

Characterizing Experimentation in Continuous Deployment: a Case Study on Bing

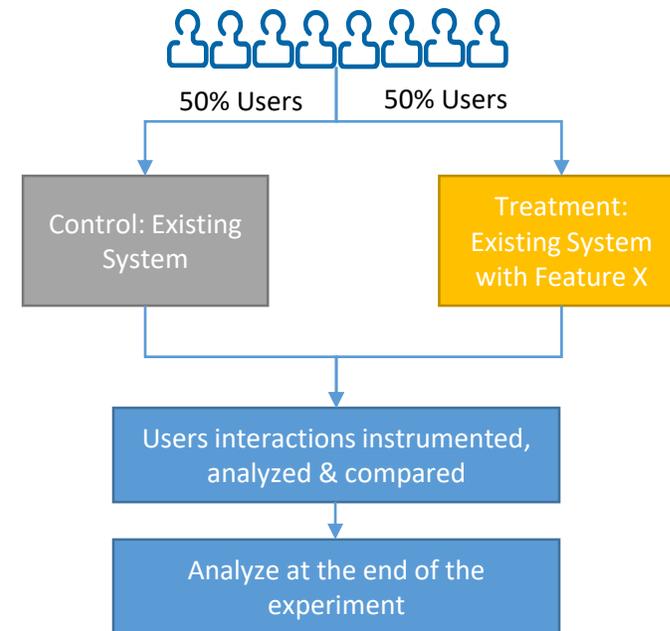
Katja Kevic, Brendan Murphy, Laurie Williams, Jennifer Beckmann



Will my users appreciate my new feature?

- Controlled Experiment

- Randomly split traffic between two (or more) versions
 - A (Control)
 - B (Treatment)
- Collect metrics of interest
- Analyze



- Must run statistical tests to confirm differences are not due to chance
- Best scientific way to prove **causality**, i.e., the changes in metrics are caused by changes introduced in the treatment(s)

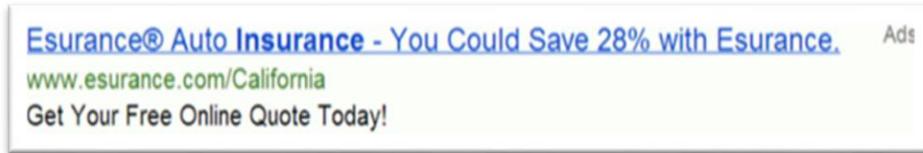


Site Links in Advertisements?

Primary Success Metric: **Revenue/User**

A

no links:



B

links:

