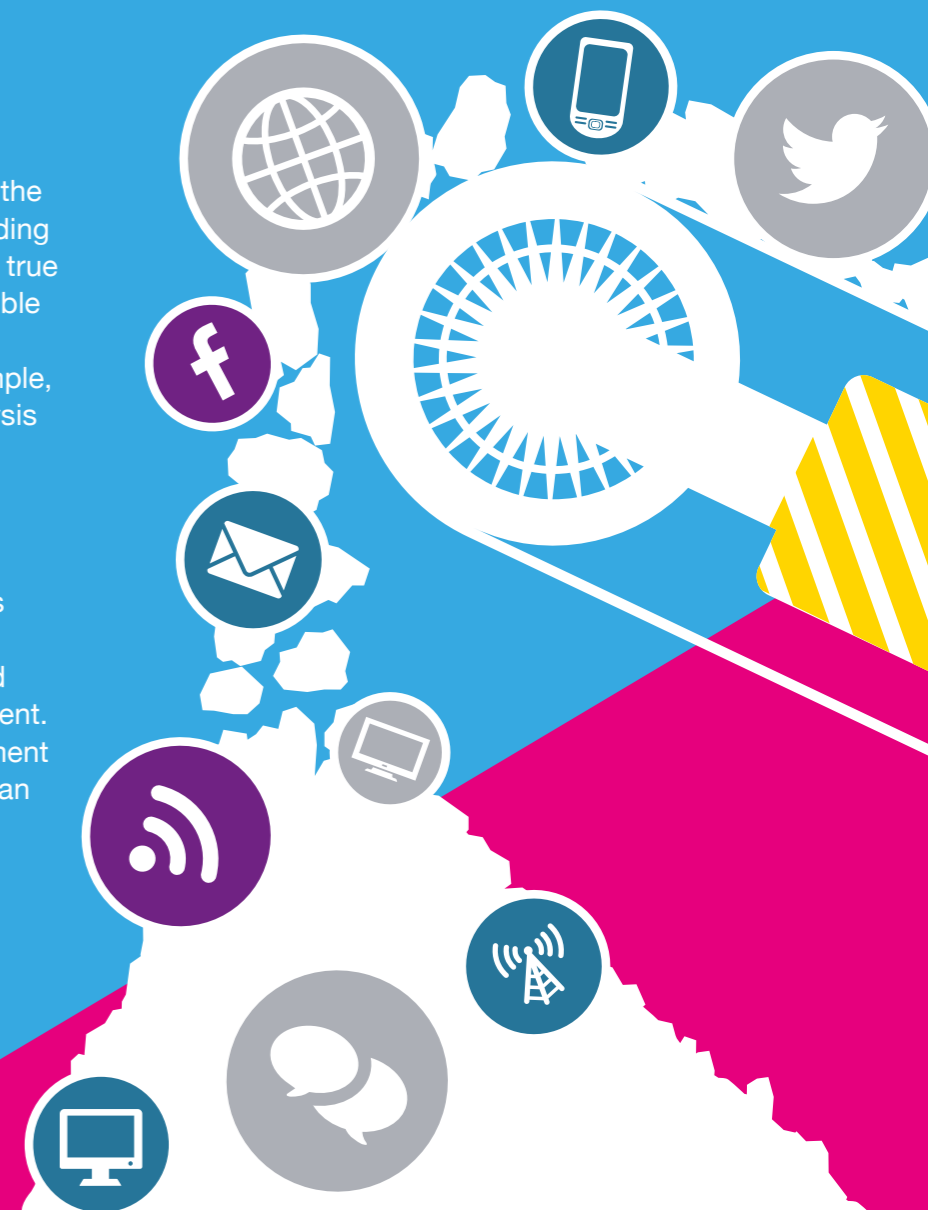
Harnessing social data at an individual customer level in a much more consistent, automated and intelligent way will provide brands with an additional lens to the existing customer view, enhancing their ability to understand and target customers. In the context of a single customer view this could be used to enrich experiences cross-channel, for example, using insight on interests or product affinity gleaned from social media to automate content personalisation on a website or in email marketing.

**There are currently several challenges which has slowed uptake, including:**

• **Data privacy concerns**
  Data privacy regulations relating to social media continue to evolve meaning that many brands are adopting a wait and see approach or are pursuing the most risk averse approach to future proof their set-up. Equally important as legal restrictions is the need for brands to develop their own position on data privacy – a position which fits with their own brand values and is acceptable to their customers. Core to customer acceptance will be developing and articulating a clear value exchange proposition, whereby brands request only those data points required and provide customers with a clear rationale for sharing the data points in question, for example, greater personalisation or more accurate recommendations.

• **Lack of clear objectives and the social data expertise to deliver against these**
  Brands will need a clear view of their objectives and the insight required by the business, matched with an understanding of what data points are available, their true incremental value in delivering actionable insight, and the ability to automate analysis and feedback loops (for example, understanding 'likes' versus text analysis and nuances of commentary).

• **Disparate data sources**
  Data is often held in disparate sources internally without a single view of the customer, limiting insight available and opportunities to personalise engagement. In order to use the social data to augment the existing customer profile, rather than hold the data in silo, these disparate sources need to be integrated with a unique identifier.

• **Technology**
  Current technology available to support the integration of social data can fall short of requirements (or expectations) and be costly to set up. This will continue to evolve and be led by first use cases and demand.

• **Organisational structure and processes**
  Fundamental across all points is the organisation's focus and set-up, including the ability to action the data and insight generated.

As the value of social as a source of insight increases and customer data becomes more central to social media strategy, this will need to be reflected in our definition and measurement of the business value of social media as discussed in this Guide.

# The Challenges of Measuring Social

As well as the challenges, it should also be noted that social presents many new and exciting opportunities. For example, the ability to measure the totality of a campaign's interactions, as opposed to those of a sample, and for that review to be based on accurately collected data, rather than unreliable recall. Similarly, social provides the chance to listen to real, unprompted conversations between consumers, allowing concepts such as sentiment to be measured.

This chapter covers the challenges of measuring social and what people are doing to address the challenges, in particular.

## How social is the same, and how it is different

The obvious measurability of social media makes it look quite different, in terms of targeting and evaluation, from other media, as does the degree of interest in earned media. Despite the differences however, there are also similarities that should be recognised and acknowledged.

### The similarities

Social has similarities to other channels and these similarities should be leveraged when planning the evaluation of a campaign.
• Paid advertising in social media is directly comparable with many other forms of advertising.
• The earned media element of social is sometimes comparable with PR campaigns,
• The direct response of social can be comparable to direct marketing.
• Measuring calls to action can be similar to couponing and some forms of POS.

As with all campaigns, the overall impact on the brand, including its perceived value and trustworthiness, need to be assessed as well as its effect on the bottom line.

### The differences

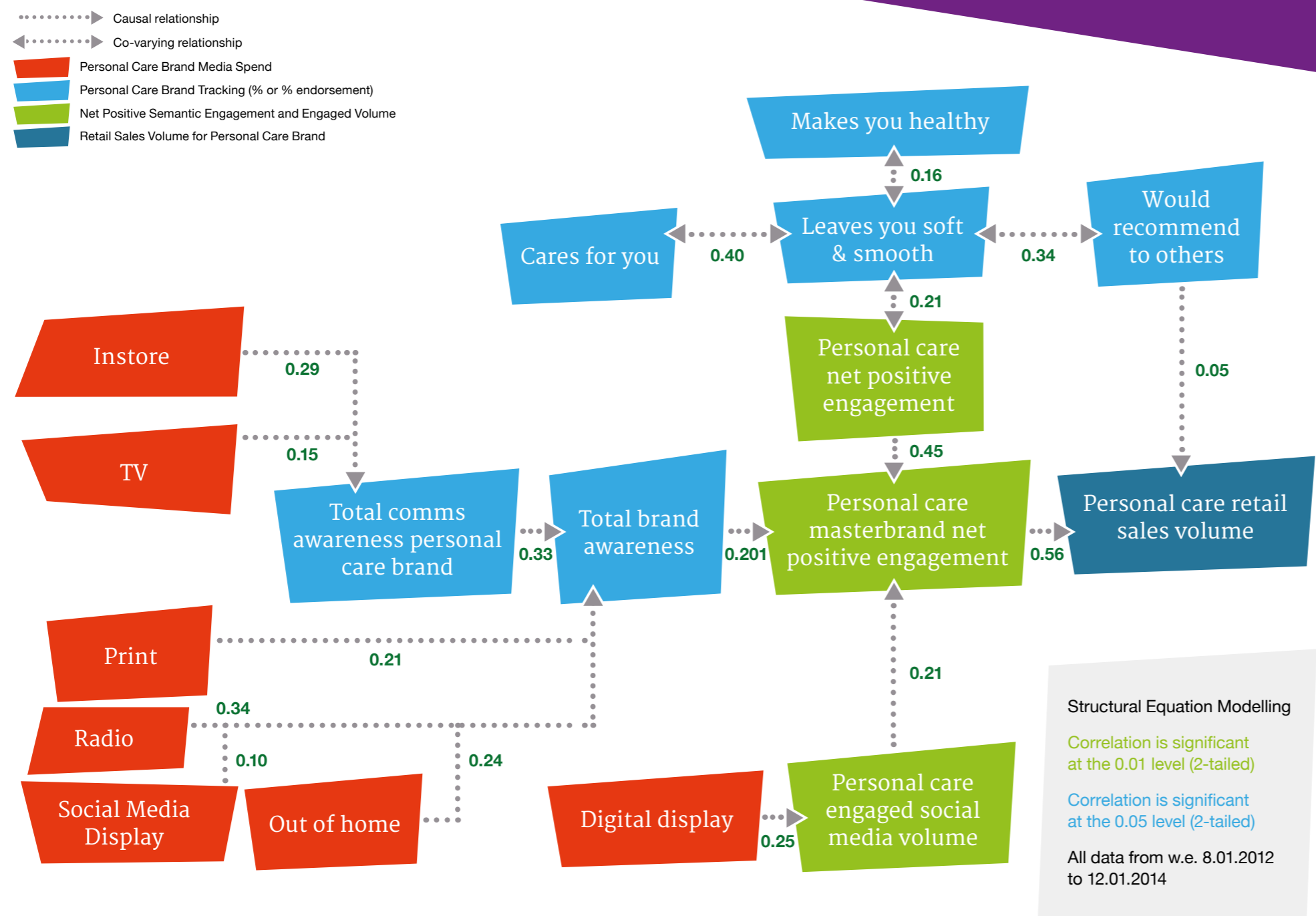The key differences in evaluating social campaigns relate to:

• The measurability of so many aspects of the campaign. Because social is such a measurable phenomenon, problems can arise from having too many choices of data rather than too few, and from having metrics which can be inconsistent over time and across platforms.

• As well as producing measurable metrics from the platforms, social produces vast amounts of naturally occurring posts and comments from consumers and citizens, creating the opportunity to use social media research techniques to explore, ideate, and measure.

• Even more than other media, social tends to be used in combination with other channels, acting directly and indirectly, and being impacted by the other channels. For example, to what extent does TV advertising impact Twitter, and to what extent does Twitter impact TV advertising?

• The need to think about where the social campaign/activity has occurred – is it in paid, owned or earned media? Exposure to earned media and cross-channel activities are often hard to control from a measurement point of view, which can make creating a control sample more difficult.

• The real-time nature of social feedback means that campaigns or activities can be modified whilst still running, which integrates measurement and metrics into the campaign management process. This integration of measurement and management makes issues such as benchmarking and the definition of pre -post much harder.

• The timeline for a campaign can be very different, for example, social can last much longer than most other campaigns or activities. Social activity can be 'always on' - especially when used in the context of customer feedback or satisfaction; or as a continuous promotion activity that is not specific to a 'campaign'. This removes the clear 'before', 'during', and 'after' that is a part of traditional measurement methods.

• Although the objectives of a social campaign/activity should always be linked to business objectives, they are not always linked directly to them. Social is often used to achieve tactical ends, such as collecting information, encouraging trial, etc. Evaluation needs to be tailored to the specific objectives of the activity.

• The budget for social activities/campaigns is often smaller than for, say, TV and the number of social campaigns/activities may be larger, which means the evaluation/ measurement approaches need to be appropriate to the circumstances and budget.

• One of the challenges for social is the lack of knowledge about the issues and the tools available. One study in France found that only 45% of relevant senior executives believed they had a good or very good knowledge of the tools available (Flores, 2013).

Figure 3 is a good example of how some of these complex interactions can be unpicked with careful analysis (see Appendix, Bottom Line Analytics).

# Consumer engagement as a lower funnel metric within the personal care insight ecosystem



········▷ Causal relationship

◁········▷ Co-varying relationship

■ Personal Care Brand Media Spend
■ Personal Care Brand Tracking (% or % endorsement)
■ Net Positive Semantic Engagement and Engaged Volume
■ Retail Sales Volume for Personal Care Brand

Makes you healthy

0.16

Cares for you — 0.40 — Leaves you soft & smooth — 0.34 — Would recommend to others

0.21

Personal care net positive engagement

0.05

Instore

0.29

TV

0.15

Total comms awareness personal care brand — 0.33 — Total brand awareness — 0.201 — Personal care masterbrand net positive engagement — 0.56 — Personal care retail sales volume

0.45

Print

0.21

Radio

0.34

Social Media Display — 0.10 — Out of home — 0.24

0.21

Digital display — 0.25 — Personal care engaged social media volume

**Structural Equation Modelling**

Correlation is significant at the 0.01 level (2-tailed)

Correlation is significant at the 0.05 level (2-tailed)

All data from w.e. 8.01.2012 to 12.01.2014

(Figure 3 Causal Map, Bottom Line Analytics)

The map demonstrates the multiple ways social can impact sales, both causally and as a proxy for brand impact, that in this instance, a paid social media display banner is being indirectly transmitted to sales through mediating brand awareness metrics. The language based consumer engagement, measured via social media commentary, is shown to be closer to the lower end of sales funnel.

In this study, the positively engaged commentary about the brand is broadly consistent with the brand's survey driven attribute scores. The cluster around a latent construct called Texture suggests that high agreement on intrinsic texture related attributes co-varies with positively engaged comment on social media (see Appendix for more details).

### Short and long-term effects

The immediate nature of social often focuses attention on short-term results and measurements. It is important to recognise that people want short-term measures, and that short-term measures are one of the key attractions of social. At the same time, it is important to ensure the long-term is also considered.

" One of the benefits of social is that it provides measurements that allow campaigns and activities to be optimised in real-time. The management of campaigns should balance long-term success with short-term success, since they tend to depend on different elements and strengths."

The key points about long and short-term effects are:

• Although there are no long-term business effects without short-term effects, the reverse is not true – there can be short-term effects without long-term effects. Long-term effects are not just the accumulation of short-term effects.

• To optimise effectiveness and efficiency, campaigns need a balance of short and long-term objectives.

The key point that Binet and Field make is that relying solely on short-term metrics could be a recipe for long-term disaster.
One of the benefits of social is that it provides measurements that allow campaigns and activities to be optimised in real-time. The management of campaigns should balance long-term success with short-term success, since they tend to depend on different elements and strengths. Adapting campaigns and activities solely to short-term metrics could sacrifice the activity's long-term prospects.

### Evaluating short-term effects

Utilising Binet and Field's, works the key messages for evaluation of short-term effects are:

• Activation effects are best measured with immediate behavioural metrics, such as direct response rates, click-throughs or immediate sales uplifts.
• These metrics point in the direction of rational advertising or price promotions, since these tend to produce the biggest short-term responses and volume growth can be quickly achieved.
• Focusing on short-term effects tends to under-perform in the longer term. Long-term success is best achieved by emotional brand building effects, which cannot be detected by measuring short-term responses.
• Short-term volume growth can initially be an efficient strategy, as share gain is achieved most efficiently for one year campaigns, but not necessarily an effective one i.e. one that drives profitability or other business metrics strongly.

### Evaluating long-term effects

Methods for measuring long-term effects are less robust and developed than those for short-term effects, but the key points are:

• The effects of emotional brand building are more subtle, but last longer and grow over months and years.
• These effects cannot be simply measured by short-term response metrics.
• Brand tracking is the established way to measure long-term brand building.
• Emotion-based and fame metrics may also help to rebalance the scorecard in favour of long-term effects.
• Other business metrics such as market share or profit that are usually measured over a period of a year are also associated with long-term effects.
• Share of voice is another, long-term metric, closely related to market share. While share of voice is harder to compute consistently in a Paid-Owned-Earned world, directionally comparing share of voice over time against market share is a good predictor of long-term growth.
• One of the most valuable business metrics is price elasticity (either in terms of being able to charge more for the same volume, or being able to sell more at the same price). The price premium is important because of its close links to profitability, and it is a metric that only tends to move over the long-term. Unfortunately price elasticity is a measurement that is complex and often overlooked. The creation of brand equity which is a key objective for most brands, would combine both a volume and a price premium.

In practice, the brand-building element of a campaign should be evaluated over a period of at least six months, whilst activation elements can be evaluated over a shorter term.

Influence has become a key topic in marketing, and in particular in terms of social. It is also a phenomenon that is being hotly debated. This section starts by presenting the case for influence that has been developed over the last 70 years, and why it can make measuring social hard. The section will then explore why influence is a hotly debated topic.

**1.**
Ideas and preferences spread though communities socially, from person to person, as opposed to individuals making independent decisions.

**2.**
Some people are more influential than others in as much as others tend to follow their practices and/or advice, these people are called influencers.

**3.**
Influence can be measured and this can be used to make influence marketing more effective.

**4.**
Marketing to influencers (also known as influencer marketing) is likely to be more effective (deliver better ROI) than marketing to everybody or marketing to, say, demographics.

**Influence makes the evaluation of social hard for three reasons:**

a. The desired effects are intended to occur amongst people who were not directly reached by the campaign/ activity.

b. There is considerable dispute about how influence works, and even about the extent to which it exists (with some people suggesting it is minimal).

c. The key issue with influence is incrementality, how many of the people who reacted were 'caused' by the influencer strategy, and how many would have bought/tried/played etc anyway.

## Influence theory

Influencers have become a hot topic in marketing over the last 10 to 15 years, particularly since the publishing of *The Tipping Point* in 2000 and *The Influentials* in 2003, although the concept of influencers can be traced much further back, for example, to Paul Lazarsfield's analysis of the 1940 US Presidential election. The appendix includes an outline of the history of interest in influencers and highlights key publications and milestones.

The key interest in influence relates to targeting (another topic which is hotly debated). Influence marketing seeks to utilise those who are most influential, and who will carry the message to the wider community. For example, a paid media campaign may be used to reach influencers, with the intention that the influencers create earned media which in turn reaches a wider group.

The subject of influence has a long history. In the 1950s Katz and Lazarsfield (1955) showed that in the 1940 US Presidential election, and in subsequent cases, influence flowed from the media to opinion leaders to the majority.

