

Behaviourlab

An Innovative Method for Predicting Markets and Improving Strategy

Executive Summary

This position paper seeks to distil twenty years of experience predicting consumer behaviour across a range of sectors and to make the case for using our proprietary randomised controlled trial (RCT) approach, Behaviourlab. Specifically, we argue:

- 1. Bad Research has an Opportunity Cost: Companies stand or fall by the quality of their decision-making. Clearly there's substantial opportunity cost if you adopt the wrong strategy. As such, you're not saving money if you pay less to get inferior advice.
- 2. Predicting People is Difficult: There are two main challenges to predicting behaviour. First, you can't ask people directly. They either don't know or will 'improve' their responses. Second, if you change the decision context, it dramatically changes their behaviour.
- 3. Traditional Research Methods Struggle: These two difficulties undermine the accuracy of traditional market research. Focus groups and surveys are based on unreliable self-report. Conjoint and max-diff are built on atypical decision environments.
- 4. Live Trials are a Partial Solution: Live trials, such as A/B testing or shelf trials, avoid these pitfalls because they involve real-world behaviour in the actual decision environment. But be careful as they are often poorly executed (e.g. not blind or missing a control).
- 5. But Live Trials have Limitations: Live trials also have inherent limits. They can be expensive and take time. Likewise, there are questions you simply can't test (e.g. big touchpoint alterations, competitor economics, unbuilt feature performance or dramatic price changes).
- 6. Behaviourlab Plugs Various Gaps: Behaviourlab is an RCT method that tracks people's behaviours as they navigate a replica of the real-world task. It therefore solves the two main problems of traditional research and circumvents the limitations of live trials.

If you're new to running experiments, this brief can be used as an introductory guide. If you're an existing practitioner, we've included observations on best practices and how to avoid some common pitfalls. But most of all, we hope that this note will persuade you of the role Behaviourlab could play in delivering better commercial strategies. In these times, when the real world is uncertain and fast changing, this is a way to simulate and foresee how consumer behaviour will respond.



1. The Cost of Poor Decision-Making

"

Lots of companies don't succeed over time. What do they fundamentally do wrong? They usually miss the future."

Larry Page

Figure 1: Common Forecasting Problems

How will my customers or prospects behave when I...



PROPOSITION

...change prices or run a promotion? ...add or remove product features? ...create a new bundle or change my range?

...launch a new product or service? ...am faced with a new competitor? ...have to implement this regulation?



...re-architect that customer journey? ...change that journey's UI or content? ...shift our approach to brokered sales? Every business tries to predict and plan for the future. But, as Larry Page points out, many fail to anticipate how customer needs and the market landscape will evolve. Why? Some of these failures are simple hubris. But where management at least attempts to inform its decision-making, our contention is that the data they use is often fatally flawed.

Companies are decision-making factories. Every day they make thousands of choices extending from a myriad of small ones through to a handful of larger tickets. Our analysis shows how shareholder value creation depends as much on the former as the latter. Intuitively, great businesses are built on optimising the commercial minutiae, as well as making the right macro calls.

This is the management philosophy of 'continuous improvement', or kaizen, an approach pioneered and popularised by Toyota. Kaizen means operating rapid test and learn cycles perpetually across all aspects of the business. This includes resolving some of the common consumer forecasting problems in Figure 1. In this paper we will argue the benefits of such trials whilst highlighting the limitations of making them 'live'.

2. The Problem of Predicting People

"

Listening to uninformed people is worse than having no answers at all."

Ray Dalio

We contend that there are two main flaws that constrain the accuracy of traditional consumer research and make RCTs preferable. First, trials are built on participant behaviour. Second, they reflect the real-world decision environment. Here's why that's important.

Behavioural: Self-Report is Bunk

Traditional market research, like focus groups and surveys, relies entirely on people accurately introspecting their motives and beliefs. However, people are extraordinarily bad at both explaining their past behaviour and predicting what they'll do in future. People post-rationalise, tend towards more socially desirable responses and, fundamentally, just don't know why they do what they do. They either can't or won't tell.

