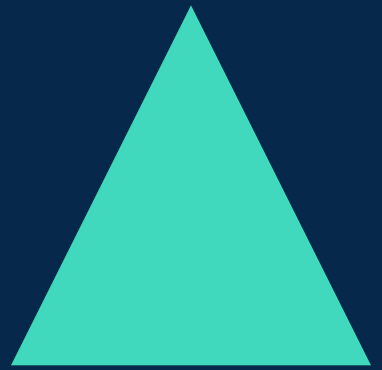
MVPs and experiments

Aleš Špetič

Lecture 6





Retrieval: the business model

Customer value

Which customer episode supports the claimed value?

Delivery

What has to work for each customer?

Payment

Which authorised buyer would pay, and for what?

Lecture route

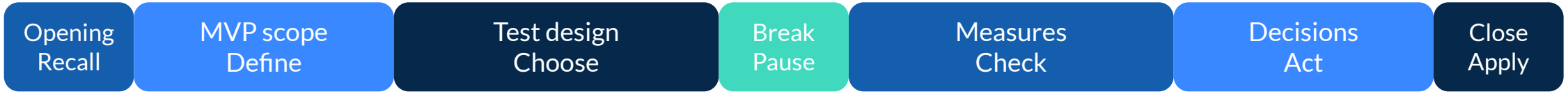
First part

Retrieval and framing
MVP scope and evidence
Hypotheses and manual tests

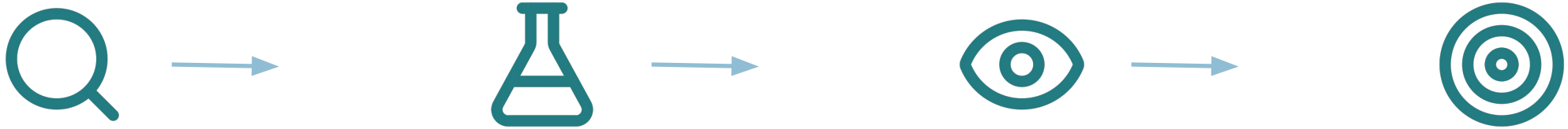
Later in the lecture

Break
Measures and misleading results
Thresholds and decisions
Experiment card and summary

Flexible lecture route



Today’s founder decision



Assumption

What must be true?

Test

What can we observe?

Measure

What counts as evidence?

Decision

What will change next?

Choose the smallest credible test of a consequential assumption.

Minimum viable product (MVP)

Learning objective

Reduce uncertainty about a specific customer or business assumption.

Minimum scope

Keep enough of the experience to make the observed behavior meaningful.

DISCUSSION

Which assumption would make further building pointless if it were false?

Prototype, MVP and pilot



Prototype

Represents or simulates an experience.



MVP

Offers enough value to investigate a key assumption.



Pilot

Runs a bounded trial in a real setting.

MVP scope: one need, one learning goal

Student venture method: use a real customer episode



Start with an episode

What did a real customer try, and where did it break?

Name the risky belief

Which assumption would change the build decision?

Keep the minimum path

What must a customer do to reveal that belief?

Leave out the rest

List attractive features that do not affect this test.

Dropbox: the demo made the idea visible

Dropbox synchronises files across devices. Its early challenge was showing people how that experience would work.

April 2007

Houston shared a short screencast on Hacker News and received feedback.

March 2008

A separate private-beta video accompanied a reported waitlist jump from 5,000 to 75,000.

DISCUSSION

What behavior did the video make observable?

Dropbox: the boundary of the evidence

Dropbox's video demonstrated file synchronisation. The waitlist recorded interest in trying the product.

Supported inference

The demonstrated experience attracted interest from the audience reached.

Still unresolved

Reliable use across devices, repeated value and willingness to pay.

DISCUSSION

Would the same video be enough evidence for a banking product?

Watch the original Dropbox demo

The 2007 screencast shows the early product interaction. Observe the workflow before judging the evidence.

Video link

Watch the 2007 Dropbox demo on YouTube.

While watching

What customer problem and product behavior become visible?

DISCUSSION

What does the demo prove, and what remains untested?