Stanley Milgram's famous six degree of separation study (1967) showed that networks of connections suggest that influence only needs to flow through a relatively small number of nodes to get from one place to another. Malcom Gladwell drew on this study to suggest that hubs were important to the flow of influence (2000). Gladwell suggested that Milgram found that about half of the experiment's key objects went through the hands of just three people.
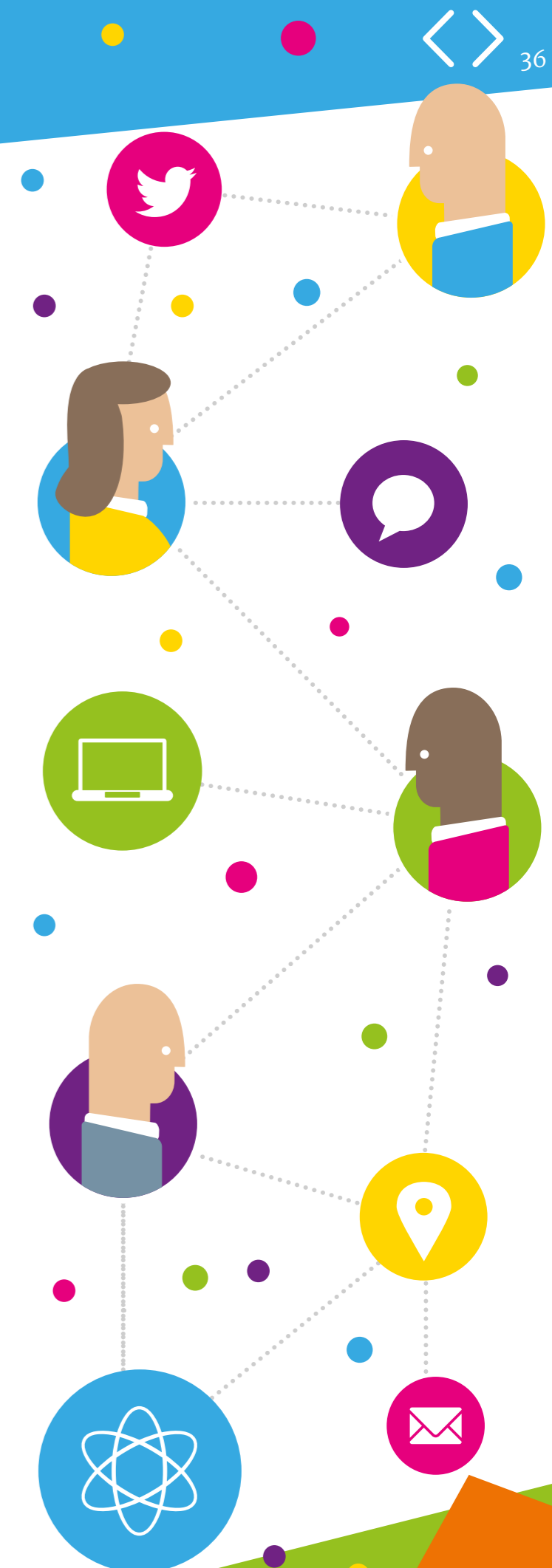
Robert Cialdini (1984) looked at the psychological mechanics of how things flow from one person to another and produced his six principles of influence: Reciprocity, Commitment, Social proof, Liking, Authority, and Scarcity. Several of these principles reappear in later writings by others, for example, Social Proof is close to the topic covered in *Herd* by Mark Earls, and Authority underpins the mavens that appear in Feick & Price, Gladwell, and Berry & Keller.

Feick and Price (1987) introduced the term Maven, or market maven, to the marketing literature, as somebody who seeks knowledge about products and services and who is seen as knowledgeable and is therefore influential. This study underpins many of the books and papers advocating influence, and is very widely cited.

Gladwell (2000), proposed that diffusion is based on three types of influencers: Connectors, Mavens, and Salesmen. He describes these as hubs in a network and says they are an essential part of the diffusion of influence.

Keller & Berry linked influence with marketing (2003), predicated on the assumption that a small group influence people, whilst the rest follow. Keller and Berry's definition of influentials is typical of the market place and is grounded in common sense: "The Influentials are active in their communities. They are highly engaged in the workplace and in their personal lives as well. They are interested in many subjects and are connected to many groups. They know how to express themselves and do so." This definition overlaps with Gladwell's three hubs.

Reichheld's Net Promoter Score (2003) is based on measuring people's propensity to say they would recommend a product or service. This measure taps into interest about word of mouth, advocacy, and influence. The NPS route does not assume that some people are more influential than others; it aggregates the NPS score across the relevant user/customer base.

## However, the concept of influence has been challenged by a number of people.

Watts (2004) brought Stanley Milgram's six degrees of separation experiment up to date, collecting over 50,000 email chains from over 160 countries, producing a more nuanced picture than Milgram's experiment. Watts' experiment suggested that six links between strangers was about right, which in turn might suggest it is a feature of networks themselves. Watts' findings led him to refute Gladwell's focus on influencers/hubs. Firstly, Watts' experiment did not produce hubs of any great size. Secondly, Watts showed that the results of this sort of experiment can be ascribed to network theory as opposed to individuals with influence.

Watts suggests the critical thing is whether people are ready to be persuaded, rather than whether somebody is persuasive. In Watts own words "If society is ready to embrace a trend, almost anyone can start one--and if it isn't, then almost no one can," (Thompson, 2008). This may well link to another explanation of shared behaviour, homophily, which suggests that people in networks connect with similar people, and their similarity is evidenced in the way they

do similar things. This copying element is picked up by Earls (2007). Earls pioneered, in the marketing world, the concept that the most important thing about people's behaviour (including what they buy) is that;

## "Our species is first and foremost a social one." This view of social leads Earls to conclude that "Word of mouth is the most powerful sales tool."

The point about influence being pull rather than push is made for forcefully by Bentley and Earls (2008), who argue that influence is not what is said or heard but more importantly what is observed. The paper asserts that the empirical evidence does not support the assumption that all (or even most) human social networks are hub-and-spoke shaped (as per the influentials hypothesis), nor are they fixed (as the network analogy tends to suggest). Bentley et al (2011) pull several strands together to suggest that people are massively influenced by what others do, but that the role of influencers is much less. They comment that celebrities have some impact "but their influence on each of us is much less than we imagine."

Keller and Fay (2012) remind readers that most WOM and influence is not yet digital, with face-to-face being the largest medium. Whilst advocating the power and utility of WOM and influence, Keller and Fay are clearly cautious about some areas, not only in highlighting the role of face-to-face, but also things like influence scoring systems such as Klout. They comment "the jury is still out on whether services like these actually provide a

true measure of influence." And "We expect that people ultimately will conclude online influence is one piece of the puzzle, but wholly insufficient on its own."

Sinan Aral is perhaps the most widely respected academic researching and publishing in the area of influence, homophily, and networks. Aral's work (2012) with both observation and controlled experiments has produced results that appear to show that influence is less common in social networks and less powerful than has been commonly supposed. For example, one study showed that not allowing for homophily led to a 700% overestimation of influence.

Aral highlights that one of the key problems with many earlier studies, and with common sense, is the lack of an examination of the counter-factual i.e. what would have happened anyway. Even when influential exist, they are likely to small in number compared to the network and hard to influence through marketing.

Research into the field of influence continues. Flores (2013) is rather dismissive of outdated PR models (e.g. number of likes, tweets, automated sentiment) and systems that assign a single value for somebody's influence (i.e. personal measures such as Klout and Kred). Finger and Dutta (2014) pull together many of the strands of the previous texts quoted here and cite a number of large scale studies. In general, they report some evidence of influence, typically from outside a network, but very limited evidence, and very little influence within networks.

## Players in Influencer marketing

Two interesting elements in the influencer marketing picture are companies providing access to influencer groups and companies providing influencer metrics.

### Influencer groups

Influencer groups are people who are supposed to have influence and who have signed up to some sort of programme to help promote products and services, typically in return for rewards. Examples include: Klout and its Perks programme, TapInfluence, and P&G's Tremor.

There are many forms of these communities. If these communities work, in the sense of producing a bigger net impact (after allowing for any extra costs) on target metrics (e.g. product trial or sales), it is likely to be a combination of two effects a) these people being influencers, b) these people being motivated (often extrinsically) to spread the message.

One of the key warnings from Sinan Aral's work is that influence marketing needs to consider the counterfactual, in particular how many of the recruited influencers would have bought the product anyway.

### Providers of influence metrics

Over the last few years, several measures of social influence have been developed and promoted, such as:

• Klout
• Kred
• mBlast
• Peer Index
• TweetLevel
• Appinions

Some people divide influence metrics into two groups, personal and contextual. Personal measures assign a score to somebody and that score is used for all contexts, for example, Kred. Contextual measures, such as Appinions, seek to measure influence within a topic, which means one individual can have different scores for different topics.

### Key debates

#### What is an influencer?

There are a wide range of definitions, but three common strands are:
1. Users/buyers of a product or service who tend to influence others, either by making recommendations or being seen to use/buy things.
2. Third parties, such as blogger, journalists, experts, analysts.
3. Members of influencer communities, people who have been signed-up by an organisation to act as the agents of influence marketing, for example, people in Klout's PERKS programme.

There are of course other definitions; the *WOMMA Influencer Handbook* describes five categories: advocates, ambassadors, citizens, professional/ occupational, and celebrity.

#### Is the notion of influence bogus?

As the review of the influence literature highlighted, there is a debate about the extent to which influence exists. Duncan Watts has challenged the very notion of influencers, asserting "A rare bunch of cool people just don't have that power. And when you test the way marketers say the world works, it falls apart. A rare bunch of cool people just don't have that power. And when you test the way marketers say the world works, it falls apart." (Thompson, 2008).

Others believe in influence, for example, PR firms such as Burson-Marsteller claim E-Fluentials can make or break a brand. According to MarketingVOX, an online marketing news journal, more than $1 billion is spent a year on word-of-mouth campaigns targeting Influentials, an amount growing at 36% a year, faster than any other part of marketing and advertising.

Sinan Aral's work suggests that influence exists, can be measured, but is much smaller than is commonly assumed, and often does not reflect a cost-efficient way of promoting a product or service.
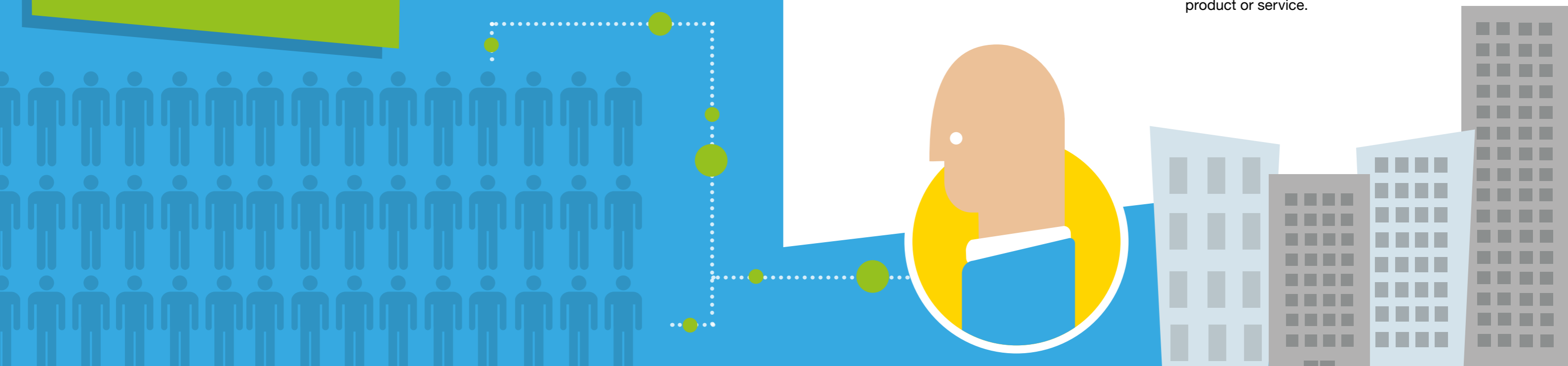
## Influence, influencers and influenced?

Is influence a push or a pull-phenomenon? Common sense tends to suggest that influence is about push, i.e. an influencer impacts other people, for example, by persuading them. However, much of the data reported in this section suggest it may be largely pull. Pull works, by people when susceptible, copying behaviour from people around them. For more on this point see Bentley et al (2011) and Earls (2012).

## Is influence largely "opportunity to see"?

If somebody is rated as influential because they have 100,000 people a month visiting their blog, or because 10,000 people typically see what they post on the Facebook page or profile, the key to their influence could be just the number of eyeballs they are able to deliver. This model would be very similar to traditional measures for conventional media.

As we can see from the Mattessons Fridge Raiders case referenced on page 21 the use of the online games celebrity 'The Syndicate Project' was used for the extent of his reach of 7m+ regular fans and followers and ability to target the right audience as much as his 'credibility' with the audience.

## What is the value of social influence metrics?

Brian Solis of Altimeter makes the point that companies who use social influence metrics, such as Klout, are looking at influence backwards. They are looking at the scores, rather than what makes the scores (2012). Also, Danah Boyd highlights the observer effect in metrics like Klout. People who buy into the system seek to improve their score, making the scoring system less valid, such that many of the high scoring people are players in the game, not necessarily the most influential. Whilst this concern is valid when looking at these metrics in comparing all customers, these metrics are probably very relevant when looking at influencer groups, where playing the game, and playing it well, are core parts of the process.

## Celebrity Vs everyday influencers

Most of the contemporary interest in influence focuses on the notion that there are influencers in a society, everyday influencers. However, there is another form of influencer, the celebrity influencer, for example, people on the Time 100 list or celebrity Tweeters like Kim Kardashian.

Whilst the majority of campaigns do not utilise celebrity influencers, they are very popular for some types of campaigns and activities, and the leading celebrities can command a high fee. For example, The Huffington Post reported that Kim Kardashian receives US $20K and her sister Khloe $13K per tweet (Kornowski, 2013).

## The utilisation of influence for marketing

There are essentially two ways of utilising influence marketing:

• Targeting marketing at people believed to be influential, with a variety of ways of identifying them.
• Working with organised groups of people identified as influencers, sometimes referred to as influence panels.

## Targeting influentials

Targeting influentials consists of identifying influencers and directing marketing at them. Here the term marketing is used in a very broad sense, for example, it might mean trying to get the influencers to be early adopters of a new product, and/or to be knowledge sharers, and/or simply to help promote awareness. Strategies for identifying influencers include the following:

• If members of a particular platform or group are felt to be influential, then media can be purchased to reach them.
• If people with certain characteristics are felt to be influential, for example, people with a certain number of followers, or people with a high influencer score, then usually a targeting system can be found to reach them. The characteristics might be influence scores (such as Klout) but they might also be derived from database scoring.

## Influencer groups/panels

Influencer groups and panels, such as members of the Klout Perks programme, are collections of people who are deemed to be influential and who have signed up to take part in advocacy and marketing programmes.

"More than $1 billion is spent a year on word-of-mouth campaigns targeting Influentials, an amount growing at 36% a year. However, the true effectiveness of this strategy is hotly debated".

## Disputes about the nature of influence

In terms of influence there are two very different schools of thought, those who consider influence to be central to how memes and fashions are transferred and there are those who consider influence to be a mirage or at best a marginal feature. Most practitioners do not fit neatly into either of the two extremes.

### The believers:

If you look at any new trend, fashion, idea you can trace it back and it will have started with a small group of people. In most cases, some of the early adopters will be connected to many more people than others, these tend to be people who appear to be influential in as much as the people they are connected to picked up the observed phenomena after they did.

Ed Keller makes the point that there are people who tend to be "in the center of the conversation." It's not that these people are the only ones who pick up on new product introductions, new ideas, etc. But, because they keep up with what's new, like to try new things, have a wider than average social network, and are often sought out for their advice and recommendations, they are people who have greater than average reach or impact through word of mouth.

The final part of the influential proposition is that this pattern repeats itself, i.e. the influencers for one trend or idea are also influential for other fashions and memes.

### The disbelievers:

The disbelievers have shown that many of the patterns observed in the data can be replicated without any assumptions of influence. This replication is normally achieved by either modelling actual data or creating artificial networks. These models tend to assume ideas are spread socially, but do not assume that some people have a special ability to 'influence' people.

This model assumes that we are all influenced by other people, but disputes that there is a class of person who can usefully be thought of as 'influencers'. For example, in a shoal of fish or flock of birds, each animal is sensitive to its immediate neighbour and will copy their movements, allowing the shoal or flock to move in unison. The role of 'influence' is solely due to position. At a different time, the fish/birds are in a different pattern, so the copying changes.

As a consequence, the disbelievers do not believe that it is better to target influencers than, say, a random set of people (or not substantially better, depending on the degree of disbelief).

### Pragmatic approaches

Professor Alex Bentley suggests a pragmatic approach:

"The definition of an 'influential' is often tautological - a person who picks up a new idea early and spreads it to many other people. The empirical question in all cases is whether these people fit any fixed categorical description, and whether influence ever travels the same path twice. For me it depends on:

- The longevity or tradition of the behaviour and the propensity for recognised experts to exist. Is this a Nobel Laureate in Economics giving a lecture or someone using a Samsung phone on the street?
- The degree of 'information overload' that may overwhelm people's ability to focus their social learning toward prestigious or expert individuals.
- The scale of the phenomenon - is this a small community or a population?

One example might be to consider the role of 'influence' in the models of traffic, flocking and pedestrian dynamics, which focus not on trying to trace the path of individual influence from one person-particle to the next, but on the architecture of the space, on adding a column in front of a doorway, the effect of merging, fluid flow, etc. This is the appropriate scale. Now, that said, you might have a situation where you can identify someone, like the person in your building during a fire drill, who wears the yellow vest and directs people out. So, I think the case-specific question is whether influentials exist in a given situation. If so, give them a yellow vest. If not, build a column in front of the doorway, so to speak."

## Homophily

Homophily is a counter proposition to that of influence, and indeed an alternative to much of the thinking behind targeting and causation. Social scientists have noted that similar people tend to cluster together, as in birds of a feather flock together. What may look like influence can in fact be lots of people responding in the same way to some external stimulus. This phenomenon is known as homophily.

### The limited influence model

The leading voice at the moment in the area of measuring influence and homophily is Sinan Aral, for example, his paper *Identifying Influential and Susceptible Members of Social Networks* published in Science, June 2012. The key conclusions of Aral's work are:

- Influence exists, is measurable, but is smaller than is generally believed.
- That models that do not take homophily into account tend to overestimate influence.
- That influencers and influence patterns for one product category may not hold for other categories.
- Seeding, giving products to 'influencers', should be restricted to a very small proportion of the universe (in one study the maximum suggested was just 0.2% of the population).
- That providing incentives to influencers tends to have a larger effect than seeding, and that combining them works best, but is still of limited effect.
- To assess the impact of using influencers it is necessary to assess what would have happened anyway. Many people would have adopted the behaviour under review without any influence being offered or assumed.

### Best practice advice when using influence marketing

- Decide on whether the campaign will target influencers or will work with one of the influencer groups to co-market/co-promote the campaign.
- Determine what you believe would happen if you selected a random group to market to, and benchmark against that.
- In terms of benchmarks, consider what the extra costs are (if any) of working with influencers.
- How you will determine success, in terms of gains over and above the non-targeted benchmark? How will the ROI be calculated?

As social media has grown, people have become more interested in how ideas propagate, through peer-to-peer networks and/or through mass media. The temptation is to see viral communication everywhere. However, clusters of similar people tend to congregate together and may look like they are influencing each other, but are in fact responding to the same external stimulus - seeing a 'Quit Smoking' ad on TV, for example. In some academic studies, over half of what looks like a cascade is actually 'homophily' (Aral et al 2009, Aral 2011). The implication is that TV and other mass media may be more important than at first sight (Watts, 2007).

The difference between influence and homophily is of particular importance when determining how to target marketing.

For example, we might notice that if one person in a social network buys something, other members of that network also buy it, then we might draw one of two inferences and design two very different marketing campaigns.

This is the difference, from a targeting point of view, between influence and homophily.

### Influencers

Some people are influencing others, so we should find and target potential influencers.

### Networks

We might assume that people with similar tastes would be connected to each other, in which case we should target people in that same network as they would have a greater propensity to purchase.

"Correlation does not imply causation" is one of the most frequently repeated statistical tropes, but whilst correlation is broadly understood, the nature, complexity, and importance of causality is less widely discussed and appreciated.

# "Digital and social media might be the most measurable media, but it can also be the easiest to misinterpret."

### Why causality matters?

Understanding 'why' something works is the key to repeatability, learning, and scaling. Most importantly, causality is a cornerstone of prediction (Gelman, 2010). This is just as true in the age of big data as it has always been, "a theory-free analysis of mere correlations is inevitably fragile" (Harford, 2014).

The key to replication is being able to tell stories, to describe how things work and to make predictions. The ability to link ideas causally is likely to be the dominant approach in marketing for many years. Machine learning approaches have been available for some time, for example, neural nets, but they are not widely used because they don't generate clear causal stories. Without causal stories it is much harder to create new material, new campaigns, and new activities.

In a world with ever more granular decisions, causality is even more important to understand, because of the need to provide advice on how to respond quickly. Historically, campaign evaluation was a 'justifying the budget' activity conducted after the campaign ended. Now, more decisions are 'live', and granular, facilitating intervention and real time management. Understanding causality is key to speed, particularly where there are complex explanations.

Causality and ROI can be very hard to prove conclusively, especially with the limited resources available to process social data and information (Lewis & Rao, 2013). The two key issues are often:

1. The need to make explicit assumptions.
2. Adopting a disciplined approach rather than adopting a fishing approach.

One approach to working with causality is to consider consumer decision journeys (Court et al, 2009). The decision journey recognises that the traditional funnel paradigm is too narrow and linear for the multiple touch points that have accompanied the explosion in product choice and digital channels. Because the decision journey is not a linear model the measurement of causality becomes much more complex. Feedback loops and compound effects need to be accommodated.

