

Testing in Production

Your Key to Engaging Customers

Session #801

Seth Eliot, Senior Test Manager



Software Test Professionals Conference – March 23, 2011

Contents

Introduction

Scenarios

Testing in Production (TiP)

The Scenario – TiP Connection

Data Mining

Online Experimentation

Active User Feedback

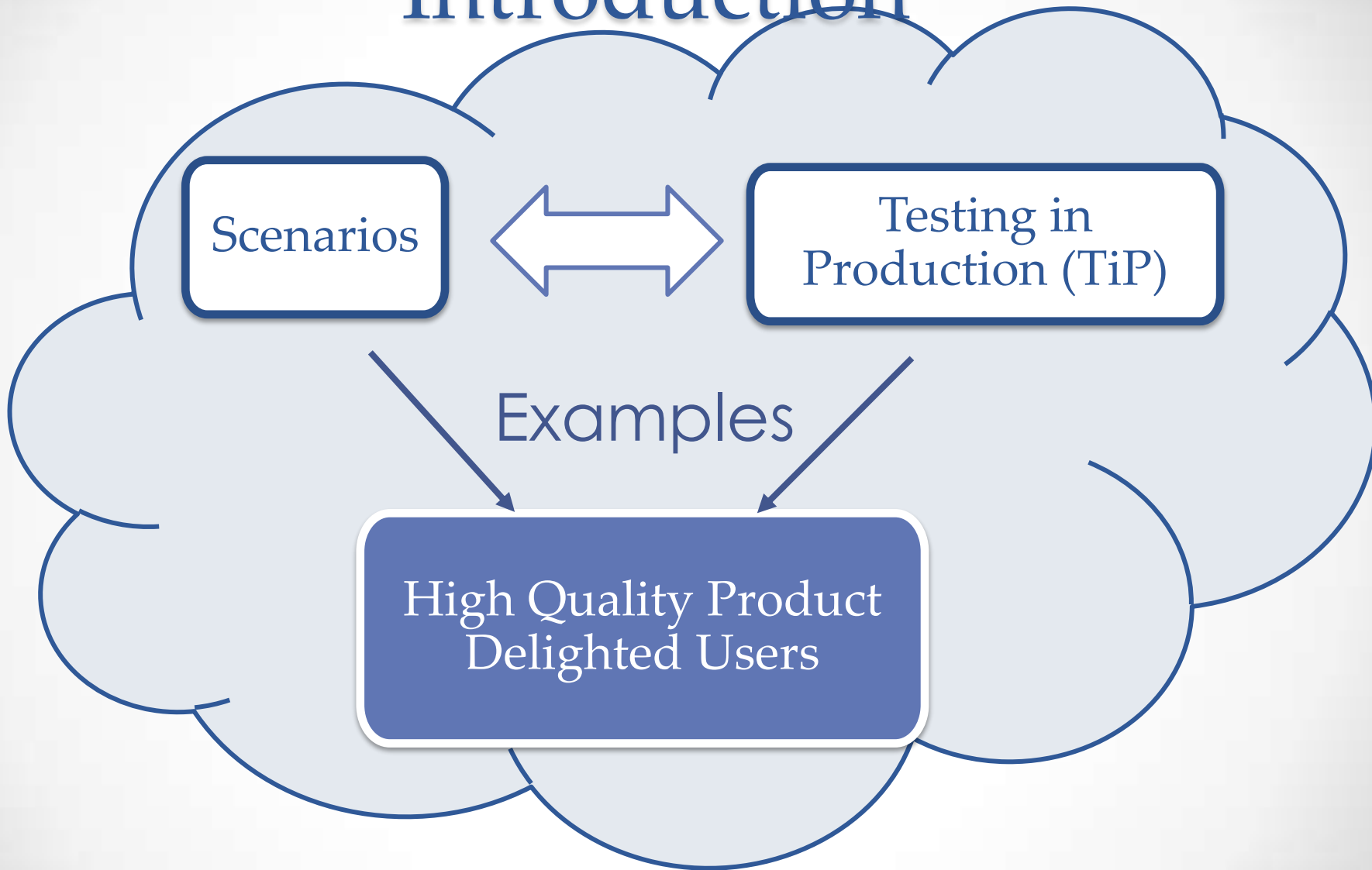
Prototypes

Conclusion

Latest version of this slide deck can be found at:

<http://setheliot.com/blog/software-test-professionals-conference-2011/>

Introduction



Who am I?

Software QA Manager

Amazon Digital Media



Microsoft®



Microsoft Experimentation Platform



Microsoft Bing

Culture Shift

- Services
- Data Driven



Scenarios

• • •

An Introduction

Ditch the Requirements – Focus on the Customer Instead



- Static requirements are fragile
- Requirements are hard: GIGO
- If they are there at all....

http://bit.ly/seth_stp_scenario

Will We Win On "Feature Wars"?