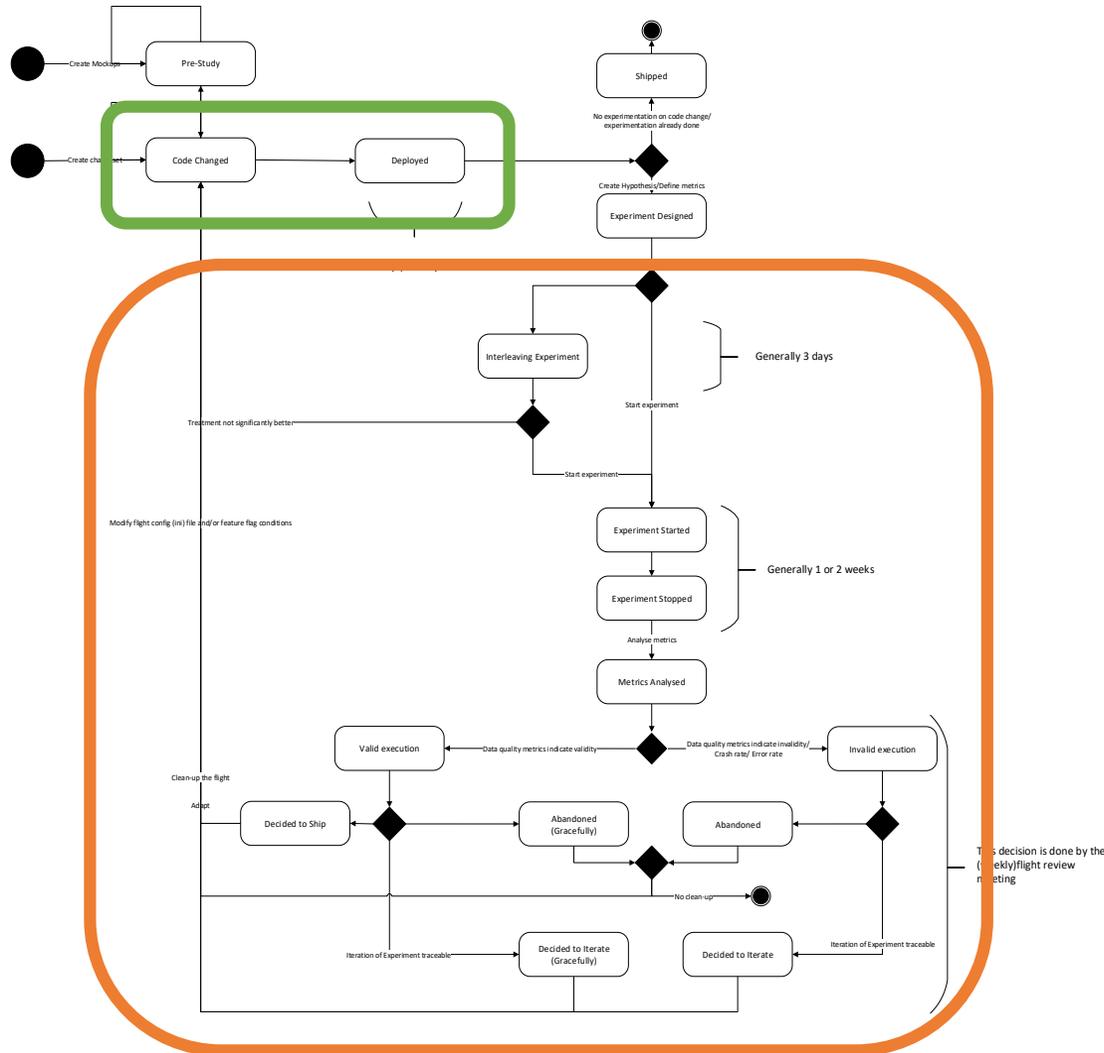


Pro: richer ads, users better informed where they land

Cons: Constraint means on average 4 "A" ads vs. 3 "B" ads

Variant B is 5msc slower (compute + higher page weight)