There is a school of thought that in the age of Big Data causality does not matter (Anderson, 2008). This view has been severely criticised (e.g. Nate Silver 2012) – and it has gained very little traction with most commentators.

### Why is causality hard to measure in social?

Digital and social media might be the most measurable media, but it can also be the easiest to misinterpret. While its interactivity – constantly asking people to make decisions and choices - creates many useful metrics. Interactivity can be the reason that it is difficult to understand what influences final decisions. Unpicking cause and effect in digital media is riddled with nuance.

One example of where measurability can struggle to identify causality is last click attribution, i.e. the faulty assumption that the credit for an action can be ascribed to the click that leads to an action. For example, Finger and Dutta (2014) quote the CEO of a media company who complained "When we publish an article on how good olive oil is for your overall health, our clients will go online and look for olive oils. They will most likely buy something we have recommended in our publication. However, the search engine will get the money, not us." Finger and Dutta assert that in cases like these "search engine advertisement has very little to no causal effect on the intent to buy."
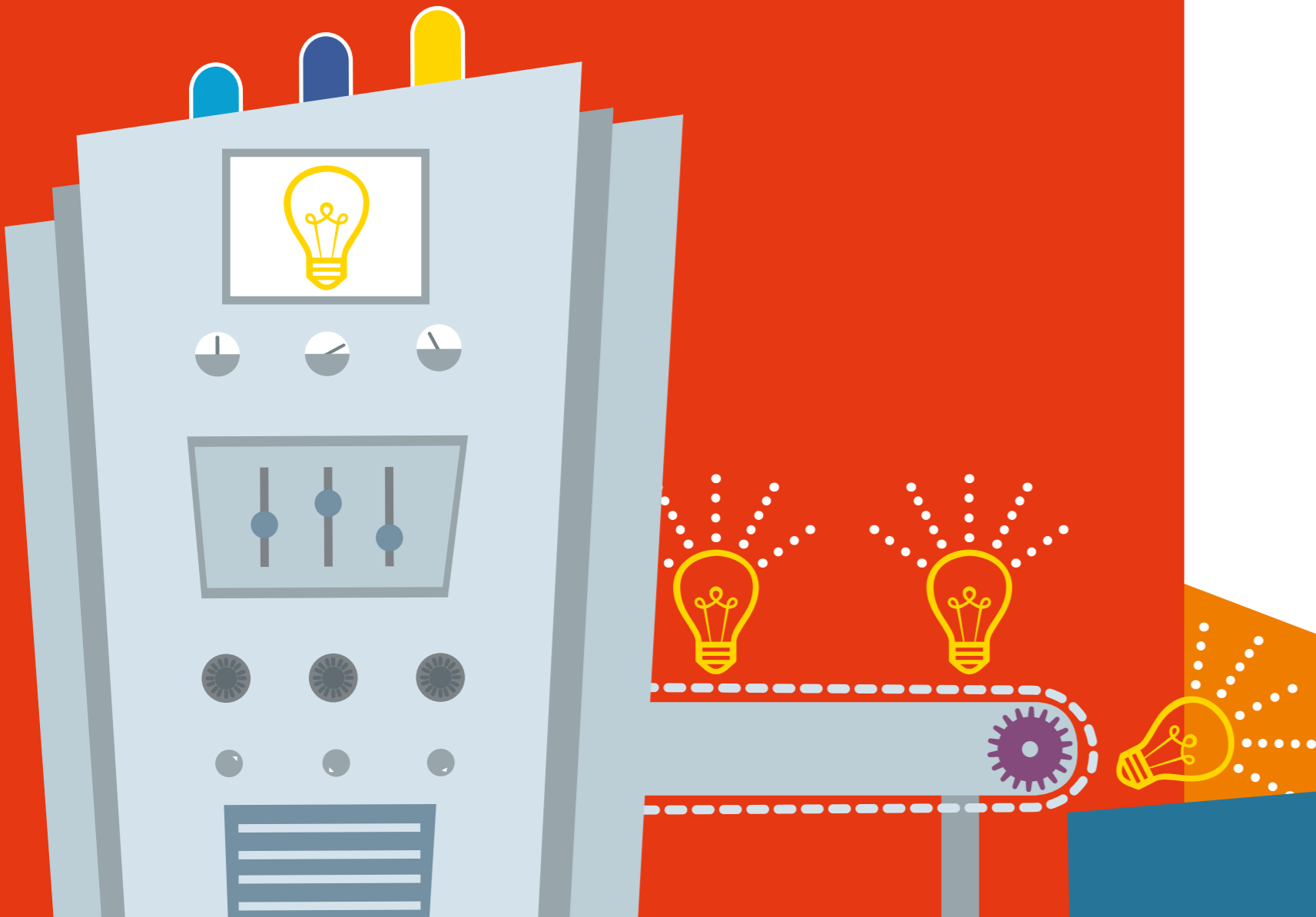
### Three reasons experiments are hard to construct in social

Establishing and measuring causality requires careful construction of counter-factual scenarios, i.e. what would have happened without the marketing intervention. The gold standard in causality is randomised controlled trials. This approach is good because it is clear that the only difference between the outcomes of different trials is the impact of the stimulus. However, clean experiments in social are hard to construct for three main reasons.

## 2.

### Synergy and interactions with other media

A/B tests try to isolate the impact of individual marketing actions. But in reality marketing often works synergistically. Marketers are interested in how social media and search or social media and TV work together rather than in isolation. For example, a Twitter interaction with a TV programme can provide re-targeting opportunities. It is important to make sure that the 'credit' is distributed to the right media channels and content so that budgets can follow accordingly. More generally, the relationship between media is seldom linear. Not only are there diminishing returns to scaling up investment, but there can be critical mass impacts (e.g. people respond better to social media campaigns when the brand achieves a 'threshold' level of salience and brand equity). Word of mouth marketing relationships are typically non-linear, include multiple variables (such as quality, exposure, and length) and involve feedback, e.g. liking something encourages buying, buying encourages liking.

## 1.

### Selection and targeting bias

Even with careful design, exposure to media and marketing is not random, especially when taking multiple channels and earned media into account (although paid media can be more controllable in social than in other media).

A deeper problem is that behaviour online and exposure to marketing is often commonly driven by intent. For example, the more somebody wants to buy a product the more likely they are to see related media about it. For example, if somebody intends to buy a Canon camera from Amazon, they are more likely to use Canon related search words, to see Canon display ads (particularly with re-targeting) and even to be a follower/fan of Canon Facebook. But none of this has influenced their decision to buy.

This 'selection bias' matters a lot. In a recent study, it was found that naïve measurement can be biased to the tune of about 1000%! (Farahat & Bailey, 2011). Variables such as clicks, reads, plays, and posts, should be treated as both inputs and outcomes. The people who respond 'virally' to campaigns are even more likely to be a biased sample. Best practice should include test and control measures wherever possible, even with econometric models or with matching techniques.
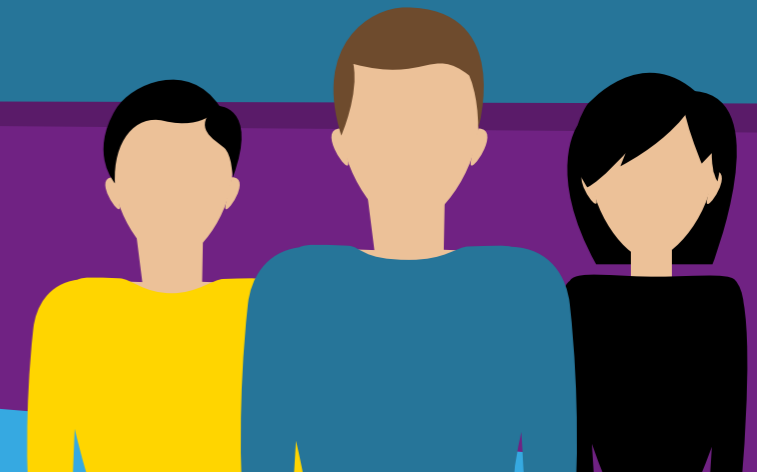
## 3.

**Similar Interests**

### Homophily

Homophily suggests that similar people do similar things and that this can look like cause and effect. In essence this comes back to the counter-factual issue, if the campaign or activity had not been run, how many people would have bought it anyway.

## Unpicking causality in social media – practical steps

Whilst there are few situations where there is a perfect solution to establishing and measuring causality, there are a number of practical steps that can be followed.

### Benchmarks and targets

The most common mistake in digital measurement is to simply quote the number of followers/fans, views, clicks or impressions. The metrics produced need to be compared with something to make them more tangible. Setting targets before a campaign is a good idea. Starting with a prediction is the first step to thinking about causality. Comparing the scores with benchmarks is also a good idea, using earlier campaigns or other current campaigns as the benchmarks. In many cases the platforms may be able to share norms.

### A/B tests

In A/B testing two cells are matched in every way bar one, and that one is the item being tested. For example, using two executions, or two targeting options, on a specific platform, such as Facebook, Twitter, or LinkedIn. If the two cells are matched, and if there is no contagion between the two cells, then it is reasonable to assume that any major differences are causally linked to the stimulus.

### Stating all relevant assumptions

In order to measure causality it is usually necessary to define the variables and the relationships that are being assumed. For example, what are the time periods, whether earned is being included, and what variables are being treated as endogenous.

### Modelling direct and indirect effects

Causality is rarely binary or linear. It is important to look for direct impacts, thresholds, and importantly interactions. Often, the best way to capture complex dynamics is through econometric modelling, a topic which is covered in the Evaluation Methods and Approaches section.

## Dynamic frames of reference

One of the challenges in measuring phenomena in social media is created by the fact that almost every feature of social is changing. Measurements tend to require consistency, to allow models to be created from observations and allowing predictions to be made and tested.

In the context of social, the number of users is changing, the competitive set is changing, the usage patterns are changing, and the ways that companies are utilising social is changing. One of the outcomes of this process is that most measurements in social do not have an absolute value, requiring comparative measurements to be utilised.

One of the key requirements created by the dynamic frames of reference is the need to create and update benchmarks. Benchmarks can be multiple activities for a specific brand or service or they can be measurements taken for several brands and services.

When a medium is changing quickly it is harder to create benchmarks, for example, as the number of people using Twitter grows and the number of tweets grows it is harder to benchmark one campaign against another. Similarly, click through rates dropped from the 1990s onwards, making comparisons difficult.

### Google Trends benchmarking

Google Trends reports the incidence of people searching for terms on Google over time. Since the number of people using Google is growing a method had to be found to standardise it. Google does this by allowing the user to specify the terms to be searched, the area (e.g. UK), and a period of time. Google then calculates for each period of time, for example, a week, the total number of searches and for the items searched their percentage of the total. Once this has been done for all of the time periods the largest percentage, for any one term in any one time slice, is set to 100 and all the other points scaled to fit.

This system has two major drawbacks:
1. The numbers only relate to that collection of terms, for that time period, for that region – there is no generality to the numbers.
2. If something else is happening in a particular month, the numbers will appear to fall, even if the absolute numbers have not fallen. For example, many terms appear to dip in the run up to Christmas, which may not reflect a dip in the number of searches for that word, but could reflect an increase in searches for Christmas related items.

### Google Flu Trends

One of the most famous uses of social media monitoring, since 2008, has been Google Flu Trends. Google Flu Trends measures the incidence of people searching for words like cough and cold. This process has been generalised by Google with tools like Google Correlate and Google Insights.

A recent article in New Scientist showed that despite Google's early success in predicting flu, its accuracy has declined over time (Hodson, 2014). The article makes the point that the changing nature and variability of social media has reduced the ability of Google to predict the likely spread of flu in the USA.

> "One of the most famous uses of social media monitoring, since 2008, has been Google Flu Trends [which] measures the incidence of people searching for words like cough and cold."

The traditional way of measuring campaigns has been to focus on ROI, and ROI remains important in the context of social. However, there are other outcomes that a social campaign might be seeking to address and these need to be identified too.

## Non-ROI goals

A social activity might have a goal such as increasing the likelihood of applying for a job. In a perfect world this goal will have some link with the final objectives of the business, but the link between the specific social activity and the final ROI may well be too long to be measured in simple ROI terms. In these cases intermediate goals are used.

For example, the assumption might be made that attracting the right prospective employees is good for the long-term profitability of the business. So, intermediate goals that measure changes in the number of good applicants are set.

Social campaigns are then evaluated against the intermediate goals.

## Efficiency versus Effectiveness

One of the key issues surrounding ROI is the difference between efficiency and effectiveness. Overemphasising either efficiency or effectiveness can generate sub-optimal results.

Efficiency focuses on the ratio of the return to the investment, which is one of the standard definitions of ROI. The key problem with overemphasising efficiency is that it tends to favour options with lower levels of spend, which can result in the most effective campaigns not being run (because they may be less efficient).

The effectiveness of a campaign refers to the extent to which the objectives are met or maximised. In terms of ROI, effectiveness tends to refer to the total return, as opposed to the ratio of the return to the investment.

## ROI – what can be measured?

The main challenge in calculating ROI is the 'R', i.e. the return. To calculate the ROI a Return has to be defined and it has to be possible to link it to the investment. A social media campaign might intend to improve engagement with a Facebook page in order to get other people to engage with the brand which will make people more aware of the brand and better disposed towards it, leading to more sales. The increase in sales is a 'Return', but there are often too many external factors to clearly link the return to just the social campaign. In terms of social, one of the factors that makes the calculation of ROI more complex is that social is usually employed as part of an integrated campaign/activity, so typically we want to understand the ROI of something like Twitter plus TV, not just Twitter in isolation.

> "There are often too many external factors to clearly link the return to just the social campaign."

Several brands have found success by creating intermediate goals. For example, if a brand can show that increases in customer satisfaction or positive sentiment are associated with increases in sales, then these measures can be taken as the 'Return' that the social activity is measured against.

As well as increased revenue, 'Return' can be accrued by reducing costs. For example, if customer services, or branding, or the collection of insight can be managed at a lower cost through the use of social, there will be a return, against which the costs can be assessed.

The I in ROI, the investment, can also present some difficulties in the context of social. Thinking about the POEM model, the cost of paid media is relatively straightforward, but the cost of owned and earned can easily be underestimated, which in turn can result in the ROI being over-estimated.

The O2 case study uses both intermediate goals (they have calculated the value of one percentage point of customer satisfaction) and cost reduction (using social to reduce the costs of customer support). It is likely that the future of ROI calculations will feature intermediate goals and cost reduction as key elements.

Social can be a great channel for customer service delivery. There is an expectation now for many brands that they will respond quickly, especially in crises, and this level of response is rewarded with loyalty and higher customer satisfaction scores.

**SOCIAL MEDIA**
RESPONSE TEAM

O2 is a leading digital communications company in the UK, with over 23 million customers. The brand realised that social media data was a valuable source of consumer insight, and that social channels would be key to cost effective customer service.

O2 partnered with social intelligence agency, Face, to develop its own platform which it called RTO2 (Real-Time O2). RTO2 was launched in 2010 across the O2 UK business with over 400 users.

## RTO2 enables O2 to monitor and instantly react to messages directed at the brand on the social media channels where it has a presence.

The platform collects the messages and augments them with 'meta data' to add contextual information which is extremely useful in customer service decision making. For example, the system collects: sentiment, customer influence level, customer location, customer history, service area & type of issue.

Using social data in this way enabled the newly formed Social Media Response team to prioritise messages by topic, negativity or influence level and ensure their replies takes into account all previous social interactions with that customer. RTO2 provides a real time reporting dashboard, with the following metrics: Response time, % Issues resolved, Team member activity, Incoming and outgoing messages, Messages by channel, and Messages by service area.

In July 2012 O2 suffered its biggest marketing and communications crisis ever, its network went down for over 10 million customers for a two day period. Naturally, customers took to social media to share their concerns, frustrations, and anger. Over 30 team members were assigned to respond to the surge in mentions. The ability to react to every mention as it appeared online significantly impacted sentiment surrounding the issue and overall brand perception. Because O2 could manage in-house much of the servicing of customers during this difficult period, using its social media systems rather than outsourcing the problem, its analytics system meant it could determine the amount of money saved per customer.

O2 uses its tool to analyse the reach of conversations arising from Marketing and PR campaigns, the brand's competitive position in the social media landscape, and to understand the role of influencers, particularly when targeting specific content areas.

Influence is measured in three main ways:

1) Who is talking about O2 the most? (Volume)

2) Whose mentions reach the biggest audience? (Visibility)

3) Whose mentions drive the biggest reactions from other people? (Engagement/Influence)

Influencers are also used for promotional offers. For example, a retweet competition was held for new handsets. O2 has calculated how much each point of customer satisfaction is worth in net profit per customer. Using this tool, it has started to establish how much social media interaction is worth: customers who have interacted with the organisation via social media deliver a customer satisfaction score of 73, whereas those who do not interact on social media deliver a score of 69.

Access the full case study from the IPA website.

RTO₂
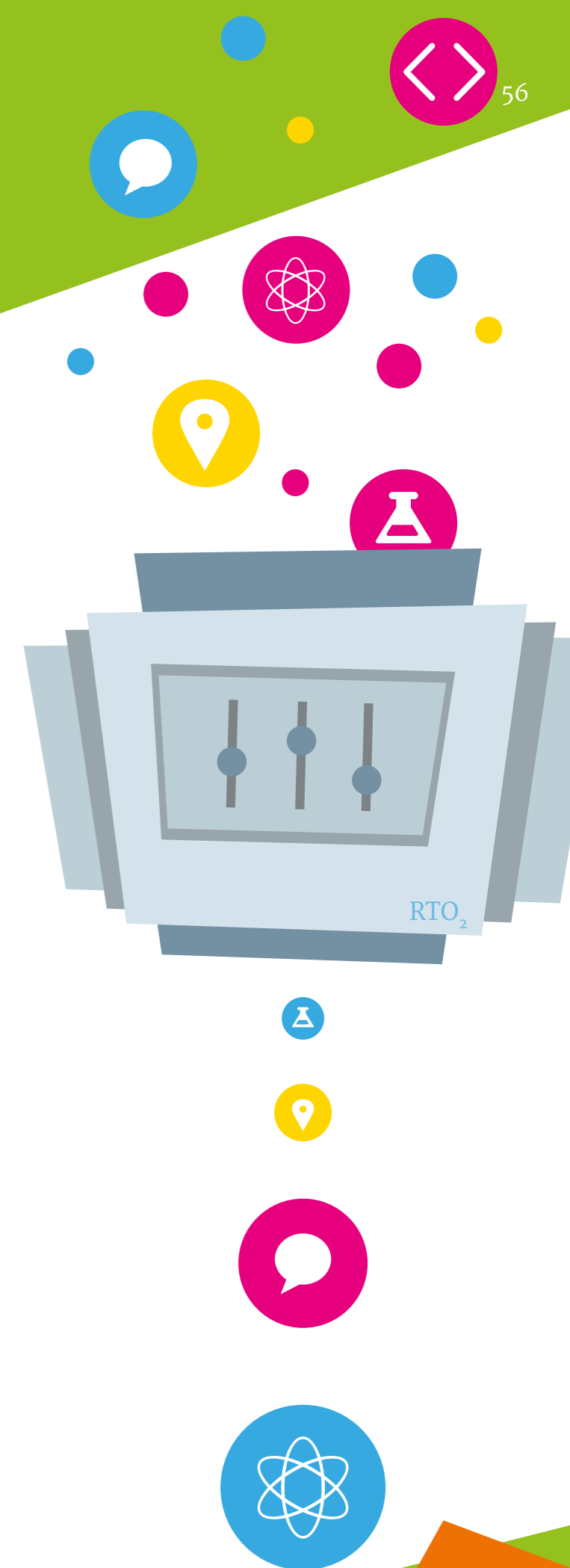
## KPI setting – how to avoid common mistakes

There are a number of errors to be alert to when evaluating social, including:

### Perverse incentives

A perverse incentive is created when pursuing a target results in an undesirable outcome. For example, pursuing more visitors can result in more inappropriate visitors and potentially fewer relevant ones, even in absolute terms.

When dealing with social many potentially interesting metrics can generate perverse incentives. For example, if a brand creates a truly engaging campaign or activity it will often acquire more followers/fans – and the growth in followers/fans is an indication of its popularity. However, if followers/fans are set as the objective, e.g. as a KPI, then brands and organisations may start to look for easy ways to boost the number of followers/fans without any regard to whether or not they help the brand at a deeper level.