Excel Excels—But Others Have Talents Too

Best Buy Excel 3.0 is a snap to use and stacked with features, while 1-2-3 3.1* gives you multipage worksheets and 3-D range names, and Quattro Pro 3.0 packs spreadsheet power into older systems.

- Yes
- No
- ◐ Delimited
- ▼ Limited only by available memory

Standard features

Price	SHI
Network file locking	●
RAM required	SHI†
Virtual memory management	○
Movable, visible windows	○
Undo multiple operations	○
Print in background	●
File viewer	●

Decision support

Sort by columns	○
Insert/delete range	○
Select, change multiple ranges	○
Select range for multiple operations	●
Outline	○
Cell and range notes	○
Trace formula references	●
Password-protect worksheet structure	○
Solver: multivariable linear	○
Solver: multivariable nonlinear	○
Multipage worksheets	○
3-D ranges	○
Autorecalculation	○
Save work space	○
Point-and-click formula linking	○
Hot links to other applications	○
Define range name with external reference	○

Extended graph ranges

Linear regression	○
Convex/concave regression	○
Arrays	○
User-definable functions	○
Paste function in formula	○
Reads/writes DBF (dBASE III)	○
Reads/writes comma-delimited ASCII	○

Presentation

Automatic column width	○
Justify numeric data	○
Variable row height	○
Restrict cell entry to text	○
User-definable formats	○

Presentation (continued)

Copy formatting to another cell	○
Movable WYSIWYG interface	○
Zoom	○
Text boxes	○
Save named reports	○
Set attributes independent of font	○
Formatting styles	○
Auto-compress print range	○
Print multiple ranges per page	○
WYSIWYG page preview	○
Display print range border	○
Draw within graphics	○
Draw on spreadsheet	○
Draw axes	○
Mixed graphs (line/bar)	○
True 3-D graphs	○
Onscreen slide show	○
Number of graph variables	○

Data management

Date entry forms	○
Extract records from closed file	○
Relational queries	○
Calculated query fields	○
External data base links	○

Programmability

Macro library	○
Execute macro with mouse button	○
DOS command access from macros	○
Debugger/trace	○
Reads/writes 1-2-3 2.2 macros	○

Service and support

Daily support (hours)	○
Unlimited free support	○
Telephone support number	○
Website support	○
BBS support	○
Fax support	○
Other extra-cost training or support	○

WYSIWYG WORD PROCESSORS			
The WYSIWYG packages are divided into two groups; the PC packages are on the left, and the Mac packages are on the right. Only a few of the packages, such as Ami Professional and Microsoft Word (● = yes; ○ = no).			
Product name	Ami Professional 1.2	BetterWorking Word Publisher 5.0	InText 1.53
Price	\$495	\$59.95	\$295
Hardware needed	PC†	PC†	PC†
Editing			
Spelling checker	●	●	○
Thesaurus	○	○	○
Outline	○	○	○
Mail merge	○	○	○
Undo	○	○	○
Math functions	○	○	○
Equation editor	○	○	○
Automatic timed save	○	○	○
Control sensitive text	○	○	○
Unit mode	○	○	○
Minimum no. of windows	13	12	1
WYSIWYG interface	○	○	○
WYSIWYG interface	○	○	○
Long document features	○	○	○
Create index and contents	○	○	○
Cross referencing	○	○	○
Page headers	○	○	○
Customization tools	○	○	○
Macro editor	○	○	○
Macro programming language	○	○	○
Customized main menu	○	○	○
Dynamic data exchange links	○	○	○
Groupwork editing	○	○	○
Document outline	○	○	○
Strike-through	○	○	○
Document structure	○	○	○
Formatting			
Style sheets	○	○	○
Conditional page breaks	○	○	○
Window/page control	○	○	○
Automatic hyphenation	○	○	○
Linking columns	○	○	○
Side-by-side columns	○	○	○
Forming and tracking control	○	○	○
Document queries	○	○	○
Print in background	○	○	○
Graphics			
File formats imported	○	○	○
Save/restore graphics	○	○	○
Move graphics frames	○	○	○
Flow text around graphics	○	○	○
Page preview	○	○	○
Lines, boxes, borders	○	○	○
Drawing program	○	○	○
Table editor	○	○	○

Minimum system requirements

- *286, 640K bytes of RAM, graphics card (Hercules, CGA, EGA, or VGA), DOS 3.0, mouse, and a hard disk drive.
- †386, 1.0MB bytes of RAM, graphics card (DOS 3.1, mouse for draw mode, and one floppy disk drive).
- ‡386, 1.0MB bytes of RAM, graphics card (DOS 3.1, two floppy disk drives, and a hard disk drive).
- §386, 1.0MB bytes of RAM, graphics card (DOS 3.1, and a hard disk drive).
- ||386, 1.0MB bytes of RAM, graphics card (DOS 3.1, and a hard disk drive).
- ¶386, 1.0MB bytes of RAM, 1 MB of expanded RAM, graphics card (EGA or better recommended, DOS 3.0, and a hard disk drive).
- ||386 Plus, 1 MB of RAM (2 MB recommended), a hard disk drive, and an 800K byte floppy disk drive.
- ||386 Plus, two 800K byte floppy disk drives, and System 6.0.



