

DO YOU KNOW
WHAT I WANT?

TOTAL RECALL

People struggle to remember accurately why they have behaved in a certain way. So tools and technologies are constantly emerging to help researchers get closer to people – to observe and interpret their behaviour, as well as to collect data. By **Tim Phillips**

It's October 1968, and Dr George Gallup is describing his faith in our ability to witness our own lives: "To prove or disprove the hypothesis that women could not remember what they had for dinner a week ago, we brought 50 women into our interviewing centre, in Hopewell, New Jersey... we took each one by the hand and led her back through the week, and ended by establishing just what she had been doing one week ago today... We found women could recall details with amazing accuracy, details that I, and everyone else, had assumed they had forgotten."

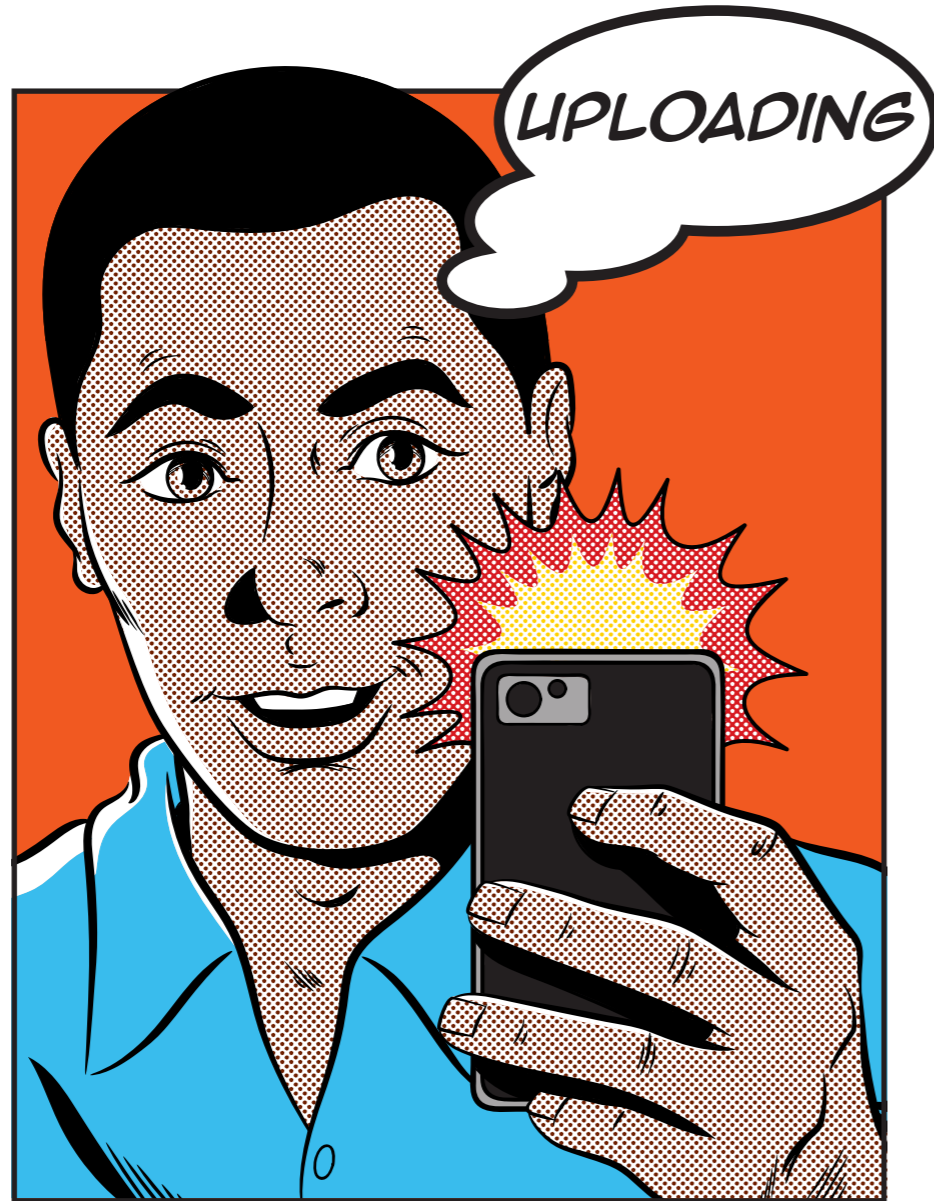
Gallup's interviewing centre was a converted theatre that he had named 'The Mirror of America'. His point was that the brain, properly stimulated, acted as a "tape recorder", in his words. The secret of accurate recall, he argued, was simply suitable stimulation: appropriate questioning to encourage the recorder to play back.

