THE NEW! SCIENCE OF PRODUCT DEVELOPMENT

JAMES BIRCHLER (@JAMESBIRCHLER)
ENGINEERING DIRECTOR, IMVU
GDC ONLINE, OCTOBER 13, 2011

QUESTIONS & COMMENTS?

@JAMESBIRCHLER

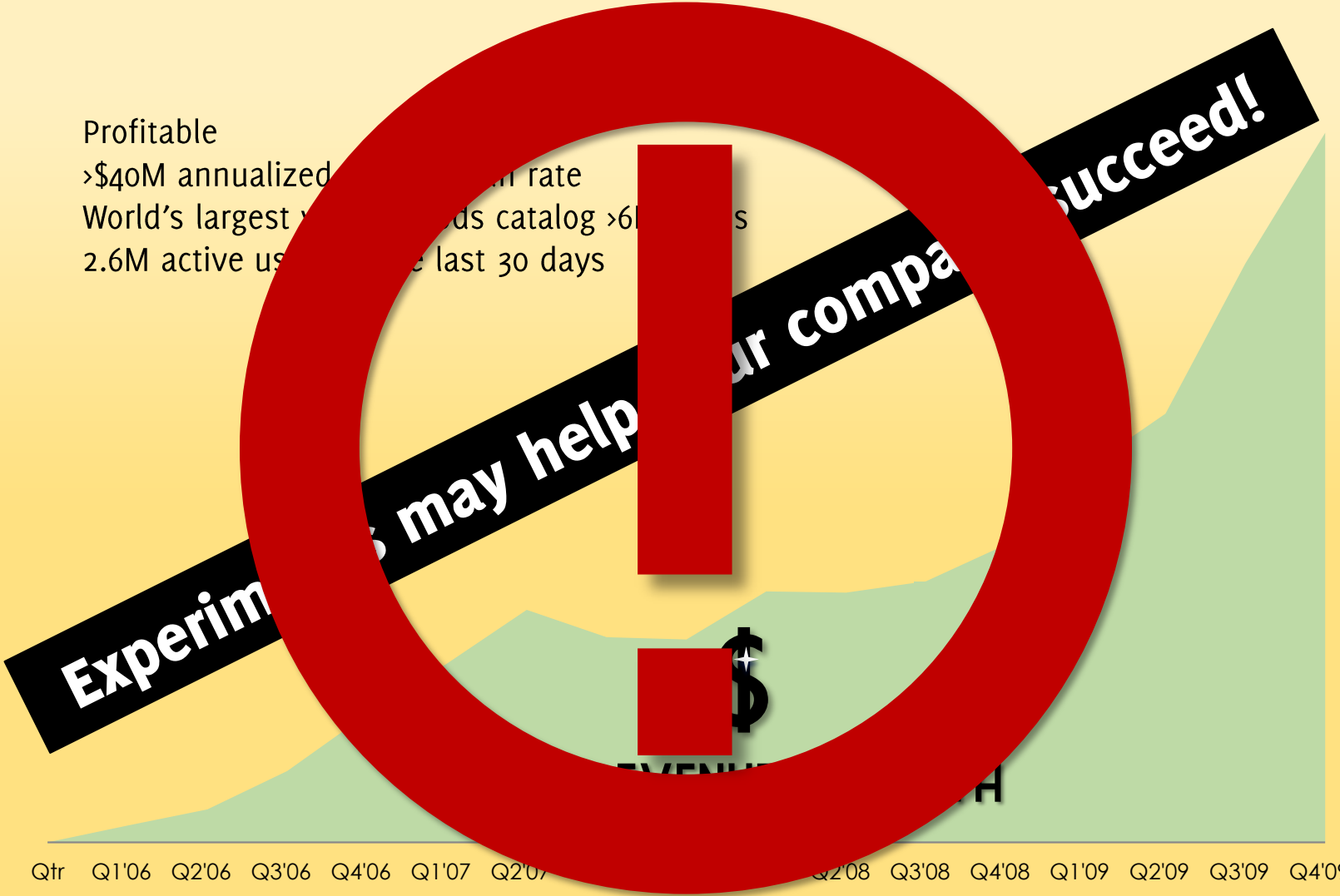
THE SCIENTIFIC METHOD IS BASED ON

EXPERIMENTATION

Inventory UI

Profitable
›$40M annualized revenue run rate
World's largest virtual goods catalog ›6M items
2.6M active users in the last 30 days

Experiments may help your company succeed!


Profitable
›$40M annualized
World's largest
2.6M active us
last 30 days
Experim
may help
succeed!
Qtr
Q1'06
Q2'06
Q3'06
Q4'06
Q1'07
Q2'08
Q3'08
Q4'08
Q1'09
Q2'09
Q3'09
Q4'09

1. Culture

2. Technique

3. Examples

Renaissance Astronomer

Heliocentrism

Experimental observation

@COPERNICUS

Renaissance Astronomer

Heliocentrism

Experimental observation

#INQUISITION

Renaissance Astronomer

Heliocentrism

Experimental observation

@COPERNICUS

Afraid to publish until on deathbed

“Copernicus FTW!”

@GIORDANO_BRUNO

Persecuted

Prosecuted

Burned at the stake

“Father of Modern Science”

Experimentation = foundation of scientific method of learning

@GALILEO

Improvement over Bruno.

Persecuted

Accused of heresy

House arrest until his death

Question: Why was it so hard for these guys to share their data and findings?

Answer: It’s complicated, but I think we can all agree that the folks in charge didn’t support hearing “bad news.”

Experiments FTW!

Share results freely!

Rapid iteration and learning!

@JAMESBIRCHLER

Profitable

>6 million virtual goods

2.6 million active users last 30 days

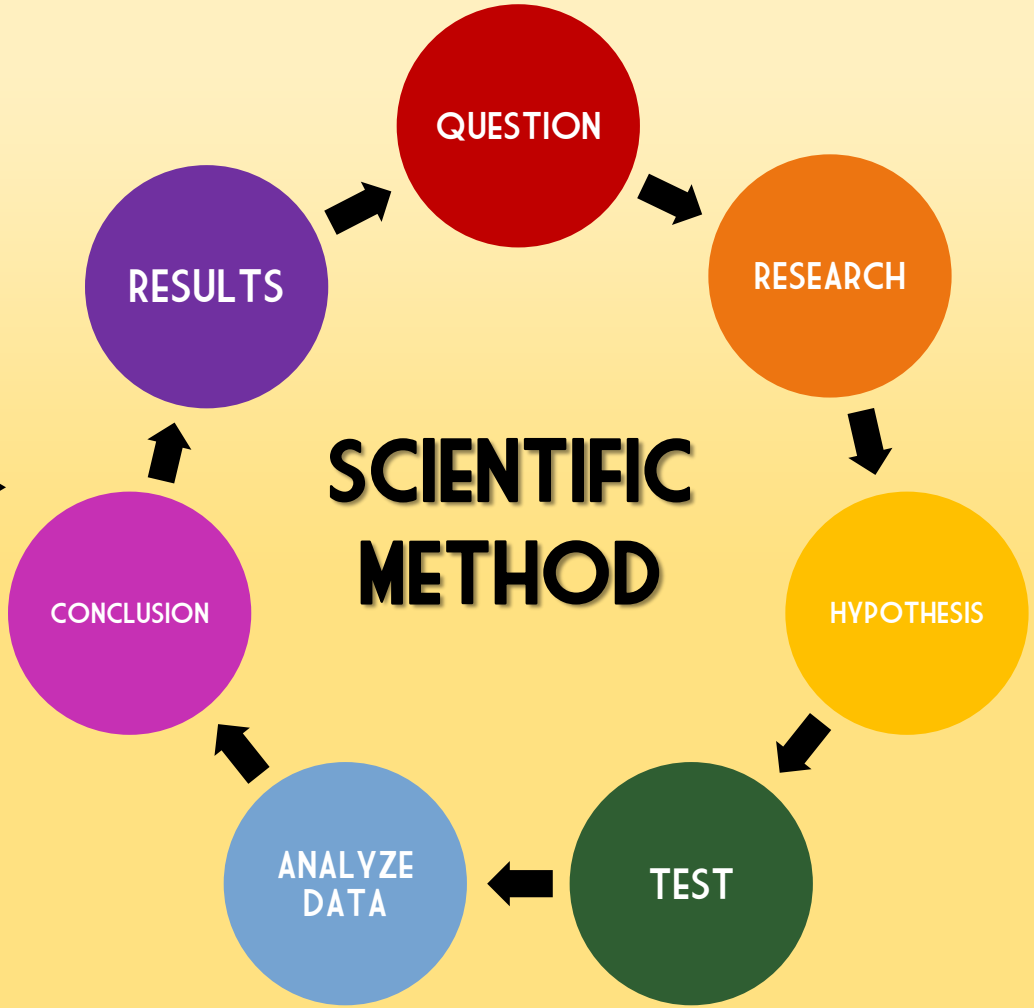
GDC Online gig!

EXPERIMENTATION
IS GOOD

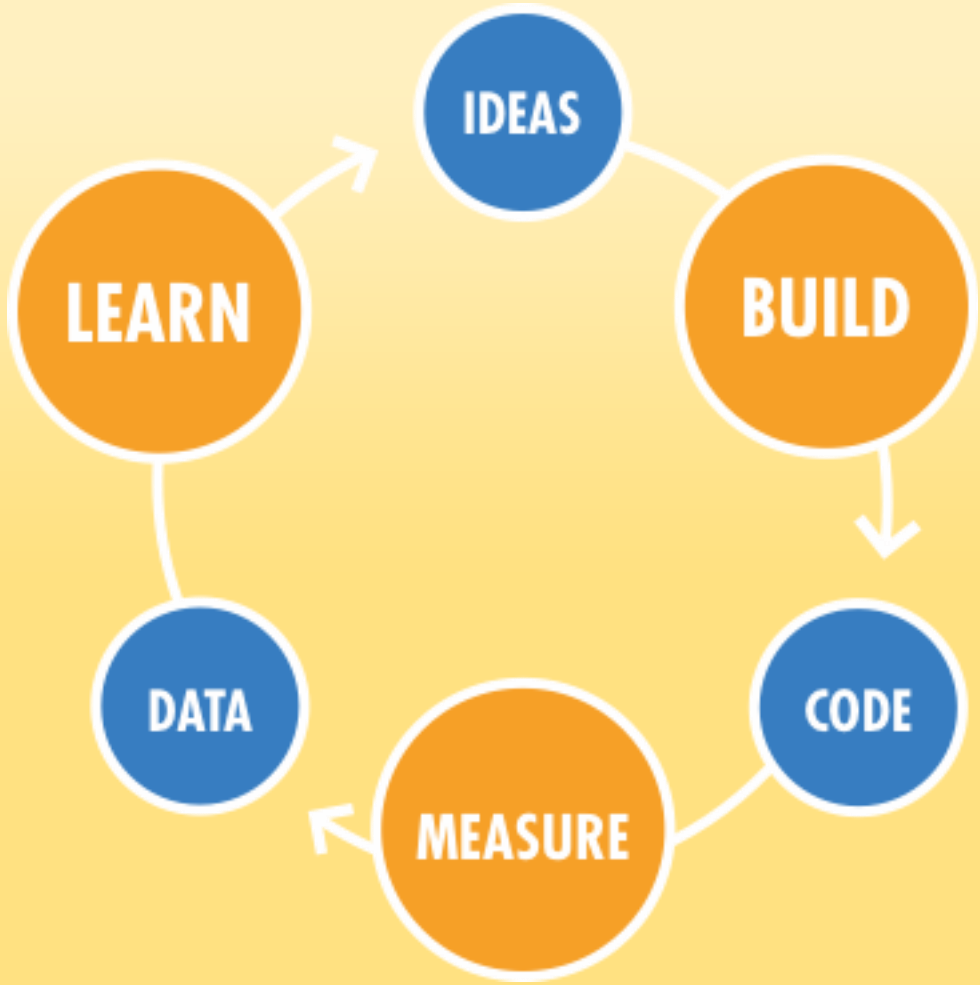
RIGHT?