September 1990

PCWorld
September 1991

Giddy Customers?

Conformance
to Requirements

Fitness for Use

Giddy
Customers



Useful

- Solves a Need



Usable

- easy and intuitive
- discoverable
- learnable
- efficient



Desirable

- Enjoy
- Crave

Scenario

- ✓ A narrative **Story**
- ✓ Told with the **Customer's** voice
 - Tells what they want to achieve
- ✓ Based on real **Research**
- ✓ Intentionally **Implementation-Free**
- ✓ That taps into **Emotion** and reveals **Insight**

An Example Scenario

- ✓ A narrative **Story**

Danielle is an aspiring author. She composes her works, consisting of short form and long form fiction, directly on her laptop computer. Her work is her livelihood so data loss is devastating. Since she is composing on the fly anything that interrupts her flow of creativity frustrates her, so meta-tasks like spelling and formatting should be handled as transparently as possible. She often shares unfinished versions of her work with peer reviewers for feedback, and ultimately needs to share her finished works with editors and potential publishers

We Are Not The Customer

✓ Told with the **Customer's** voice

ACE
The helpful place.

Dear Consumer:
Thank you for participating in this promotion
Unfortunately we could not honor your
request due to the following reason(s):
Ace January 2011 Offer
- Ace SKU not found

I get the following in response to logging into the HP software support site.

`java.lang.Exception: Sign in again too soon, may be due to browser cookie issue`



This operation returned because the timeout period expired.

OK

Cancel

Dreamweaver



Names can only contain low-ASCII letters and numbers and cannot begin with a number.

OK

Who is the Customer?

- ✓ Based on real **Research**

What Do Real Users Need?

Ask / Watch / Study them

- For proposed website changes at Microsoft

1/3	1/3	1/3
Positive Ideas	No Statistical Difference	Negative Ideas

[Kohavi, et al; 2009]