Gallup's mirror, using this method, was likely to be distorted.

“Groundbreaking work on the fallibility of recall was done between the wars. Bartlett established memory is reconstruction, not reproduction”

But we know that; much of the groundbreaking work on the fallibility of recall had been done between the wars. Sir Frederic Bartlett, the pioneer of modern cognitive psychology, had published *Remembering* in 1932, which established that memory is reconstruction, not reproduction. In 1913, Marcel Proust didn't need a cognitive psychologist to tell him that "remembrance of things past is not necessarily the remembrance of things as they were".

The article to which Gallup was responding in 1968 had been



published in the forerunner of the *International Journal of Market Research (IJMR)*. It laid out the industry's best response to what it called 'The Paradox of Memory in Market Research'.

Commercial pressures

This paradox, argued the article's authors, Henry Durant and Martin Simmons, was that techniques that reduced the reliance on memory in market research made its insights more accurate, but they also made the research process less efficient. Hence, commercial

pressures and convenience force us to do research at a different place and time. Out-of-context research is sometimes necessary, but it is rarely optimal.

Durant and Simmons gave an obvious example: a comparison of the relative accuracy of surveys done once a week and once a day. Surveying once a day generated seven times as much data, but experimental methods had already shown that this was far more likely to be inaccurate – Gallup's argument notwithstanding – especially for unremarkable activities, such as

eating dinner. The paradox could be resolved, the authors speculated, if researchers "get the results [they] want by other means", which meant measuring in context. For them, the means did not involve taking respondents by the hand. Instead, they listed the techniques that might improve the quality of the data by measuring closer to the appropriate place, time and mood.

Durant and Simmons explored the potential for what was colloquially known as the 'Dustbin Audit' – passively collecting data from used packaging; they recommended interviewing customers at the point of sale (POS) rather than waiting, which had been introduced to the UK by the authors in 1962; they discussed "the measurement of pupil dilation and its applications", a new scientific insight that had been discussed at an Esomar conference three years previously; and they even noted a paper on an innovative camera – called the DynaScope – which had taken 1.5m pictures of Americans supposedly viewing television advertisements in 1967,

and had discovered that, for 19% of the time, no-one was watching.

Almost 50 years later, market research still struggles with the paradox, and still considers large-scale transactional data, point-of-sale observation – what we have learned to call 'neuro' – and life-logging to be innovative. In the

“When humans recall feelings and actions, we are far from the tape recorders that Gallup assumed. We forget, misremember and create”

trade-off between efficiency and accuracy, clients have often decided that measurement in context, using these technologies, is not worth the expense, the methodological problems or the time.

For Peter Mouncey, editor of the *IJMR* – who chose 'The Paradox' as the first in a series of classic papers that are still relevant today – the

trade-off is often ill-judged: "You have to compromise, but many buyers of research now do not have research experience. They are not educated, informed buyers... We knew, back in 1968, that we weren't measuring what really happens, but the methodological rigour being applied was well thought through, and was in the public domain."

Context capturing

More optimistically, technologies and innovative methods are better placed to allow us to capture context in our data in 2015. We have the opportunity to do a better job of resolving Durant's and Simmons' paradox. However, as the past 50 years have shown, it's easier said than done.

Neurological, psychological and behavioural research has shown us that, when humans recall feelings and actions, we are far from the tape recorders that Gallup assumed we were. We forget, misremember, and create convincing, though untrue, narratives of our lives and experiences. In extreme

LOOKING INSIDE PEOPLE'S MINDS

A combination of eye tracking and skin-conductance response was used by COG Research to try to determine people's 'brain arousal'

"We were doing pure-play eye tracking, which was objective in that it allowed us to see exactly where people were looking," says Rob Ellis, founder of COG Research, "Sometimes, that was incredibly helpful – if you were testing commercials, for example. But we're increasingly getting briefs or requests from clients that say, 'Can you tell us what's actually going on in people's minds when they're having different experiences?'"