As an example of a perverse incentive, Finger & Dutta (2014) cite a client who made CPO (cost per order) their key metric. Initially, this seemed to work, the CPO which was already low fell even further. The problem was that they found their sales were falling too. The campaign was being targeted mostly at existing customers, who were likely to click, which helped achieve lower CPOs. The unintended consequence of prioritising CPO was to reduce the size of the pool the campaign targeted to an adverse level.

### Metrics that are easily gamed.

In this context, gamed means that a high score can be achieved without the underlying phenomenon of interest increasing. For example, if followers/fans/plays can be acquired easily/cheaply from a third-party supplier, the metric is easily gamed. There is clearly a link between perverse incentives and metrics that are gamed. The difference relates to focus, with perverse incentives the focus is on how it will impact the way the business is run, with easily gamed metrics the focus is on the integrity of the measurement.

### Vanity metrics

Vanity metrics is a term that is used for metrics that can make a brand or campaign look good, but which may have little impact on the performance of the brand (Ries, 2009). The number of visitors, followers/fans, or downloads can be considered vanity metrics if they are not targeted at a specific purpose and linked to key outcomes. In general, it is not the metric itself which makes it a vanity metric, it the way a metric is selected and used that creates a vanity metric.

One problem with measures like plays, followers/fans, or likes is that they may not reflect current levels of interest or engagement. Few people unfollow, unlike, or remove comments, which means that the historic number of connections does not necessarily relate to what is happening at the moment.

### Fake followers/fans

There is growing business in creating false followers/fans, typically utilising low-cost economies and/or bots. These make measurement problematic. Measures such as influence and advocacy can be thrown badly off track if fake followers/fans are included. In November 2013 StatusPeople.com published an investigation of Katy Perry, who has nearly 50 million followers, and estimated (they use a sampling method) that a majority of her followers were either inactive or fake. It should be noted that the picture for one celebrity does not necessarily give a representative picture and other studies have produced lower estimates, with Business Insider quoting figures for fake followers of about 5% from both Facebook and Twitter (D'Onfro, 2013).

The providers of metrics and in particular the platforms themselves are aware of the problem and are constantly updating their systems to try to eradicate the bogus elements.
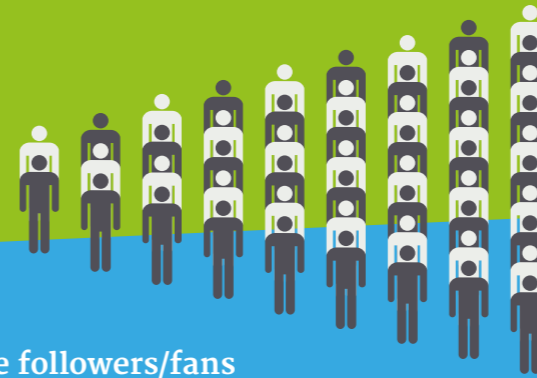
### The wrong number of KPIs

If too many KPIs are selected the real meaning can become obscured. A large number of KPIs will usually mean that there is too much correlation, which can lead to double-counting. Also, with too many KPIs something in the mix is likely to look good, leading to people highlighting it. Having too many KPIs is more like trawling than measuring - and suggests the organisation does not understand the real drivers of the business.

Too few KPIs is likely to result in something being missed or produce a finding which is so narrow that it struggles to explain the wider picture.

### Benchmarking problems

Few activities in social media have an absolute value. It is hard to say what 1000 views, 10,000 Tweets, or 20,000 shares means, unless the numbers are benchmarked against other similar brands, campaigns, and activities. The benchmarking process should be started at the campaign planning stage, by identifying relevant comparators.

Failure to set a target for the key metrics is a common failing.

# Strategies for Measuring Social

This section consolidates the advice and input received from industry leaders in terms of what is currently seen as the best way to measure social campaigns and activities.

Evaluation methods and approach, including the use of market research to evaluate social and the use of social media research, is the subject of the next chapter.

## 4.1
### Designing an evaluation strategy and process

Measurement does not just happen; it requires a framework, it requires planning, and it requires the selection of the right metrics. The measurement of traditional media has been developed over decades and has been subject to a high level of testing. By contrast, social is relatively new and highly dynamic, which means that there are few established rules and practices.

Because social is less well developed, there is a greater need to design a measurement strategy, rather than adopting an existing approach or plan. Most social campaigns and activities operate in conjunction with other media and this factor needs to be an integral element of any evaluation.

## 4.2
### Learning from established practices

Social may be new, but the planning process is not and the best way to utilise the new opportunities presented by social is to ground them in what is already known about campaigns and activities. In this context it is useful to think about the planning cycle and the role of short-term and long-term effects.

Steven King's planning cycle, can be very helpful in asking the key questions that marketers need to think about - not only in terms of their marketing communications activity, but also about the position of the brand as a whole.

In the context of social it is important to be clear about the objectives and the role that social will play. In a perfect world, social will be evaluated with the same tools and approaches as other elements of campaigns, although there are sometimes pragmatic reasons for recognising differences.

One key issue for the evaluation of marketing is that the long-term effects are not simply the aggregation of short-term effects, and focusing on short-term benefits, such as numbers of clicks, can make it harder to achieve the long-term benefits. Short-term benefits are often specific actions, whereas the long-term effects are more likely to be related to being able to charge more and/or sell more. This presents a real challenge for social where there is a substantial focus on the short-term and an abundance of short-term measurements.

(Fig 2)

Where are we?

Why are we there?

Are we getting there?

Where could we be?

How could we get there?

"In the context of social it is important to be clear about the objectives and the role that social will play."

## Establishing expectations and a framework

As with the measurement of any other form of media, measurement in the context of social media requires a framework and it requires an understanding of what can and can't be done. Although some elements of social media are different, the fundamental rules still apply. The practices for social media should build on what has already been learned about how media works.

The expectations and framework should define the goals of the campaign/activity and check they are aligned to the business strategy/objectives. Because social is seen as undeveloped and evolving campaigns are often undertaken because there is a perceived need to 'be doing something'. Even in these cases targets and predictions should be created, to allow the campaign or activity to be adequately assessed.

Benchmarking and targets are key to the evaluation of social media campaigns and activities. Few social media metrics have any independent, absolute meaning.

Fifty thousand 'likes' or tweets is hard to interpret, but when compared to two thousand or two million 'likes' or tweets there is, at least, some context. As the IAB have said, "Without comparative benchmarking most metrics are completely meaningless."

The use of KPIs requires a target, which should be set in advance. Benchmarks are common method of setting targets, but other methods can also be used, for example, forecasts based on market mix modelling. The key feature is to set realistic expectations and performance measures.

## Planning ahead

As with any measurement of campaigns or marketing activities, the campaign objectives and the items to be measured must be set before the campaign starts. Specifically, the metrics need to be identified and the methods of collecting the data should be agreed.

With social media there can be a temptation, since the data for earlier time periods often exist, to assume it is OK to wait until the campaign is underway or even finished before looking backwards to find evidence of its impact. This is flawed and will often lead to the wrong interpretation. The presence of external/unrelated factors mean there will usually be something in post-reviewed data that looks positive, and trawling for information will generate spurious findings and correlations.

Try to plan the media to facilitate measurement, for example, by not scheduling everything to happen at the same time. The ideal is to create experimental cells, where different people will be exposed to different executions and combinations.
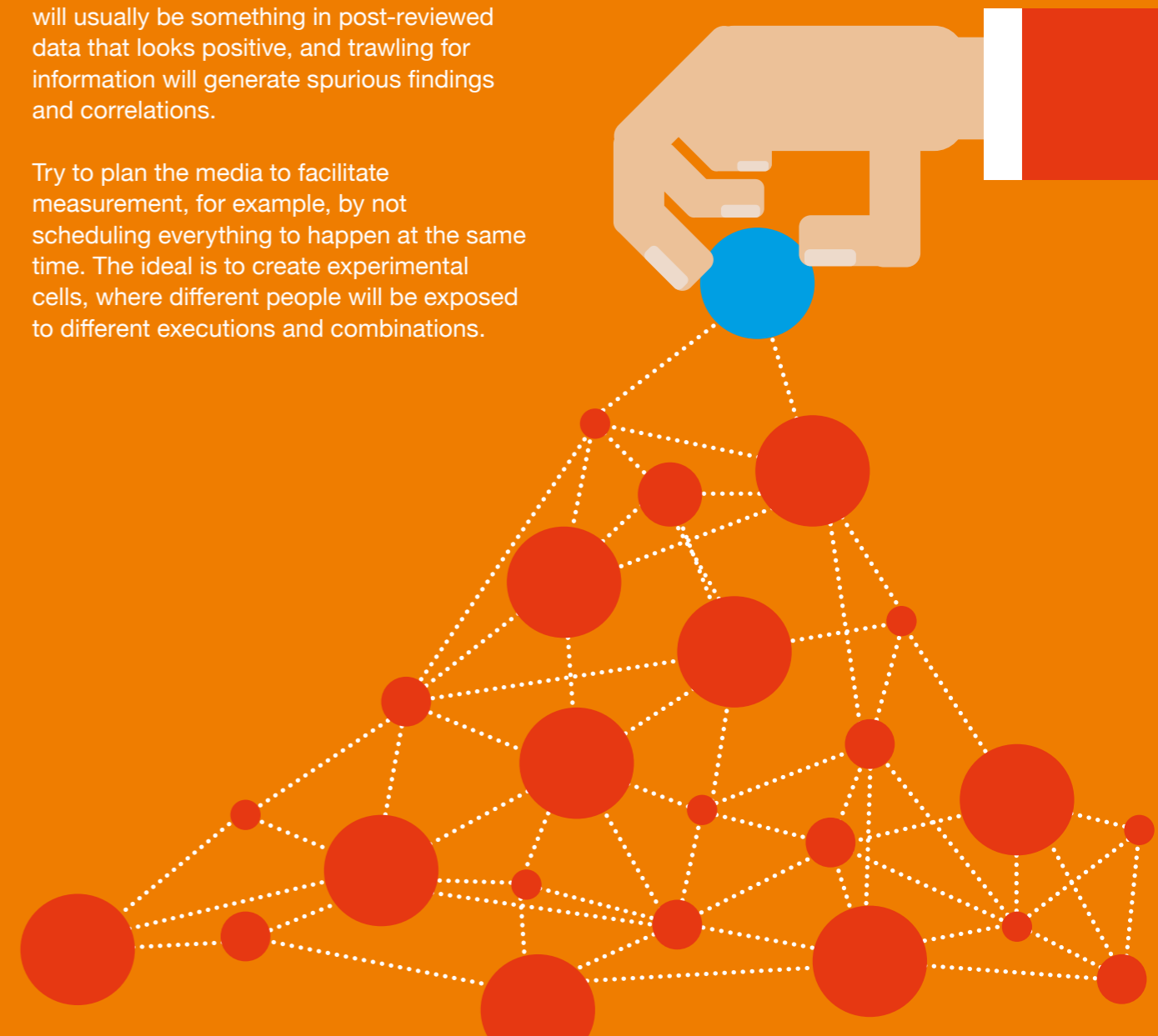
## Being SMART

The evaluation process should embody a SMART approach, i.e. Specific, Measurable, Achievable, Relevant, and Time bound. The evaluation should not confuse objectives with metrics or KPIs. The objectives are what the campaign or activity is aiming to achieve.

### The metrics

The metrics/KPIs are measurements that should indicate the extent to which the objectives are being met. It is important to recognise that there is no single best metric and that the best metrics for the last project may not be the best for the next project.

Select measurements that are quantifiable, facilitate action, and linked to business outcomes. One key element in making the metrics useful (for both action and evaluation) is to limit the number of KPIs, ensuring they are easy to understand and efficient to collect.

Set aside time, budget (perhaps 1% to 2% of the campaign spend, or 5% for smaller projects), and resources for analysis. There is a perception that the data from social is free. Whilst the acquisition of data is sometimes free (and sometimes not), the proper analysis of the data takes time and money.

### Investigate what analysis options are available from the platforms and providers.

Look beyond just pre-post. Look for relationships over time, using data from before, during, and after the campaign/activity. Think about other factors, such as: seasonality, homophily, other brands, other campaigns, ongoing activities.

## Efficiency

Efficiency is a ratio approach, typically with the desired outcome (e.g. sales, leads, registrations) as the key number, and the cost (e.g. money, resources or time) as the base. Efficiency is, in effect, the same as one of the measures of (Revenue) ROI.

## Effectiveness

If you obtain £4m of value for a spend of £1m, then the efficiency/ROI is four million over one million, or four. Effectiveness is a measure of how much the activity delivered.

In the example above, the effectiveness would be £4m (or alternatively, £4m minus £1m, i.e. £3m if the organisation includes costs in their calculations). A good campaign is both efficient and effective.

## Integration with other elements of the campaign

Whilst some social campaigns or activities are 'pure play' (i.e. they are only social) many are part of a broader picture, utilising a number of channels. Where social is being integrated into a broader campaign the evaluation needs to operate at three levels:

| THE IMPACT OF | THE TOTAL IMPACT | CONTRIBUTION OF SOCIAL |
|---|---|---|
| of the social media. | on the overall campaign. | to the total picture. |

The third element in this process usually employs an attribution model or market mix modelling - which is covered more fully in the appendix.

The metrics for social have two roles, the first is to help optimise the campaign, and the second is focused on evaluation and measuring. Some metrics will be used for both roles, others will be chosen for just one role.

## Continuous measurement versus ad hoc campaigns

As with other marketing activities it is possible to conduct continuous measurement and/or measurement targeted at specific social campaigns/activities. The key issues for these two types of measurement are:

### Continuous metrics

Continuous metrics should seek to measure expected changes and at the same time have the capacity to pick up unexpected changes and effects. One example of a metric that can help identify changes that were not expected is sentiment.

### Ad hoc campaigns

Ad hoc campaigns typically require that the key outcomes are identified in advance. Once the key outcomes are identified, baselines can be created. In some cases the objectives might need to be tweaked to ensure that the results are measurable.

One similarity between continuous and ad hoc campaigns is that the real-time and granular nature of social media measurement means that the management of campaigns can often be tweaked or amended in light of social media feedback.

# 4.9
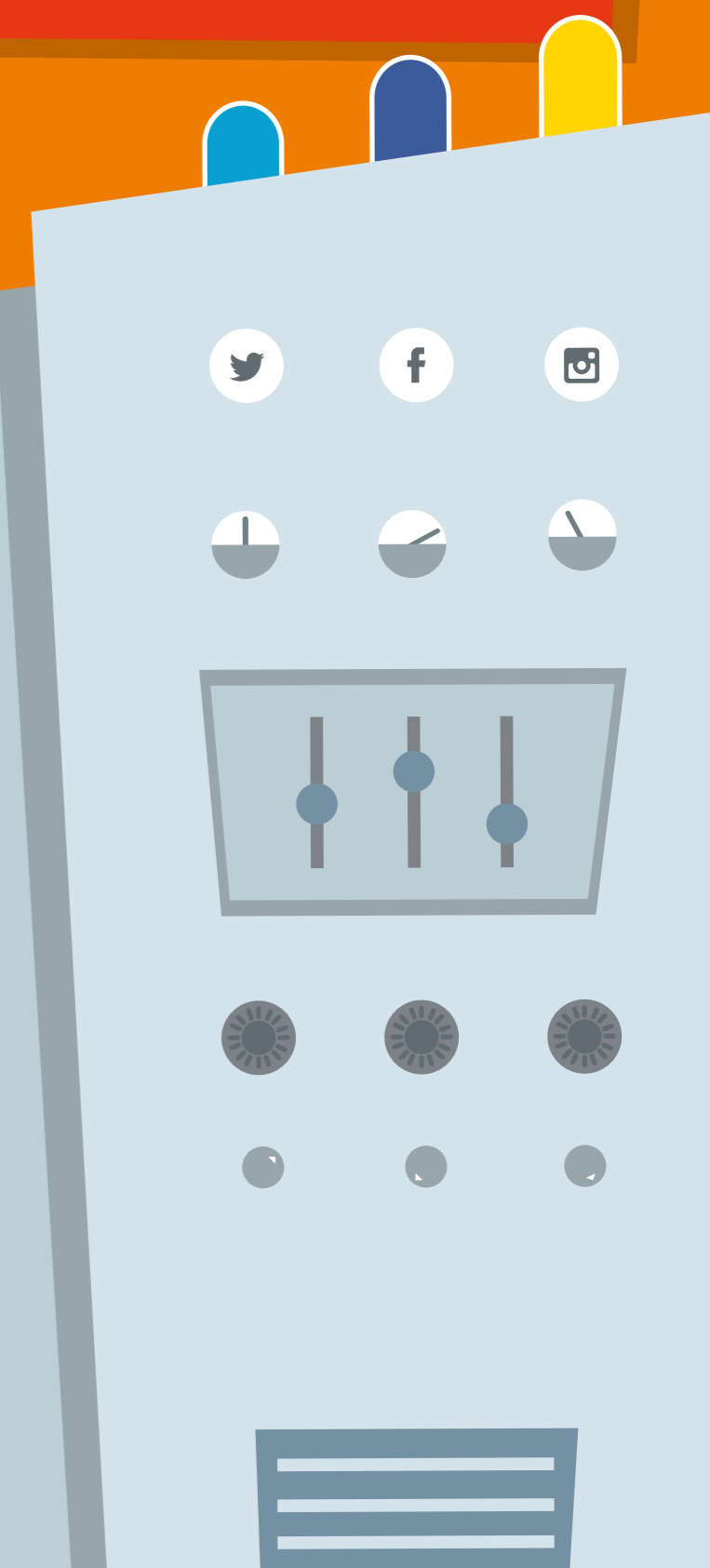
## Advertising versus services (outgoing versus incoming messages)

The measurement of campaigns, for example, advertising, will focus on the goals the campaign was designed to impact, plus the measurement of core brand/organisational attributes. By contrast the measurement of incoming messages relating to service provision and the general social media buzz about the brand needs to be more open-ended.

### Incoming messages

The measurement of services and general buzz will often include the number of people talking about a brand in social media, sentiment analysis, reviews, and referrals. The BT case study quoted provides a good example of how service delivery and social can be integrated, how they can be part of a wider campaign, and how they can be measured.

The measurement of services and general buzz will often include the number of people talking about a brand in social media, sentiment analysis, reviews, and referrals. The BT case study provides a good example of how service delivery and social can be integrated, how they can be part of a wider campaign, and how they can be measured. It is important to note that customer satisfaction measurement by social is a developing field and needs careful analysis. Twitter, for example, may achieve low scores for initial satisfaction, because it can be used as a "complaint" route. Customers who are unhappy and who Tweet can, of course, be converted to brand champions if handled in the right way, but the conversion may not be handled or expressed through the same channel and therefore the channel's effect may be misattributed.

# 4.10

## Case study: BT

Use of social channels in customer service can make that service easier, cheaper and quicker to deliver for the brand and more effective for the customer – improving loyalty and overall returns. Since 2009 BT has been developing a multi-faceted social approach, including:

### Social monitoring

A social monitoring tool (Debatescape) to listen to what customers are saying about its service experience and the brand.

### Customer service

A customer service team to proactively and reactively offer help to customers complaining in the social web.

### Embedding social

Embedding social media activity within call centres, using trained, existing technical help advisors.

### Social sresences

Setting up social media presences in the places where conversations are taking place, using Twitter to enhance communications and creating opportunities for self-help and community support.

BT divides its metrics into Soft ( e.g. Facebook likes, followers on Twitter, positive mentions, views on YouTube, etc.) and Hard (e.g. customer satisfaction, customer effort, likelihood to churn and call deflection). The calls deflected, for example, show impressive returns. Over the time period measured (12 months) Twitter deflected 38,023 calls, saving £98,000; YouTube deflected 55,000 calls, saving £72,000; and the Forums deflected 504,000 calls, saving £942,000. The ROI model has been built on volume of contacts, unique customers, effectiveness (full resolution of query) and cost of contact.

In 2013 BT stated that the lower effort, lower cost, and positive brand impact have resulted in a £2 million saving. In terms of satisfaction scores, BT has revealed detail of the initial impact after the inclusion of using social media to offer customer service. A comparison of three month rolling average satisfaction scores revealed that in October 2010 the score was 63% and by Oct 2011 this had improved to 87% (a 37% increase).