Experimentation process within Bing



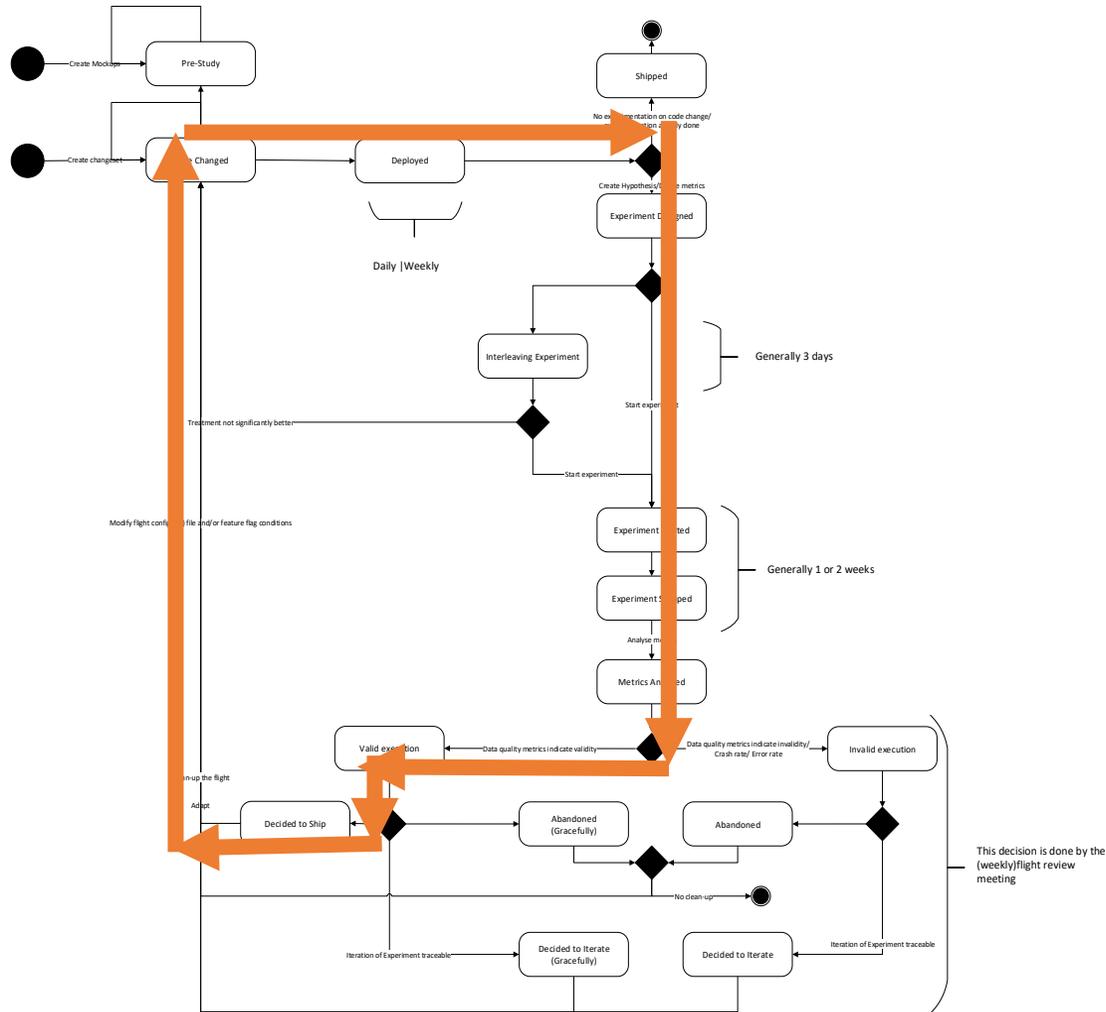
Experiments run for at least 1 week

Statistical validity for trustworthy comparison

Segmentation of users

Interactions between experiments

Experimentation process within Bing



Experiments run for at least 1 week

Statistical validity for trustworthy comparison

Segmentation of users

Interactions between experiments

Research Questions



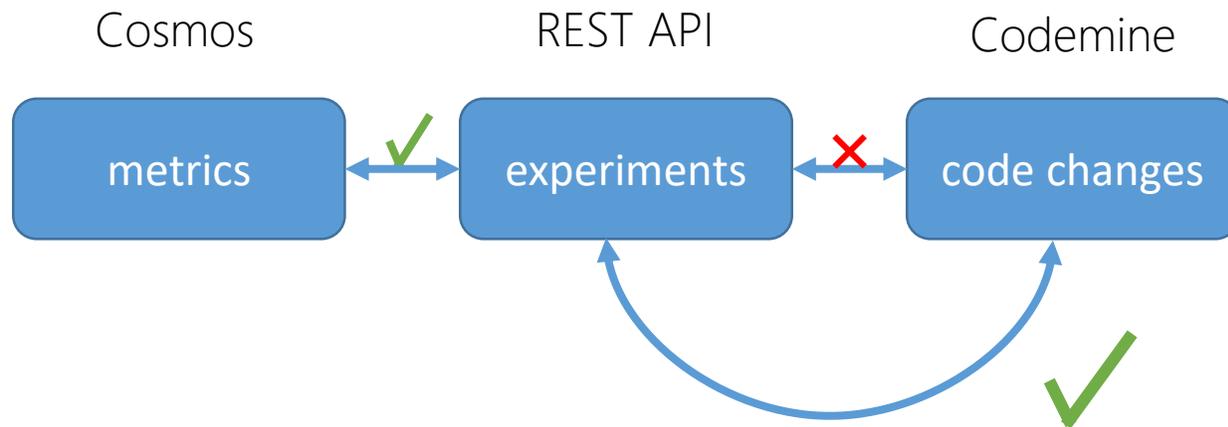
RQ1: What are the characteristics of experiments and their development efforts?