Last year, Ellis did work for the Outdoor Media Centre for which the glasses provided real-time evidence of where the gaze was, but no measurable interpretation of the brain activity connected with visual attention – what the person was 'feeling'. "The outdoor industry has always claimed that people are more alert in the outdoor space. Some

evidence of this would be a good thing to provide to the media buyer, who is notoriously tough on claims such as this, and who loves figures."

Ellis contacted Durham University's cognitive psychology team, led by Dr Amanda Ellison, and developed a way to measure people, over a day, using a combination of eye tracking and skin-conductance response (SCR), the technology used in a polygraph. This gave the opportunity to measure their level of 'brain arousal' in the outdoor space. COG has since used it for other clients, including a retailer and Channel 4.

There were two reasons for picking SCR. On the one hand, it was practical to use in context: users could walk around and do their usual daily tasks with just a small sensor on their wrist, rather

than being placed in an MRI scanner, or having electrodes connected to their head. SCR is also directly triggered by the autonomic nervous system – the part of our body that controls the instinctive fight-or-flight response – so it is a quick and accurate measure of arousal.

Ellis admits this is a work in progress: "There are two levels of frustration, one of which is that people love to invent a new technological solution to something, and we're all suckers for that. Often it detracts from the credibility of good work being done. But my big beef isn't with the technology; it's with interpreting how we think this technology works."

As with any neuro or biofeedback technology, the SCR response is not sufficient on its own to determine how the

subject was feeling: only that there was some feeling. The SCR data becomes a way to edit that day's events selectively – to pick out moments when the data signals arousal, and to ask the subject to recall what he or she was thinking and feeling at that point. But the highlights are stripped of their data on gaze or arousal, so as not to feed cues to the respondent. If the subject can tell a story that matches the pupillometry and the SCR data, then this is a credible insight.

Ellis emphasises that for this and similar experiments to succeed, the research design has to include a naturalistic context. "The massive downside to most technology is that it's such an artificial environment. We found clearly, through experimenting, that nothing trumps a naturalistic ecological test."

circumstances, this can lead to completely misleading recollections.

On 22 July 2005, witnesses recalled seeing Jean Charles de Menezes, wearing a heavy padded coat, vaulting the barriers at Stockwell tube station while running from the police, who pursued and shot him. However, CCTV showed a man – wearing a light, denim jacket – pick up a free newspaper, then calmly use his ticket at the barriers and stroll down the escalator.

'Squishy' memory

We know that, with the passage of years, memories – especially the 'flashbulb' recall of important events – can create an entirely new narrative. *NBC Nightly News* anchor Brian Williams was suspended after he 'remembered' being in a military helicopter that was shot down by rocket fire over Iraq in 2003. It would have been terrifying – if the events he recalled had actually occurred.

"Memory is squishy, malleable, ever-changing, and, sometimes, even invented," says Jason Brooks, vice-president of marketing science at Lieberman Research Worldwide, who has a PhD in applied cognitive

psychology. "As market researchers, it is our responsibility to design approaches and instruments that can simultaneously probe squishy consumer memory, without 'jiggling' that squishy stuff so much that it gives us poor information."

For some, this means research that relies on recall is always flawed. Philip Graves, author of *Consumerology*, a book that claims "market research is a myth", says methods that rely on the belief "that asking people questions leads to revealing answers" is not measuring appropriately. For Graves, behaviour

“ For Graves, the only sound contextual evidence is a contemporaneous record of what we do, which is now easier to measure ”

is what matters, because a survey is "not a conversation about evidence, it's a conversation about belief".

His view is that the only sound contextual evidence is a contemporaneous record of what we do, which is now easier than ever to

measure precisely, with ever-increasing volumes of data. The loyalty card replaces the Dustbin Audit; geolocation records where we are; and web analytics provide precise records of online behaviour in real, or near-real, time.

This method of measurement captures accurate context (purchases, location, search records) in the data, but few market researchers believe this, alone, is enough to create rich insight. 'Customer science' company Dunnhumby was collecting big data for Tesco Clubcard before the category had a name, but it doesn't use purchasing data in isolation. Zakaria Haeri is research development lead for its Shopper Thoughts activity, which investigates the value of combining transactional data with survey data – compiled at a later date – from the same sample.