BT identified that 'ease' of customer service was a factor in reducing churn and also had positive effects on customer loyalty and advocacy which could help it grow in new areas. BT asked the question "How easy was it to get the help you wanted from BT today? Effort became therefore a more commercially effective metric to track than the 'Net Promoter Score' used by many other organisations. So BT developed a "Net Easy Score" instead. And use of social media was clearly seen to be driving the Net Easy Scores higher.

See the full case study on the IPA website.

## Five point plan for baking measurement into social

The #IPASocialWorks team has generated a five point plan to help ensure that measurement is baked into social campaigns and activities. Note, these are not five linear steps. Organisations and their uses of social, are very varied, so different projects might enter the process at different points and may travel through the list more than once to ensure that the opportunity to evaluate the campaign is maximised.

**2**

### Why social?
### What is the role of social?

Is it being used on its own or in addition to other components? Which channels and content are going to be used? Including a consideration of the role of paid, owned, and earned media.

List the channels to be used. Paid and owned media channels tend to be self-evident, but earned media can be quite wide.

**1**

### What is the campaign/activity designed to do?

How does this link to the wider business or organisational needs? While not all social campaigns are intended to link directly to broader business objectives, it is important that there is a series of links, model or hierarchy of effects that can trace the investment back to value for the business.

List the macro and micro objectives. Macro objectives tend to be business objective such as sales. Micro objectives relate to how the campaign is intended to work, such as downloads, registrations, or sharing. Each needs to be covered by relevant metrics.

**3**

### What decisions will be made on the strength of the evaluation?

- Are the metrics to be used at the end of the campaign to assess it, or are some to be used dynamically during the campaign/activity to manage the implementation?

- How would these decisions be made otherwise?

- How much resource should be invested in evaluation?

The decisions drive the selection of metrics, and the timing of the decisions is a major factor.

**4**

## What are the most appropriate datasets and metrics?

How will they be collected? The key considerations are:

**Audit Metrics** **A**

e.g. what was sent out and who was reached, including virality.

**Resonance Metrics** **B**

e.g. engagement, attitudinal, customer satisfaction and reputation effects.

**Responce Metrics** **C**

e.g. sales, customer acquisition, and other calls to action.

**D** Metrics that differentiate between the channels and between social and non-social components

e.g. TV and Twitter, Search and Social.

For each metric, use historic data to form a prediction or target for what you expect the activity to achieve. As ever, KPIs should be SMART – Specific, Relevant, Attributable, Relevant & Time-bound. Create a matrix of what is possible with the objectives and decisions, to determine an optimal set of metrics. Small enough to be manageable, large enough to cover the needs.

**5**

### Designing the evaluation process

Determine how the data, metrics, and objectives are going to be analysed to measure the incremental effectiveness of the campaign or activity. Plan for actionable granularity & learning. Evaluations should ideally measure not only the incremental impact of the campaign but how success can be repeated. Because measuring causality for social can be complex, plan in media and content tests and also how they will be evaluated.

Design the evaluation, including the method, budget, timing, and reporting. With social an integrated dashboard showing real-time feedback is often expected. But this step should also include the potential for 'test and learn' approaches as the campaign continues – or 'adaptive planning'.

**Step 1 & 2**

Steps 1 and 2 help determine the nature of the campaign and the balance between the channels, for example how much of it will be social, and within that what specific channels are going to be used. Steps 1 and 2 should also generate the hypotheses that the evaluation will seek to test and evaluate.

**Step 3**

Step 3 determines the priorities and timeliness of the reporting.

**Step 4 & 5**

Steps 4 and 5 determine what will be collected and how it will be analysed.

Section 5

# Data, Metrics and KPIs

With the growth in platforms and the proliferation of metrics the challenge is to select the right metrics for a specific purpose. There are four key sources of metrics, each with merits and issues:

- The platforms, such as Facebook, YouTube, and Twitter.
- The client, when owned media is being used.
- Third parties, such as ComScore, Klout, and Google along with market research companies and social monitoring companies.
- Participants, for example location data gathered from mobile device users.

No single book or document can list and review all the metrics available, although Stephen Rappaport's The Digital Metrics Field Guide does a good job of reviewing 197 key metrics on 12 key fields. Selecting the metrics for a specific campaign or activity is likely to be a bespoke process. There is currently no simple template that can be applied. The five point plan outlined earlier in this chapter will help ensure the key items are covered.

The key issues surrounding the choice of data, metrics and KPIs are set out in the sections below.

## A taxonomy of metrics

The metrics used with social can be divided into the following broad categories:

### Followers/fans

Including fans, contacts, followers, for example, people who have 'followed' a Twitter profile or 'liked' a Facebook page. As a metric these tend to be easy to count and have initial face validity. However, they are sometimes gamed, and tend to reflect the sum of historic interactions as opposed to current activities.

### Reach

Reach is a measure of how many people were able to see or interact with the content produced by a campaign. Its strength is that it produces a measure of breadth to complement a measure of volume. A key weakness is that opportunity to see does not necessarily imply impact.

### Time spent

A useful measure for content that requires longer to consume, such as videos, games, or longer posts. The key strengths of time spent include that it can often relate to one of the objectives for a campaign and/or be a component of engagement. There are relatively few weaknesses with this measure, except where the material was intended to be used quickly and/or used once.

### Referrals

Some tools, for example, Google Analytics, suggest where traffic is coming from. The key strength of referrals is that it indicates which links or actions are associated with people interacting with the target material. The key weakness is that it can encourage analysts/users to attribute the final outcome to the last step in what might have been a long chain of causes and effects.

### Volume

A measure of the amount posts, including tweets, shares, and re-tweets. Volume is relatively easy to measure, but may not reflect the effectiveness of activity. The key weakness is that the amount of material produced is not necessarily directly related to its impact, and even the total amount of buzz created is not the same as the quality or effectiveness of the messaging.

### Engagement

Usually a derived measure, based on who is sharing, linking, commenting on, re-tweeting, liking, playing, viewing, or favouriting. The concept of engagement is widely accepted as a key measure of social. The key challenge is defining which elements should be combined to form a specific engagement metric. Another challenge is creating measurements that are easy to understand and consistent across platforms and over time. Most of the platforms have a preferred measure, as do many third parties, for example, Klout.

### Sentiment analysis
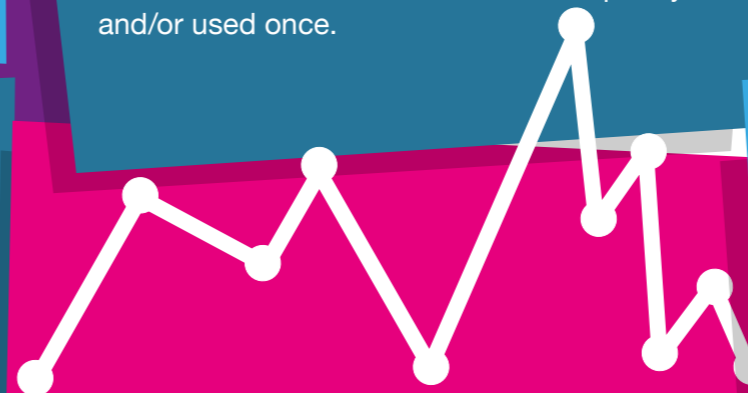
Tends to measure whether comments are positive or negative. Sentiment analysis can be used more subtly, for example, evaluating whether a campaign is producing responses that are on message. The key advantage of sentiment analysis is that it goes beyond how many people are talking about something towards what they are saying and what they mean. The main negative is the wide degree of scepticism about automated sentiment analysis – although this is being addressed in a number of ways, explored later in the next chapter.

### Share of voice

Reflects the relevant mentions divided by all mentions in the category. The term has a least two distinct uses. When talking about ad revenue models, share of voice is the ratio of the brands relevant impressions to the total number of relevant impressions. Amongst social media researchers it tends to be the ratio of the relevant brand mentions as a proportion of all the relevant mentions. The key strength of share of voice is that it is not based on absolute values. Share of voice benchmarks the activity against a wider category. The key weakness is similar to its strength, a good score can be a reflection that there was little to compete with at that time, and a less good score could reflect that there was something else being talked about, such as a major news story or seasonal issues. Share of voice is best used in combination with measures such as reach or volume.

## Validity and reliability

The chosen KPIs need to have the appropriate reliability and validity, covering the following key points:

### Validity

Does the measurement correctly reflect what has happened, including internal and external validity.

### Reliability

The same measurement is run twice, measuring the same phenomena, will it provide the same results?

### Accuracy

Can the data be collected with relative accuracy?

### Internal validity

The change being measured caused by the underlying phenomenon we are seeking to assess, for example, is the change in 'liking' caused by the campaign?

### External validity

Can the finding be generalised to other situations, for example, if the campaign is rolled out on a larger scale would the same effects be observed?

### Cost

Can the data be collected at a reasonable cost?

A measure can be reliable without being valid, if it is measuring the wrong things, or with a systematic bias. And it can be valid without being reliable if it is unstable or prone to measurement error.

## Working with the platforms

There are a wide range of platforms, for example, Facebook, Twitter, LinkedIn, Pinterest, and YouTube. Each of the platforms has its own strengths in terms of the metrics it can offer. The metrics from the platforms can be enhanced in two ways:

### 1.

Working with the platforms to track behaviour, for example, by using cookies cookies and log-on ID's. The range of tracking options changes over time and the different platforms have different options, some going well beyond cookies and email addresses.

### 2.

Third parties, such as Datalogix, comScore, Nielsen and others, offer additional options, in terms of metrics, identification, and tracking.

One of the key things to keep in mind in terms of the platforms and the measurement of social is the potential to gain detailed inputs from paid and owned material – from the POEM model. When working with a single platform a variety of options are available to target who sees what and to track exposure and in many cases behaviour. Earned media is much less amenable to control and targeting.

# 5.4 Selecting the metrics

Selecting the metrics for a specific campaign or activity is likely to be a bespoke process. There is currently no simple template that can be applied. The five point plan outlined in the previous section can be adapted to help ensure the key items are covered.

## 1.

**List the macro and micro objectives.**

Macro objectives tend to be business objective such as sales. Micro objectives relate to how the campaign is intended to work, such as downloads, registrations, or sharing. Each needs to be covered by relevant metrics.

## 2.

**Why social? What channels?**

List the channels to be used. Paid and owned media channels tend to be self-evident, but earned media can be quite wide.

## 3.

**What decisions?**

The decisions drive the selection of metrics, and the timing of the decisions is a major factor. Also, collecting metrics and evaluation costs resources. Consider how the decision would be impacted with the next best alternatives.

## 4.

**Select the metrics**

Create a matrix of what is possible with the objectives and decisions, to determine an optimal set of metrics. Small enough to be manageable, large enough to cover the needs. Give prominence to metrics that can, over time, be used to set meaningful targets. Include non-social metrics (eg TV ratings) where these influence the performance of social.

## 5.

**Design the evaluation and reporting**

Design the evaluation, including the method, budget, timing, and reporting. With social an integrated dashboard showing real-time feedback is often expected. Build in tests and experiments wherever possible.

# Evaluation Methods and Approaches

The evaluation of social can be broadly divided into two categories:

- Formal evaluation of causes and effects.
- Pragmatic measurement.

Evaluating the impact of social media is not easy. When dealing with traditional media, there are usually clearly defined campaigns, often with control over who was exposed. This is less the case with social media, especially earned media. Consumers responding to social campaigns and activities are more self-selecting, are likely to have a predisposition towards the brand, and their engagement is often driven by external factors (including paid advertising). Within a single channel there can be a high degree of targeting and control in terms of owned and paid media, the complexity arises between channels and with earned media.

When evaluating social media we are often interested in how it works in combination (e.g. Twitter and TV or Search and Facebook), rather than how it works in isolation from other factors. Consequently, the choice of measurement technique is very important. As with any aspect of measuring communication effectiveness, it is important to remember that while some approaches are certainly better than others, the best advice is to look at a variety of methods in the context of clear hypotheses of how we think the communication is working.

## 6.1

### Market mix modelling

Market mix modelling (MMM) has been the conventional 'gold standard' approach to measuring communication effectiveness for some years. Typically, it seeks to isolate the impact of advertising and other marketing impacts through a statistical analysis of aggregate weekly sales and marketing data. Its key advantage is that it is a flexible technique that can quantify quite complex relationships in the context of the overall marketing mix.

As with any technique, MMM has limitations when measuring conventional marketing, but there are additional challenges with social:

- Social media (and digital media more generally) is hard to measure consistently and so capture in a model.
- The impact can be relatively small (compared to price and promotions), so is lost amongst the background noise.
- Social activity is often to reinforce the perceptions of brand 'loyalists', so the impact is longer-term.
- It works in combination with other media along the consumer journey, so needs to be measured 'synergistically' with other media.
- It is not always clear what is causing what. Great marketing might lead to online conversation and buzz, which is indicative of sales but not predictive. This can, in particular, be a problem with weekly or monthly aggregated sales, where there is insufficient time granularity to separate which effect comes first.

Model design is very important when using MMM to measure social. For example, the Nielsen's Digital Media Consortium has suggested that poorly designed marketing mix models can understate the ROI from Facebook by as much as 48% and from

Google search by up to 39% (Neff, 2014). The same study also found that 25% of the value ascribed to paid search should actually be shared with TV, print, or other digital media. One reasons for these sorts of effects is that conventional MMM typically measures short to medium-term response in sales, but advertising often works over the longer term by reinforcing existing habits. If long-term effects are being sought then the MMM needs to be appropriately configured (Cain, 2008).

## 6.2

### Data inputs

The following data input issues need to be addressed, each of which is explored in more detail below:
- Issues such as cross-media comparability, synergy, and granularity.
- Interactions between social and other media.
- An assessment of causality using, for example, Granger causality or Hausman tests.
- Consider whether to use VAR (Vector Autoregression).
- Consumer journey modelling builds a model of the consumer journey as a series of links, with functions linking each step, for example, paid search might be a function of TV and earned media, and sales a function of paid search.

The approach Reckitt Benckiser used to measure their Facebook marketing activity for Lysol, their market leading cleaning brand in the US, is one example of this. As an established FMCG brand the direct impact of Facebook and other social media platforms on purchases was expected to be low. As a result the team worked with marketing analytics agency Ohal to focus on understanding the role of social media across all stages of the purchase funnel (which for Reckitt Benckiser breaks down as Awareness, Involvement and Active Consideration) and how these work together to drive sales.

Ohal used a structured least squares approach to enable them to model the complete path to purchase, including brand metrics, and to show the flow-through from each of these to sales. The analysis determined that Facebook activity was primarily driving the Awareness and Active Consideration stages, and furthermore that viral activity from Facebook users liking, sharing and commenting was generating the greatest incremental Awareness, whilst targeted paid advertising was driving Active Consideration. So, whilst the analysis confirmed that Facebook had a minimal direct impact on sales for the brand, the impact of these customer journey stages on sales shows that Facebook accounted for circa 0.2% to 0.5% of total sales for the relevant period.
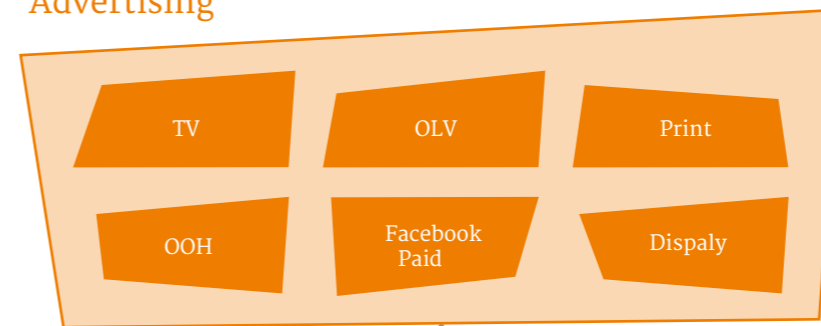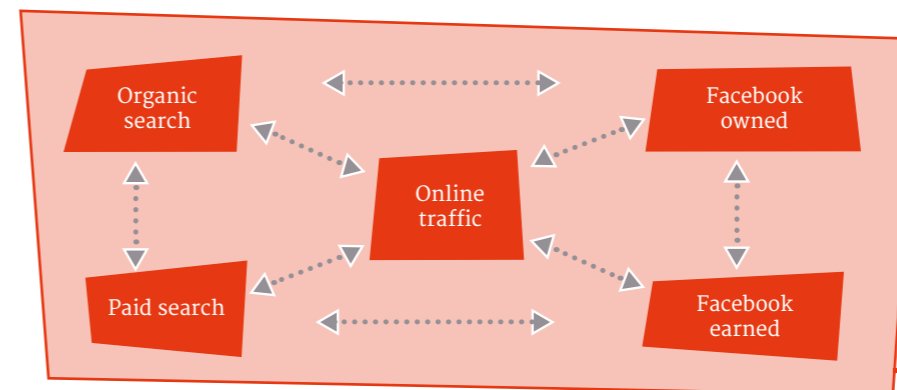


Figure 1: Reckitt Benckiser Path-to-Purchase analysis: The path analysis examined the interactions between advertising and "other drivers" of sales and metrics that can be both outputs and inputs into later models.

**Legend:**
- Other Factors
- Consideration
- Involvement
- Awareness
- Facebook paid
- Facebook owned
- Facebook earned

Figure 2: Facebook Impact on Path to Purchase: Paid and Earned have 0.9% and 1.5% impact on Awareness, respectively, with a 1.2% and 0.3% impact on Consideration. These stages in turn impact later stages in the journey including sales.
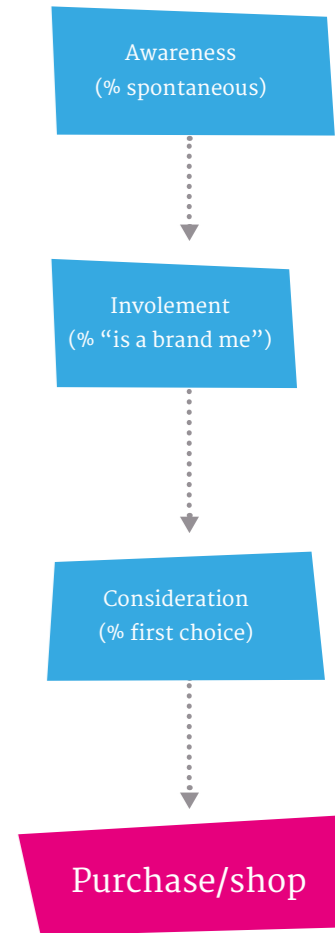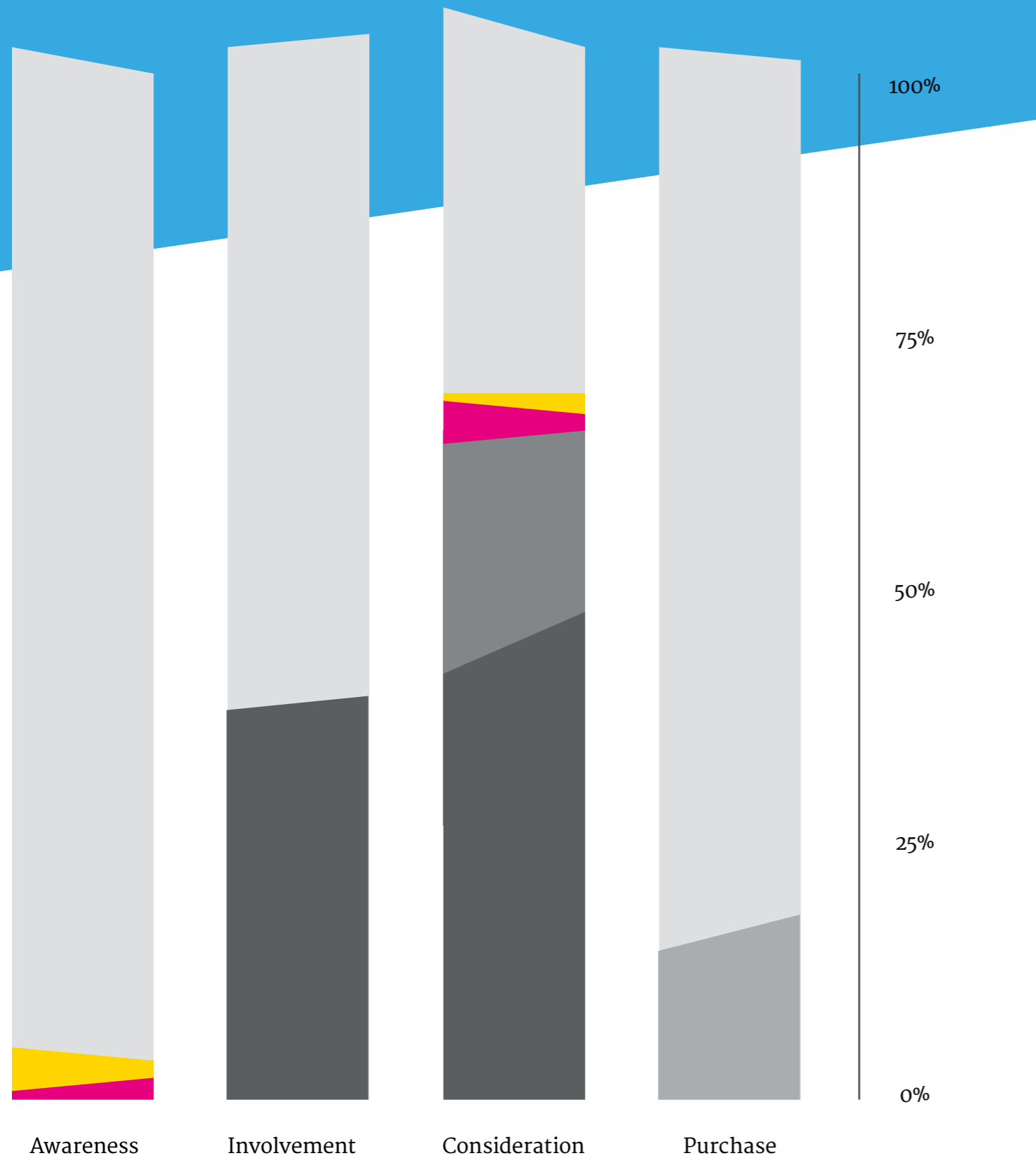
**MMM Data Inputs - Best Practices**
When organising the data for a MMM project there are a number of key considerations and broadly accepted best practices including the four points below.