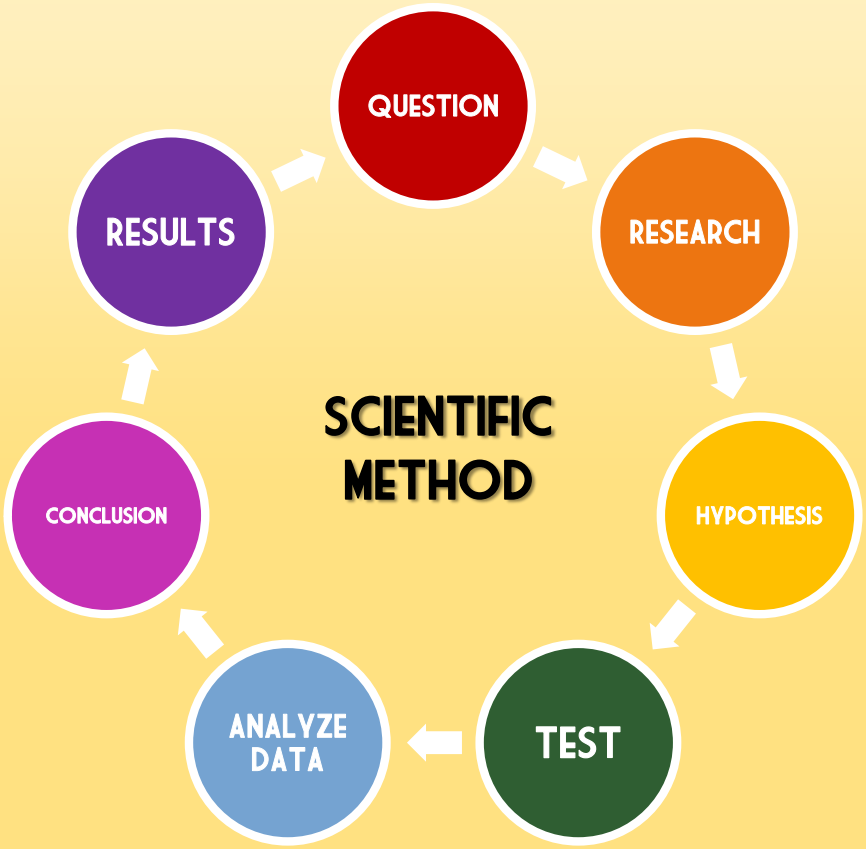
ASK QUESTION
DO RESEARCH
HYPOTHESIS
TEST
ANALYZE DATA
CONCLUSION
REPORT RESULTS
GOT SCIENCE?

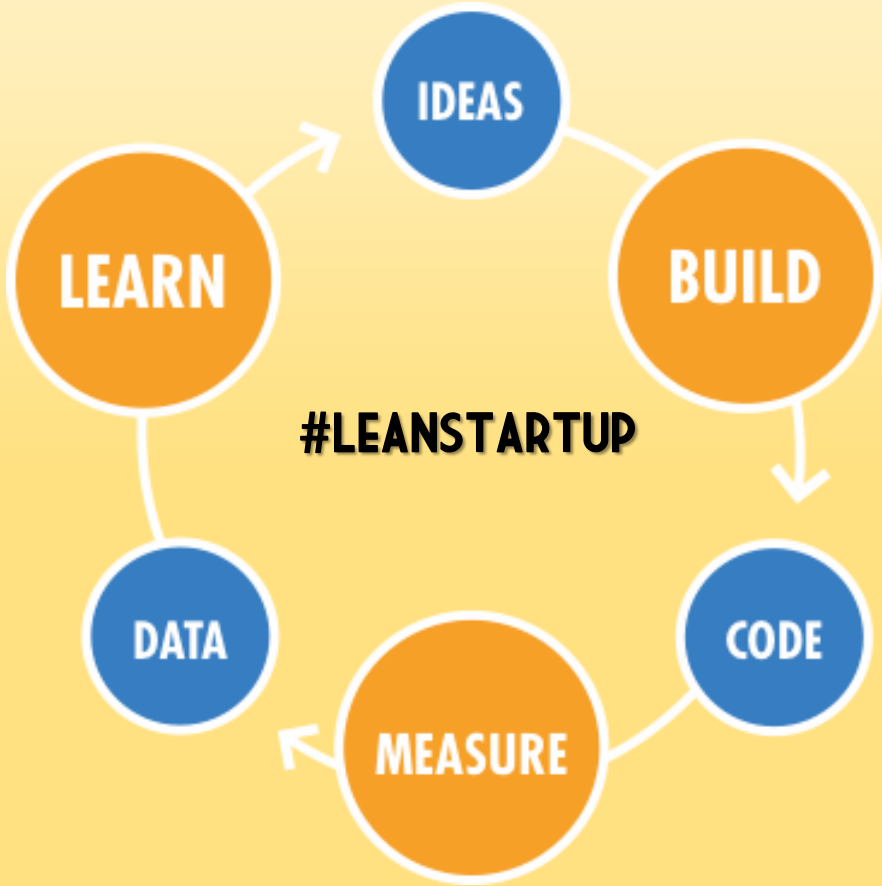
Rapid iteration = Rapid learning

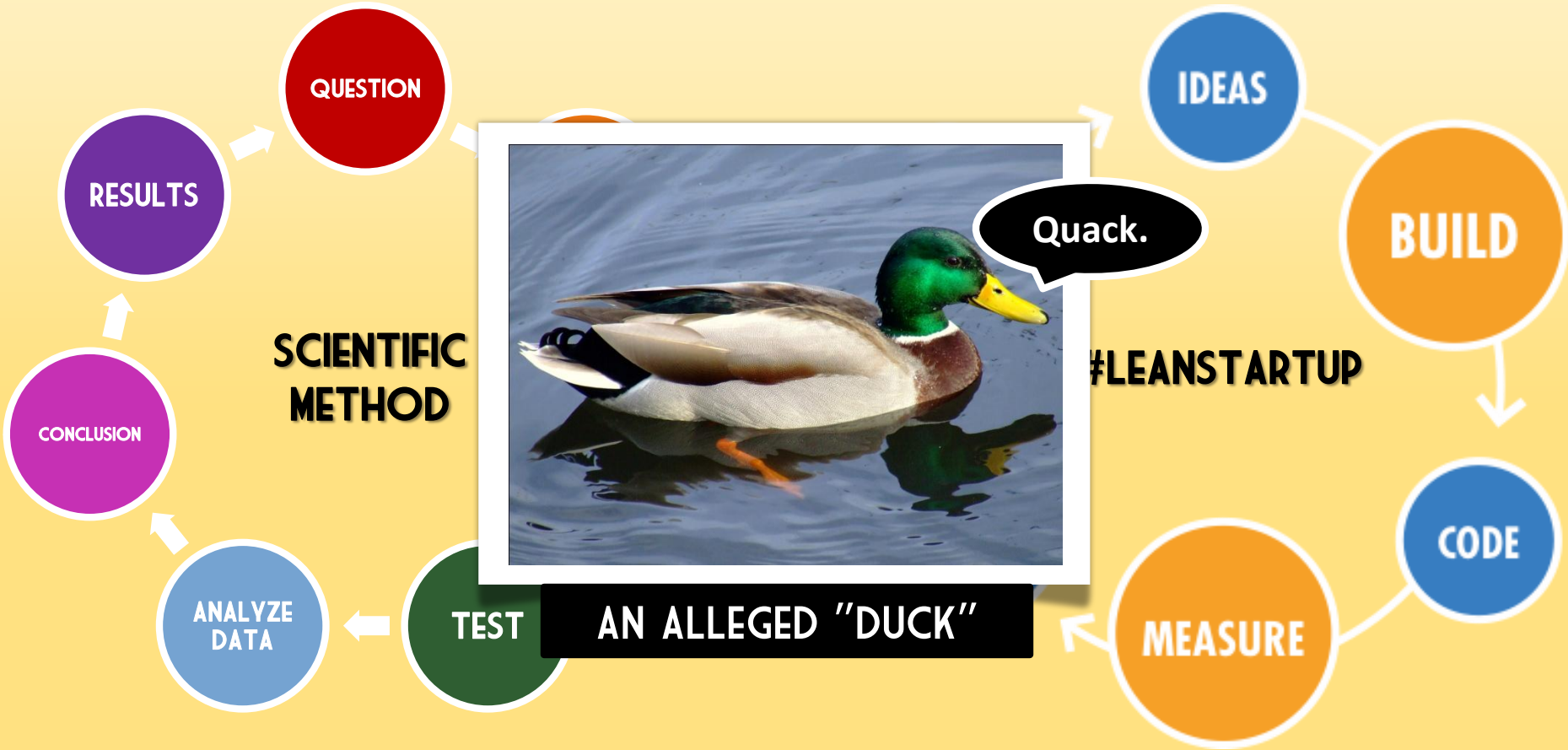


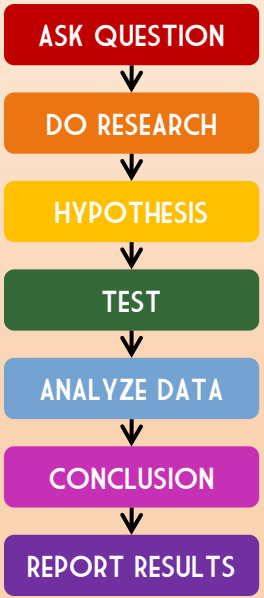
“Lean Startup”




QUESTION
RESEARCH
HYPOTHESIS
TEST
ANALYZE DATA
CONCLUSION
RESULTS
SCIENTIFIC METHOD


IDEAS
BUILD
CODE
MEASURE
DATA
LEARN
#LEANSTARTUP


QUESTION
RESULTS
CONCLUSION
ANALYZE DATA
TEST
SCIENTIFIC METHOD
Quack.
AN ALLEGED "DUCK"
IDEAS
BUILD
CODE
MEASURE
#LEANSTARTUP



IMVU Feature Experiment "How-To"

1. Talk to customers--use cases!
2. Form hypothesis to test
3. Write code, test on dev machine
4. Test in production as QA/admin
5. Roll out to a % of customers
6. Results, conclusion
7. Share learning

LOOKS A LOT LIKE THE SCIENTIFIC METHOD (IT IS!)