Figure 2 illustrates this point. People were asked to state their preference for ethically sourced goods. Their actual purchase behaviour was then observed. The self-report bears absolutely no relationship to actual shopping behaviour. To understand people's choices, you simply cannot ask them outright. Some clients like focus groups because they help generate ideas, which is fine. But some clients use focus groups to filter ideas when they are too inaccurate and unreliable for that task.

Figure 2: The Divergence of Self-Report and Actual Behaviour





Immersive: Context is King

At the annual brand review of a leading British grocer we sat in on the store-choice conjoint debrief over many years. The output was utterly confusing – the importance of stocking organic produce was over-stated, convenience was under-stated, and other features lurched around counterintuitively year on year. Sure, this research was cheap, but the client wasn't saving money because, as Ray Dalio says, the opportunity cost of getting the wrong advice is large.

Why were those conjoint studies so misleading and unstable? Because people's behaviour is remarkably susceptible to context. People can be nudged by changing the default options, re-framing decisions and increasing task complexity. Figure 3 shows an example. People's credit card statements either make it easier or harder to find the minimum payment. Foregrounding that small number re-frames the decision. The average payment when the minimum is less noticeable is 62% higher. These context effects are why clients often report unhappy outcomes with max-diff and conjoint. These methods of extracting people's preferences are entirely alien to any decision environment a customer would ever encounter in practice. Hence, the resultant data is heavily distorted. Moreover, the approach is actually quite limited. For example, you can't test path dependent choices, journey re-designs and other important preference related problems.



Figure 3: The Malleability of a Credit Card Payment

3. The Partial Solution of Live Trials

"

If you can look into the seeds of time, and say which grain will grow and which will not, speak then unto me."

William Shakespeare

Because they are based on consumer behaviour in a real-world decision environment, live trials don't suffer from the serious shortcomings outlined above. In this section we discuss live trials in more detail, along with our observations on best practices drawn from client experience.

Anatomy of a Live Trial

Live trials are a type of RCT. The appendix explores how RCTs work, but a key benefit is that they identify causal direction and precisely measure the size of any effects. This contrasts with the statistical techniques used by data scientists that can't resolve causality and therefore unavoidably over-state predicted effects by an unknown amount.

A/B Testing is the most common example of a live trial. But there are others, such as regional advert tests and retail store or shelf trials. All these different types of live trial involve the five main steps described in Figure 4. Figure 4 also includes a real-world example to help illustrate the steps and generally bring the methodology to life. An American multinational wanted online customers to upgrade their laptop specification at checkout. Uptake was low so a new screen with a tailored recommendation and voucher was created. The design was targeted at higher value gaming and business markets. After three months of randomly allocating traffic between the two screens, the new design had raised conversion 12.0%, revenue per customer 11.3% and skewed sales towards higher margin products.

Figure 4: Running a Live Trial

Step	Description	A/B Test Example		
1. Scope	What activity are we seeking to optimise?	Laptop customisation stage of online sales process		
2. Conditions	What changes to the activity should we test?	Existing screen vs. tailored upgrade recommendation		
3. Outcomes	What are the outcomes of interest?	Conversion rate and revenue per visitor		
4. Analysis	How can we tell if a change has had an effect?	Large trial volumes mean no statistical analysis is needeo		
5. Forecast	What's the implementation business case?	Both outcomes must increase by a pre-agreed amount		

Live Trial Limitations

Great as they are, live trials also have their limitations. Here we discuss some common execution errors and some more fundamental constraints that preclude using live trials for a range of commercial problems.

