CON VERS ION.

Challenge assumptions, beliefs and doctrine

1

Experimentation should not be limited to optimising website landing pages, funnels and checkouts. Use experimentation as a tool to challenge the widely held assumptions, ingrained beliefs and doctrine of your organisation. It's often by challenging these assumptions that you'll see the biggest returns. Don't accept "that's the way it's always been done" — to do so is to guarantee you'll get the results you've always had. Experimentation provides a level playing field for evaluating competing ideas, scientifically, without the influence of authority or experience.

"Like anyone else, I've built up a number of beliefs and biases about who our customers are, what they want, how they behave, and what will resonate with them. But over time, left unchallenged, those positions proved themselves to be holding my organization back from achieving its potential — in marketing, in product design, in overall company strategy.
It was only when we were willing to question our core assumptions through interviews, data collection, and rigorous experimentation that we found answers to why growth had slowed, why a new product wasn't working, or why messaging didn't resonate. In the future, I'll be working to actively prove my assumptions wrong, knowing that if I fail, my path will be even more clear and worthwhile."

Rand Fishkin
CEO & Co-Founder
SparkToro

Always start with the data

2

It sounds trite to say you should start with data. Yet most people still don't. Gut-feel still dominates decision making, and experiments based on gut-feel rarely lead to meaningful impact or insight. Good experimentation starts with using data to identify and understand the problem you're trying to solve. Gather data as evidence and build a case for the likely causes of those problems. Once you have gathered enough evidence you can start to formulate hypotheses to be proven or disproven through experiments.

"At the end of the day, tests are either people building something new or asking a question of something that already exists. But how do we know we are asking the right questions? That is the power of data as a starting point for experimentation. Data can tell you what currently isn't working, then experiments tell you if new things would work better instead.

In my experience I have found more rewards and winners using data to look into small user journey/funnel improvements than totally new features. If you are approaching a full page redesign, use data to tell you which parts of the page are most used but produce a negative effect on CTR or CVR. This will give you the right component to start from and then develop around.

Use all possible data points and do not forget about heatmaps, attention maps, hovers and scrolling especially for mobile users. Build an experience model, aka a tree of the possible dimensions you can use to split your data for a specific page or feature, populate it and you will easily find what needs to be tackled first."

Andrea Mestriner
Head of Analytics, Visualisation and Experimentation
Just Eat

Experiment early and often

In any project, look for the earliest opportunity to run an experiment. Don't wait until you have already built the product/feature to run an experiment, or you'll find yourself moulding the results to justify the investment or decisions you've already made. Experiment often to regularly sense-check your thinking, remove reliance on gut-feel and make better informed decisions.

"A great example of the value of experimenting early at Facebook was when we were launching a feature to allow page admins to increase visibility of posts. Our planned button copy for this was "Promote post". Before the launch I shared this feature with a few people outside Facebook for a final sense-check. The feedback was that "Promote post" as button copy did not make it clear what this feature does. I went back to the team with the feedback and said we should totally test what this button says.

This was controversial. We hit this big point of friction inside the company and people were like "no we can't test this". We had just done a press release and announced to the world that we were launching the "Promoted Post" product. And so that's the name of the product. We can't change it now.

We had to challenge this. Ultimately I said to the team it's our product and we can change the name if it's confusing and people don't understand. I pushed the team to just do the test. And we agreed to let the test decide. The result was a 16% lift in people promoting posts just by changing one word on one button — "Promote post" to "Boost post".

The test proved that "promote" was jargon. It didn't sound like jargon to us because we were used to it. We had been calling it the "Promoted Post" product for months before we launched it. We all work on something for months before launch and the name and the words are so natural to you that you don't realise that people don't understand what the hell you're talking about."