CULTURE OF
EXPERIMENTATION

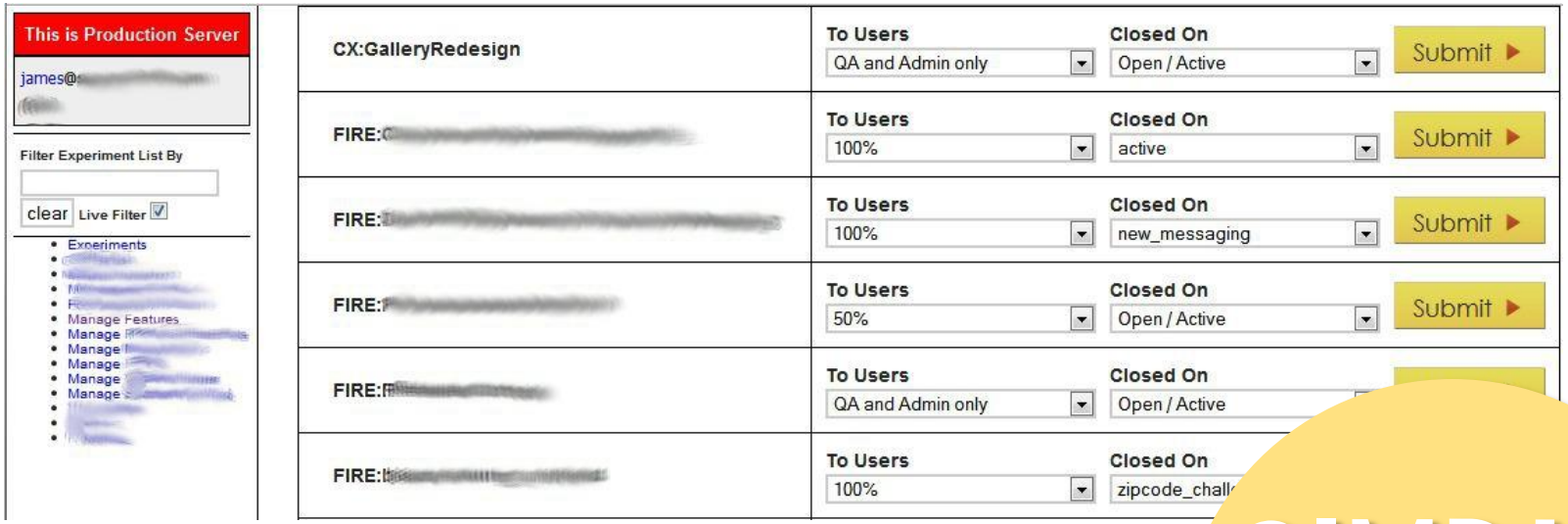
Okay, I'll run an experiment!
Okay, I'll run an experiment!
Okay, I'll run an experiment!
Okay, I'll run an experiment!
Okay, I'll run an experiment!
Okay, I'll run an experiment!
Let's keep this simple!
Prove your idea is the best!
Okay, I'll run an experiment!

KEEP IT SIMPLE

```
if( setup_experiment(...) == "control" ) {

  // do it the old way

} else {

 // do it the new way

}
```

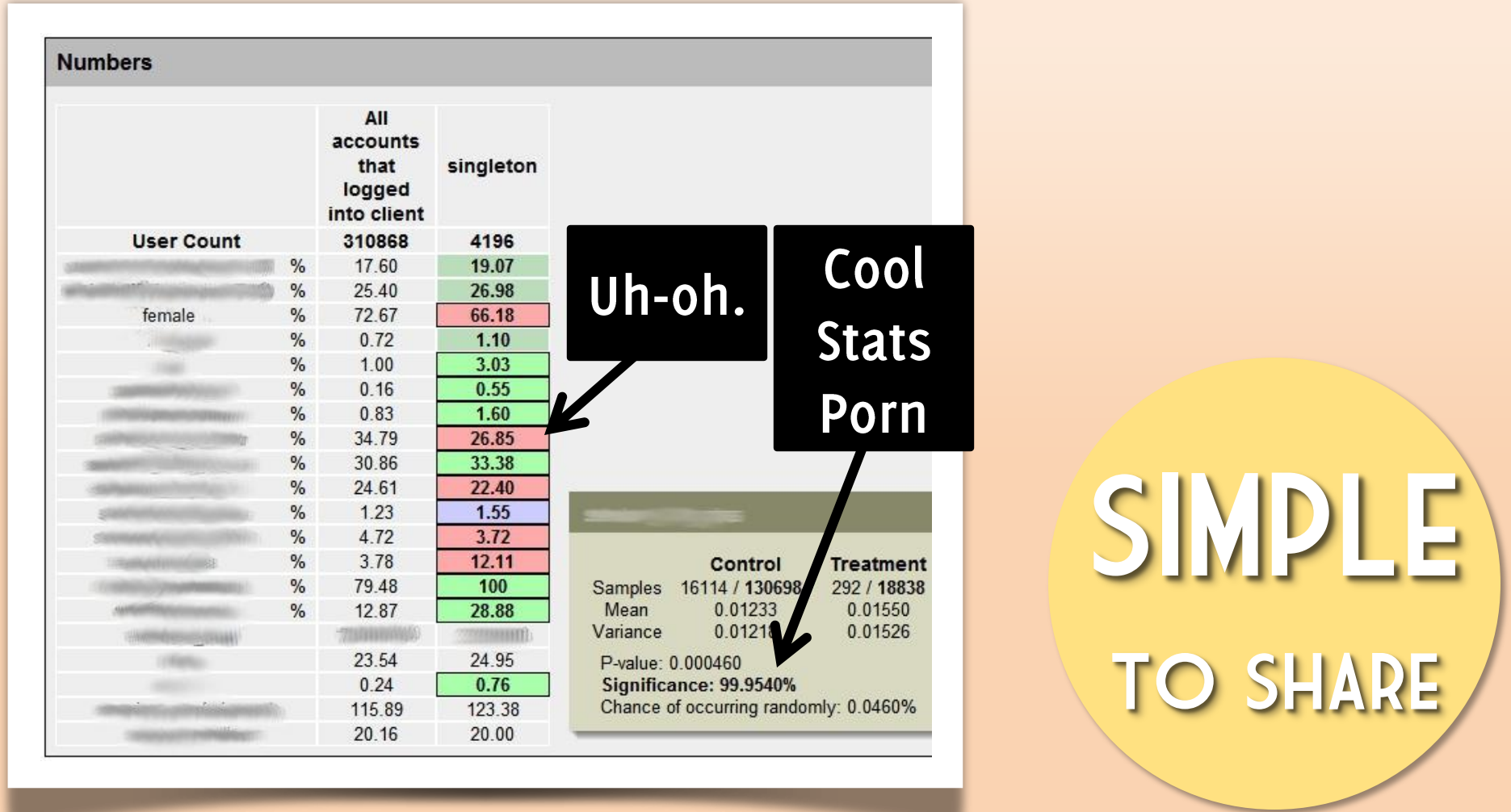
SIMPLE

TO CODE



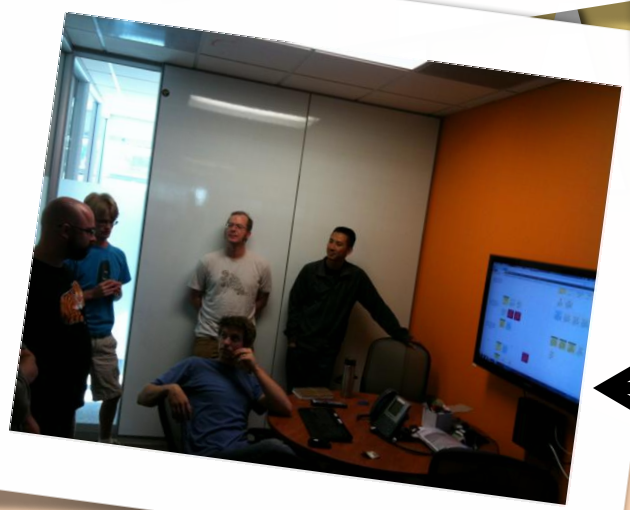
SIMPLE

TO MANAGE


Numbers
All accounts that logged into client
singleton
User Count
310868
4196
17.60
19.07
25.40
26.98
female
72.67
66.18
0.72
1.10
1.00
3.03
0.16
0.55
0.83
1.60
34.79
26.85
30.86
33.38
24.61
22.40
1.23
1.55
4.72
3.72
3.78
12.11
79.48
100
12.87
28.88
23.54
24.95
0.24
0.76
115.89
123.38
20.16
20.00
Uh-oh.
Cool Stats Porn
Control
Treatment
Samples
16114 / 130698
292 / 18838
Mean
0.01233
0.01550
Variance
0.01526
P-value: 0.000460
Significance: 99.9540%
Chance of occurring randomly: 0.0460%
SIMPLE
TO SHARE