Zapier: payment before launch

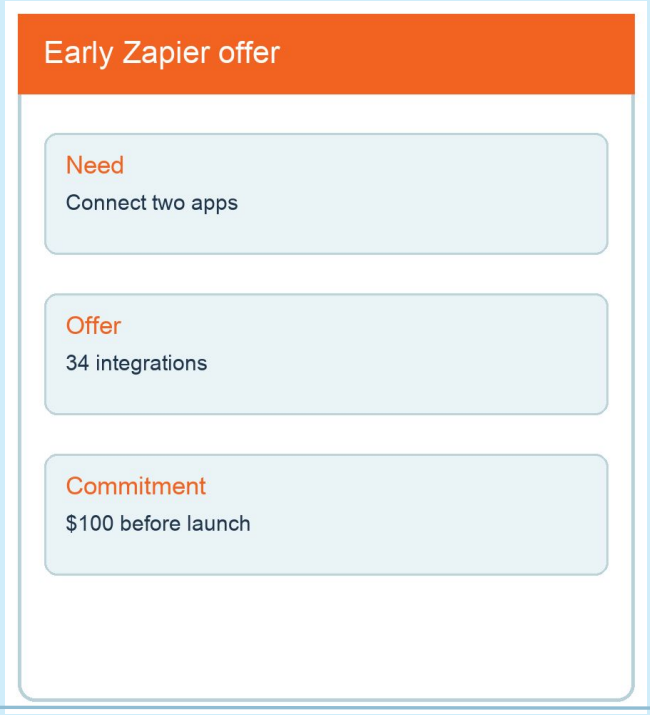
Zapier connects software tools. Its team began with a small integration set and pursued specific customer needs.

Narrow first version

It launched with 34 app integrations.

Paid commitment

One customer paid $100 before the public launch.



Teaching diagram based on Zapier's account

DISCUSSION
What does a prelaunch payment prove?

Zapier: payment and repeat use differ

A prelaunch payment is stronger evidence than a click, but it does not settle repeated use or economics.

Paid signal

An early customer paid $100 for a specific need before public launch.

Still unresolved

Can the team deliver reliably, retain customers and support the service at a viable cost?

DISCUSSION

What would you measure in the first month?

A test method answers a particular question

Method choice depends on the claim being tested

Uncertainty	Low-cost test	Observe	Still unknown
Comprehension	Demo or storyboard	Explains the benefit	Real use
Usability	Task prototype	Completes the task	Demand
Value in use	Manual service	Uses the outcome	Scale economics
Paid demand	Clear paid offer	Authorised payment	Repeat value

A working prototype can answer several questions, but each inference needs evidence.

A hypothesis can fail a specific test

People + setting

Name the customer and the situation.

Prediction

State the behavior or outcome expected.

Conditions

Specify the offer, channel and time window.

Decision boundary

Say what result would change the next step.

Groupon: a deal before a platform



Groupon sold local deals. Its first version paired a simple site with manual coupon handling.

Simple first site

The first Groupon site used WordPress to present local deals.

Manual fulfilment

The team prepared and emailed PDF vouchers.



DISCUSSION

Which parts had to work for this test to be credible?

Groupon: a manual fulfilment loop



Case: Groupon; observation plan is teaching analysis

01

Request

Customer buys a deal.

02

Fulfilment

Team creates and emails the PDF voucher.

03

Observation

Track redemption, support requests and refunds.

Groupon: the hidden work in a sale



Groupon handled early coupon delivery manually. A sale is only useful if fulfilment and merchant value can work.

Learning gained

Paid deals reveal demand and fulfilment problems.

Missing economic proof

Manual voucher creation and support can hide the true cost per deal.



DISCUSSION

What would you record before scaling deal volume?

Warby Parker: try before buying online

WARBY PARKER

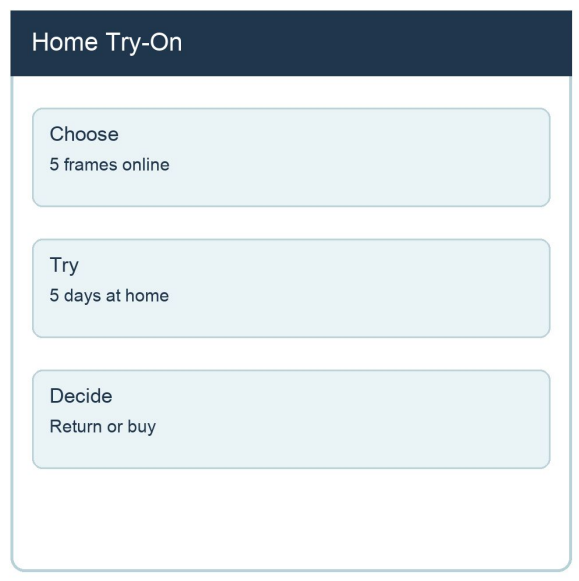
Warby Parker made eyewear an online purchase and introduced Home Try-On to help buyers assess fit.

Online offer

The company began selling glasses online in 2010.

Home Try-On

Customers could try five frames at home for five days.



Teaching diagram of Home Try-On

DISCUSSION

Which buying uncertainty did try-on reduce?

Warby Parker: isolate the buying barrier

WARBY PARKER

Teaching comparison; not a reported Warby Parker experiment

01

A: product page

Photos, dimensions and price.

02

B: fit guidance

Advice and frame recommendations.