Brian Hale
Vice President - Growth Marketing
Facebook

One provable hypothesis per experiment

4

Every experiment needs a single hypothesis. That hypothesis statement should be clear, concise and provable — a cause-effect statement. A single hypothesis ensures the experiment results can be used to evaluate that hypothesis directly. Competing hypotheses introduce uncertainty. If you have multiple hypotheses, separate these into distinct experiments.

"The objective of experimentation is to learn. Consider the following two hypotheses:

#1: "Let's change the button from yellow to blue and see if it increases the magic number."

#2: "We observed in user research that some people have difficulty finding the "buy now" button. We suspect this is caused by the low contrast between the font and the background. To solve this user issue, we will change the button from yellow to blue. If this solution works, we expect to see more users hover and click, and eventually purchase."

Both are describing the exact same change to the product. The key difference is the presence of a theory and a mechanism. The focus is on the how and the why.

Consider what we have gained by adding a more detailed hypothesis:

- We can reason about why this specific implementation will work, and perhaps not another. (e.g. making the button green will not increase contrast, but making the text black will)
- We can think of potential follow ups when the result does support our hypothesis. (e.g. we would try to increase contrast in other places)
- We can think of potential follow ups when the result does not support our hypothesis. (e.g. perhaps we should increase visibility even more; and make it blink)

In other words: we are not guessing which colours are optimal, we are collecting evidence for theories and learning to solve customer problems."

Lukas Vermeer
Booking.com

CONVERSION.

Define the success metric and criteria in advance

5

Define the primary success metric and the success criteria for an experiment at the same time that you define the hypothesis. Doing so will focus your exploration of possible solutions around their ability to impact this metric. Failing to do so will also introduce errors and bias when analysing results — making the data fit your own preconceived ideas or hopes for the outcome.

“Imagine you are walking around the Texan countryside. You encounter a big red barn. On the sides of this barn are drawn several targets: big white concentric circles on a red background. When you look more closely, you notice that in the dead center of each of these targets sits a single neat bullet hole. This is the barn that belongs to the Texas sharpshooter.

If we take our observations of the barn at face value, we might conclude that the Texas sharpshooter is a very good shot. But what is missing from the description above is an explanation of how the scene came to be; and this detail is quite important.

The Texas sharpshooter likes to shoot first, and draw the target after. He fires his gun at the barn, finds the bullet hole, and then draws neat circles around it. By reversing the order of events, he guarantees success. It would be wrong, then, to conclude from only the data that this man is a good shot. To support such a conclusion, we would want to draw the target first, and then check if the Texas sharpshooter is able to hit that predefined target.

The same applies to experimentation. Any targets drawn after the experiment is run should be called into question. The evidential value of an experiment comes from targets that were drawn before we started the test. Order matters.”

Lukas Vermeer
Booking.com

Start with the minimum viable experiment, then iterate

6

When tackling complex ideas the temptation can be to design a complex experiment. Instead, look for the simplest way to run an experiment that can validate just one part of the idea — the minimum viable experiment. Run this experiment to quickly get data or insight that either gives the green light to continue to more complex implementations, or flags problems early on. Then iterate and scale to larger experiments with confidence that you're heading in the right direction.

"It's easy to want every experiment to be "perfect".

Instead of testing a quick solution, you test the perfect solution — even if it takes days or weeks to develop the creative and functionality.

And that means committing a huge amount of resource to a single experiment — then hoping that it's successful. (Of course, most experiments fail — especially, it seems, the big ones.)

That's why we borrow an idea from product development — we start with the "minimum viable experiment".

Instead of looking for the perfect solution, we ask ourselves, "What's the quickest way to validate (or invalidate) our hypothesis?"

That way, we don't spend multiple weeks on one experiment — instead we run multiple experiments, validating multiple hypotheses in the process.

This is the fastest way to grow: spread your experiments across multiple concepts, and only increase your investment when you see those initial concepts working.

(Don't forget — it's easy for other people to criticise minimum viable experiments as "bad for the customer". That's not true — it's 100x better to test minimum viable concepts and kill off the bad ideas, than it is to invest in something your customers don't want.)"