**Cross-media comparability.**
Metrics should be comparable to other channels and provide a measure of exposure. These should be based on both reach and impressions to guard against 'excessive' frequency. Engagement metrics (clicks, likes, shares, comments, number of followers/fans, etc.) tend to be a poor measure of brand activity, are not closely linked to sales (e.g. Nielsen, ComScore, DataLogix research), and potentially suffer from bias in interpretation because the causality can be in the 'reverse direction'.

**Granularity.**
Models are best built at the most disaggregated geographic level possible (i.e. weekly or finer), to maximise the chance of measuring small differences e.g. TV region, postal sector etc.

**Synergy.**
Model connections and feedback inherent in and between social and other digital media, do not just model direct impacts (e.g. TV + search + social).

**Media channel granularity.**
Model Paid, Organic, and Viral streams separately where possible. Similarly, model PC social separately from mobile.

**Modelling interactions.**
A more fundamental issue with traditional models is that there is a complex pattern of interaction between the media and sales variables. For example, TV helps drive search, which in turn drives social exposure. Similarly, social helps build the brand, but, also, stronger brands will have more people who are likely to engage with the brands social activity. This means there is a feedback loop from social to sales and from sales to social. In the parlance of econometrics, social exposure and engagement isn't exogenous (determined outside of the system) but largely endogenous (determined by other marketing variables).

There are several different techniques to improve conventional MMM and to overcome the issue of 'endogeneity'. It is important to remember that there is no single 'best' approach. As with all models, the choices and decisions largely depend on the most important features of the model and the data.

## Granger causality

At the heart of the issue of measuring media linkage and synergy is how to measure 'causality'. The simplest situation is when we want to know whether a social metric can be a causal indicator for sales or another behavioural KPI. This might be part of the model checking process or an issue in its own right. For example, Twitter is often used as a lead indicator of TV ratings, but does Twitter 'cause' higher viewing or is it 'caused' by programme popularity.

The standard technique for establishing causality is based on a test originated by Clive Granger, and known as 'Granger Causality'. WPP's analysis of how Twitter leads to larger programme audiences is a good example of Granger causality (Kantar Media 2014). The study found that Twitter caused significant changes in live viewing for 11% of TV shows with a marked variation by genre, adding about 1-2% incremental audience. Conversely, TV viewing increased Twitter conversation for just over half the programmes analysed.

Granger causality is based on 'predictive causality' i.e. if A causes B, then knowing A will help make a better prediction of B. More formally, a variable X is said to Granger-cause Y if knowing about the history of X and Y allows a better prediction of Y, compared with only knowing the history of Y.

The concept seems straightforward, but is typically only part of an analysis. Granger causality can help identify whether e.g. Twitter leads to higher TV ratings, but it doesn't quantify how big the impact is. It is directional, so it is usually used as a step before a more formal model is built. One key point is that Granger causality is based on past values - it does not help determine causality between two events that happen at the same time.

One problem to look for is the post hoc ergo propter hoc fallacy (after this, so because of this) – which describes the problem of thinking that if A keeps following B it is caused by B. Two variables can look like they are interrelated but are actually determined by a third. This can be a particular issue in social and viral marketing because of homophily – i.e. that similar people group together. For example, the success of a product launch can look like it is driven by viral or word of mouth marketing because sales adoption is clustered. However, the cause might be broadcast media. This can be a particular problem if there is time aggregation bias e.g. analysis is done at aggregated weekly level, but the 'time-ordering' happens faster. For example, both social buzz responds to a promotion and sales respond to promotion.

'At the heart of the issue of measuring media linkage and synergy is how to measure 'causality'

## Vector autoregression (VAR)

A more comprehensive method of accounting for interactions between media variables and feedback from sales is based on vector autoregression (VAR). VAR is an econometric methodology which seeks to measure interdependence between time series variables. Its key strength, according to its supporters, is that it accounts for complex interactions between variables in a relatively 'theory free' way, i.e. the analyst is not imposing a structure on the data, or deeming some variables to be exogenous. It has been popular in macroeconomics for some time and is now being applied in marketing.

Besides being able to capture interactions between marketing variables, VARs can to some extent capture long-term dynamics. Given that the economics of advertising frequently rests on its long-term impact, being able to trace the impact on base sales is important. Conventional MMM will typically use adstocks as a measure of the long-term impact on base sales, where VARs are unconstrained and can detect the impact of any marketing driver that influences habit.

While VARs have many benefits in principle, they face a number of practical criticisms:

1. Prioritising data mining over 'theory' puts a great emphasis on data quality and data collection. The impact of measurement error or missing variables can be amplified through the models. And, if too many variables and time slices are added then the model will be 'overfitted', producing a great description of past data, but at the cost of less accurate about the future.
2. Measuring causality in social is very difficult. VARs often use weekly data - and this can be insufficient to tease out whether A is causing B, or B is causing A, for example, whether earned media generated sales interest or sales stimulated online conversation among the brand's user base.
3. VARs use deep correlations in the data to measure long-term marketing impacts. This can prove problematic for measuring social when there is a long-term trend to using social media (driven by technology) and a long-term trend in sales (for example, driven by marketing or the economy).

An example of VAR in use is a study reported by Stacey et al (2014) which found that 62% of store traffic for an electronic retailer was driven by paid marketing, and the predominant channel was achieved through generating consumer conversations about the products. Within this mix, Twitter was effective at helping to stimulate conversations. interactions is a study looking at Twitter & TV advertising (Rubart, 2013). A Twitter-sponsored study for UK mobile operators found that Twitter improved the efficiency of TV advertising by 35%.

## Consumer journey modelling

Granger causality and vector autoregressions have deep traditions in economics and time series econometrics, but are often not chosen by practitioners because the approaches can magnify data errors through amplified feedback loops (so leading to overestimated impacts of media ROI). The simplest alternative approach is to model elements of the consumer journey directly in a linked and chained series of models. For example, modelling paid search as a function of TV and earned media and then including the search terms in a sales model.

Modelling systems of equations has been a long tradition in econometrics. An important requirement with this approach is to make sure that variables that are determined simultaneously are treated as endogenous. For example, if search leads to more social exposure, but social leads to more search, then the models should be treated as simultaneous. The extent of simultaneity often depends on time aggregation: there is more unobserved feedback in weekly than in daily data.

There are a number of techniques for measuring systems of equations that are used by practitioners besides conventional

simultaneous modelling (e.g. Two Stage least squares). For example, Structural Equation Modelling is useful when there is a latent (a hidden or unobserved) variable that has an impact on sales. For example, 'word of mouth' is not observable but is partly driven by earned digital media. Another option is to use Bayesian Networks to model interrelationships. One of the advantages of Bayesian approaches is that they can provide more control of interrelationships, so they can be more stable when there is noisy and incomplete data.

For more information on consumer journey modelling see Peter Cain (2014a & b).

## Experiments and A/B testing

Marketing mix modelling (MMM) is the conventional method of measuring advertising effectiveness for above the line media. However, one of the benefits of most digital media is that it is possible to construct experiments, and online brand tracking studies have made use of this approach for some time.

An increasingly common way to measure social media ROI is to fuse or link social data to household panel or purchase data. For example, both Facebook and Twitter have worked with DataLogix to trace online exposure to offline sales. Data is typically either linked through cookies, email, and/or other identity measures, and then analysed anonymously.

There are two elements to experimental studies:
· Controlling the exposure.
· Measuring the effect (see Kohavi, 2008).

In terms of the exposure, the aim is to control it at the individual level. Options that allow people to have multiple accounts, such as cookie tracking, can be flawed as they can mean some double-counting.

## Controlling exposure
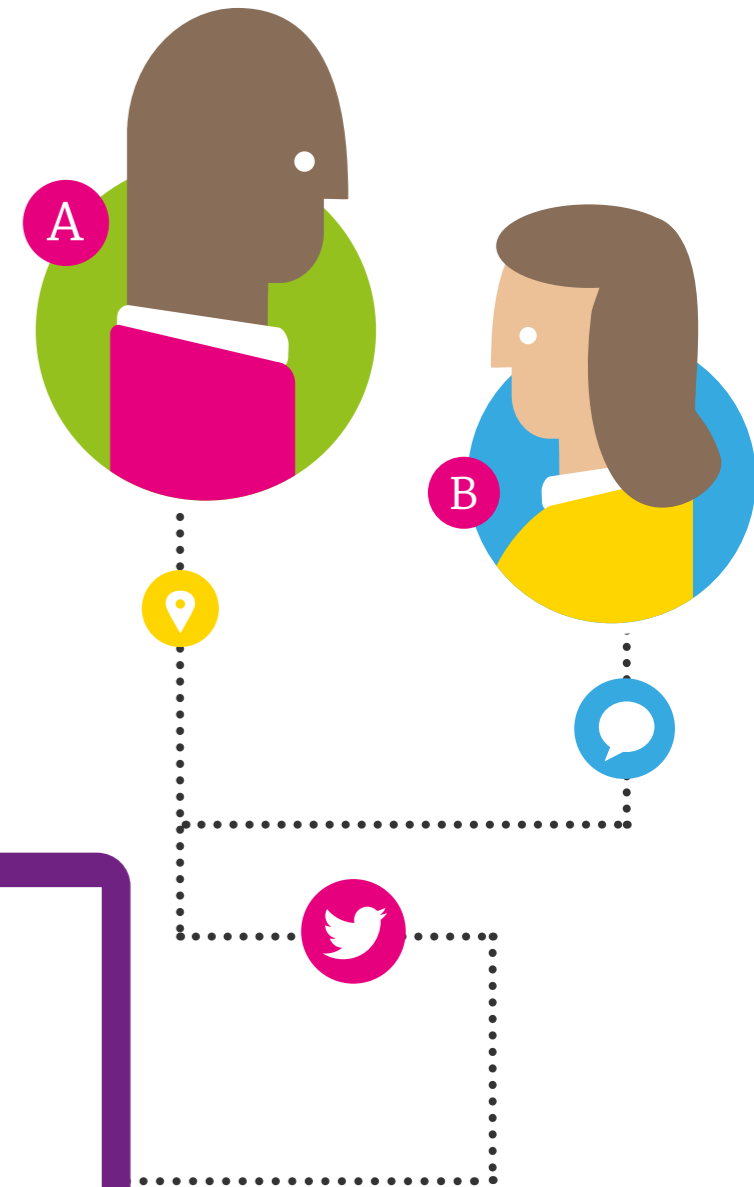
The gold standard approach to testing/experiments is randomised controlled trials, where people are randomly exposed to media and the only difference between the groups is their exposure. For example, serving the test ad or survey in random rotation.

While some academic studies may achieve this high bar, true random experiments are expensive and are not routine. In particular, many social campaigns are not random by design. So, while it is possible to create randomised exposure to paid social advertising (within a specific channel), exposure to the social/earned element is not random (i.e. the interactions between consumers). Importantly, the consumer-to-consumer interactions are a key element of social effectiveness.

When it is not possible or economic to create control exposure groups, the exposed/non-exposed groups are created after the activity has run. This technique is called matching. Matching is widely used in some social sciences (e.g. politics) and is also known as propensity matching. The idea is to create a quasi-matched sample based on observable characteristics of the sample. For example, after the activity has run, the data provider might create two samples from the data set, each with identical demographics, social media usage, etc. The only difference should be that one group was exposed to the advertising, and one wasn't.

Whilst it appears easy to simply compare site A vs site B, or treatment A vs treatment B, the research should try to control for the counterfactual – i.e. what would have happened anyway, for example, what would have happened if neither A or B had occurred.

### Exposed Audience

### Un-exposed Audience

## Measuring experimental effects
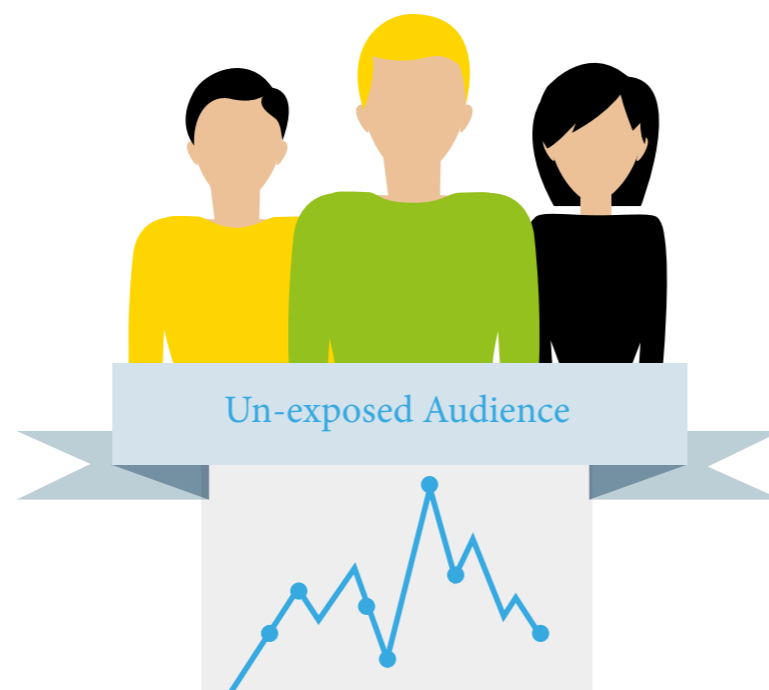
There are lots of potentially confounding factors that can frustrate the accuracy of the analysis (time of day, site, offline media). Without careful control, each element can significantly bias the results. Social commerce, for example, might provide an easier way for brand loyal consumers to purchase the brand - but many would have done so anyway, so the incremental sales are not as high as they might appear.

With experimental approaches it is still important to measure uplift, not just increased propensity, i.e. the difference in likelihood of responding given being in the treatment group versus being in the control group (mathematically, $P(O = 1 \mid x; Treatment) - P(O = 1 \mid x; Control)$). Quite often, approaches just measure propensity through a binary response model (e.g. logit, probit).[1]

Similarly, care has to be taken with conventional methodologies for online tests. It is easy to over-estimate the power of the test because the number of events/impressions look large, but the number of distinct people is much smaller (Bakshy 2013).[2]

### Event or individual/household level modelling

Inevitably, there are a variety of techniques that mix elements of the modelling and experimental approaches. Typically, they use consumer or respondent level data. In the context of social, there is a strong advantage to some of these approaches because they can capture or model the non-random way that social messages diffuse through the population.

Three key approaches, consumer mix modelling, attribution models, and a combination of consumer mix and experiments should be considered.

An Evaluation of Bootstrap Methods
View >

> "It is easy to over-estimate the power of the test because the number of events/impressions look large, but the number of distinct people is much smaller."

[1]Bias can be further mitigated using difference-in-differences estimators, i.e. looking how changes between the treatment and the control group change over time.

[2]Bakshy, E, Eckles, D

## Attribution models

'Attribution modelling' is a term that is applied to many different techniques but is typically used to refer to how digital media work together across the digital consumer journey[3]. Most approaches are based on modelling converter versus non-converter paths using sophisticated analytical techniques (e.g. machine learning, Shapley value etc).

A good example of how attribution modelling can be used to measure social media is provided by attribution modelling specialist Visual IQ (2012). The study finds that Facebook is undervalued by traditional methods (typically, by around a third) because it helps drive demand at the 'top of the funnel'. This finding is reasonably consistent from other attribution models.

## Consumer mix modelling

Consumer mix modelling is econometric modelling at the consumer level, typically using sales panel data. The advantage of modelling at the individual respondent level is the ability to exploit variation in exposure across the data but controlling for exposure variables (similar to matching), which means that social can be measured in the context of price, promotions and other marketing instruments. This also means that communication can be measured against consumer behavioural segments (brand loyals, trialists, etc) and not only capture ROI, but also how the advertising is working.

Consumer mix modelling relies on good single source data or data links, for example, the GfK Media Efficiency Panel (see Cadbury/GfK 2010). Its weakness is usually that the media exposure data, particularly for above-the-line media, is not single source and is estimated/fused.

An extension of the Consumer Mix Modelling approach can be used to measure the impact of social on media products (e.g. films, TV programmes, games etc.). Traditional modelling techniques are problematic because there is both a lot of variation in the underlying quality across products and the 'shelf-life' of the product is short, with most interest in the impact of social on launches. Modelling the data as a panel, with sophisticated approaches that allow social media effectiveness to vary both across products and over time, is an effective solution (Deloitte, 2013).

## Combination of consumer mix and experiments

The most sophisticated approaches combine experiments and very advanced analytical techniques. This allows the study to control and understand the interplay between viral transmission (i.e. non-random exposure), homophily (i.e. non-causal response), and message. Currently, these approaches are used in academia but are not yet available at scale in the commercial sector.

A good example of these sophisticated analyses and the level of granularity of the findings is Sinan Aral's paper on the role of influentials and susceptible individuals (Aral, 2011). The paper uses an experiment and a model to unpick how a piece of social content (a movie rating app) spreads in a social network. Of interest is the finding that typical estimates would over-estimate contagion by 300-700% because they don't account for similar people responding in the same way to external factors. In the study over 50% of the effect was attributed to homophily.

## Example of MMM incorporating social media research

Bottom-Line Analytics supplied the project team with an example looking at a specific category of personal care. The inputs to the media mix model included: digital (including display, paid search, mobile, and social media ads), traditional media (including outdoor, radio, TV, print, and cinema), retail (including distribution and price), a factor to account for periodic seasonality, and two metrics based on measuring open ended comments in social media.

The social media measures were created using a proprietary approach utilising stance-shift analysis to generate scores for the target brand, topics and competitors. The study looked at weekly UK data from January 2012 to April 2014.

The scored social media engagement metric showed a strong association with retail sales volume, i.e. when consumer discourse around the brand was engaged and positive it tended to be followed by an increase in sales and when commentary was less positive, lower sales followed shortly afterwards. The study suggested that social commentary acted as a leading indicator.

## Advice when conducting formal evaluation of campaigns

The following four points should be considered when undertaking the formal evaluation of campaigns, for example, when using MMM or A/B testing.

1. Use more than one measurement technique, seeking to measure the entire consumer journey No single approach is perfect. Combinations of modelling and experimental methods work well, particularly as part of a test and learn process.
2. Biases in aggregate modelling or A/B testing approaches can be severe, resulting in spurious findings. Always be sceptical about very high ROIs unless there is a clear consumer-driven reason. Check the scale of the impact makes sense with supporting bottom-up calculations.
3. With all modelling approaches, use granular data to avoid social signals being drowned out by other elements of the marketing mix.
4. If using aggregate sales data, model social as part of a 'system'. Think first of the consumer journey, then introduce more sophisticated statistical approaches where these are warranted.