RQ2: What percentage of experiments are ultimately deployed to all users?

RQ3: How do the experiments which are deployed to all users differ from the experiments which were not deployed to all users?

Historical Experiments

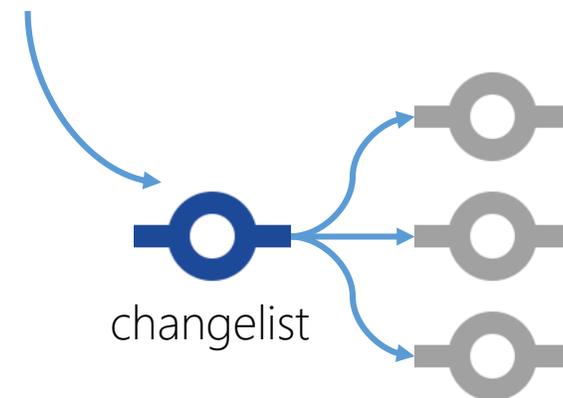
21,220 experiments over the past 2.5 years



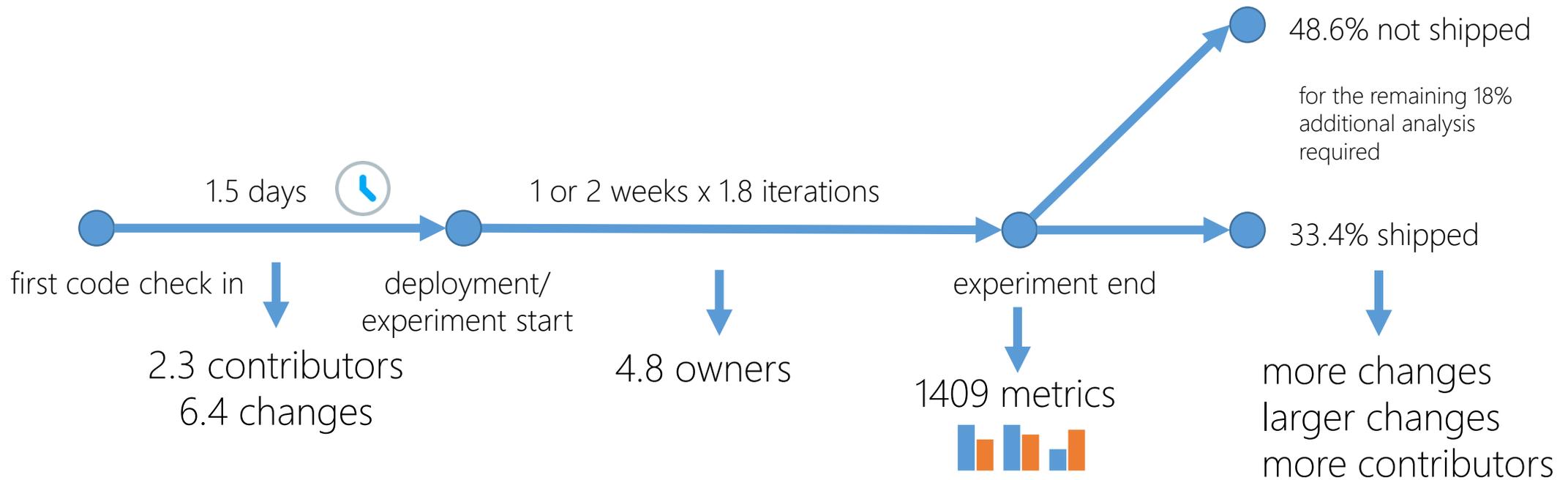