On the one hand, Haeri's investigation supports the criticism of survey data. He argues that it is often internally consistent, but functions relatively poorly when used to predict intention to purchase in the real world.

"For a strong correlation between responses on a survey alone, you'd look for 0.7 and above. But correlate responses and how people behave – generally, we find there's consistency, but it's quite a weak relationship," he says.

"One explanation is that people have a bias against giving paradoxical responses in surveys. If they have said they like the product, they are more likely to say they would buy it later in the survey."

Survey quirks

Nevertheless, Haeri has found that survey data still improves past purchasing data as a predictor of future activity; though, again, not consistently. Haeri's data shows interesting quirks: for example, affluent respondents are more likely



to disconfirm something in a survey that the existing data shows is factually true. It's easier to know this bias exists than to decide what to do about it, other than to use it as a caution about taking surveys at face value.

At Research Now, passively collected data is most often used to prompt and validate, rather than as an end in itself. "I'm a big advocate for measured data; we do a lot of it. But it doesn't tell us why," says Ben Hogg, managing director for EMEA Portfolio Companies. "All it shows us is what's happening. But it

“ Haeri's data shows affluent respondents are more likely to disconfirm something in a survey that the existing data shows to be true ”

encourages people to do quantitative research that they wouldn't have done before, because they are using behaviour data. We can discover what is happening, then we can conduct further research to understand why."

It also gives Research Now confidence that it is measuring the 'what' better, because it establishes relevance using passive measurement, especially tracking mobile behaviour.

Rather than ask panel members if they have done something at a precise location within the past day, passive measurement allows the firm to preselect a panel that it knows fits this profile, without asking would-be respondents to qualify themselves. Data both validates the responses and shortens the survey.

Another way of resolving the

DATA

▶ 'paradox' is to make it easier to capture the survey response close to the time and place, so the respondent's memory is fresh and their mental state is close to what it was at the relevant time.

Timely and accurate

Mobile data collection has made a dramatic difference to our ability to collect information in a timely, accurate way, and is now pervasive in market research.

The *GreenBook Research Industry Trends Report*, released in November 2014, showed that 64% of its respondents used mobile surveys. One year previously, the figure had been 41%. Four out of five respondents said that mobile devices were encouraging them to review the way they collected data.

The combination of the availability of devices, plus the potential to use geolocation as a vital piece of contextual information, makes mobile an important collection tool, says Hogg.

"Rather than rely on what may have happened in a bank or retailer a week ago, we can get subjects as they are leaving, as close to the moment as possible," he says. Research Now does this by building a virtual 'fence' around a location, triggering a survey when a potential respondent crosses that fence.

The mobile device can also shorten and validate responses – for example, people can photograph their cinema ticket and upload the picture. Nowadays, Hogg says, some of the most innovative work in research design work is done in collaboration with software developers, who would not consider themselves researchers, but who understand the interactivity that mobile research demands.

Mobile also provides an important 'life-logging' dimension: for example, a recent project by Research Now, to measure outdoor advertising, used a small sample that was geolocated and tracked, thereby

eliminating entirely the need for recall. "We can see people's journey to work – we don't need to ask 'did you pass through Paddington station today?', Hogg says.

"We can measure what you do at home on your PC, your path to the store, what was in the Facebook feed you checked on the path to the store. We could actually see people responding to the survey in the car park outside the retailer."

At Opinium Research, research director Steve Looney has been using mobile technology to gain instant responses. "On a current project, we are using iPads to get feedback in real time. We are evaluating the display format of outdoor advertising – so not the creative, but the display format: a six-sheet against static digital against moving digital.

“Availability of devices, plus the potential to use geolocation as contextual information, makes mobile an important collection tool”

"The study is relatively simple; we are asking a bank of questions on each display type and then some overall preference questions. We use iPads because this allows the respondents to be in front of the stimulus while answering, and also enables us to use visual stimulus when we come to the overall preference section."

In 2006, Dr Tim Snaith, chief research officer at OnePoint Global, designed a mobile-survey platform for the company that has delivered more than 40,000 projects in 74 countries, for clients including Barclays, Total, Tesco and Nielsen.

