User Conversion X Qubit.

What type of experiments have the biggest impact?

It is perhaps a rather candid and subjective view, but experiments that address usability, in other words facilitate an existing behaviour, do not have as great an impact as those that willingly seek to change a behaviour or attitude.

January 2021

In association with

Qubit.

Report Overview and Purpose

Whilst the purpose of an experiment is to prove or disprove a hypothesis, it inherently has commercial gain associated with it. A "winning" experiment - although it is rarely that binary - should drive significant change to ultimately drive commercial KPIs.

Subjectively, the current state of the industry, either through spurious case studies or wider discussions, highlights that commercial gain can be achieved through the smallest of changes; be that a 'back to top' button or a 'sticky add to cart' button.

These types of experiments lack impact because they don't seek to actively change a user behaviour or attitude, instead, they facilitate an existing behaviour (what we are noting as a "usability improvement").

At User Conversion, we undertook a piece of research on some 789 experiments that identified twhat types of variable change, as controlled experiments, have a demonstrable impact.

1 | Those experiments that were rooted in an anxiety had x34 times the impact of usability improvements

2 | Those experiments that were rooted in motivation had x100 times the impact of usability improvements

Qubit, similarly, undertook research on some 6700 experiments, finding that 90% of experiments had an effect of less than 1.2% on revenue, noting that most simple UI changes are ineffective.

After identifying similar behaviours, Qubit and User Conversion decided to collaborate to write this guide. The objective, of which, is to help guide and educate on which types of experiments have more of an impact on commercial metrics.

1. Do your experiments address an actual user problem?

When you're running an experiment, a question for you; why are you running it? A hypothesis always has a "why". It's a purpose, belief or an assumption that if you change a variable, you will see a resulting expectation or action. Whether it's written down or not, and we would recommend it is, your hypothesis should always contain a why.

Our hypothesis structure at User Conversion looks like this.

Whilst it doesn't address the why within the hypothesis statement, you'll note that it does ask for it in a different section. This ensures that all hypothesis have a purpose behind them - or evidence aligned to why the original hypotheses exists.

1.0 Summary

1.1 Hypothesis

We believe that <<Change>> for <<User Segment>> will result in <<Primary Metric>>

1.2 Why?

Using <<Research Methodology>>, we have observed <<User Problem - sell the user problem, really highlight why it's a good idea>>

1.3 Objectives and KPIs

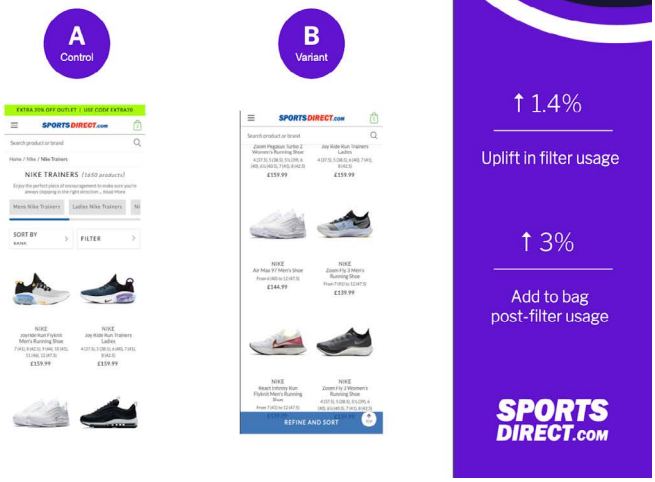
This test is aligned to our quarterly objective to <<Objective>> and will be measured by <<KPI(s)>>

Here is an example of an experiment that doesn't have a why or purpose.