Procedure adopted by a section of Bing



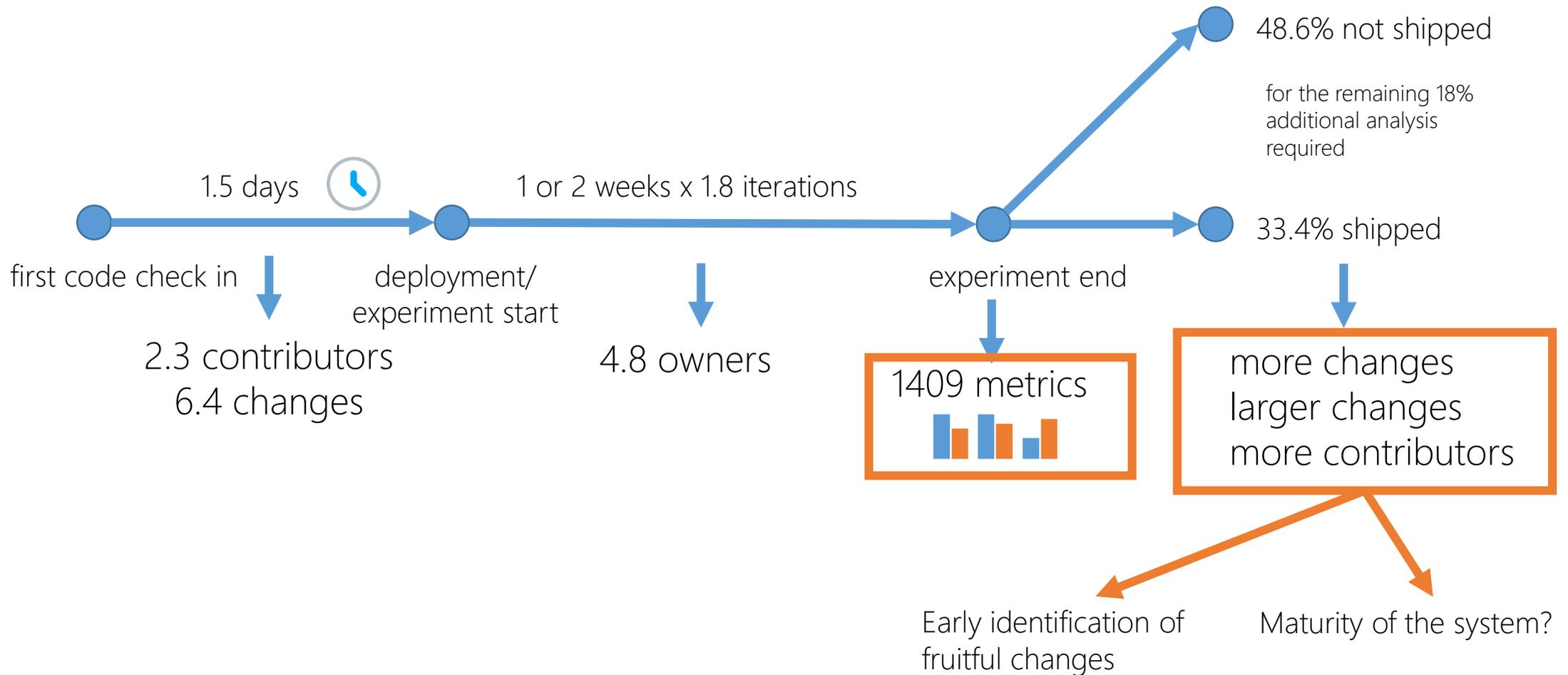
Parser based on pattern matching



Findings



Discussion and Future Work



Summary

Experimentation radically changed how software is deployed

Empirical characterization of an experimentation process in a mature and large-scale product

Performing an experiment is a substantial investment of time