03

C: home try-on

Five frames for five days.

Warby Parker: who requests try-on?

WARBY PARKER

Home Try-On reduces fit uncertainty, but recruitment and actual purchases determine what it proves.

Useful signal

Customers requested frames and tried them at home.

Inference limit

A free try-on request does not prove a purchase or represent every buyer.

DISCUSSION

Who would you recruit for a purchase test?

Manual tests and technical prototypes

Scope follows the uncertainty

Method	What stays real	What is simplified	Useful for
Concierge service	Customer outcome	Founder does the work	Value and workflow
Wizard of Oz	Visible interaction	Human runs the backend	Interaction design
Clickable prototype	Task and interface	No full operations	Usability
Technical spike	Hard technical task	Narrow environment	Feasibility

A simulated backend needs appropriate disclosure and data protection.

A bounded test still needs clear promises

Scope and resources

Define participants, duration, budget and what the test will deliver.

Participant expectations

Explain availability, payment terms, data use and how to get help.

DISCUSSION

What would make a cheap test too costly for the participant?

Before a live MVP test

A short checklist before approaching external customers



Task and value

Can a participant attempt one meaningful job?

Offer and consent

Are terms, limits and data use clear?

Observation

Can you record the action, result and denominator?

Feedback and decision

Who will review the evidence, and when?



Break
10 minutes

Exposure, action and outcome

01

Exposure

Who actually encountered the offer?

02

Action

What did they do under those terms?

03

Outcome

Did they receive value and return?

Pebble: crowdfunding measured commitment

pebble

Pebble proposed an e-paper smartwatch for iPhone and Android users. Its 2012 Kickstarter sought funding to bring it to market.

2012 campaign

About $10.27 million pledged by 68,929 backers.

What this records

Commitment to a proposed watch and its campaign terms.



Original-generation Pebble watch; product photograph

DISCUSSION

What does a funded campaign leave untested?

Pebble: demand and delivery are separate

pebble

Pebble's 2012 backers committed to a connected watch. The campaign result addresses only part of a hardware business.

Demand signal

Backers accepted the proposition and the campaign's uncertainty.

Delivery questions

Can the team produce reliable watches at the required cost and volume?



DISCUSSION

Which delivery risk would a larger advertising budget fail to resolve?

Pebble: the next test follows the risk

pebble

Pebble combines hardware and software. Consider a hypothetical next test after demand appears promising.

Proposed technical test

Run a small batch through realistic battery, connection and failure conditions.

Proposed decision rule

Agree acceptable reliability and cost before committing to a larger batch.

DISCUSSION

Why would a click-through rate be irrelevant to this decision?

A metric needs a denominator and a window

Counted event

Define the action, the eligible person or account, and how duplicates count.

Time and comparison

Fix the observation window and record how people entered the test.



DISCUSSION

What is missing from “30 people signed up”?

Bing: pooled clicks can mislead

Teaching reconstruction inspired by Bing's ad headline experiment

Sessions	A clicks	B clicks	What changed?
Returning	8 / 100 = 8%	63 / 900 = 7%	B gets more returning sessions
New	18 / 900 = 2%	1 / 100 = 1%	A gets more new sessions
All	26 / 1,000 = 2.6%	64 / 1,000 = 6.4%	Different audience mix
Within each group	Higher click rate	Lower click rate	No random assignment

Invented teaching data, not Bing results. Traffic mix prevents a causal comparison.

Bing: test the headline change fairly

Bing's reported A/B test; measures shown as teaching summary

Concurrent assignment

Randomly assign eligible search sessions to the old or new headline display.

Outcome and guardrail

Compare revenue per eligible search and check user-experience metrics.



DISCUSSION

What could still invalidate the comparison?

A small numerical lead can be uncertain

Observed result

Version A: 3 of 10. Version B: 4 of 10.

Practical interpretation

One extra person changes the result. More evidence may be needed.



DISCUSSION

Would you spend a month rebuilding the product on this result alone?

Measurement failures that change the story

Selection

Only enthusiastic friends join.

Instrumentation

A click event fires twice.

Changing conditions

The price and channel change together.

Premature stopping

The team stops when the result first looks good.

Success thresholds and decision criteria

Economics

What minimum behavior makes the next step affordable?

Customer value

What observable improvement matters?

Service guardrail

What cost or failure level is unacceptable?

Evidence quality

Was the test delivered and measured as planned?

SlotFill: the pilot context



Case: SlotFill, fictional salon waitlist service

Customer

Independent salons with last-minute appointment cancellations.

Offer

€199 per salon per month for waitlist slot-filling.

Service

Text offers, manual confirmation and calendar updates.

Next decision

Whether to expand a paid pilot.