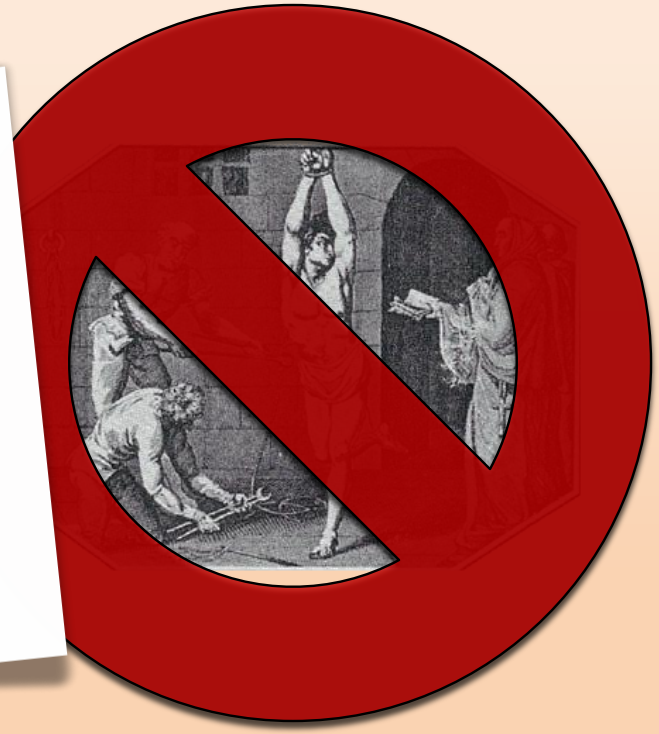
Suits

Company

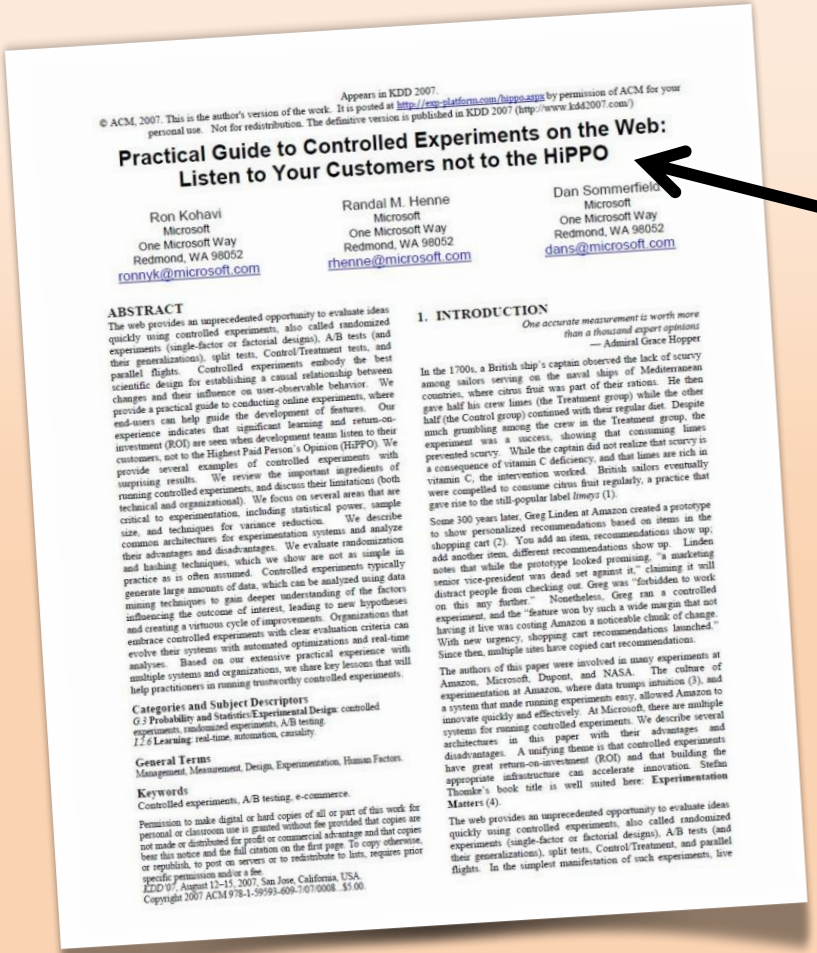


Teams

“Embrace failure as an opportunity to learn.”



HOW YOU RESPOND TO FAILURE SAYS A LOT ABOUT YOUR CULTURE



Appears in KDD 2007.

© ACM, 2007. This is the author's version of the work. It is posted at http://exp-platform.com/hippo.aspx by permission of ACM for your personal use. Not for redistribution. The definitive version is published in KDD 2007 (http://www.kdd2007.com/)

Practical Guide to Controlled Experiments on the Web: Listen to Your Customers not to the HiPPO

Ron Kohavi
Microsoft
One Microsoft Way
Redmond, WA 98052
ronnyk@microsoft.com

Randal M. Henne
Microsoft
One Microsoft Way
Redmond, WA 98052
rhenne@microsoft.com

Dan Sommerfield
Microsoft
One Microsoft Way
Redmond, WA 98052
dans@microsoft.com

ABSTRACT

The web provides an unprecedented opportunity to evaluate ideas quickly using controlled experiments, also called randomized experiments (single-factor or factorial designs), A/B tests (and their generalizations), split tests, Control/Treatment tests, and parallel flights. Controlled experiments embody the best scientific design for establishing a causal relationship between changes and their influence on user-observable behavior. We provide a practical guide to conducting online experiments, where end-users can help guide the development of features. Our experience indicates that significant learning and return-on-investment (ROI) are seen when development teams listen to their customers, not to the Highest Paid Person's Opinion (HiPPO). We provide several examples of controlled experiments with surprising results. We review the important ingredients of running controlled experiments, and discuss their limitations (both technical and organizational). We focus on several areas that are critical to experimentation, including statistical power, sample size, and techniques for variance reduction. We describe common architectures for experimentation systems and analyze their advantages and disadvantages. We evaluate randomization and hashing techniques, which we show are not as simple in practice as is often assumed. Controlled experiments typically generate large amounts of data, which can be analyzed using data mining techniques to gain deeper understanding of the factors influencing the outcome of interest, leading to new hypotheses and creating a virtuous cycle of improvements. Organizations that embrace controlled experiments with clear evaluation criteria can evolve their systems with automated optimizations and real-time analyses. Based on our extensive practical experience with multiple systems and organizations, we share key lessons that will help practitioners in running trustworthy controlled experiments.

Categories and Subject Descriptors
G.3 **Probability and Statistics/Experimental Design**: controlled experiments, randomized experiments, A/B testing
I.2.6 **Learning**: real-time, automation, causality.

General Terms
Management, Measurement, Design, Experimentation, Human Factors.

Keywords
Controlled experiments, A/B testing, e-commerce.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
KDD'07, August 12–15, 2007, San Jose, California, USA.
Copyright 2007 ACM 978-1-59593-609-7/07/0008...$5.00.

1. INTRODUCTION

One accurate measurement is worth more than a thousand expert opinions
— Admiral Grace Hopper

In the 1700s, a British ship's captain observed the lack of scurvy among sailors serving on the naval ships of Mediterranean countries, where citrus fruit was part of their rations. He then gave half his crew limes (the Treatment group) while the other half (the Control group) continued with their regular diet. Despite much grumbling among the crew in the Treatment group, the experiment was a success, showing that consuming limes prevented scurvy. While the captain did not realize that scurvy is a consequence of vitamin C deficiency, and that limes are rich in vitamin C, the intervention worked. British sailors eventually were compelled to consume citrus fruit regularly, a practice that gave rise to the still-popular label *limeys* (1).

Some 300 years later, Greg Linden at Amazon created a prototype to show personalized recommendations based on items in the shopping cart (2). You add an item, recommendations show up; add another item, different recommendations show up. Linden notes that while the prototype looked promising, "a marketing senior vice-president was dead set against it," claiming it will distract people from checking out. Greg was "forbidden to work on this any further." Nonetheless, Greg ran a controlled experiment, and the "feature won by such a wide margin that not having it live was costing Amazon a noticeable chunk of change. With new urgency, shopping cart recommendations launched." Since then, multiple sites have copied cart recommendations.

The authors of this paper were involved in many experiments at Amazon, Microsoft, Dupont, and NASA. The culture of experimentation at Amazon, where data trumps intuition (3), and a system that made running experiments easy, allowed Amazon to innovate quickly and effectively. At Microsoft, there are multiple systems for running controlled experiments. We describe several architectures in this paper with their advantages and disadvantages. A unifying theme is that controlled experiments have great return-on-investment (ROI) and that building the appropriate infrastructure can accelerate innovation. Stefan Thomke's book title is well suited here: **Experimentation Matters** (4).

The web provides an unprecedented opportunity to evaluate ideas quickly using controlled experiments, also called randomized experiments (single-factor or factorial designs), A/B tests (and their generalizations), split tests, Control/Treatment, and parallel flights. In the simplest manifestation of such experiments, live

The "highest paid person's opinion" (HiPPO) is not assumed to be correct.

Recap:
Culture is key.

FREEDOM TO EXPERIMENT

WHAT COULD POSSIBLY GO WRONG?

SIMPLE
TO
SCREW UP

1. Use cases

2. Test hypotheses

3. Rinse, repeat

“A lot of times, people don't know what they want until you show it to them.”

–Steve Jobs

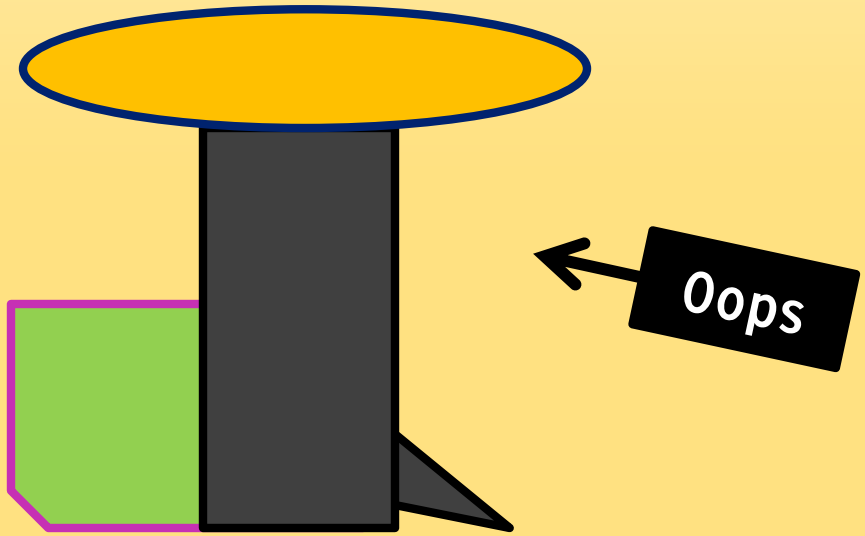
Start with use cases.

Don’t ask customers what they want.

Ask customers what they are trying to do.

WHAT NOT TO DO

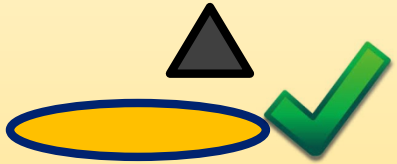
Build what customers say they want.



Experiment 1

A

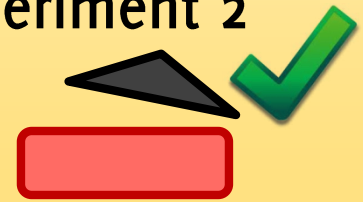
B



Experiment 2

A

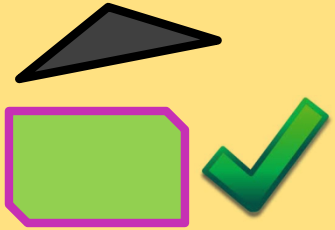
B



Experiment 3

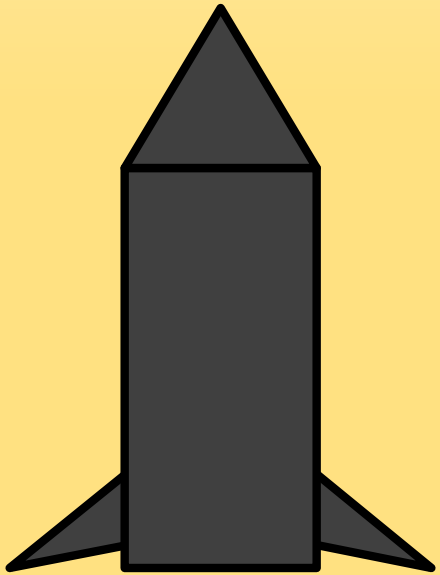
A

B



START WITH USE CASES

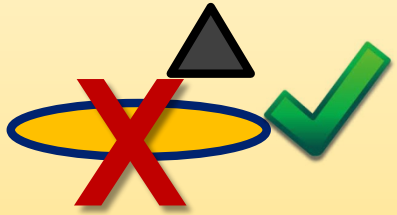
1. Start with use cases.
2. Test hypotheses to learn best ways to fulfill them.