COMMON EXECUTION ERRORS

Over the past twenty years we've witnessed clients struggle to interpret poorly designed live trials as well as made our own fair share of mistakes. Here's our top ten blunders to try and avoid:

- 1. Subscale: The trial isn't run for long enough or on a large enough sample to find an effect.
- 2. Unrepresentative: The study uses an atypical sample or environment (e.g. a trial store).
- 3. Un-Blinded: Customers (and staff) know they're in a trial and behave differently.
- 4. Recursion: The results are confusing and simpler, better designed trials are required.
- 5. Fruitless: None of the conditions are impactful because they are too timid or uninspired.
- 6. Exogeneity: Something important, like rising vs. falling markets, isn't part of the trial.
- 7. Uncontrolled: There's no control condition, so the benefits vs. the status-quo are untested.
- 8. Confounded: Randomisation is either missing or uses a poor proxy (e.g. time of day).
- 9. Weak Metrics: The tracked outcomes exclude some key commercial effects (e.g. CSat).
- 10. Politics: The trial flushes out internal strife over commercial trade-offs and becomes mired.

INTRINSIC SHORTCOMINGS

Whilst the above problems are avoidable, there are also some inherent constraints that mean live trials aren't suitable for every type of commercial question. Those limitations are shown in Figure 5. It was to address these limitations that we developed Behaviourlab.

Figure 5: Limitations of Live Trials

v Markets straints	You can't test people who aren't customers or prospects The functionality to field certain
straints	The functionality to field certain
	changes isn't yet available
lability	The systems only have capacity to test 5-10 conditions
ed	The trial is delayed by other roadmap priorities
-Averse	You can't test anything too radical or controversial
fidentiality	You might tilt your hand to competitors or suppliers
r-Gaming	You can only test your operations, not competitors' or brokers'
comes	You're limited to what is already captured by your systems
se	Fieldwork has lots of noise and other effects (e.g. sunny day)
	ability ed Averse fidentiality r-Gaming comes

4. Our Behaviourlab Approach

"

If you double the number of experiments you do per year, you're going to double your inventiveness."

Jeff Bezos

What is Behaviourlab?

Behaviourlab is our solution to harnessing the power of RCTs whilst overcoming the limitations of traditional live testing. It is an immersive, online behavioural approach that replicates, as closely as possible, the real-world environment that people experience in practice. Figure 6 shows how Behaviourlab tackles the five main steps of designing an experiment.

Figure 6: Running a Behaviourlab Experiment





Behaviourlab's Advantages

Behaviourlab is a purpose-built, immersive, and flexible method that can be used to study a wide range of consumer behaviours. As such, it confers various benefits, particularly compared to existing research techniques. Based on client feedback, these are the five most useful:

Accurate: To quote a client, it's "extraordinarily accurate". Because Behaviourlab replicates the real-world decision environment and is based on behaviour, not self-report, it avoids all the distortions cited earlier. Moreover, it's a cleanroom that excludes all the noise that comes with live trialling.

Omnipropriate: Behaviourlab is custom-made and can be extended to questions it's impossible to live trial. For example, you can test extremes (e.g. big price or feature changes), things that don't exist (e.g. dev roadmaps, new products), or competitor economics (e.g. what they can afford to bid).

Informed: Behaviourlab has an explicit condition creation stage which draws on existing behavioural research. The extensive list of effects identified in the literature helps both generate better ideas for testing and focus on those with a higher chance of success.

Panoptic: Whilst conjoint only measures preferences and live trials measure only what's observable with existing systems, Behaviourlab can be used to capture all aspects of participant behaviour, together with their knowledge, perceptions and relevant personal details and real-world behaviours.

Rapid-Scalable: Since it's not run on live systems, Behaviourlab can be executed quickly and won't be delayed by competition for resources. Being cheaper than live trials, it can be used to test more conditions. Taken together, Behaviourlab delivers the kind of rapid iterative cycles needed for kaizen.

5. Behaviourlab Case Study

To bring the approach to life, this section details a Behaviourlab case study. A Big 6 energy firm was losing 10k accounts per week through a combination of government policies designed to boost churn and cost-advantaged small suppliers. The brief was to identify a strategy to stem the flow. We ran a Behaviourlab experiment to find the best solution.