Filter prominence

Increasing filter usage by making filters more prominent

Experiment worth up to £42k per month incremental revenue



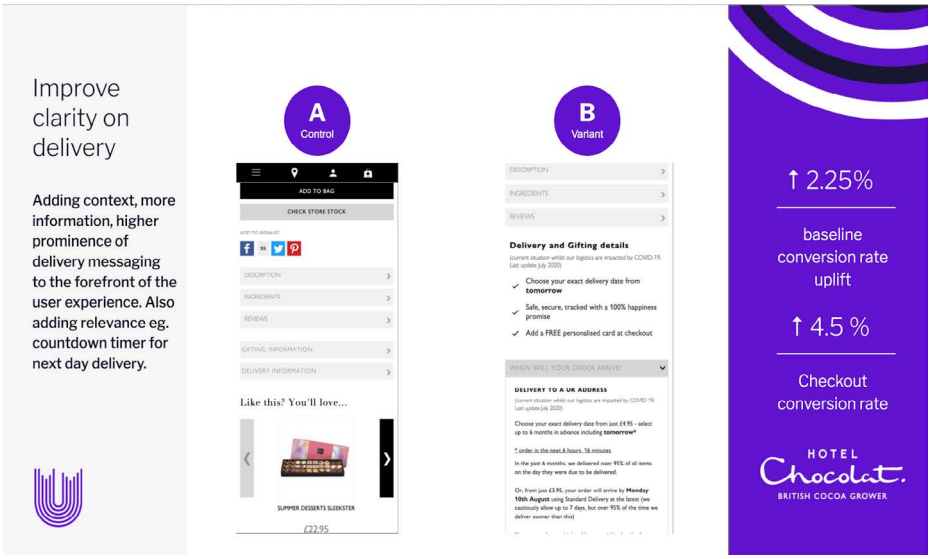
They are generally known as "best practice" and in some instances, teams assume they should just be "done" rather than "tested". Yes, the results suggest an increase, however, one could argue that a natural increase in exposure to 'filters' would create an uplift in its usage anyway. This experiment didn't impact behaviour enough to create a significant commercial impact.

Ultimately, it's a question of priority. Which of your experiments holds a greater priority - or we should ask - which of your user problems holds the biggest opportunity?

1. Do your experiments address an actual user problem?

The following experiment was one rooted in an anxiety - specifically, the lack clarity of delivery cost and convenience to the user. Not only was the impact greater (yes, it's a completely different experiment with different metrics), but it drove an increase in baseline conversion rate.

Changing baseline averages such as conversion rate demands a significant shift in a series of behaviours which, in turn, adds impact.



It's vital to always align the solution with a user problem, otherwise, we would question, why are you testing?

"Just the way every problem needs a solution. Every solution also needs a problem. At Frasers Group, alongside User Conversion, we believe if we don't tie our solutions and problems together then what and who are we designing for?"

Perri Savage, Senior Digital Design and UX Lead, Frasers Group

2. Do your experiments address a user anxiety or a motivation?

There are three categories of changes that you can make on a website. Two of these categories mould and shift user behaviour, and one of them does not - at least not significantly. These three changes are usability, anxiety and motivation. Only the latter two address a core user problem meaning, by solving it, we can see a direct change in user behaviour.

The 3x types of changes:

1 **Usability.**
A change that supports a user to complete a task or action. For example, removing navigation links within a checkout flow to create a closed checkout

2 **Anxiety**
A change that addresses a user anxiety; removing a roadblock to allow them to complete a task or action. This might be a test to pricing or the clarity of content for a certain uncertainty like delivery. For example, adding trust logos to reduce the anxiety of payment cautiousness

3 **Motivation**
A change that facilitates or plays off against the motivation of a user to persuade them to complete a task or action. This might be a change to the proposition of product or new feature. In the above examples of checkout, this might be using pyschological techniques. For example, scarcity or urgency to push users through a checkout process

From our experience, the majority of experiments tend to be usability-focussed experiments - not that there's much wrong with that, but those are the type of experiments that don't make too much of a difference to the journey of the user because they don't inherently affect the behaviour or the attitudes of the user.

When we talk about anxiety i.e what prevents a user from taking an action, and motivation i.e what persuades a user to take action - they are behavioural or attitudinal changes for the user journey. Pricing. Proposition. Product. Features. Journeys. Content. They actively encourage an action, opposed to a usability change which passively facilitates an existing action.