- Research enables us to weed out the 2/3rd of ideas that do not work

✓ Intentionally Implementation-Free



automatically file share
photos directory with
grandparents

print photos and give
to them

print photos and put
in album

put photos
on a website



John wants to
share the
latest kid
pictures with
grandparents

easy template for
emailing



email
slideshow

online
slideshow

send photo via instant
messaging



see pictures for
people in my
social network



show pictures on
TV screen

put photos on a
CD

easy to use kiosk
in a retail outlet

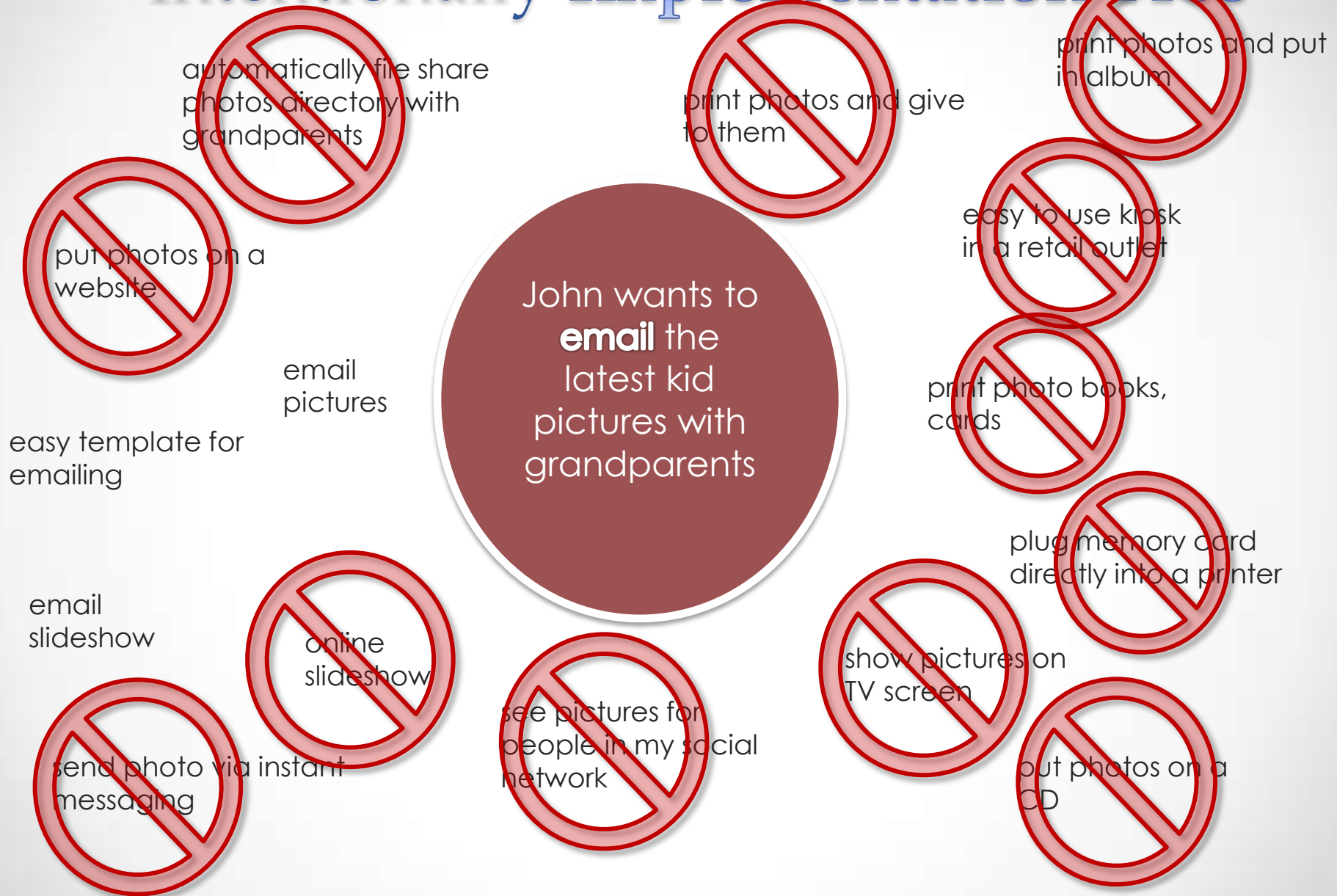


print photo books,
cards

plug camera
directly into
a printer



✓ Intentionally Implementation-Free



An Example Scenario - Revisited

- ✓ That taps into **Emotion** and reveals **Insight**

Danielle is an **aspiring** author. She composes her works, consisting of short form and long form fiction, directly on her laptop computer. Her work is her **livelihood** so data loss is **devastating**. Since she is composing on the fly anything that interrupts her flow of creativity **frustrates** her, so meta-tasks like spelling and formatting should be handled as transparently as possible. She often shares unfinished versions of her work with peer reviewers for feedback, and ultimately **needs** to share her finished works with editors and potential publishers

- Why Emotion...?

This is Emotion!



Microsoft Store Launch,
Scottsdale, Oct 2009

Ship products
customers crave



iPhone Launch, San Francisco,
June 2007



Testing in Production (TiP)

• • •

An Introduction

Testing in Production (TiP)

TiP is primarily for **Services** (this includes Websites)

Production

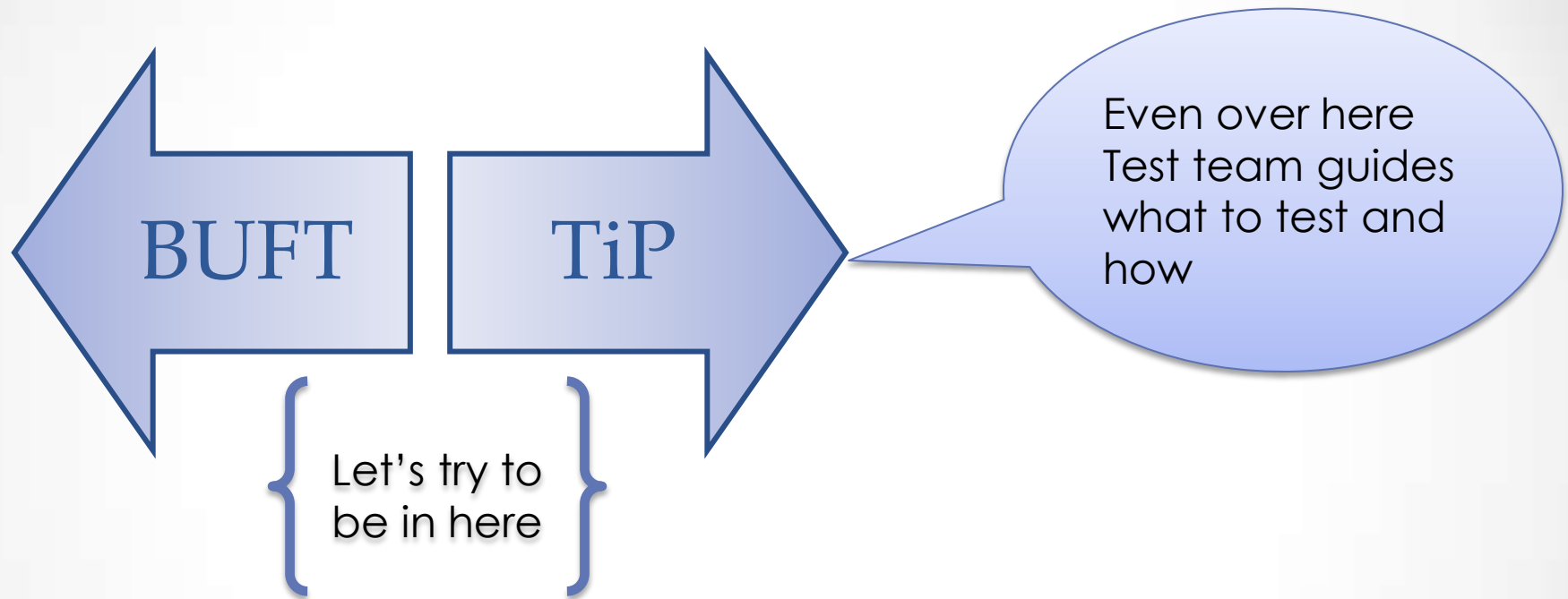
- Data Center where V-curr runs
- Real world users / Real world traffic