Snaith claims that, even for simple surveys (70% of the company's work is still based on SMS feedback), the platform delivers results without a

recall bias, and without the presence bias that skews results when talking to a researcher.

However, he warns that many researchers destroy the contextual insight by overloading the survey.

Asking less

Snaith was inspired to set up OnePoint Global because, increasingly, he saw research as something that was being "done to" respondents. His advice for researchers who want to use mobile technology to improve research quality is: you'll get more by asking less. Don't make your respondents work too hard – use the technology to capture contextual information passively, if possible.

"There's still a temptation simply to move your online survey to mobile," he says. "Respondents are self-trained, so you don't have that cost. But, on the other hand, they'll be less giving of their time, so make it short and sharp."

There are situations, however, in which using techniques that eliminate recall may squeeze out exactly the information you want to capture, argues Steve Ogborn, director at ICM Unlimited.

"As an insurance brand, you might want to know how your customers feel about a claims process. You want to know the lasting impression," he says.

"On the other hand, if you want to understand the claims process – what was said, when it was said – you don't want to rely on recall. But one isn't necessarily better than the other."

Net Promoter Score (NPS) trackers are an example of where both methods have a function, Ogborn says. "There's relationship NPS, a slice of brand health – and that means we are collecting recall; the respondents are not thinking too hard about it because the point is, we want to measure how they feel. The other side of this is touchpoint NPS, when we collect in real time or

READING LIFELOGGING DATA

It might not be difficult but categorising tens of thousands of photos from lifelogging data is laborious, so could an expert system to take over the role of an intern be the answer?

"A lot of what we have been doing in recent years has been dealing with the loss of context from research," admits Dr Bob Cook, director of innovation and inspiration at Firefish. "The research process has evolved to become convenient for the researcher and also for the clients. It happens on appointed occasions, which are in specific places, which makes it very easy to package it up, execute it and sell it. You end up doing research way out of context, so we spend a lot of time talking to people about ideas written on bits of paper and in a viewing studio after work about a breakfast cereal."

In 2012, Firefish did a study for the Internet Advertising Bureau (IAB) called RealView, to understand how people used digital media, but without the problems of trusting their recall. It needed context, Cook says: "We wanted to measure how much screen attention the devices were going to get? How much noise

and everything else is going on around that person as they're commuting? Are they squashed half to death in the Central line? You want to understand all of the opportunities, which is all of the moments our subject has a digitally-enabled device on them. So you would ask them to catalogue, in detail, every single moment of their day, or at least every interaction they have, and it's very obvious it's a completely impossible task for them to do."

Therefore 20 subjects wore cameras to capture four days of their lives. Subjects checked phones during romantic dinners, for example – the sort of behaviour they would be unlikely to recall to a researcher.

But, to discover this, there were tens of thousands of photos, which had to be categorised: if a mobile device was in the picture, it represented device use. It was a long and boring process but not difficult. In this case, Firefish's intern did the job, but the struggle

made Cook wonder if it would be possible to build an expert system to 'read' lifelogging data. He turned to Cathal Gurrin at Dublin City University, the leader of a team developing cutting edge visual algorithms that can 'read' pictures.

Does it work? "It's significantly less effective than the human eye at the moment," Cook admits, "It depends what you're looking for. Objects like PCs and smartphones and tablets look fairly similar so they're quite easy for the computer with the algorithm to learn."

The trade-off is that the results are less accurate, but potentially have far more impact, because the volume of data ceases to be a constraint. Longer, larger lifelogging projects would be possible, analysed in near real time automatically.

"I don't think it'll be in the next year, but in the next five years the algorithm recognition technology we're using will become genuinely fit for purpose," Cook says.

near-real time. We can get high volume, closed-loop recovery, and it helps with processes.

"But just because we have real-time information, it doesn't mean we understand what's going on. We still need to do longitudinal research to understand the key drivers that affect that score.

"A lot of the time we collect real-time information, but clients don't know how to use it because they don't know why the score is moving."

Also, in specific situations, recall can be a much better

predictor of intention to purchase, says Ogborn. For example, a financial services company that uses ICM to track customer experience discovered that, in the first two years of a customer's relationship with the firm, touchpoint NPS successfully predicted the firm's ability to cross-sell or upsell. After two years, it lost its predictive ability. It seems the customer's accumulated recollections of the brand's performance overwhelmed the relationship. In the long run, potentially unreliable recall trumped



**Up close
& personal**

**Research Now.
Powering your
decisions**

► ‘objective’ measurement as a business driver.