2. Do your experiments address a user anxiety or a motivation?

We undertook research on some 789 experiments that identified what types of solutions, ran as experiments, have more of a demonstrable impact over others.

Those experiments that were rooted in an anxiety **had x 34 times the impact of usability improvements**

Those experiments that were rooted in motivation **had x 100 times the impact of usability improvements**

For example, at User Conversion, to ensure that all our tests are focussed on addressing user anxieties or facilitating a user motivation, we request that all experiments are categorised as such (which is how we were able to create the data sets above).

Theme of test	Usability / Reduce User Anxiety / Increase User Motivation
Primary Metric	What is the main goal of the idea? This should always be the same as your hypothesis
Bahavioural Metrics	• What are the things you want to find out from this test? • If you were to list 10 or so things you wanted to learn within the final analysis, what would they be? • Are there any segments to consider eg. device, user type?

Similarly, Qubit published a paper[1] in which they found that "Of 6700 experiments, 90% had an effect of less than 1.2% on revenue. Most simple UI changes are ineffective. Of the 29 common categories of treatment included in this paper only 8 have a greater than 50% probability of having a positive impact on revenue per visitor."

Those experiments that did make an impact were those that were rooted in psychological and emotional behaviour - scarcity (+2.9%), social proof (+2.3%), urgency (+1.5%) - opposed to those changes that are simply 'variable changes' like colour (+0%), buttons (-0.2%), calls to action (-0.3%).

Qubit (see table below) has provided us with recent scarcity and social proof experiments that have positively impacted key KPIS including revenue per visitor and conversion rate.

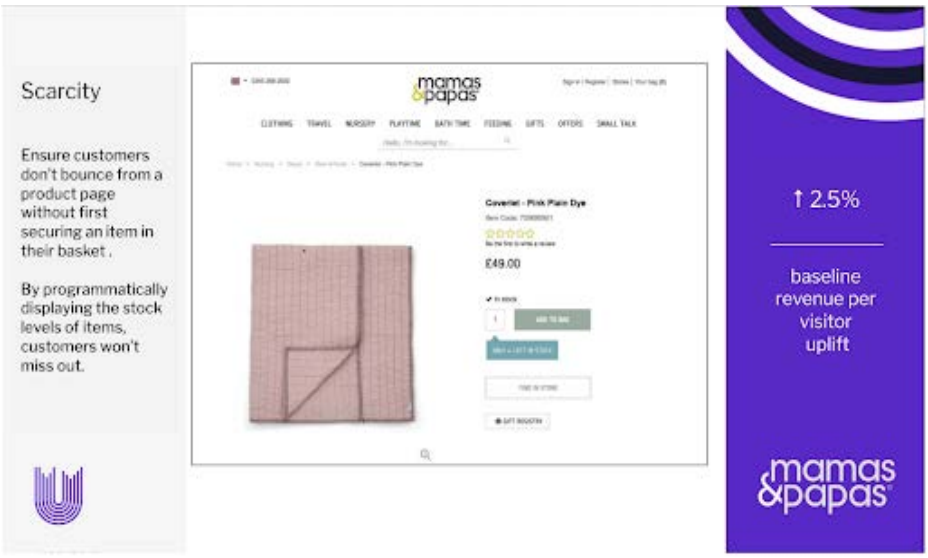
[1] https://drive.google.com/a/userconversion.com/file/d/1FnSCrvLezC8scDmGZHdyLcNVVNZyVvxt/view?usp=sharing