Experiment 1

A

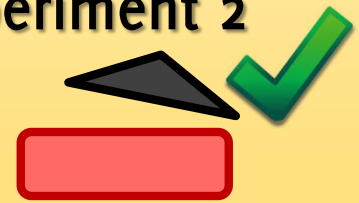
B



Experiment 2

A

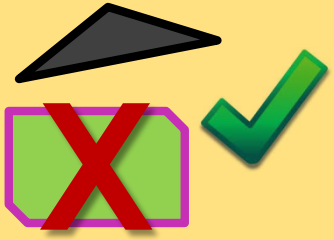
B



Experiment 3

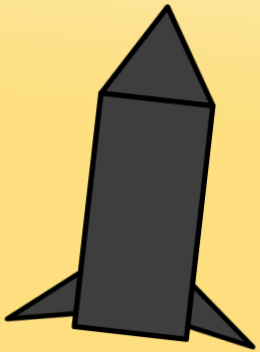
A

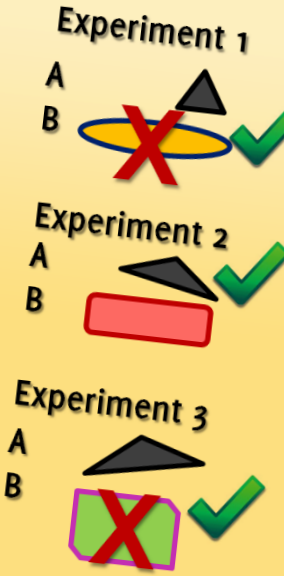
B



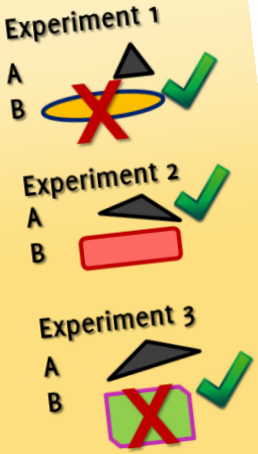
"Learning happens when you have the courage to validate—or invalidate—your hypothesis."

Systematic Experimentation: validate results against your hypothesis.

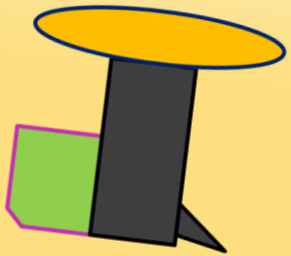


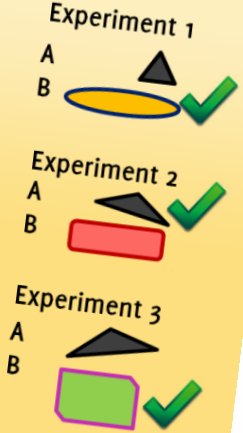


Systematic Experimentation: validate results against your hypothesis.



Build what customers say they want.





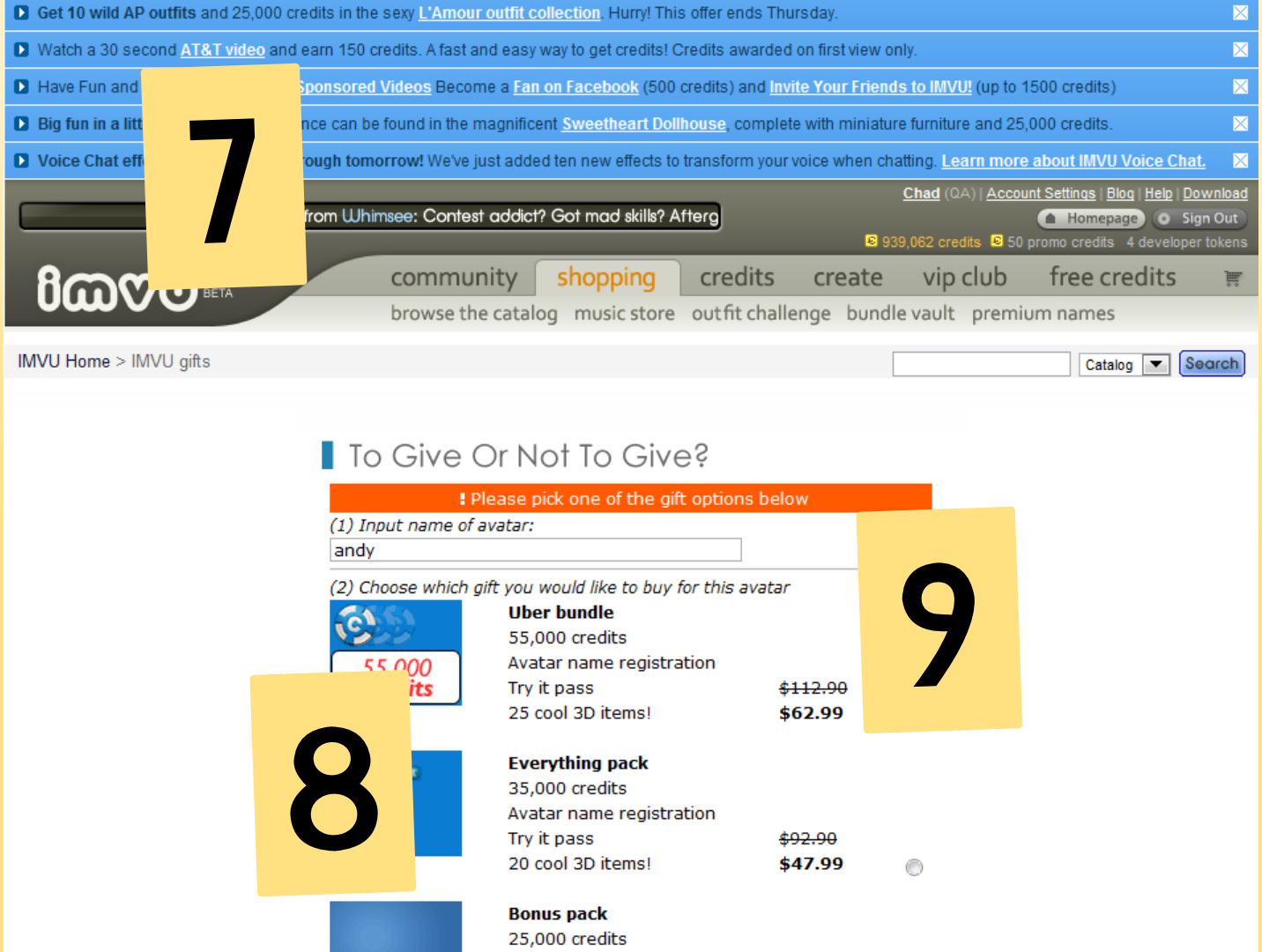
“Balancing your experience, instinct, and imagination against your experiment results is hard.”

“If you design a toaster oven and need to include directions for making toast, you have failed at designing a toaster oven.”

—Laura Klein, Principal, Users Know

My Room
1
Chat Now
2. VISIT YOUR HOMEPAGE
Design the hottest page on IMVU!
Homepage
3. FIND YOUR FRIENDS
Are your friends online?
Find Friends
Help | Tutorial
3
5
4
Report abuse
Chad
Text chat
Rooms
Inventory
Music
889,026 credits
0 promo credits
dress
decorate
tops
2

IMVU Messenger
longerrrr name
longer name
hi
Credits: 500
Refresh
TestName
home page
Buddies
Inventory
Male
catalog
Manly pants 1
put on
nosub2
Shoes
Womanly shoes 1
Womanly shoes 1b
take off
Woshoes 2
take out
Woshoes 2b
put back
Woshoes 3
apply
Woshoes 3b
get info
Scenes
View deleted items
Chat now!
Add a buddy
Find a friend
Shop catalog
3D window
My inventory
New IM
Buy Credits
6


Get 10 wild AP outfits and 25,000 credits in the sexy L'Amour outfit collection. Hurry! This offer ends Thursday.
Watch a 30 second AT&T video and earn 150 credits. A fast and easy way to get credits! Credits awarded on first view only.
Have Fun and
Sponsored Videos Become a Fan on Facebook (500 credits) and Invite Your Friends to IMVU! (up to 1500 credits)
Big fun in a litt
nce can be found in the magnificent Sweetheart Dollhouse, complete with miniature furniture and 25,000 credits.
Voice Chat eff
ough tomorrow! We've just added ten new effects to transform your voice when chatting. Learn more about IMVU Voice Chat.
7
from Whimsee: Contest addict? Got mad skills? Afterg
Chad (QA) | Account Settings | Blog | Help | Download
Homepage
Sign Out
939,062 credits 50 promo credits 4 developer tokens
imvu BETA
community
shopping
credits
create
vip club
free credits
browse the catalog
music store
outfit challenge
bundle vault
premium names
IMVU Home > IMVU gifts
Catalog
Search
To Give Or Not To Give?
! Please pick one of the gift options below
(1) Input name of avatar:
andy
(2) Choose which gift you would like to buy for this avatar
Uber bundle
55,000 credits
Avatar name registration
Try it pass
$112.90
25 cool 3D items!
$62.99
9
8
Everything pack
35,000 credits
Avatar name registration
Try it pass
$92.90
20 cool 3D items!
$47.99
Bonus pack
25,000 credits

My Room
Get 10 wild AT outfits and 25,000 credits in the sexy L'Amour outfit collection. Hurry! This offer ends Thursday.
Watch a 30 second AT
Have Fun and
Big fun in a litt
Voice Chat eff
IMVU! (up to 1500 credits)
ure and 25,000 credits.
arn more about IMVU Voice Chat.
Account Settings | Blog | Help | Download
Homepage
Sign Out
50 promo credits
4 developer tokens
free credits
mium names
IMVU Home > IMVU gifts
Catalog
Search
Credits: 500
Refresh
home page
catalog
put on
apply
get info
1. M
2. VISIT YOUR
Design the h
page on IM
3. FIND YOUR
Are your frie
Help | Tutoria
Text chat
dress
deco
GLOVES
10!
25,000 credits

Home - IMVU
65,702 people online now from 86 countries
james
47,841 credits
Settings | Help | Sign Out
Friends Only
Home
Games Lobby
Chat Rooms
Shop
Dress Up
Home
PULSE
Dress Up
Shop
Credits
Chat Now
Chat Rooms
My Room
7
Inbox
Games
Friends
How do you like the latest 3D chat? Please send feedback

Experiment Design

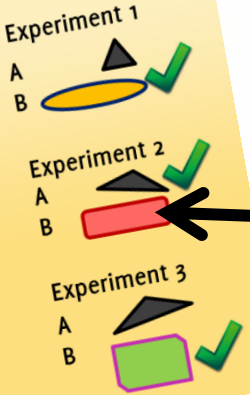
1. Start with use cases
2. Have a specific hypothesis
3. Ask the right questions
4. Ask the right people
5. Ask enough people
6. Randomized control trial

Start with use cases

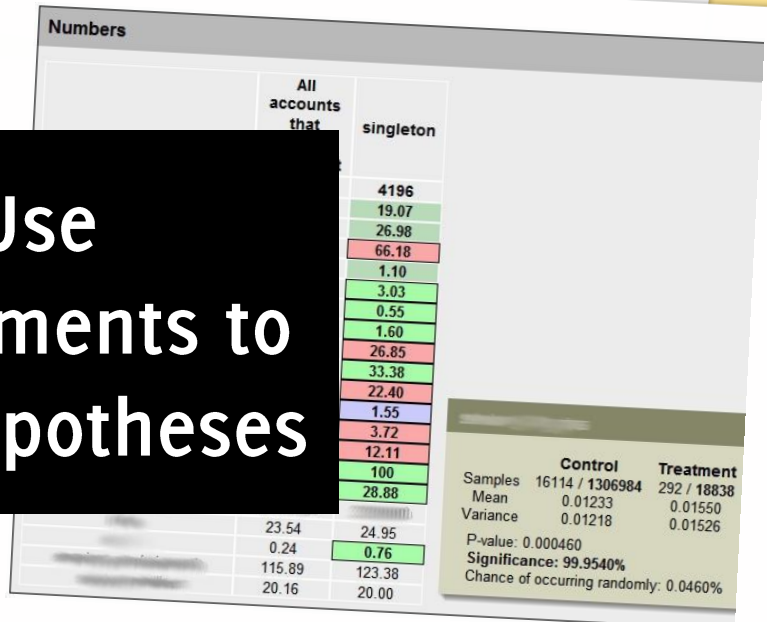
Experiment Interpretation

- Use the right *p* test
- Don't rely only on *p* test
- Don't confuse correlation & causation

Build what customers say they want.