[3]In digital analytics, attribution modelling has a wide range of meanings. For example,
- Fractional attribution, the credit for an action (e.g. download or sale) is divided across multiple factors, according to a fixed ratio. - Multichannel attribution, these models typically apply to multiple digital channels, and require that each channel employ tagging/tracking. - Algorithmic attribution, weights are given to different factors in a dynamic process, based on an algorithm, usually proprietary. - Last click, a single source approach, based on the last click before the action. - Last non-direct action, this the basis of the standard Google Analytics conversion. - First interaction, the first thing that happened gets all the credit. - Linear attribution, all the steps in the path get equal weight, by definition this can't be as good as the right model, but will be better than most bad models. - Time decay model, the first step gets a low value, the second interaction more, through to most for the last step.

## 6.9 Pragmatic measurement

In many cases social campaigns/activities may be too numerous or have too small a budget to warrant formal evaluation. In these cases pragmatic approaches are adopted. In particular:

**1.**

Identifying chains or connections. For example, if evaluating a campaign to raise the intention to visit a destination, we might measure: the number of posts, the number of people engaging with the posts, and the number visiting tourist websites. In a formal approach there would typically be an attempt to check that the people visiting the websites had experienced the campaign, but in a pragmatic approach the check is often at the aggregate level.

**2.**

Focusing on outcomes. For example, a social campaign may be run to promote a snack. If the sales go up the campaign is assumed to be successful, if the sales do not go up it is assumed not to have been successful.

Pragmatic approaches adopt an informal Bayesian approach, if I do A and X happens, and I do B and nothing happens, then, all other things being equal, there is a better chance that A is more beneficial than B. However, this thinking can be flawed.

The main two reasons it may be flawed are:
1) It does not consider the counterfactual, what would have happened anyway, i.e. if A and B had not happened.
2) It does not consider the long-term effects, such as brand affinity.

Even when using pragmatic approaches benchmarks should be used, targets should be set before the campaign takes place, and some effort should be made to consider the counter-factual.

## 6.10 Using market research to evaluate social

Social media tends to be widely dispersed and often very transient and as a result most conventional market research tracking studies may not have a big enough sample to effectively detect and measure it. Nevertheless, traditional market research often has a valuable contribution to make, for example, in the measurement of long-term benefits, such as intention to buy, likelihood of recommending, and affinity.

Measuring the impact of social media via native metrics gives an indication of the strength and reach from the perspective of the campaign. Conversely, traditional research (e.g. tracking studies) tends to be based on a representative or relevant target group and can measure the phenomena from the orientation of the audience.

**Key challenges for market research include:**
• The sample size may not be large enough to identify impacts among the audience.
• Measurements often depend on recall, which can be unreliable.
• Challenges in terms of identifying exposed versus not-exposed.

**Key advantages include:**
• Focusing on target groups, as opposed to everybody reached by the campaign.
• The ability to measure multiple channels, especially when passive data collection is included in the mix.
• The ability to measure longer term issues, such as changes in intention, beliefs, and attitudes.

Being able to measure people who have not interacted socially with the campaign, i.e. have not clicked, played, shared etc. Over recent years there has been a growth in the ability to create research panels which combine the passive tracking of activities (including exposure to advertising and marketing) with surveys. This addresses some of the challenges faced when using market research to evaluate social.

In addition to using traditional market research to evaluate social campaigns and activities market research has developed a new/additional approach, social media research, which is covered in the next section.

## Social media research

Social media research is a growing field and can be used to assess marketing campaigns and activities. The use of social media research includes assessing campaigns in traditional media, in social media, and the evaluation of integrated campaigns and activities.

## Using social media to evaluate campaigns

Whilst the details of using social media research to evaluate campaigns can be complex the basic steps are fairly straightforward to describe:

### Social media research

Measuring the effectiveness of communication campaigns through traditional media such as TV advertising has long been the remit of quantitative researchers across the globe. Representative sample surveys aimed at measuring the public's awareness of a campaign, recall of its messages and more importantly whether it has shifted the needle are the norm. The advent of social media, and the unprompted brand mentions it can yield, offers an additional opportunity to get a read on the effectiveness of campaigns and activities.

Social media research is a relatively new discipline within market research and one that offers a vast range of uses, including ideation, qualitative research, PR evaluation, and customer satisfaction. As mentioned earlier, this Guide focuses solely on the use of social media research to measure and evaluate marketing campaigns and activities.

The sorts of questions that social media research is used for include:
• Has a TV campaign led to people talking about the brand in a different way?
• Are people posting messages about interactions with a new outdoor advertising campaign?
• Have people noticed a change in service?
• Are people engaged or cynical about a new social media campaign?

### Definitions

Social media research tends to be described in two ways, using a narrow and a broad definition.

• *Narrow definition*. Quantitative research based on locating, collecting, and analysing naturally occurring discourses within social media. Examples include counting mentions of particular terms, measuring the sentiment of posts on a particular topic, and monitoring the volume of key phrases.
• *Broad definition*. Qualitative and quantitative research that either use naturally occurring social media discourses as their source material or socially created research situations. Examples of this broader category of social media research include: netnography, insight communities, and research about social media usage, as well everything covered by the narrow definition of social media research.

More information about social media research is available from *The Handbook of Online and Social Media Research.*

This Guide focuses on the narrow definition of social media research, i.e. the quantitative evaluation and measurement of campaigns and activities through the collection of naturally occurring discourses in social media.

### 1.

Collect mentions from social media, for example, from Twitter, blogs, Facebook (from public locations) etc. Create a corpus, i.e. a database of mentions, tagged with date and other metadata. The data is often collected via third-party services, such as Brandwatch or Radian 6.

### 2.

Clean the corpus of irrelevant mentions. For example, if evaluating the brand Apple, mentions of fruit are likely to be irrelevant. Remove or separate mentions from non-target groups. Examples of non-target groups might be people from other countries or mentions from PR companies.

### 3.

Evaluate the volume and content of the mentions. The volume tends to be simply the number of mentions, or the number of people commenting. The content of mentions can be measured in terms of sentiment (positive, negative, and neutral), or a wider/deeper level of analysis can be applied.

In social media research there is limited meaning to absolute values, therefore most researchers prefer to monitor several brands or campaigns, in order to benchmark the brand/campaign of interest.

The key areas of discussion in terms of measuring brands and campaigns via social media are:
a. Whether to use automated content/sentiment analysis, manual analysis, or some combination of the two.
b. The extent to which mentions in social media are representative of the wider population.

## Strengths and weaknesses of social media research

One of the key strengths of social media is its immediacy, making it an excellent way to get an early read on what people think of campaigns, potentially within the first hours of its launch. The fact that posts are self-generated and can be mined retrospectively is also a key asset for social media as a data source.

"Social can represent unprompted, unpremeditated, and unmediated responses from people going about their everyday lives."
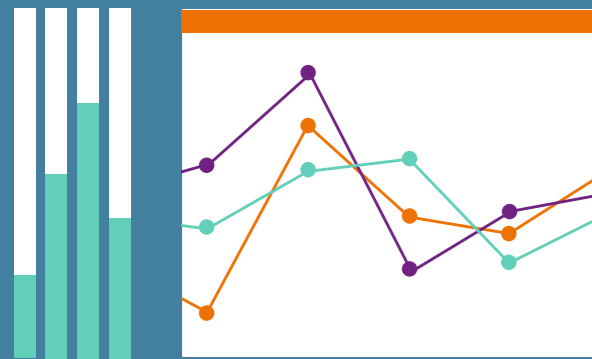
Social media research means that researchers do not have to rely on respondents' recall, as with more traditional methods, and can potentially measure unprompted awareness from the level of mentions the campaign receives in social media. It also means that benchmarks of awareness and perceptions prior to the campaign can sometimes be derived after the campaign has ended, as there are fewer time constraints. This is a key advantage that traditional research does not have. Although, misuse of this phenomenon can lead to fishing rather than measuring.

Social media can reveal which aspects of a campaign are the most salient to the public in their own words and contexts. In traditional research, respondents are prompted by the research process, which may frame or change their responses. It has been argued that social media research provides a purer reflection of consumer perceptions and attitudes towards campaigns and activities and ultimately how they affect brand image/beliefs/attitudes than those derived through traditional research techniques.

Social media does not just enable measurement; it can also provide an in-depth understanding of initial reactions to a campaign, which could only be replicated through qualitative research techniques.

This is because social can represent unprompted, unpremeditated, and unmediated responses from people going about their everyday lives.

Social media research is often perceived as a low cost option, which it can be sometimes. However, the use of the better software platforms and the amount of analyst time required can make social media research more expensive that it might initially seem.

One key concern is about social media's representativeness (or some would say lack thereof) of the public's opinion. Despite the fact that the reach of social media is expanding daily and that in the UK Facebook has a reported active user base of over 31 million and Twitter 15 million, the demographic representativeness of this audience is sometimes questioned. A more nuanced concern is that not everybody who uses social media uses it to post comments about brands and services – which could result in some voices being given too much importance.

Some researchers argue that as long as the lack of representativeness is recognised when contextualising and interpreting the content of conversations, it is a secondary issue. These concerns reinforce the need for social media not to be used in isolation from other data collection techniques to provide context. The key question is to assess whether the attitudes and perceptions expressed in social media conversations reflect those of a wider audience.

The increasing use of hashtags by brands which serve as prompts to the campaign can to some extent remove the candid nature of social media conversations about these activities, generating something akin to prompted mentions. This should be considered when analysing results and analysed separately if appropriate.

### Using social media research in conjunction with other approaches and metrics

As well as being used on its own, social media research can be used in conjunction with other approaches. For example, in the evaluation section of the appendix there is an example from Bottom-Line Analytics that shows how social media scores can be used in conjunction with metrics from digital, mass media, and retail to conduct media mix modelling.

## Social media advice

As with other aspects of this Guide, the term 'Best Practice' has been avoided as there is not yet a settled and validated consensus on the best way to conduct this form of research. The advice presented here has been gathered from the leaders in the field.

• Run a benchmark analysis prior to the campaign. This step is key to measuring any shifts in levels of conversation about the brand, but also existing attitudes and perceptions. This will also be a useful exercise to determine which metrics the campaign should utilise. Using a three month time frame before the campaign is likely to smooth out most of the spikes driven by other events, activities, or campaigns.

• Build an intelligent search query. Using the campaign strapline or title will not be enough to gather relevant content. Use key words which relate to key elements of the campaign, for example, central characters and premises and also key words associated with the themes or topics covered. This will ensure that the range of content gathered is in consumers' own words.

• Remember that volumes and share of voice can hide rich insights. While volumetrics are sometimes useful, they are not the be-all and end-all of social media analysis. The key to understanding the effectiveness of social is about measuring and not counting. This is why human analysis is often important in this context.

• Decide on the balance of human and automated analysis. There is a broad consensus that fully automated solutions are not yet capable of producing the required level of sensitivity. Analysis that is entirely based on human analysis requires sampling and can be expensive and lengthy. The most attractive options tend to combine human analysis with automation, often via some form of machine learning/training.

• Where appropriate, explore what research and data is available directly from the platforms.

## Sentiment analysis

Sentiment analysis, in particular automated sentiment analysis, is a cornerstone of utilising social media to evaluate marketing campaigns and activities. However, it is a cornerstone that attracts a large amount of concern and indeed criticism. Sentiment analysis has attracted criticism in terms of the accuracy of the sentiment analysis, the representativeness of online comments, and lack of established models that translate online sentiment to specific business outcomes.

### What is sentiment analysis?

Sentiment analysis relates to a wide variety of approaches that seek to analyse text (i.e. qualitative information) and to ascribe quantitative values to it. For example, a sentiment analysis system might take a large number online comments about a brand and try to determine the proportion that were negative, positive, and neutral (but it can also be used to ascribe more complex codes). In terms of campaigns and activities the aim of sentiment analysis is often to monitor performance over time.

The typical process for sentiment analysis is:

1. Collect a body of text, called a corpus.

2. Clean the corpus. Cleaning the data can include steps such as removing references to irrelevant brands (for Coke this would include removing references relating to the drug coke), possibly removing duplicates, and filtering to restrict the text to comments from an area or from a specific group (for example, from consumers but not from marketing partners).

3. Apply analysis to code or score the text, for example, into positive, negative, and neutral.

Whilst sentiment analysis can be conducted manually, either on a whole data set, or on a sample of a data set, most of the interest in sentiment analysis, in the area of social media, is about automated sentiment analysis.

### Sentiment analysis – a two-dimensional approach

A campaign should be measured on two dimensions, the strength of the response to the campaign (typically the volume of mentions) and the sentiment of the response to the campaign. The strength of the response to the campaign can be further refined to look at: how many people made comments, how many comments each person made, and how influential are the people who made comments.

The queries about sentiment analysis tend to stem from the measurement of the sentiment, rather than the measurement of volume.

## The different approaches to sentiment analysis

There are a variety of approaches to sentiment analysis.
The key variations being:

• **Manual coding:** In manual coding a human coder, or a team of people, review the statements and assign codes to them. The code can be as simple as positive, negative, and neutral, but it can also be more complex. Unless the number of statements/comments is small, manual coding requires a sampling approach. When sampling statements/comments, care must be taken to sample them in a way that does not introduce bias.

• **Fully automated coding:** In fully automated coding all of the work is done by the software. The software can utilise a wide range of approaches. For example, some are based on natural language processing, some code on key words, some apply content analysis, whilst others are more dictionary/database driven.

• **Machine learning:** Machine learning approaches are a hybrid of manual coding and automated coding. In a typical case the process starts with some of the comments being coded manually (for example, perhaps one thousand comments). The next step is for the software to evaluate the manual coding, allowing it to create rules and then to code the rest of the corpus. Some systems have a process to identify text that is 'hard to code', allowing human coders to deal with these cases, which in turn improves the learning process.

• **Mechanical turk:** MTurk (Mechanical Turk) is a crowdsourced service offered by Amazon. Using MTurk is a form of manual coding, but one where the costs tend to be lower, and the level of oversight is less. Using MTurk is often accompanied by screening for skills and by coding the same item more than once to increase the reliability.

Whilst no system is perfect, the general consensus is:

1. Manual coding is the most meaningful, but is less consistent than automated coding, and is usually slow and expensive.

2. Fully automated coding has a few fans and many critics.

3. Machine learning is an approach that tends to balance the criticisms and costs best, and is used by the case studies quoted in this Guide.

### Increasing complexity of sentiment analysis systems

The earliest systems of automated sentiment analysis were based on identifying and scoring key words, for example, scoring "love" +1, and "hate" -1. As systems became more complex the next step was to score combinations of words, for example, "in the pink" as good, "in the red" as bad.
The next step was systems that decomposed phrases and attached them to specific objects. For example, the phrase "I hate coffee but I love Coke" identified that the post had negative sentiment for coffee and positive for Coke.

The next level of complexity comprises a wide range of approaches, most of which are designed for purposes much more complex than simply scoring sentiment; their main purpose tends to be text analytics, automated services, and translation.
The key ones are:

• NLP (natural language processing) seeks to interpret the meaning of the text, applying rules that have been generated from machine learning. Typical uses include translation and writing summaries.

• Statistical scoring of phrases, for example, the way that Google Translate operates. Probability values are assigned to patterns that are observed.

• Content analysis seeks to decompose the text into codes, which are grouped into concepts and from there an overarching structure can be created.

• Discourse analysis seeks to understand what people do when they make meaning and employs approaches such as conversation analysis, psycholinguistics, and stance-shift analysis to work out what the people making the posts and comments were seeking to do.

The rate of development in this field is rapid and new options regularly appear on the market.

**Issues surrounding sentiment analysis**
Automated sentiment analysis of social media posts and comments arrived on the research and marketing scene about ten years ago with a rush of hyperbole, claiming that it was going to fundamentally change the way that consumers' voices and views would be assessed and measured. However, time and experience has identified a number of problems and limitations with social media listening in general and sentiment analysis in particular.

**How representative are the comments?**
Not everybody uses social media, and not everybody who uses social media posts comments. Consequently, people have queried how representative online comments are. This is an evolving discussion and will not be resolved soon. Whilst concerns about representativeness should be taken into account so should the ability of sentiment analysis and social media research to reach places that other research can't reach.

**How good could sentiment analysis be?**
The upper limit of how 'good' machine coding can be is determined by the agreement that can be achieved by human coders, which in turn determined by the complexity of the tasks. In a typical campaign evaluation, the inter-coder agreement is typically about 80%, so it is hard to envisage of machine coding ever being better than 80%, in these sorts of cases. If the machine was 100% in agreement with one of these two hypothetical coders, then it would only be 80% in agreement with the other.

However, machine coding is much faster and much cheaper than using human coders, and it can be much more consistent.

In general, the consensus is that if machine coding were to be 70% accurate then it would provide a really useful input into the evaluation of campaigns and brands.

**Do we mean positive or 'on message'**
Simple sentiment analysis tends to classify text as positive or negative. For some campaigns that simple approach is not an appropriate classification.Dr Stuart Shulman (National Conference on Health Communication 2012) has been mining Twitter to assess an American anti-smoking campaign. The ads are shocking, showing people who have had surgery after cancer, so most of the people for whom the campaign would appear to be working are using words that would be associated with negative sentiment. Shulman's analysis of these campaigns tended to identify who noticed the campaigns, and whether the post exhibited fear arousal (for example, "Man that smoking commercial where the lady has a hole in her throat scares me every time") or fear rejection (for example, "That smokefree commercial is bull shit. My grandma has been smoking cigarettes since she was 10 and she doesnt have a hole in her damn neck.")

It is likely that as sentiment analysis develops it will move beyond positive and negative, as a number of systems already do.

## 6.17

### Other uses of sentiment analysis

Sentiment analysis can be used for much more than the analysis of marketing campaigns and activities, for example:

#### Customer satisfaction

Social media monitoring and the analysis of sentiment can be used to assess customer satisfaction.

#### Reaction marketing

Analysing social media, and assessing the sentiment of those discourses, can allow an organisation to react to opportunities. The response might be to offer help, to promote a solution, or to create a campaign.

#### Service delivery

Sentiment analysis can be used to assess the reactions of service users and facilitate real-time management of services, from traffic management, to retail, to users of social services. One of the key ways that modern customers request support or express anger is to shout it out in social media, which allows brands to listen and act.

#### Commercial information

Thompson Reuters has incorporated Twitter sentiment analysis into its Eikon trading platform - 'the first mainstream financial platform to provide Twitter sentiment'. This follows an SEC decision to allow companies to communicate news to investors on Twitter.

#### Competitor analysis

Social media discourses, and the sentiment they reveal, can be used to assess competitors as well as an organisation's own brands. This is useful in terms of competitive intelligence and essential in benchmarking the discourses relating to an organisation's brands.

#### Semantic analysis

Semantic analysis (such as discourse analysis or content analysis) seeks to understand the language being used on social, often fitting the meanings expressed into a context, rather than the more narrow and limited scoring provided by sentiment analysis. Sentiment analysis tends to focus on positive and negative mentions, or some other evaluative scoring. In many cases about 90% of social comments are neither positive nor negative, being classed as neutral.

## 6.18

### Sentiment analysis guidelines

**Key advice includes:**
- Sentiment analysis is more valuable in combination with other approaches.
- The results for the target brand, campaign, or activity should be benchmarked against other brands, campaigns, or activities.
- The corpus should be checked and cleaned.
- Determine whether the analysis is to be completed on the whole data set or on a sample – manual analysis usually requires sampling.
- Manually check some of the statements assigned as positive and some assigned as negative, to gauge the level of accuracy/confidence.

## 6.19

### The merging of research, service provision, and marketing

Social media research is one of the research areas at the front of the move towards a more integrated picture of marketing and market research. A social media listening tool might be used simultaneously to:

1. Track sentiment.
2. Identify individuals with problems and offer solutions.
3. Identify sales interest at the individual level.

The ethics and regulatory framework for this research and marketing integration is not clear or complete and is likely to evolve over the next few years.

## In summary

This study has recommended that people use the following five point plan to bake measurement into their social campaigns and activities. It is likely that the formal evaluation process will need to be tailored to this broad outline.

### 1.

**What is the campaign/activity designed to do?**

What is the campaign/activity designed to do? How does this link the wider business or organisational needs? Not all social campaigns are intended to link to broader business objectives, which is fine as long as the aims are defined. The measurement of a campaign should be linked to the aims of that campaign.