2. Do your experiments address a user anxiety or a motivation?

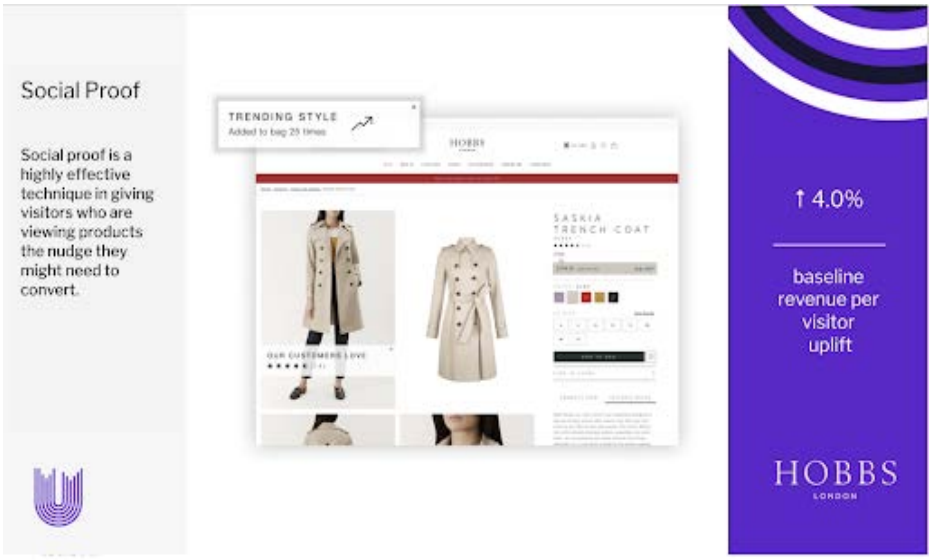
Treatment	Uplift mean	Uplift s.d.	Uplift probability	Median impact	Number of treatments
Scarcity	2.9%	2.8%	84%	38%	125
Social proof	2.3%	2.5%	82%	63%	119
Urgency	1.5%	2.8%	70%	36%	119
Abandonment	1.1%	1.9%	71%	18%	105
Product recommendations	0.4%	0.5%	76%	74%	119
Welcome message	0.2%	0.6%	64%	44%	78
Page redesign	0.2%	0.9%	59%	67%	83
Banner	0.1%	0.3%	63%	44%	212
Popup	0.0%	2.0%	50%	34%	91
Colour	0.0%	0.4%	49%	81%	81
Nudges and pointers	-0.0%	0.3%	48%	44%	105
Resizing elements	-0.0%	1.1%	49%	85%	36
Filters	-0.0%	0.9%	48%	57%	126
Upsell	-0.1%	0.6%	41%	49%	99
Product badging	-0.2%	0.8%	42%	64%	39
Buttons	-0.2%	0.4%	33%	75%	197
Image	-0.2%	0.4%	34%	40%	105
Free delivery	-0.2%	1.3%	44%	50%	65
Navigations	-0.2%	0.7%	35%	62%	216
Search	-0.2%	0.3%	20%	60%	219
Default setting changes	-0.2%	2.0%	45%	50%	58
Landing page	-0.3%	0.9%	36%	39%	55
Calls to action	-0.3	0.5%	24%	71%	172
Back to top	-0.4%	0.3%	12%	78%	54
View all	-0.7%	2.2%	36%	34%	30
Sticky navigation	-0.7%	1.7%	32%	45%	40
Mobile search	-1.0%	0.5%	5%	33%	30
Weather	-1.1%	0.9%	13%	43%	27
Mobile navigation	-1.7%	1.9%	17%	30%	33

[1] https://drive.google.com/a/userconversion.com/file/d/1FnSCrvLezC8scDmGZHdyLcNVVNZyVvxt/view?usp=sharing

In the first example, Mamas and Papas wanted to avoid customers missing out on products they had previously demonstrated intent for. For any returning visitor who had viewed a product in the last 60 days, Qubit deployed a pop-up on the homepage that highlighted any previously viewed products that were now low in stock. While simple scarcity experiments display the stock levels to all visitors, Mamas & Papas incorporated an element of personalisation that drove meaningful impact.



In the next example, Hobbs incorporated social proof in its abandonment layer. While social proof, a behavioural heuristic that leverages the 'wisdom of the crowd' to encourage purchase, can be used at various stages of the user journey including the PLP and PDP, Hobbs experimented by incorporating social proof to visitors further down the purchase funnel. Qubit deployed a message to abandoning visitors that indicated that items in their basket are selling fast, thereby highlighting the fear of missing out, to great success. Both examples demonstrate experiments that facilitate a motivation to purchase.