Testing

- Functional and Non-Functional



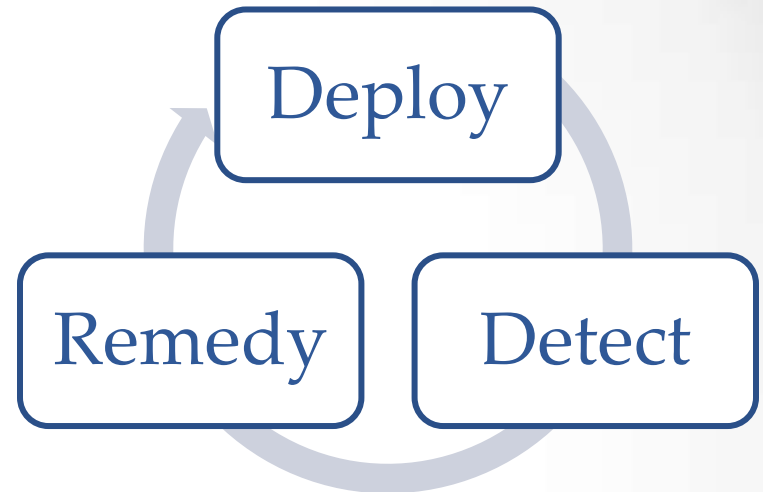
Fire the Test team and just put it in production...???



- Leverage the diversity of production,
 - Provide coverage to find bugs that our internal tests (BUFT) aren't finding.
- But be smart and mitigate risk.

What is a Service?

- You control the deployment independent of user action.
- You have direct monitoring access.



*"It is not the strongest of the species that survives, nor the most intelligent, but the one most **responsive** to change."* - Charles Darwin

Examples:

- Google: All engineers have access to the production machines: "...deploy, configure, monitor, debug, and maintain" [Google Talk, June 2007 @ 21:00]
- Amazon: Apollo Deployment System, PMET Monitoring System - company wide supported frameworks for all services.
- Microsoft: Big Red Switch



TiP is Not New

amazon.com
Hello, Seth Eliot. We have recommendations for you. (Not Seth?)
Seth's Amazon.com Today's Deals Gifts & Wish

Shop All Departments ▾ Search Electronics ▾

All Electronics Brands Bestsellers Audio & Home Theater Camera & Ph

Prime
Member: Seth Eliot

One-Day 1-Click®—\$3.99 Two-Day 1-Click®—FREE

Note: This item is not immediately available to ship.

Electronics Test ASIN - 207
by test pubcode

★★★★☆ (3 customer reviews) | Like (6,592,029)

Price: **\$99,999.00** & eligible for free shipping with Amazon Prime

Temporarily out of stock.
Order now and we'll deliver when available. We'll e-mail you with an estimated delivery date as soon as we have more information. Your account will only be charged when we ship the item. Ships from and sold by Amazon.com. Gift-wrap available.

Quantity: 1 ▾

Add to Cart
or Buy now with
Two-Day 1-Click®—FREE

Ship to:
Seth Eliot ▾

☐ Add gift-wrap/note
or
[Buy with PayPhrase](#)