We re-created a price comparison website environment. Figure 7 shows an example screenshot. The conditions involved varying brands, prices, savings, features and so forth. To assist with the fidelity of the experiment, the task was tailored to each participant, based on their incumbent provider and current spend. Provider choice and likelihood of switching were the two main outcomes, though other variables were captured, such as decision latency, post-purchase product knowledge and so on.

Figure 7: Example Screenshot of the Behaviourlab Experiment

Selected fuel type Gas and Electricity ased on the details you have given us we estimate you spend £996.00 valiable tariffs in our database, not including your own. In some cases			Your current spend £996.00 per year per year. The following is a list of the best you may already be on the best available tariff:		
Provider	Tariff Features		Cost	Saving	
irst:utility	Variable	No minimum contract length Paperless billing	£870.86 per year	£125.14 per year	Select O
eon	Fixed	24 month contract No cancellation fees	£880.52 per year	£115.48 per year	Select O
GreenSTAR energy	Fixed	12 month contract £30 cancellation fee	£891.67 per year	£104.33 per year	Select O
	Variable	No minimum contract length Paperless billing	£892.58 per year	£103.42 per year	Select O
Reitich Gas	Fixed	24 month contract No cancellation fees	£901.37	£94.63 per year	Select

Figure 8: Rank Elasticity of Demand



The two main outcomes were statistically modelled (specifically a multinomial logistic regression and ordered probit, with experimental conditions as independent variables and other inputs, such as demographics, as covariates). The models were then calibrated to real-world data to produce forecasts like Figure 8. As providers cut their prices their acquisition rates increase. But this effect isn't smooth. Conversion rate is driven more by rank than absolute price – another example of a strong framing effect.

The work quantified other effects, such as the role of incumbency, brand, tariff features and so forth. Based on these findings, the client launched a white-label, onlineonly tariff, with stripped down features and aggressive rank-focused pricing. The tariff was designed to target proactive switchers and minimise cannibalisation of the existing, reactive, customer base. The new tariff performed as forecast and, as shown in Figure 9, reversed the client's weekly account losses to stabilise the base.

Figure 9: Account Acquisitions and the New Tariff Launch





Appendix: What is an Experiment?

Experiments, or randomised controlled trials (RCTs), are the scientific gold standard. Dating back to 1747 and James Lind's discovery that lemons cure scurvy, RCTs are now a regulatory requirement in medicine and standard practice across the biological and social sciences. Their great merit is that they definitively prove the impact of making a change, as in the case of administering a new drug.

As shown in Figure 10, RCTs achieve this by randomly allocating participants into groups. One group, the control, continues as before. Other groups experience alternative treatments, in Lind's case lemon juice supplements. Outcomes of interest are recorded for both groups and compared statistically. The control group sailors didn't recover, but the lemon juiced sailors did. Thus, an RCT establishes a causal relationship between an innovation and an outcome. And it precisely measures the size of the effect. Because of the random allocation, the change in a treatment group relative to the control group is the only thing that's different. Hence it is the only thing that could have caused the different outcome. And the size of the observed change is what will occur if we implement that change at scale.

Over-estimating a strategy's real-world impact is a common commercial disappointment. Businesses often rely on the correlative statistics used by data scientists where causation is unclear. So, the size of an observed effect is always going to be muted in practice. People who use a service more might be less likely to churn. But incentivising usage will have less retention impact than forecast since cajoling non-adopters to use the product is different to noticing that un-incentivised fans are more loyal.

Figure 10: The Stages of a Randomised Controlled Trial





Who are Dectech?

Dectech provides the most accurate and best value forecasts available on how people will behave in new situations. As you can tell, we are enthusiastic proponents of doing more experiments. Founded in 2002, we ran our first Behaviourlab study a few years later. We are based in London and staffowned. We regularly publish articles and briefs. Follow us on social media to receive them or stop by our website to sign up for email notifications.







Contact us

Telephone: +44 (0)20 7193 9812 Email: enquiries@dectech.co.uk

Network Building 97 Tottenham Court Road London W1T 4TP

www.dectech.co.uk