SlotFill: one pilot, four decision gates

SF

Case: SlotFill; fictional design and provisional thresholds

Hypothesis

Some salon owners will pay €199 and renew after recovered appointments.

Test

Offer the same one-month pilot to 10 qualified salons; enrol up to 3.

Measure

Paid acceptance, filled slots, value, renewal and service cost.

Decision rule

At least 3 pay; at least 2 renew; each value and cost check passes.

SlotFill: the operating record

Case: SlotFill; fictional measurement plan

Measure	Definition	Window	Evidence
Paid acceptance	Paid enrolled salons / 10 offers	Offer period: 2 weeks	Offer + payment record
Filled slots	Paid bookings filled from waitlist	Pilot month	Calendar + reply log
Customer value	Incremental slots × contribution	Same pilot month	Bookings + finance log
Service cost	Text, staff time, setup, help	Same pilot month	Time and cost ledger

A message sent, a reply and a filled paid appointment are different outcomes.

SlotFill: value and cost checks

Case: SlotFill; all quantities and criteria are hypothetical

Check	Working assumption	Pilot criterion	Reason
Salon value	4 slots × €50 = €200	At least €199 / salon	Potential value vs fee
Service cost	€90 / salon / month	At or below €90	Preserve €109 contribution
Renewal	2 of 3 enrolled salons	Same €199 offer and scope	Repeat commitment
Service delivery	Timely text + booking	No serious errors or opt-outs	Acceptable experience

Incremental slots need careful attribution. €109 is before fixed costs.

SlotFill: a plausible mixed result

Case: SlotFill; invented pilot results for discussion

Observed item	Fictional result	Planned criterion	Decision question
Paid salons	3 / 10	At least 3	Is paid demand enough?
Renewals	2 / 3	At least 2	What does renewal add?
Service cost	€72, €96, €130	Each at or below €90	Can delivery improve?
Estimated value	€250, €200, €100	At least €199 / salon	Which salons benefit?

Estimated value = 5, 4 and 2 filled slots × assumed €50 contribution.

SlotFill: decision after the pilot

SF

Case: SlotFill; fictional pilot decision and next test

A

Decision

Do not expand under the agreed value and cost gates.

B

Diagnosis

Check baseline cancellations, staff time and true incremental bookings.

C

Next test

Test salons with frequent cancellations and simpler reply handling.

An inconclusive test needs a diagnosis

Business uncertainty

The right people saw a functioning offer, but evidence remains weak or mixed.

Test failure

Recruitment, fulfilment or measurement broke, so the intended question remains unanswered.



DISCUSSION

When would you repeat the test, and when would you change the hypothesis?

The MVP experiment card

Follow-up project output: one card for one consequential assumption



Claim and decision

State one main assumption and the decision that depends on it.

Test and constraints

Method, recruitment, offer, owner, duration and budget.

Measure and threshold

Defined events, denominators, time windows and guardrails.

Result and next step

Record evidence, limits, revised belief and next action.

Closing discussion

Case-based questions; SlotFill's results are fictional.



Dropbox

A waitlist grows. What would you test before scaling use?



Warby Parker

Try-on requests arrive. What would prove purchases?



SlotFill

Salons renew, but costs exceed the cap. What happens next?

MVPs and experiments: key takeaways

The assumption sets the scope

Build enough to observe relevant behavior.

The measure limits the claim

Interest, payment and delivered value differ.

The rule comes first

Choose thresholds and guardrails before results.

The result changes a decision

Continue, revise, pause or investigate an inconclusive test.

Next: product, activation and retention

Outside class

Prepare an experiment card. Gather an external signal where feasible; record limits and the next decision.

In Lecture 7

Examine first value, onboarding, cohorts and repeated use after the initial test.

Sources: demonstrations and manual services

Dropbox

Drew Houston: Startup Lessons Learned

Zapier

Zapier: first million users account

Groupon

Andrew Mason: founder interview

Warby Parker

Warby Parker: SEC filing on Home Try-On

Sources: methods and hardware

Experiment design

Strategyzer: Test Card

Pebble

Original 2012 Kickstarter campaign

Teaching data

SlotFill and the Bing traffic table are fictional.

Bing A/B experiment

Kohavi: controlled experiment case

Reported 12% revenue lift; guardrails held.

Visual sources

Company marks

Teaching marks and company icons

Original illustrations

Zapier: teaching diagram of early offer

Teaching diagram

Warby Parker: Home Try-On teaching diagram

Product photograph

Mobile Fun: original Pebble watch

Homework 6: MVP experiment

Experiment

Name the customer, riskiest assumption and smallest MVP or manual test.

Measures

Define exposure, action, outcome, time window and a pre-set success or guardrail rule.

Result

Show external behavior, limits and next decision; label an unrun test as a plan.

Submit

Add 1–3 exam slides; save the test card and raw evidence in the folder.