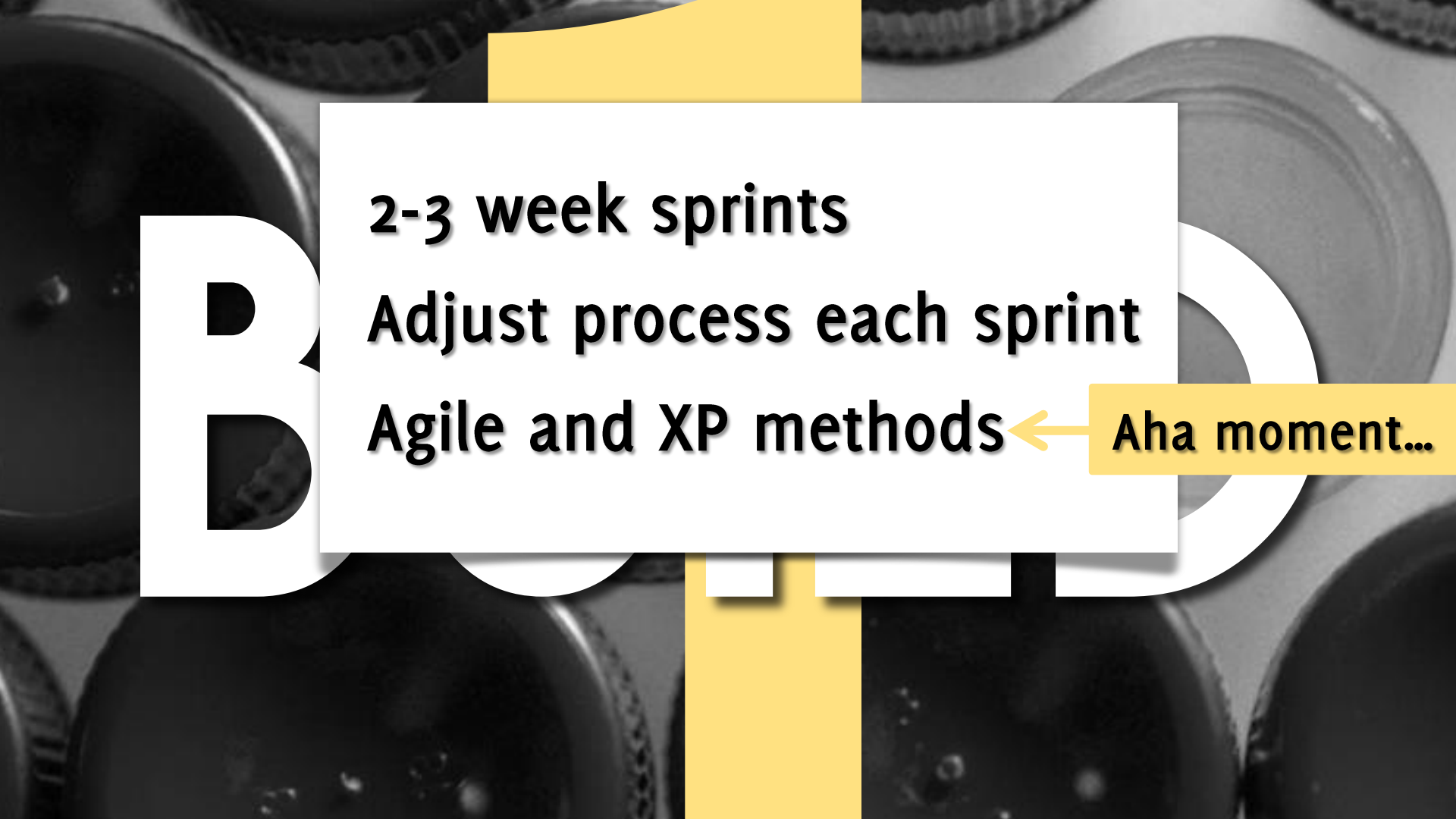


Use experiments to test hypotheses



"Experiments are a great way to test hypotheses; not to form them."

PRODUCT DEVELOPMENT


2-3 week sprints
Adjust process each sprint
Agile and XP methods
Aha moment...

Agile & XP Methods

1. Embrace change (and try to make it painless)
2. Prefer flexibility to perfection
3. Iterate, improve continuously
4. Feed data and in...
5. ...ment (TDD)
6. ...
7. Use defects to drive improvements
8. Rapid, continuous code integration and deployment
9. Software cluster immune system: prevent defects
10. Pair Programming
11. Just-in-time scalability
12. Refactoring

"Change, flexibility, iteration, continuous improvement..."

Learning:

Agile and XP support experimentation.

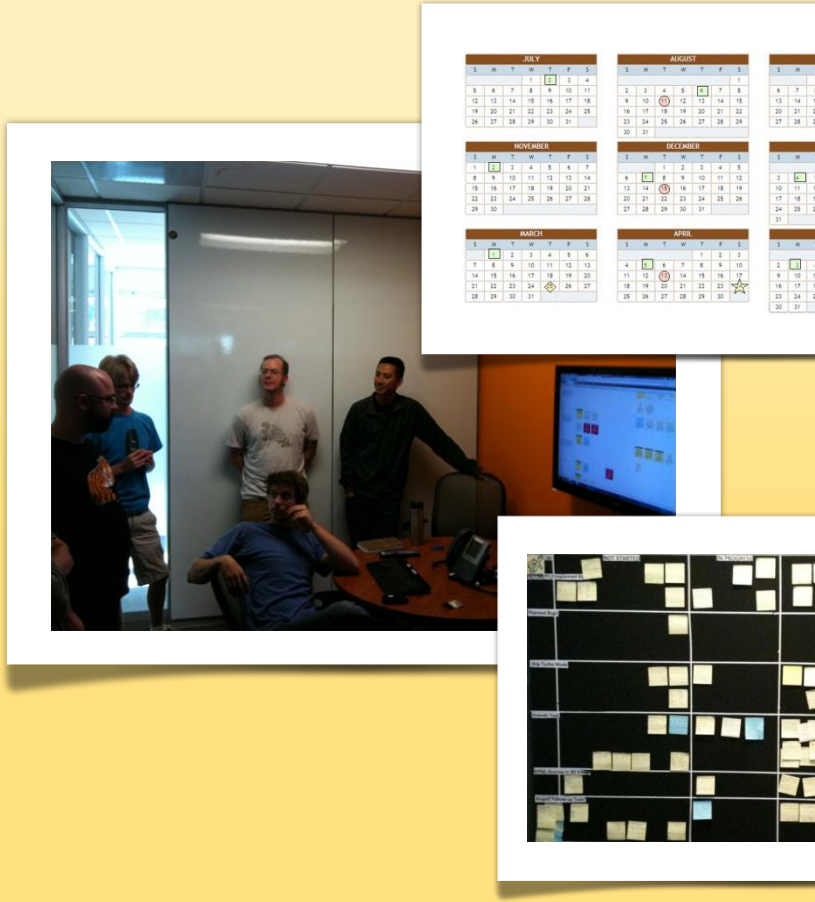
Scrum: A project management methodology, works well with Agile/XP.