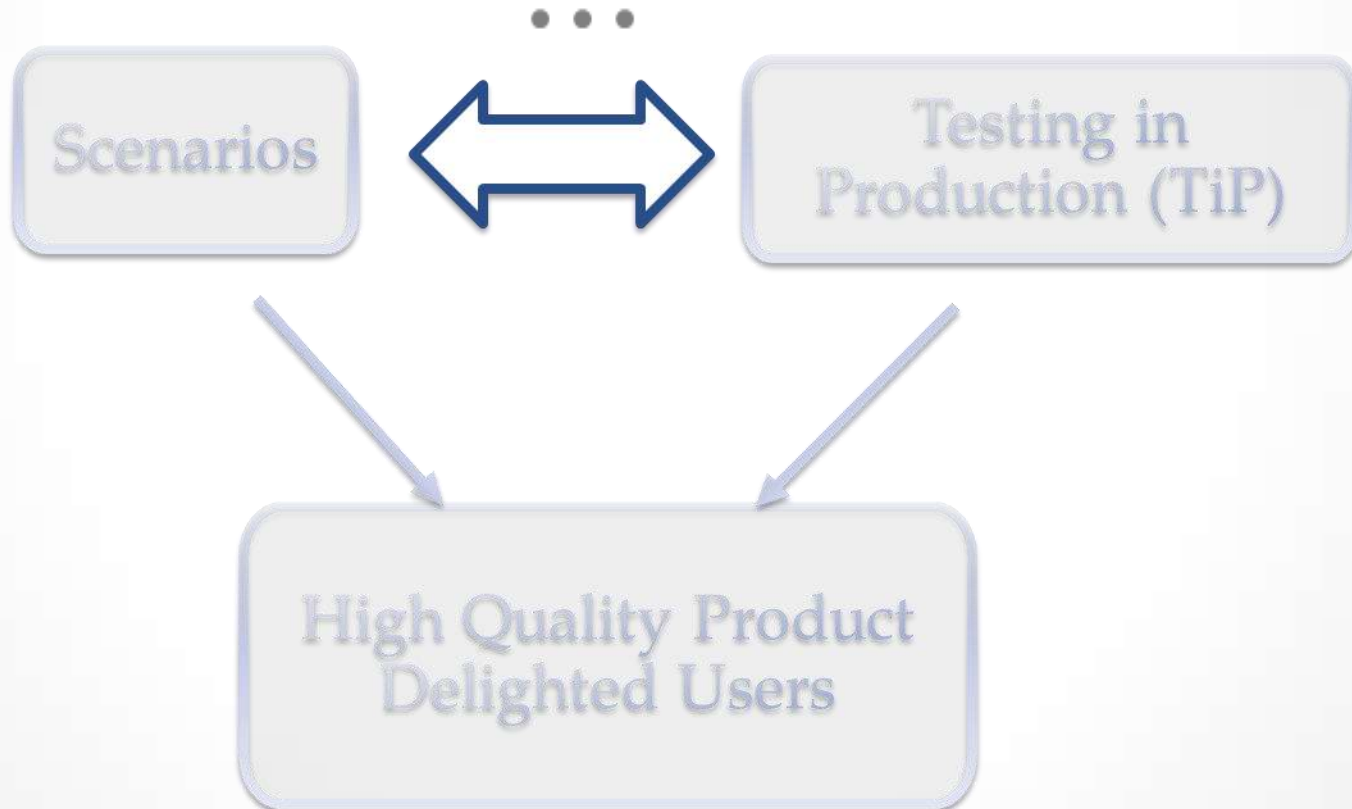
Add to Wish List ▾
Add to Shopping List
Add to Baby Registry

Customer image from Jason Dickinson "Jason"

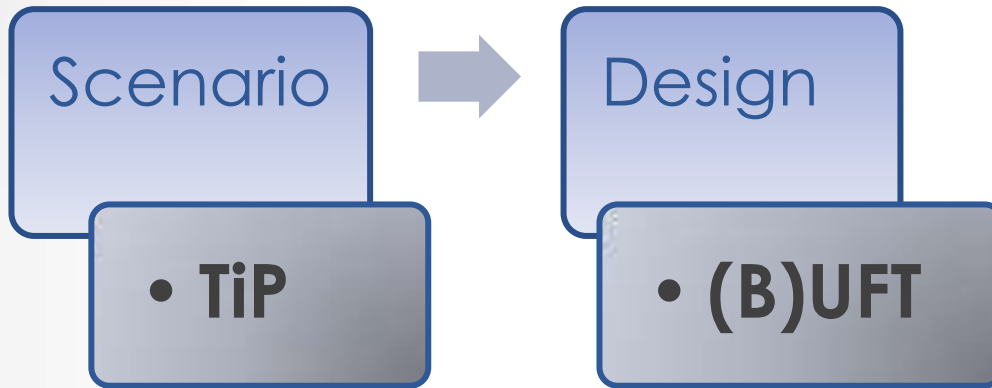
But leveraging it as a legitimate Test Methodology may be new.
Later... many examples how we do this right

