Part I Introductory Topics

**Online Controlled Experiments Terminology**

A/B test, A/B/n tests (to emphasize multiple variants), field experiments, randomized controlled experiments, split tests, bucket tests, and flights (controlled experiments and A/B tests interchangeably)

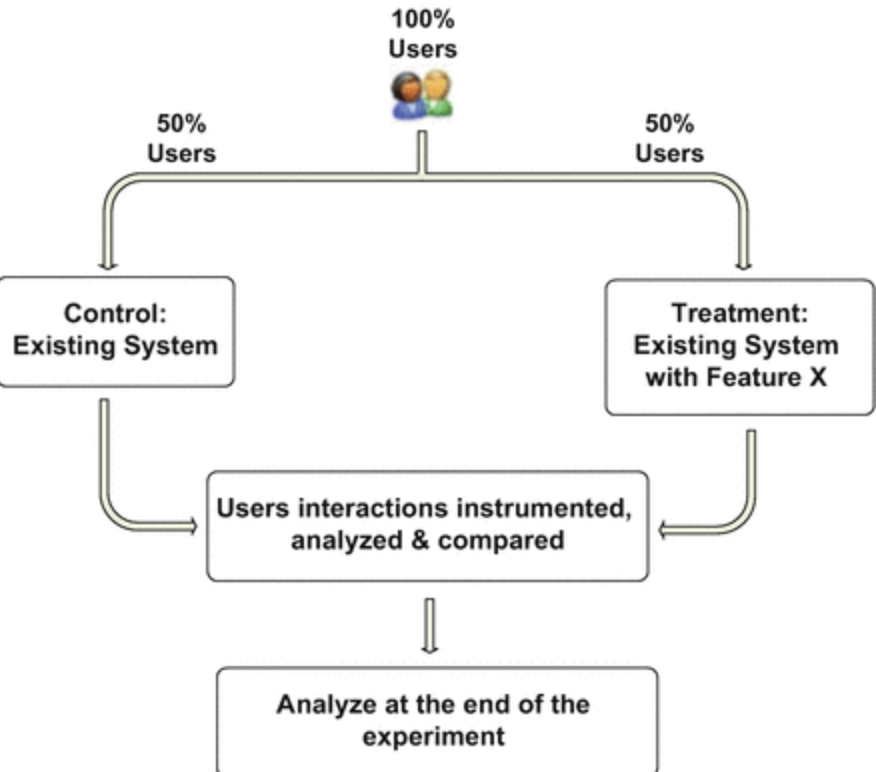


Figure 1.2 A simple controlled experiment: An A/B test

In the simplest controlled experiments, there are two variants: Control(A) and Treatment(B).

Overall Evaluation Criterion(OEC) must be measurable in the short term ( the duration of an experiment) yet believed to causally drive long-term strategic objectives.

In statistics, this is often called the Response or Dependent variable.

Parameters: A controlled experimental variable that is thought to influence the OEC or other metrics of interests. Parameters are sometimes called factors or variables. Parameters are assigned values, also called levels. In simple A/B tests, there is commonly a single parameter with two values. In the online world, it is common to use univariable designs with multiple values (such as, A/B/C/D). Multivariable tests, also called Multivariate Tests(MVTs), evaluate multiple parameters (variables) together, such as font color and font size, allowing experimenters to discover a global optimum when parameters interact

Variant: A user experience being tested, typically by assigning values to parameters. In a simple A/B test, A and B are the two variants, usually called Control and Treatment. Treatment: A variant; Control: A special variant.

Randomization Unit: A pseudo-randomization (e.g. hashing) process is applied to units (e.g. users or pages) to map them to variants. Proper randomization is important to ensure that the populations assigned to the different variants are similar statistically, allowing causal effects to be determined with high probability. You must map units to variants in a persistent and independent manner (i.e., if user is the randomization unit, a user should consistently see the same experience, and the assignment of a user to a variant should not tell you anything about the assignment of a different user to its variant). It is very common, and we highly recommend, to use user as a randomization unit when running controlled experiments for online audiences. Some experimental designs choose to randomize by pages, sessions, or user-day)

**Why Experiment? Correlations, Causality, and Trustworthiness**

Randomized controlled experiments are the gold standard for establishing causality. Systematic reviews, that is, meta-analysis, of controlled experiments provides more evidence and generalizability.

We believe online controlled experiments are:

* The best scientific way to establish causality with high probability.
* Able to detect small changes that are harder to detect with other techniques such as changes over time(sensitivity)
* Able to detect unexpected changes. Often underappreciated, but many experiments uncover surprising impacts on other metrics, be it performance degradation, increased crashes/errors, or cannibalizing clicks from other features.

**Necessary Ingredients for Running Useful Controlled Experiments**

1. There are experimental units( e.g., users) that can be assigned to different variants with no interference( or little interference).
2. There are enough experimental units (e.g., users). For controlled experiments to be useful, re