Extract meaning

Collecting data about what respondents choose, do or buy, using passive techniques, is as old as the Dustbin Audit – the Attwood Survey first started rooting through the bins of British consumers every week in 1948; the struggle to extract meaning from the data has been going on just as long.

Collecting data using biofeedback presents methodological problems that are at least as difficult, because we need to both measure the phenomenon and interpret it.

Measurement is a technical problem, and interpretation risks reintroducing bias and post-rationalisation through the back door. For example, we can measure blood flow to areas of the brain, or conductance (see Pg35); however, understanding what that may imply is much more complex.

The problem, again, is one of inferring intention from actions. In 1949, the philosopher Gilbert Ryle first wrote of the “ghost in the machine” as a way to criticise our inadequate knowledge of how the mind and body interact.

In market research terms, we can measure a person’s brain activity, sweat, pupil dilation or where they are looking and for how long – but this is not the same as knowing what this activity means.

Visuality has tested shopper experiences and POS for, among others, Morrisons (for which it filmed, processed and analysed 10,000 full customer journeys), Asda and GSK. Its culture has always been to observe shoppers in context – stopping them to ask their opinions during the shopping trip, rather than relying on recall without context. It has recently started to use eye tracking, but with caution, says research director Nicola Scrafton.

“It needs to be used alongside other methodologies and also the

experience of the researcher. I’m very sold on that. In the wrong hands, with limited knowledge, it can be quite misleading,” she says,

“You could look at fixations, and at the communication on a piece of point of sale, something that seems to be attracting a lot of attention. It could hold someone’s fixation for a long period of time. When I say long period, I’m talking two, three seconds – or people could keep going back to it. You have to interpret whether they find it

“Using technology to capture context can bring us closer to the consumer or respondent, but never eliminate subjectivity completely”

totally engaging, or it’s confusing and they are trying to work out what it is. They are two very different messages that you could feed back to the client.”

While eye-tracking glasses provide a rich source of data, for Visuality, this is both to complement insight and originate it.

“Our bank of shopper experience and knowledge

was very important to us when taking on eye tracking, because we thought about how to use the tool, rather than just thinking about how to generate heat maps,” says Scrafton.

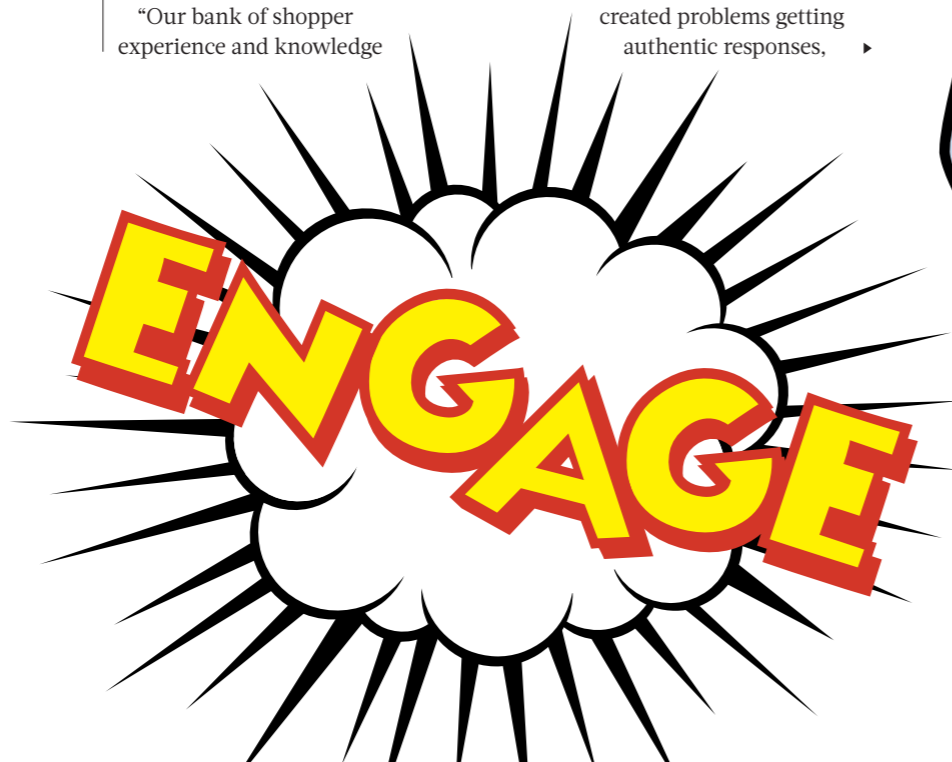
Using technology to capture context can bring us closer to the consumer or respondent, but never completely eliminate subjectivity: even the choice of what to measure is subjective. Therefore, the jump from the ‘what’ to the ‘why’ will never be straightforward.