Stephen Pavlovich
CEO & Founder
Conversion.com

CONVERSION.

Evaluate the data, hypothesis, execution and externalities separately

7

When faced with a negative result, it can be tempting to declare an idea dead-in-the-water and abandon it completely. Instead, evaluate the four components of the experiment separately to understand the true cause:

1. The data — was it correctly interpreted?
2. The hypothesis — has it actually been proven or disproven?
3. The execution — was our chosen solution the most effective?
4. External factors — has something skewed the data?

An iteration with a slightly different hypothesis, or an alternative execution could end in very different results. Evaluating against these four areas separately, for both negative and positive results, gives four areas on which you can iterate and gain deeper insight.

"Most experiments lose. But what caused an experiment to lose isn't always obvious. It could be that your hypothesis has been disproven. But it could just be your execution sucked. Perhaps your copy was not persuasive enough, or the usability was confusing. An undervalued skill in optimisation is being able to figure out why things went wrong. It's crucially important to spend the time reflecting on what caused the result — good or bad! To be an effective experimenter you need to become an expert in experiment diagnosis.

In experimentation we're exploring complex ideas. We're formulating hypotheses from often incomplete or imperfect data, then validating these hypotheses by testing multiple solutions on unpredictable users. At each step of this process there are opportunities for missteps that can impact the end result significantly. If we abandon an idea completely at the first sign of a disappointing result, the only thing we can be confident of is that we're going to miss out on a lot of opportunities."

Kyle Hearnshaw
Head of Conversion Strategy
Conversion.com

Measure the value of experimentation in impact and insight

8

The ultimate judge of the value of an experimentation programme is the impact it delivers and the insight it uncovers. Experimentation can only be judged a failure if it doesn't give us any new insight that we didn't have before. Negative results that give us new insight can often be more valuable than positive results that we don't understand.

"One thing I love about experimentation is the impact it can have on how teams work, and how the people in those teams think. Experiments give us a controlled way to take risks. This means teams that embrace an experimentation mindset are bolder and less afraid of failure. These teams embrace failure as a necessary and valuable step in problem solving.

Teams that are confident with experimentation (and can launch experiments quickly) start to see experiments as an extremely valuable source of actionable data and insight. They run experiments with the goal to learn things and answer questions, rather than just always to increase the conversion rate. Start measuring the value of experimentation by the impact it is having on how you work and solve problems. You'll find you won't even need to worry so much about the conversion rate impact you're having — that'll come naturally as a consequence of experimenting effectively."

Kyle Hearnshaw
Head of Conversion Strategy
Conversion.com

Use statistical significance to minimise risk

Use measures of statistical significance when analysing experiments to manage the risk of making incorrect decisions. Achieving 95% statistical significance leaves a 1 in 20 chance of a false positive — seeing a signal where there is no signal. This might not be acceptable for a very high risk experiment with something like product or pricing strategy, so increase your requirements to suit your appetite. Beware of experimenting without statistical significance, that's not much better than guessing.

"Be Skeptical of Great Results

A common bias is to accept good results and investigate bad results. If the results of an experiment are statistically significant and very positive to key metrics, the inclination is to celebrate; if the results are negative, the desire is to investigate and find any flaw with the methodology or process, so we can discredit the result and avoid having to change our beliefs (especially if it was a long and expensive project). Avoid this confirmation bias and think about Twyman's law (http://bit.ly/twymanLaw), which states that any figure that looks interesting or different is usually wrong. Are there suspicious metrics and anomalies associated with the experiment? Was the experiment replicated to improve the trustworthiness of the result? The best data scientists are skeptics that double-check, triangulate results, and evaluate the positive and the negative results with the same scientific rigor."

Ron Kohavi
Distinguished Engineer, General Manager, Analysis and Experimentation
Microsoft

CONVERSION.

Experimentation principles

To find out more, please email
marketing@conversion.com