### 2.

**Why social?**

Why has social been chosen? Is it being used on its own or in addition to other components? Which channels are going to be used? Including a consideration of the role of paid, owned, and earned media.

### 3.

**What decisions will be made on the strength of the evaluation?**

Are the metrics to be used at the end of the campaign to assess it, or are some to be used dynamically during the campaign/ activity to manage the implementation? How would these decisions be made otherwise?

### 4.

**What are the most appropriate datasets and metrics? How will they be collected?**

The key considerations are:
a. Audit metrics, e.g. what was sent out and who was reached, including virality.
b. Resonance metrics, e.g. engagement, attitudinal, customer satisfaction, reputation effects
c. Response metrics, e.g. sales, customer acquisition, and other calls to action.
d. Metrics that differentiate between the channels and between social and non-social components e.g. TV and Twitter, Search and Social.

### 5.

**Designing the evaluation process.**

Determining how the data, metrics, and objectives are going to be analysed to understand the effectiveness of the campaign or activity, and to assess incremental value, including assessing where the value of social comes from.
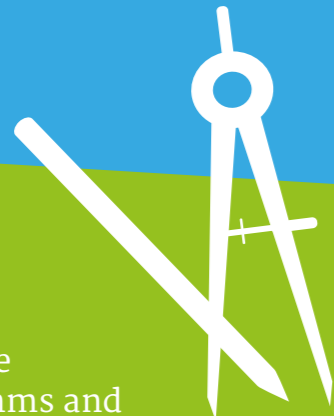
## Seven Key Messages

Use of social is still in its infancy and social tools, data and methods are fast-changing. Currently, there is no one best approach to measuring social activity. We are still learning. This Guide marks a moment in time, and a step on a journey. However, seven key messages have emerged from the project to date.

## 1.

### Social is more than marcomms and is challenging organisations

Social is helping to bring the voice of the consumer to the heart of the organisation. Not only is it broadening the definition of media, but it is also blurring the traditional lines of responsibility for marketing and insight. It is operating as a communication channel, a service delivery platform and a source of insight. It is challenging the concept of a campaign with a clear start and end, as it is always on. For many organisations social data sets are now becoming part of their company-wide digital transformation. However, this can bring with it challenges for creating the sorts of reliable data sets suitable for accurate, predictive and attributional modelling. It is also shifting the balance of the organisation from collecting data to interpreting and analysing signals from multiple sources.

## 2.

### Social is changing the way we measure – its evaluation is more than a dashboard

Social is a new and powerful source of insight for advertisers. For evaluation, it provides new ways to understand not only 'what' happened but 'how' and even potentially 'why'. Beyond generating new metrics, social is also changing the nature of measurement. Because its feedback is real-time, the evaluation process is being integrated with each stage of campaign management from strategy, targeting, content development, delivery and evaluation. Increasingly, faster learning will require a greater emphasis on predictive benchmarks and testing and not just metrics and dashboards. This 'always on' aspect should force organisations to adopt a much broader culture of test-and-learn than is currently evident, increasingly in collaboration with external data partners, agencies and platforms.

## 3.

### Avoid a siloed approach to social measurement

Social tends to work in conjunction with other media. It cannot be measured in isolation. Social needs its planning and evaluation to be integrated with other channels in order to maximise its benefits, establish its value, and be more trusted as a mainstream option. Further, in the majority of cases, the success of owned and especially earned tends to be a product of paid and interaction with other media. The learning objective for social evaluation is to understand how it works with other marketing at all stages of the consumer journey.

## 4.

### It is easy to overestimate the value of earned media and influencers

Accurately measuring causality for earned media is hard. Even with some of the most sophisticated statistical techniques, it is easy to see a causal link when in reality there is only correlation. Another reason to cultivate a broad 'test and learn' culture.

## 5.

### The commercial value of social will increasingly lie in the richness of its data

Current methods of collection and analytics are not fully mature. Two areas in particular have further potential: sentiment analysis and Social CRM. Sentiment analysis will never be 100% accurate, but improvements in algorithms and data collection, will allow the signal to be stronger and more reliable. For Social CRM, given the potentially clearer value exchange for customers in offering personal social data, these data sets could be part of a gateway into much richer insight across an organisation.

## 6.

### Social can learn from traditional planning

Social may be new, but the planning process is not, and the best way to make use of the new opportunities presented by social is to ground them in what is already known about campaigns and other communication activity, e.g. linking to objectives, based on clear assumptions, using comparable metrics. Social needs to adhere to the strategy and planning disciplines used across other marcomms activity and to be designed in from the start, not added retrospectively.

## 7.

### Even short–term results need a long–term context

One of the benefits of social is that it provides measurements that allow campaigns and activities to be optimised in real-time. However, the management of campaigns should balance long-term success with short-term success, since they tend to depend on different elements and strengths. The IPA has shown that key factors such as profitability and loyalty result from long-term effects, not simply cumulatively from short-term successes.

# Appendix

- List of contributors
- Glossary
- References

## #IPASocialWorks

- Professor Patrick Barwise, London Business School (academic advisor)
- Celina Burnett, Ohal
- Jenny Burns, RSA
- Fran Cassidy, Cassidy Media Partnership
- James Devon, MBA
- Simeon Duckworth, Mindshare
- Mark Earls, Herdmeister
- Jane Frost, MRS
- Gemma Greaves, The Marketing Society
- Nigel Gwilliam, IPA

- Jed Hallam, Mindshare
- Janet Hull, IPA
- Chris MacLeod, TfL
- Stephen Maher, MBA
- Matt Taylor, Twitter
- Christian Walsh, MRS
- Christopher Wellbelove, BT
- Jake Steadman, Twitter
- Leila Woodington, Facebook
- Sarah Woodley, The Marketing Society

## Project team

The core project team for The Guide was:

- Ray Poynter (lead author), The Future Place
- Fran Cassidy, Cassidy Media Partnership
- James Devon, MBA
- Simeon Duckworth, Mindshare
- Mark Earls, Herdmeister
- Jed Hallam, Mindshare
- Christian Walsh, MRS
- Christopher Wellbelove, BT
- Celina Burnett, Ohal

The project would like to thank the following people for their contributions to the project, in terms of copy, comments, and/or reviewing material.

- Masood Akhtar, Bottom-Line Analytics
- Patrick Barwise, London Business School
- Malcolm Bell, Zaggora.com
- Alex Bentley, University of Bristol
- Gaelle Bertrand, Precise
- Patrick Bray, stream:20
- Helen Bryant, RSPB
- Peter Cain, Marketscience Consulting
- Francesco D'Orazio, FACE
- Edwina Dunn, Starcount
- Paul Edwards , Non Exec Director, IPSOS
- Laurent Flores, INSEEC Business School
- Alison Henderson, TfL

- Ed Keller, Keller Fay Group
- Tom Kerr, Tesco Bank
- Jake Steadman, Twitter
- Stan Sthanunathan, Unilever
- Matthew Stockbridge, Mondelez
- Rory Sutherland, Ogilvy Group
- Jennifer Thompson, The Walt Disney Company
- Ben Tuttle, Insight Catalyst
- Sue York
- Tom Ewing, Brainjuicer
- Tony Regan, Brand Performance
- Lucy Whitehead, First Group

The fact that people have made contributions to the project does not mean they necessarily endorse or agree with the entire report.

### For more information

Whether you want to submit case studies, collaborate, or share your/your company's expertise in another way, we want to hear from you.

- Kelli Nelson, IPA (kelli@ipa.co.uk)
- Sarah Woodley, The Marketing Society (sarahw@marketingsociety.co.uk)
- Christian Walsh, MRS (christian.walsh@mrs.org.uk)

# Glossary

| Item | Description |
|---|---|
| A/B Testing | A/B Testing is a method of identifying which elements perform best. In a typical case, two randomly assigned cells are shown executions that differ and the results compared. A/B testing can be used to measure the effect of different treatments, or it can be used to select a best performing execution. |
| AVE | Advertising Value Equivalent, the cost of buying the space taken by a piece of 'earned' content, if it had been purchased. AVE has been used by the PR industry to assign a value to earned media, but it is generally seen to be an overestimate. |
| Attribution Models | Attribution models seek to determine which factors contributed to an outcome and to assign values to the weight for each element.<br><br>Key terms and models include:<br>**Single source attribution**, all the credit is given to a single factor, such as last click or first click.<br><br>**Fractional attribution**, the credit for an action (e.g. a download or sale) is divided across multiple factors, according to a fixed ratio.<br><br>**Multichannel attribution** models typically apply to multiple digital channels, and require that each channel employ tagging/tracking.<br><br>**Algorithmic attribution**, weights are given to different factors in a dynamic process, based on an algorithm, usually proprietary.<br><br>**Last click**, a single source approach, based on the last click before the action.<br><br>**Last non-direct action**, this is the basis of the standard Google Analytics conversion.<br><br>**First interaction**, the first thing that happened gets all the credit.<br><br>**Linear attribution**, all the steps in the path get equal weight, by definition this can't be as good as the right model, but will be better than most bad models.<br><br>**Time decay model**, the first step gets a low value, the second interaction more, through to most for the last step.<br><br>**Customised allocations**, different parts of the pathway to the final clicks are given different weights depending on some exogenous factor (such as belief). |

| Item | Description |
|---|---|
| CPI | Cost per impression, to evaluate how widely something might have been seen. |
| CPE | Cost per engagement, where engagement might have been liked, commented on, shared etc. |
| CPL | Cost per lead, where sales leads are tracked back to campaign elements. |
| CPO | Cost Per Order, the cost of the marketing/activity divided by the number of orders. |
| CPR | Cost per referral, for example, where shares are tracked, or the use of discount codes etc. |
| CTR | Click-Through Rate, the percentage of people who see something who click on it. |
| Granularity | Granularity refers to how fine the detail is. One key issue with MMM is the time granularity. If data is supplied as monthly aggregates then it will often be impossible to determine what caused what, because during the month there will have been multiple changes and multiple outcomes. Weekly data is better than monthly, but may not always be granular enough to identify different actions and consequences. |
| MMM | MMM, Market mix modelling, is an advanced statistical technique that takes multiple measures of marketing activity, such as advertising, and seeks to estimate the contribution of each element to the outcome (e.g. to sales). MMM is considered the 'gold standard' in terms of assessing the impact of marketing activities, such as advertising. |
| ROI | Return on Investment. At one level the term ROI relates to a generalised concept, expressing what a project delivers. More formally it is the value of the return divided by the cost of the investment – for example, ROI% = (Net Profit/Investment)*100. |
| ROMI | Return on Marketing Investment, is a subset of ROI, linking the marketing investment and the measured return. |
| Share of Voice | In the ad business this tends to mean impressions for a brand divided by all impressions. In social media research it tends to mean mentions of a brand divided by all mentions of other relevant brands. |

• Anderson, C, 'The End of Theory: The Data Deluge Makes the Scientific Method Obsolete', Wired, 2008, http://www.wired.com/science/discoveries/magazine/16-07/pb_theory

• Aral, S, Content and Causality in Influence Networks, 2011, http://videolectures.net/icwsm2011_aral_influence/

• Aral, S, Muchnika, L, & Sundararajana, A, 'Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks', 2009,http://www.pnas.org/content/106/51/21544.full.pdf

• Bakshy, E & Eckles, D "Uncertainty in Online Experiments with Dependent Data:

• An Evaluation of Bootstrap Methods" http://arxiv.org/abs/1304.7406

• Barwise, P & Farley, J, 'The State of Interactive Marketing in Seven Countries: Interactive Marketing Comes of Age', Journal of Interactive Marketing 19, 1 (Summer 2005), 67-80.

• Barwise, P & Meehan, S, 'The One Thing You Must Get Right When Building a Brand', Harvard Business Review, 2010, http://hbr.org/2010/12/the-one-thing-you-must-get-right-when-building-a-brand/ar/1

• Bentley, A, et al, 2011, I'll Have What She's Having: Mapping Social Behavior (Simplicity: Design, Technology, Business, Life), MIT Press

• Bentley A and Earls M 2008: Forget Influentials: copying is how brands spread, Admap October

• Bughin, J, Doogan, J, & Vetvik, O, '', McKinsey & Company, 2010, http://www.mckinsey.com/insights/marketing_sales/a_new_way_to_measure_word-of-mouth_marketing

• Cain, P, 'Limitations of conventional market mix modelling', Admap, April pp 48-51, 2008

• Cain, P, 'Advanced Methods in Marketing Econometrics', http://www.itqanalytics.com/community/blog/advanced-methods-in-marketing-econometrics#.U-H6iPldXe6, 2014a

• Cain, P, 'Marketing-mix Modelling Big Data', http://www.itqanalytics.com/community/blog/marketing-mix-modelling-big-data#.U-H6svldXe4, 2014b

• Carl, W, 'What's all the buzz about? Everyday communication and the relational basis of word-of-mouth and buzz marketing practices', Management Communication Quarterly, 2006

• Charman, S, 'Fake Reviews: Amazon's Rotten Core', Forbes, 2012, http://www.forbes.com/sites/suwcharmananderson/2012/08/28/fake-reviews-amazons-rotten-core/

• Cialdini, R, 1984, Influence: The Psychology of Persuasion: Issues for social marketing and social evaluation, William Morrow

• Corcoran, S, 'Defining Earned, Owned And Paid Media', Forrester, 2009, http://blogs.forrester.com/interactive_marketing/2009/12/defining-earned-owned-and-paid-media.html

• Court, D, Elzinga, D, Muder, S, & Vetvik, O, 'The Consumer Decision Journey', McKinsey & Company, 2009, http://www.mckinsey.com/insights/marketing_sales/the_consumer_decision_journey

• Deloitte, 'Tweets for Sales – Gaming', 2013, http://www.deloitte.com/view/en_GB/uk/industries/tmt/media-industry/tweets-for-sale/index.htm

• DeMers, J, 'How To Execute A Converged Media Content Strategy (And Why You Should)', Forbes, 2013, http://www.forbes.com/sites/jaysondemers/2013/09/11/how-to-design-and-execute-a-converged-media-content-strategy-and-why-you-should/

• Divol, R, 'Demystifying social media', McKinsey & Company, 2012, http://icwww.mckinsey.com/insights/marketing_sales/demystifying_social_media

• D'Onfro, J, 'Twitter admits 5% of its users are fake', Business Insider, 2013, http://www.businessinsider.com/5-of-twitter-monthly-active-users-are-fake-2013-10

• Earls, M, 2007, HERD: How to change mass behaviour by harnessing our true nature, Wiley

• Farahat, A, & Bailey, M, 'How effective is targeted advertising', Yahoo, 2011, http://research.microsoft.com/en-us/um/beijing/events/ia2011/ayman%20sigir%20_mike_v2.pdf

• Feick, L & Price, L, 1987, The Market Maven: A diffuser of marketplace information, Journal of Marketing

• Finger, L & Dutta, S, 'Ask, Measure, Learn', O'Reilly Media, 2014

• Flores, L, 'How to measure digital marketing', Palgrave, 2013

• Gladwell, M, 'The Tipping Point: How little things can make a big difference', Little Brown, 2000

• Gelman, A, 'Causality and Statistical Learning', Columbia University, 2010 http://www.stat.columbia.edu/~gelman/research/published/causalreview4.pdf

• Harford, T, 'Big data: are we making a big mistake?', The Financial Times, 2014, http://www.ft.com/cms/s/2/21a6e7d8-b479-11e3-a09a-00144feabdc0.html#axzz2yHpWXRpX

• Harvesting the Twitter Firehose for Measurement and Analysis, August 2012, National Conference on Health Communication, Marketing,and Media, Atlanta, USA. Link http://www.healthmediacollaboratory.org/wp-content/uploads/2012/11/Szczypka_hcmmconf2012.pdf

• Hodson, H, 'Google Flu Trends gets it wrong three years running', New Scientist, 2014, http://www.newscientist.com/article/dn25217-google-flu-trends-gets-it-wrong-three-years-running.html

• Kantar Media (2014) http://www.kantarmedia.co.uk/news-views/our-latest-news/kantar-media-study-confirms-link-between-twitter-activity-and-tv-viewing/

• Kantar Media (2014) "TV and Twitter Causality Study" http://www.kantarmedia.co.uk/businesses/social-tv/a-year-in-the-life-of-tv-twitter-in-the-uk/

• Katz, E & Lazarsfield, P, 'Personal Influence: The part played by people in the flow of mass communications', The Free Press, 1955

• Keller, E & Berry, J, 2003, The Influentials: One American in Ten Tells the Other Nine How to Vote, Where to Eat, and What to Buy, The Free Press

• Keller, E, Fay, B, & Berry, J, 'Leading the conversation: Influencers' impact on word of mouth and the brand conversation', WOMMA, 2007

• Keller, E & Fay, B, 2012, The Face-to-Face Book: Why real relationships rule in a digital marketplace, InkWell Publishing

• Kohavi, R, Longbotham, R, Sommerfield, D, & Henne, R, 'Controlled experiments on the web: survey and practical guide', Springer Science+Business Media, 2008.

• Kornowski, L, 'Celebrity Sponsored Tweets: What The Stars Get Paid For Advertising In 140 Characters', Huffington Post, 2013, http://www.huffingtonpost.com/2013/05/30/celebrity-sponsored-tweets_n_3360562.html

• Lewis, R & Rao, J, 'On the Near Impossibility of Measuring the Returns to Advertising', 2013, Google and Microsoft, http://www.aeaweb.org/aea/2014conference/program/retrieve.php?pdfid=985

• Lutz Finger and Soumitra Dutta, 2014, Ask, Measure, Learn: Using Social Media Analytics to Understand and Influence Customer Behavior, O'Reilly Media

• Milgram, S, "The Small World Project", Psychology Today, 1967

• Mutinga, D, Moorman, M, & Smit, E, 'Introducing COBRAs - Exploring motivations for brand-related social media use', International Journal of Advertising, 2011, http://www.tandfonline.com/doi/abs/10.2501/IJA-30-1-013-046#.VM-h42isXng

• National Conference on Health Communication, Marketing and Media, "Harvesting the Twitter Firehose for Measurement and Analysis", 2012 http://www.healthmediacollaboratory.org/wp-content/uploads/2012/11/Szczypka_hcmmconf2012.pdf

• Neff, J, 'Digital Is Selling More Soap Than It Gets Credit For: Nielsen Study', Advertising Age, 2014, http://adage.com/article/digital/study-digital-return-investment-undervalued/291772/

• Nielsen Global Trust in Advertising  (2013) http://www.nielsen.com/us/en/insights/reports/2012/global-trust-in-advertising-and-brand-messages.html

• Reichheld, F, 2003, One Number You Need to Grow, Harvard Business Review

• Ries, E, 'Why vanity metrics are dangerous', 2009, http://www.startuplessonslearned.com/2009/12/why-vanity-metrics-are-dangerous.html

• Rubart, C, 'New MarketShare Analysis Reveals Twitter Impact on TV Ads', MarketShare, 2013

• Silver, N, 'The Signal and the Noise - Why So Many Predictions Fail-but Some Don't', Penguin, 2012

• Solis, B, 'The Rise of Digital Influence: A "how-to" guide for businesses to spark desirable effects and outcomes through social media influence', Altimeter Group, 2012

• Stacey, E, Pauwels, H, & Lackman, A, 'Beyond Likes and Tweets: Marketing, Social Media Content, and Store Performance', The Center for Measurable Markting, 2013

• Thompson, C, 'Is the tipping point toast?', Fast Company, 2008

• Watts, D, 2004, Six Degrees: The Science of a Connected Age, WW Norton

• Watts, D, 'Viral Marketing for the Real World', 2007, http://hbr.org/2007/05/viral-marketing-for-the-real-world/ar/1

• Cadbury/GfK, 'Marketing Mix Evaluator: Cadbury's Chocolate Charmer Campaign', 2010, http://www.scribd.com/doc/45468533/Cadbury-Campaign-Results-Dec-2010

• Nielsen, 'The Follow-Back: Understanding The Two-Way Causal Influence Between Twitter Activity And Tv Viewership', Nielsen, 2013, http://www.nielsen.com/us/en/insights/news/2013/the-follow-back--understanding-the-two-way-causal-influence-betw.html

• Visual IQ, 'A Visual IQ Perspective - Four Crucial Takeaways on Attribution & Facebook ROI', 2012, http://www.visualiq.com/sites/default/files/four-crucial-takeaways-on-attribution-and-facebook-roi.pdf

# IPA Social Works