Go to Amazon.com and search for {test ASIN} ☺

The Scenario – TiP Connection



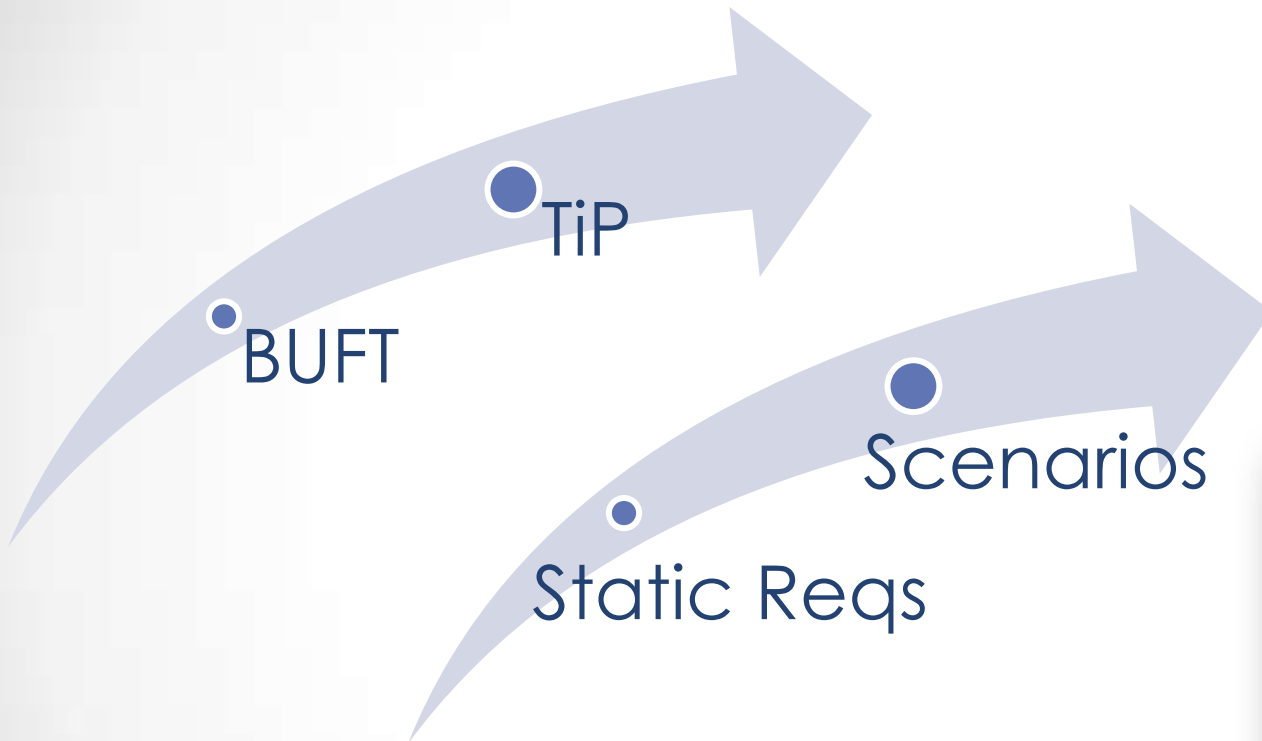
How Do You Test Scenarios?



- Scenario
 - defines user intent
 - TiP is user driven
- Design
 - defines system function
 - UFT is function driven

TiP is How You Test Scenarios
UFT is How You Test Design

Leverage the Long Tail



The Long Tail of real world usage
Products -> Use Cases



[Anderson, July 2006]

Leverage the Long Tail

After missing a connecting flight Jane wants to notify colleagues and expedite her arrival at her destination meeting by finding a new flight....

Mobile device shall support e-mail

The Long Tail of real world usage
Products -> Use Cases



[Anderson, July 2006]

Testers Are Not the Customer



- TiP uses real users / Scenario uses real research

Examples of TiP and Scenarios

- Real Users
 - Data Mining – Passive User Feedback
 - Active User Feedback
- Experimentation (A/B Testing)
- Iterative Prototyping



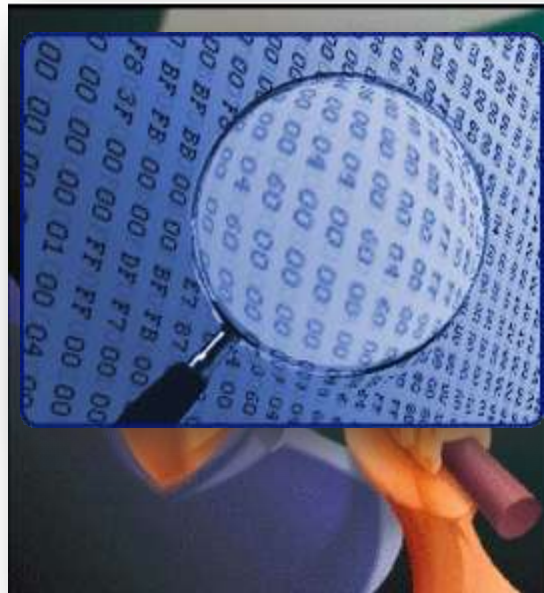
Data Mining

• • •

With Examples

Data Mining

- Looks at Large Data Sets
- In this case Data from
 - Real Users
 - Systems Used by Real Users



Example: Favor Long Tail Over Test Cases

$q = \{\text{amazon stock price}\}$



???