In a series of deliberately provocative videos posted on YouTube in 2014, Thinktank partner Sabine Stork demonstrated the tool that she believes is still best for measuring people in context: conversation.

Thinktank researchers set up some tables in Brixton Market, and then offered glasses of wine to passers-by, who often sat down in groups of two.

The method deliberately ignored the usual researcher-respondent relationship. It was a free-flowing conversation in which the researcher was an equal participant. For this group of young people, Stork says, the formality of the research process – which

created problems getting authentic responses, ►



► especially when researching the opinions of young people – was cut back. In the videos, one interviewee gleefully reveals his deep loathing of some of his Facebook friends using a four-letter word.

Another says that she would “rather be dead than be a man”.

Stork says the experiment was a response to what she sees as an over-emphasis on data and structure, arising from the use of technology in research.

“There is this emerging consensus that we need to get as close to consumers as possible, which I

“I was reading about how we should stop asking questions and just observe. That seems perverse – treating consumers as lab rats”

completely and utterly have no problems with,” she says. “But I was reading a lot last year about how we should do this by stopping asking questions completely, and we should just observe.

CONTEXT

“To me, that seems perverse; it means we are treating consumers as lab rats.

“I was wondering whether it all came from the data people, who really don’t want to engage – they are much more comfortable observing people from the safety of a computer screen.”

Interpretation

The evolving technology that allows us to measure accurately – but also

efficiently – in context, may often deliver insight quickly, inexpensively, and at scale.

However, Stork warns that data-acquisition tools will only be one aspect of intimate, context-aware research.

“The danger is, I think, that there is a suggestion that the data doesn’t lie,” she says, “but there’s always going to be the question of who’s interpreting that data.” ■

FACIAL CODING

In 2014, Nickelodeon wanted to measure its viewers’ screen life, their perceptions of ads, and their level of engagement. One problem: the viewers were kids

How would children articulate their thoughts on advertising? “The majority of existing research methods require children to think and categorise their thoughts, which can result in a claimed post-rationalised view, rather than a true representation of how they feel,” says Alison York, research director, Nickelodeon UK & Ireland.

Nickelodeon turned to John Habershon, director at Momentum Research, who had developed a facial-coding

application for testing response to video in adults. “Asking people how they feel about an ad is unreliable. People can’t accurately recall their feelings; facial coding measures the emotional response as the ad is being viewed,” Habershon says.

Momentum records the tiny, split-second emotional responses, which are the most powerful signals of when the children are engaged. “In its computerised form, this is a well-established technology in

the US, used for ad testing by many of the major advertisers,” Habershon says. “This is done by webcam, and analysed using computer algorithms, to identify key points on the face and track their movement, to identify an emotion.”

Instead of computerised recognition, Momentum films smaller samples on HD and analyses them using trained human-facial coders. This means the client can determine which emotions they want to look for.

Nickelodeon interviewed pupils at their primary school, asking two at a time to watch ad breaks on a laptop. At the beginning and the end, they were asked about their viewing habits.

The technique had never been tried with children before, but Habershon says it “was surprisingly easy... the client knows precisely when content is engaging the viewer and when they switch off. They can see at a glance how successful each ad is in comparison to others.”



Up close & personal

Permission to get closer to consumers is no longer an obstacle. Research Now connects you to your audiences’ behaviours & opinions with our suite of permission-based digital, mobile and social technologies.

Research Now. Powering your decisions

www.researchnow.co.uk