3. Are your experiments too isolated?

Within the industry we have respective jargon that can confuse, isolate and even detract from the original purpose. Websites are carved up and often spoken about in templates. Homepage, PLP, PDP, Basket, Checkout. This isn't how users navigate or purchase and it's also not how cohorts or segments of users shop, so it begs the question, why do we use it too?

The best way to describe this is "if you take a generic or isolated approach to optimisation, you'll only get generic or isolated results."

Templates are sometimes the effect, not the cause. Their individual problems are generic in nature, and aren't fit for a personalised approach where real impact can be made. In other words, user problems and motivations transcend templates, they aren't isolated to them.

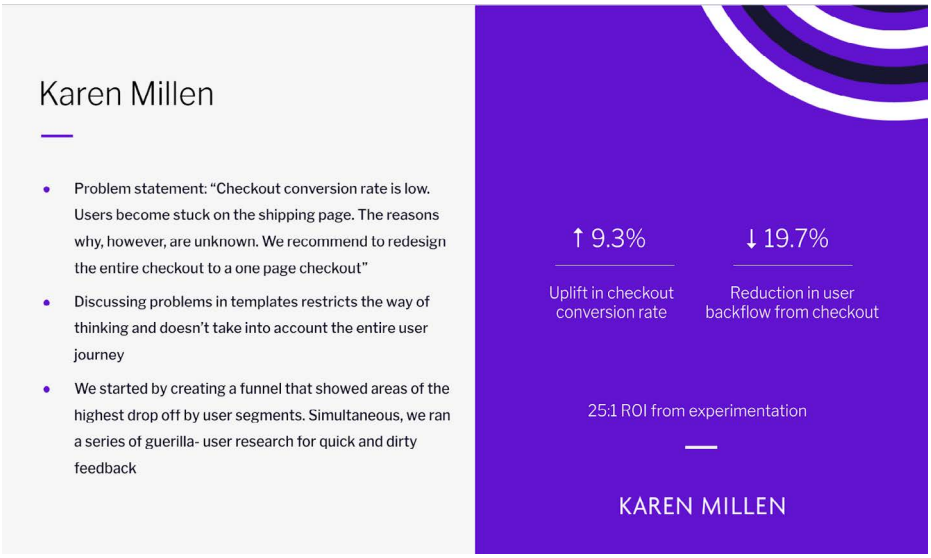
"The guys at UC didn't talk about a generic approach, but one that was personalised to our audience and specific to their needs and behaviours. The result? An immediate uplift in some of our most valuable segments"

Neil Sumner, General Manager, eCommerce, BrewDog

In the example of Karen Millen, the highest exit rate in the checkout funnel was the shipping page. When analysing why this was the case, it was identified that the unexpected shipping costs resulted in a 'shock' factor and therefore more users exited at this page.

This is an example of the shipping page i.e. an isolated view of a template, being the effect not the cause.

After setting user expectations effectively prior to checkout, conversion rates increased as checkout exit rates dropped.



4. Do your experiments focus on the user at a personalised level, or generically address them all?

Optimisation without focus lacks impact. User journeys are unique - customers have distinct preferences, category affinities, price sensitivities. They browse at different times of day, in various orders, and with varying levels of purchase intent.

Generic solutions that treat all customers the same end up with generic results.

- Personalisation is important because it tailors a journey to a customer's unique tastes and preferences.
- Personalisation can be best done by matching the right product to the right customer, at the right time.
- Product recommendations are designed for effective product selection, i.e. they surface relevant products a customer may not have otherwise found.

We know product recommendations are common across e-commerce. On average, product recommendations typically increase site-wide revenue by 3-4%.

However, experimenting with product recommendations provides additional impact, i.e. impact over impact. In partnering with Google Cloud, Qubit has deployed deep-learning recommendations across a number of retailers and tested against standard recommendation engines (powered by collaborative filtering). On average, deep-learning recommendations generate a 5-6% increase in revenue on top of the impact standard recommendations already drive.

The impact over impact can be explained by the fact that most standard product recommendations lack true elements of personalisation.