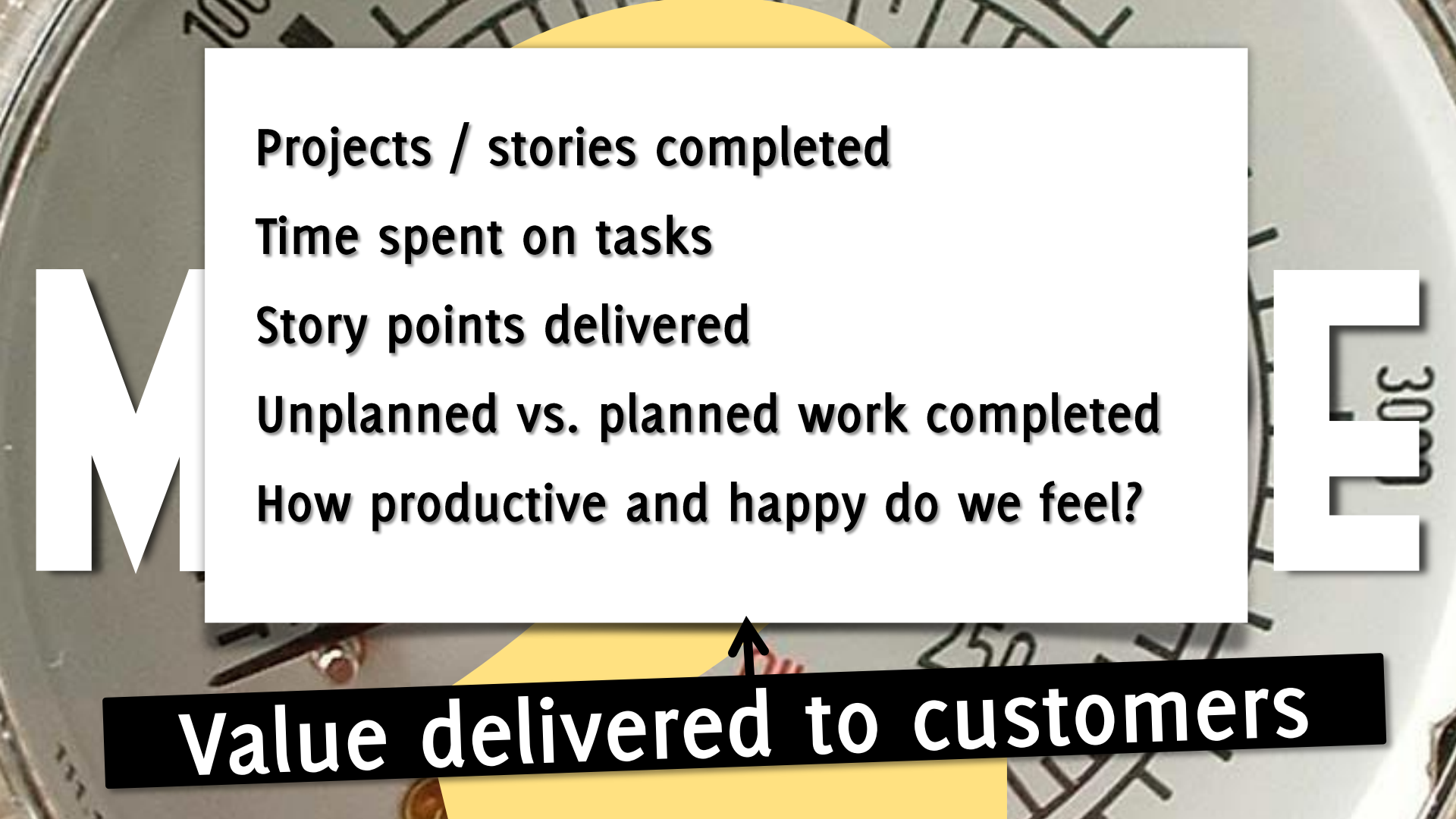
IMVU SCRUM V1.0

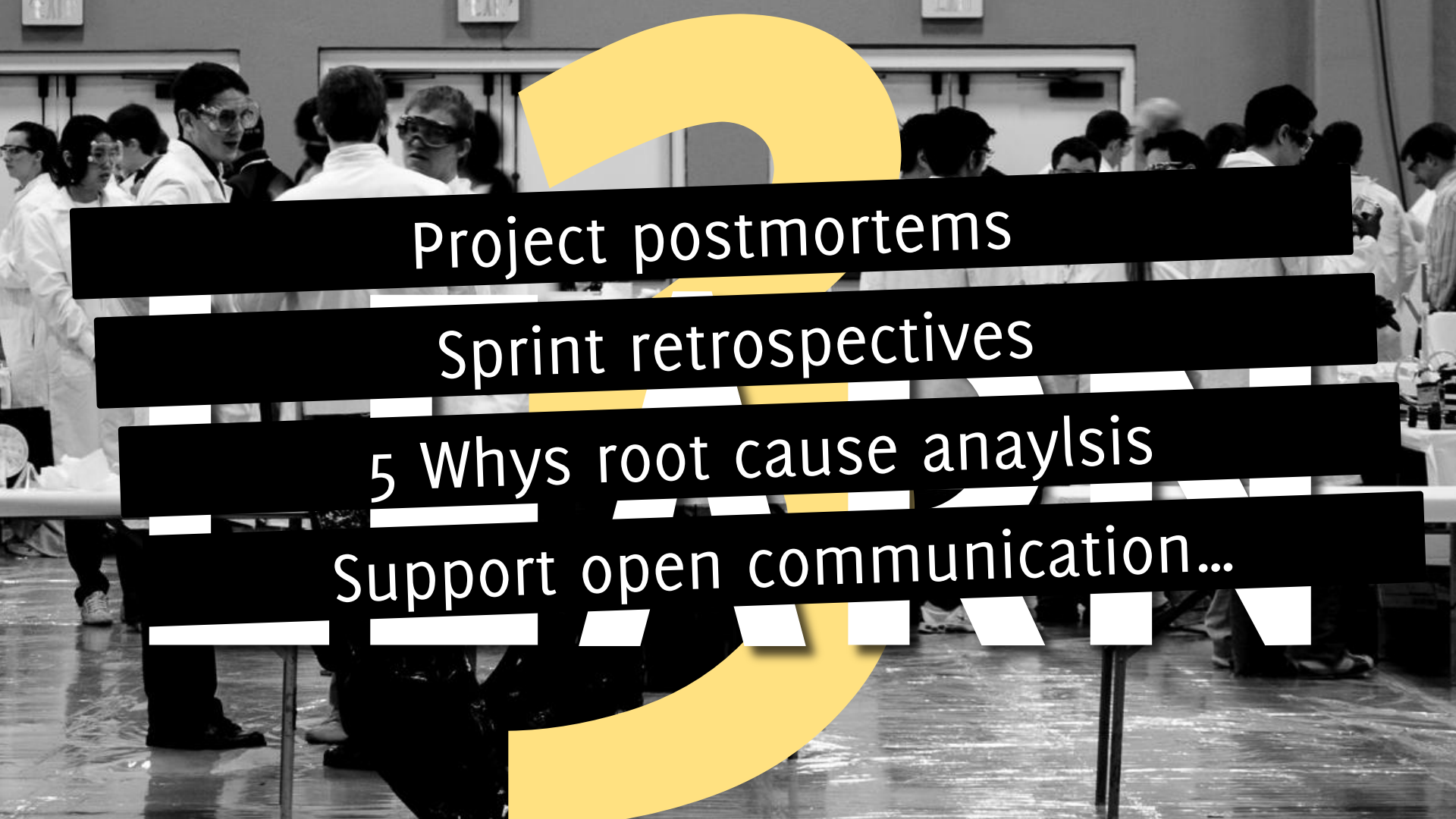
1. Self-organizing teams, short sprints, daily standups FTW!

2. Clear roles+responsibilities: Product Owner, Tech Lead

3. Stay flexible, don't be dogmatic.




Projects / stories completed
Time spent on tasks
Story points delivered
Unplanned vs. planned work completed
How productive and happy do we feel?
Value delivered to customers


Project postmortems
Sprint retrospectives
5 Whys root cause anaylsis
Support open communication...

Engineering project managers
Matrix management
Scrum of Scrums
Team swaps
Open floor plan
Place value on communication

Postmortems & Retrospectives:

1. Meeting roles
2. Metrics
3. Action items

Let's see it in action!

Appoint a skilled facilitator.

Foster communication and engagement.

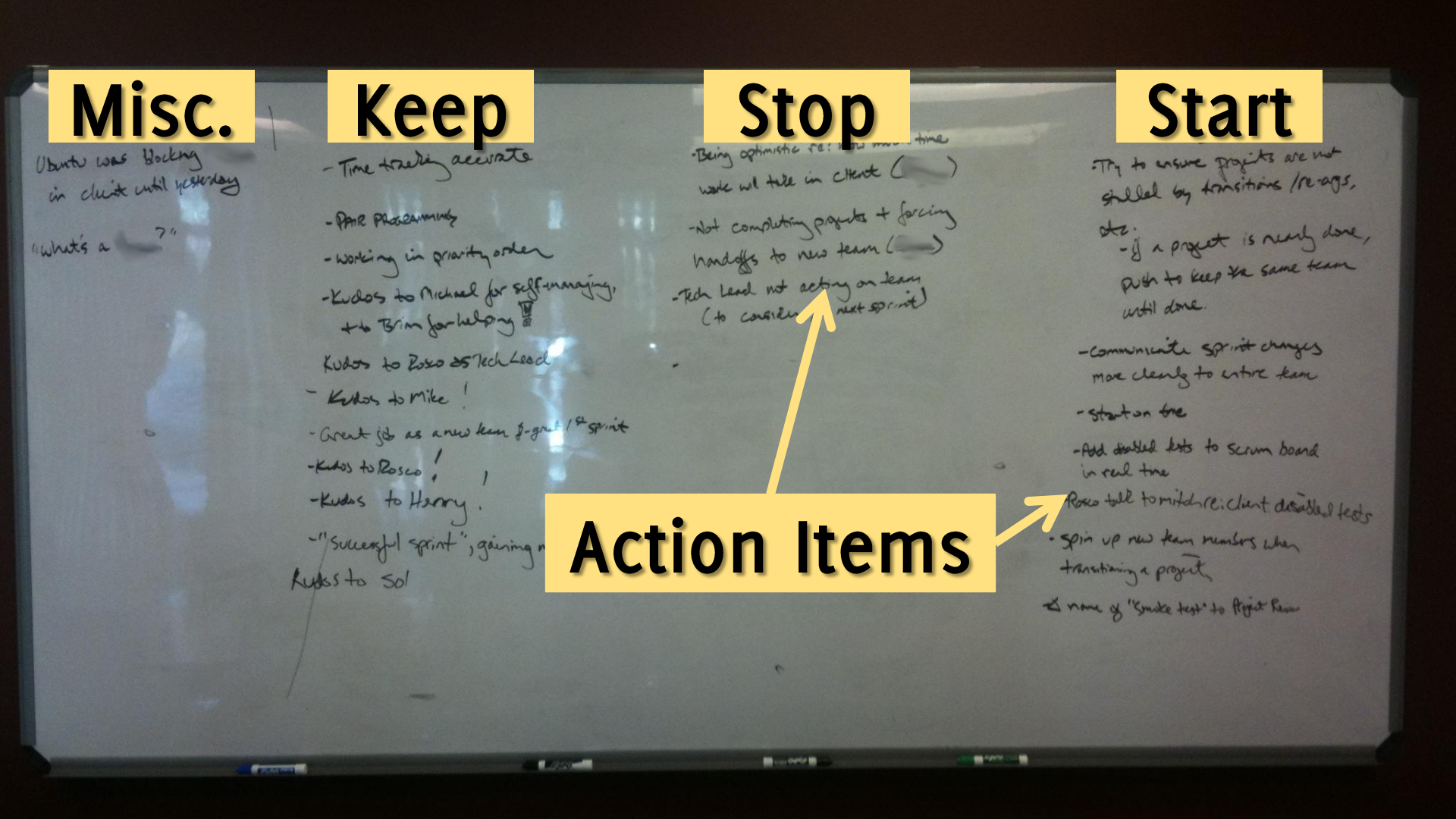
C4 2010 S14 Retrospective

Story Points

Days	Project	Story Points
5 + 12	Interrupts + Pulled Builders	0/0
1	P1 & P1 Followup	0/0
1	Farmandia	2/2
0	FTUX 1	0/0
3	FTUX 2	1/1
20	Pulse Phase 2	5/5
5	Pulse 2 Spam Features	5/5
3	Bugs	3/3
1	Backyard Monsters	2/2
2	2nd FTUX Experiment	2/2
10	Project Follow-up	0/0
	Total Days is 53, taking out for days already accounted for in planning	20

Misc

- Loaned Jinsuck to Loco Team for 4 days
- Added work mid-sprint (Pulse)
- BB better starting mid-sprint
- Postmortems: Pulse Phase 2


Misc.
Keep
Stop
Start
Action Items
Ubuntu was blocking in client until yesterday
"What's a ?"
- Time tracking accurate
- Pair programming
- working in priority order
-Kudos to Michael for self-managing, to Brian for helping
Kudos to Rosco as Tech Lead
- Kudos to Mike !
- Great job as a new team & - great 1st sprint
-Kudos to Rosco !
-Kudos to Harry .
-"Successful sprint", gaining
Kudos to Sol
-Being optimistic re:
work will take in client ( )
-Not completing projects + forcing handoffs to new team ( )
-Tech Lead not acting on team
-Try to ensure projects are not stalled by transitions / re-orgs, etc:
- if a project is nearly done, push to keep the same team until done.
-communicate sprint changes more clearly to entire team
- start on time
-Add detailed tests to scrum board in real time
Rosco talk to mitch re: client detailed tests
- spin up new team members when transitioning a project
name of "Smoke test" to Project Review

5 WHYS

5 WHYS

Y1: Why were logins failing?
A1: The database has too many connections.

Y2: Why did the Database have too many connections?
A2: A new feature had slow-running queries.

Y3: Why were there slow queries?
A3: They were implemented with a cache time to zero.

Y4: Why was cache time set to zero?
A4: James didn't know that slow queries must be cached.

Y5: Why didn't James know about caching slow queries?
A5: He's new, and we didn't cover that in our training.