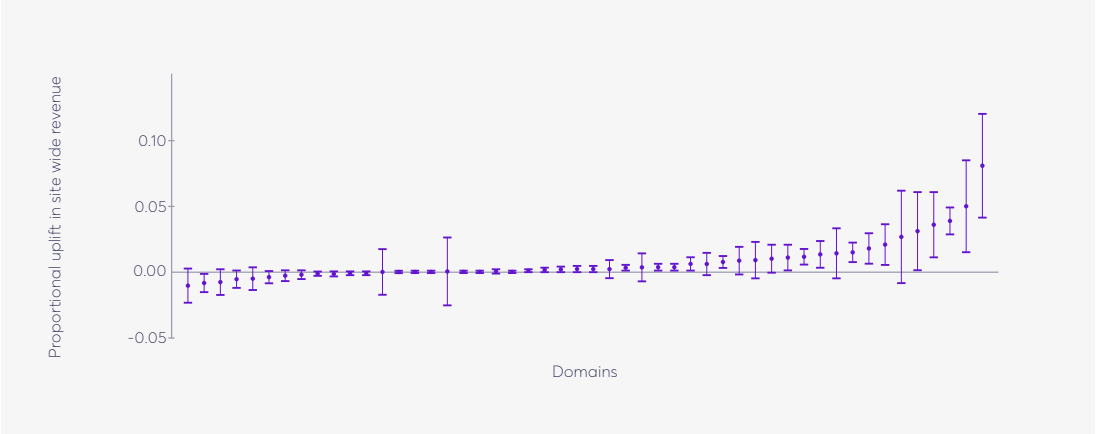
This is because most product recommendations today rely on patterns found in loads of data through the lens of a product to product relationship. Put simply through an example - a customer is browsing a blue shirt - they see a "You may also like" carousel of selected similar products (to the blue shirt), based on browsing behaviour of the masses. The products selected in the "You may also like" carousel by standard recommendations are typically limited to those that are most viewed or purchased within the catalogue and therefore may not be the most relevant to the customer. Instead, recommendations should be pulling the best products from the entire breadth of the catalogue to effectively inspire customers.

4. Do your experiments focus on the user at a personalised level, or generically address them all?

Personalised or deep-learning product recommendations adapt to real-time in-session behaviour of the customer and are anchored on the individual's shopping journey. They use key context such as time spent on a product, the order in which products are browsed, and adapt to changes in signals of buying intent. Deep learning recs also put the whole catalogue to work, constantly testing long-tail products to customers. This is because it uses additional product features such as description, product longevity and product image. The same technology that seeks out new viral videos on YouTube (recommendations power 70% of YouTube's video views) is used to find the next best product for each customer.

Experimenting with product recommendations can demonstrate the current engine's effectiveness - is there opportunity for impact over impact?

Experimentation can impact revenue. In the below figure by Qubit, in the same paper, we observe businesses that do not see a significant proportional increase in site-wide revenue through experimentation. However, some businesses receive uplifts of over 5% through personalisation and experimentation strategies.
The overall effect is positive.



Summary

Some experiments lack impact. The reason for their lack of impact can be attributed to one of four issues:- a lack of why, focusing on usability improvements, isolated solutions and addressing users as a whole, not as segments.

1 | Ensure that your experiment addresses a user problem. This will ensure that your solution should always have some form of outcome or action that addresses the initial problem

2 | Ensure that your experiment addresses a user's anxiety or facilitates a motivation. Rather than focussing on usability, layout or UI changes, focus on the improvements that will move behaviour

3 | Consider whether your experiments are too isolated. An isolated experiment, leads towards isolated results

4 | Consider whether you want to have a greater impact on a smaller segment of users, or a lesser impact on a greater impact of users. Focus is the key to optimisation

Thank you

If you have any questions, please feel free to get in touch.

In association with

Qubit.

User Conversion

www.userconversion.com

@userconversion

User Conversion Limited, Jutland House, Lower Ground Floor, Jutland St, Manchester M1 2BE
Company Number: 8151145