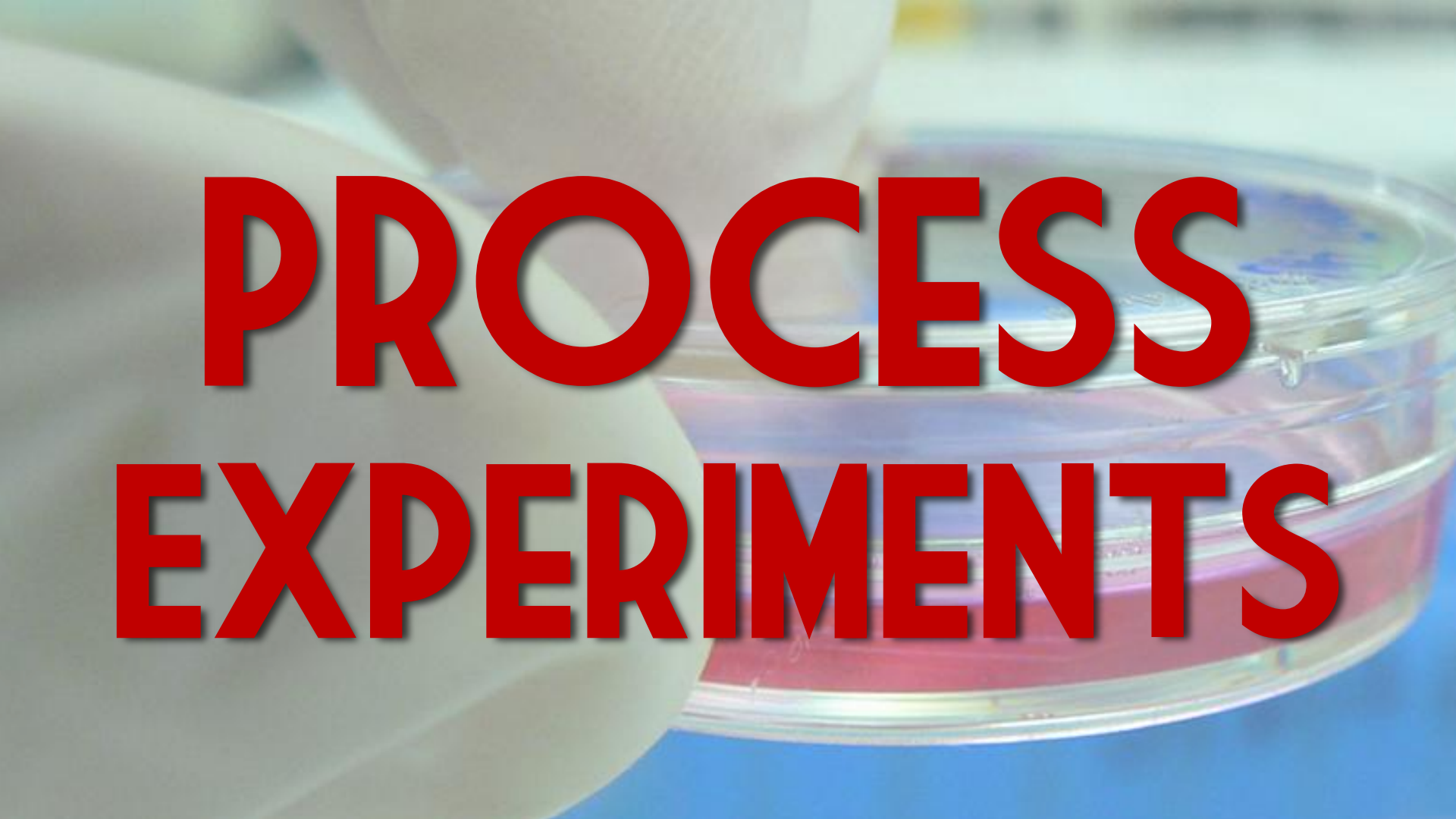
ROOT CAUSE:
FIX THIS!

A couple notes about action items:

1. Fix root causes.

2. Make the size of the fix commensurate with the size of the problem.

Recap:
Learning is key.


PROCESS
EXPERIMENTS

DOGMA

HANDLING INTERRUPTS

1. Share interrupts among teams

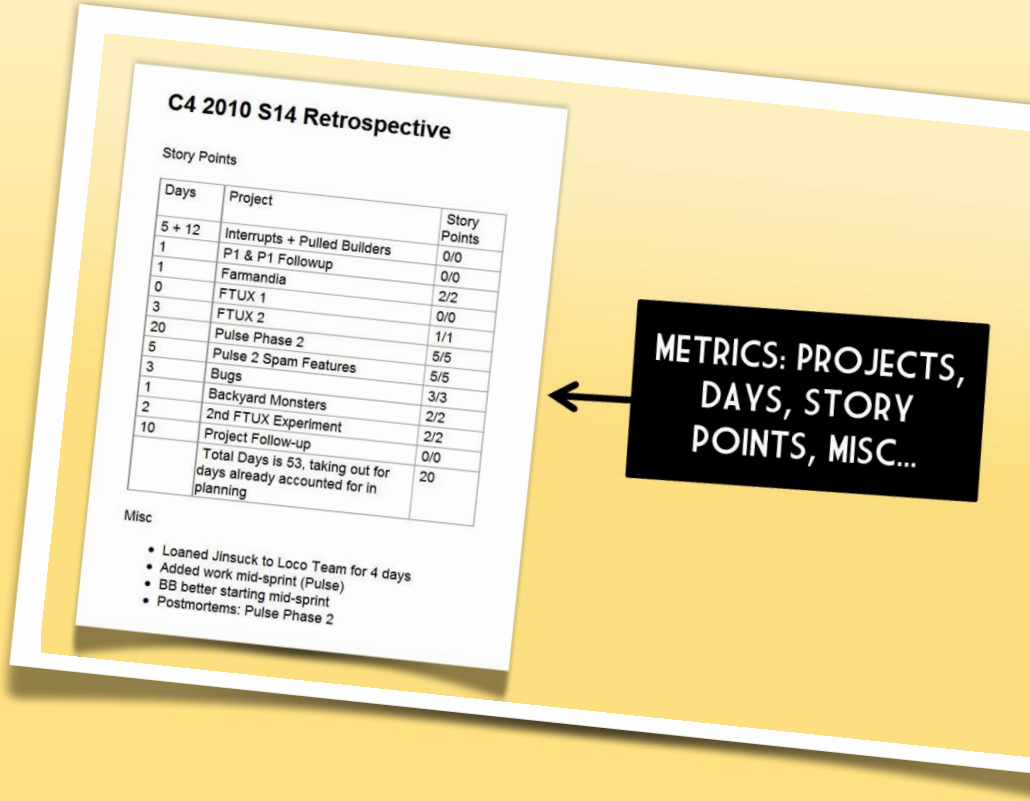
← Everyone feels the pain, all team schedules at risk

2. Create “Interrupts” team

← Establish “home” team for ICs, and rotate them here; reduces risk for all other teams

MEASURING VALUE TO CUSTOMERS

Are we working on the right projects?



C4 2010 S14 Retrospective

Story Points

Days	Project	Story Points
5 + 12	Interrupts + Pulled Builders	0/0
1	P1 & P1 Followup	0/0
1	Farmandia	2/2
0	FTUX 1	0/0
3	FTUX 2	1/1
20	Pulse Phase 2	5/5
5	Pulse 2 Spam Features	5/5
3	Bugs	3/3
1	Backyard Monsters	2/2
2	2nd FTUX Experiment	2/2
10	Project Follow-up	0/0
	Total Days is 53, taking out for days already accounted for in planning	20

Misc

- Loaned Jinsuck to Loco Team for 4 days
- Added work mid-sprint (Pulse)
- BB better starting mid-sprint
- Postmortems: Pulse Phase 2

METRICS: PROJECTS, DAYS, STORY POINTS, MISC...

1. Focus on features and value to customers , not time.

2. Short planning meetings.

3. Caution: Reduced ability to predict progress.

KEEP YOUR TEAM TOGETHER

1. Embed designers, QA, data analyst, engineering and product: location matters!

- Yes, make your team a “team”
- Communication easier
- Issues/questions resolve quickly

TECHNICAL PROJECT REVIEWS

1. Organized by team technical lead—this is a technical review of the code: improve quality, reduce “tech debt”

- In-depth review of code
- All engineers welcome
- Note follow-up action items
- Prioritize actions, and do them!

SELF SELECTING TEAMS

1. Let people self-select to the teams they want to work on.

- Let people follow their passions
- Happy people are more engaged and productive

SCRUM 2.0: CONTINUOUS PLANNING

1. Based on successful process experiment
2. Just-in-time planning: Kanban
3. Are we working on the right stuff?

HAVE A "FIXER"

1. Remove blocks: communication, technical, planning, anything else!
2. Remain objective
3. Team happiness = team productivity

SMALL TEAM + NO RULES=BIG WIN

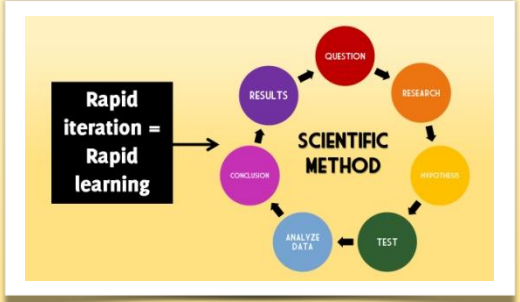
1. Break out of routine
2. Minimize process overhead
3. Focus on value to customers

RECAP

1. Scientific method and experimentation FTW!

2. Culture is key.

3. Learning is key.



CREDITS

1. flickr/boule de plasma/abdallahh
2. flickr/Mouths/The Wandering Angel
3. flickr/Culture Tubes/Hey Paul
4. flickr/Test Tubes/Håkan Dahlström
5. flickr/Tablet packet macro/nick@
6. flickr/11-08-09/idovermani
7. flickr/Gauge/Hey Paul
8. flickr/Cell Culture/kaibara87
9. IMVU Top Secret Garage/Photo courtesy of Robert Otani
10. dukenostalgia.com/Mad Scientist Game
11. flickr/Acceleration of a bouncing ball is constant/dullhunk
12. commons.wikimedia.org/blue ribbon/Gothika
13. commons.wikimedia.org/ Ebers7766.jpg/ignacio icke
14. Vito's Dancers/Photo courtesy of Ken Patterson
15. Build-Measure-Learn courtesy of Eric Ries http://www.startuplessonslearned.com/
16. flickr.com/books/rodrigo galindez
17. flickr.com/cowboys/anyjazz65
18. slideshare.net/jamesbirchler Scaling Product Development at a Lean Startup: Agile at IMVU
19. flickr.com/neanderthal/hamed saber
20. flickr.com/Clock/Jörg Weingrill
21. flickr.com/Open Tattered Book/playingwithbrushes
22. Practical Guide to Controlled Experiments on the Web: Listen to Your Customers not to the HiPPO, by Ron Kohavi, Randal M. Henne, and Dan Sommerfield, @Microsoft, KDD 2007.

THANKS

Jesse Imbach, Josh Adlin, Joe Durand, Lene Harbott, and Debora Wrathall provided excellent advice, ideas, laughs, and gourmet food while I was preparing this talk—thanks you guys!

Brett Durrett, Roland Blanton, Steven Peterson, and Matt Danzig provided amazing support and useful feedback that greatly improved this talk.

Eric Ries provided invaluable guidance and input—thanks!

Finally, thanks to all of my colleagues and friends at IMVU who make all this possible—you rock!

James Birchler
Engineering Director, IMVU
@jamesbirchler
james@imvu.com