A screenshot of a search engine results page for the query "amazon stock price". The search bar at the top shows the query and a magnifying glass icon. Below the search bar are tabs for "Web" and "More". The results section is titled "ALL RESULTS" and shows "1-10 of 60,400" results. The first result is "AMZN: Summary for Amazon.com, Inc. - Yahoo! Finance". A blue callout box is overlaid on the right side of the results, containing the text: "Any scenarios begin with 'Larry needs to find the price of his favorite stock'?"

pattern = {[company_name] stock price}
where (*Finance Answer Did Not Fire*)

This works...

{expedia stock price}

{Comcast stock price}

{Bed Bath and Beyond
stock price}

...but this does not

{expedia.com stock price}

{Comcast Corporation
stock price}

{Bed Bath & Beyond stock
price}

Example: Speller Split Screen

fasfa 

Web Videos Images More ▼

ALL RESULTS 1-11 of 455,000 results · [Advance](#)

Were you looking for: [fafsa](#)

[Home - FAFSA on the Web-Federal Student Aid](#)
www.fafsa.ed.gov
Electronically submit the **Free Application for Federal Student Aid (FAFSA)**. All students interested in financial aid for college will need to complete this form.
[Contact Us](#) [Help](#)
[Espanol](#)

Quick Access
Customer service 800-433-3243

[Professional help filing the FAFSA - Federal Student Financial Aid ...](#)
Looking for assistance with the **FAFSA** application for federal student financial aid? We provide a wide range of services for a low fee.
<https://www.fafsa.com> · [Mark as spam](#)
[See more results](#)

Results for fasfa

[FASFA or FAFSA - Applications](#)
Looking for the **FASFA**? You may want FAFSA. We offer assistance with the FAFSA application for federal student financial aid at a low fee.
www.fasfa.com · [Mark as spam](#)

[FAFSA - Free Application for Federal Student Aid](#)
<https://fafsa.ed.gov/FOTWWWebApp/complete013.jsp> · [Cached page](#) · [Mark as spam](#)

[FAFSA - Wikipedia, the free encyclopedia](#)
[Eligibility](#) · [Types of financial aid](#) · [Filing options](#)
The Free Application for Federal Student Aid (known as the FAFSA) is a form that can be prepared annually by current and prospective college students (undergraduate and graduate) in ...
en.wikipedia.org/wiki/FAFSA · [Mark as spam](#)

[FASFA](#)
What is **FASFA**? The Frozen at Sea Fillets Association (**FASFA**) was formed in the autumn of 2000, to create awareness of the very high quality of Frozen at Sea Fillets of cod and ...

Top Pane

Result 1 and Result 2 from
Corrected Query

Bottom Pane

Result 1 through Result 10
from Original Query

Example: Speller Split Screen



If (TopPane Clicks > 95%) →
???

Just alter query...

Learn from user action to
Anticipate user intent



If (BottomPane Clicks > 95%) →
???

Leave query alone...

Example: Speller Split Screen

Some Top Results

Query	Correction	Top Clicks	Bottom Clicks	Make Correction	Keep Original
googlecom	google.com	99.7%	0.3%	✓	
yah	yahoo	99.8%	0.2%	✓	
fasfa	fafsa	90.5%	9.5%	✓	
utube music	youtube music	90.4%	9.6%	✓	
facebookcom	facebook.com	98.4%	1.6%	✓	
imbd	imdb	96.3%	3.7%	✓	
evony	ebony	0.5%	99.5%		✓
century link	centrelink	2.0%	98.0%		✓
yout	youth	3.8%	96.2%		✓

Automated Failure Analysis

- Systems can run fine in the lab
 - TiP Replaces imperfect lab with rich production.
- Find Faults in Production
 - Crashes
 - Error Aggregation
- Auto File Bug with...
- Telemetry Captured
 - Logs
 - Crash Dumps
 - User actions that lead to problem



Failure blocks ALL Scenarios
from being fulfilled

AFA Example: Cosmos AutoBug

- Microsoft Cosmos
 - >10K Machines per Data Center
 - More info: http://bit.ly/cosmos_blog
- AutoBug: Error Aggregation and Crash Detection

Service	Aug Release	AutoBugs Fixed	Sept Release	AutoBugs Fixed	Nov Release
S1	495,952,928	1	4,215,476		2,396,548
S2	10,107,128		4,130,825		2,765,910
S3	47,960		38,782		0
S4	39,495,893		41,507,524		28,517,958
S5	293,042		403,239		568,981
S6	62,029,022		108,220,120	4	76,796,685
S7	547,276,342		1,034,032,213	3	417,558,505
S8	154,046,954		300,026,351		110,671,176
S9	189,924,567		417,903,048	3	56,429,723
S10	278,999		661,982	1	357,056
S11	403,643,186		4,995,264,188	2	494,706,072

Online Controlled Experimentation

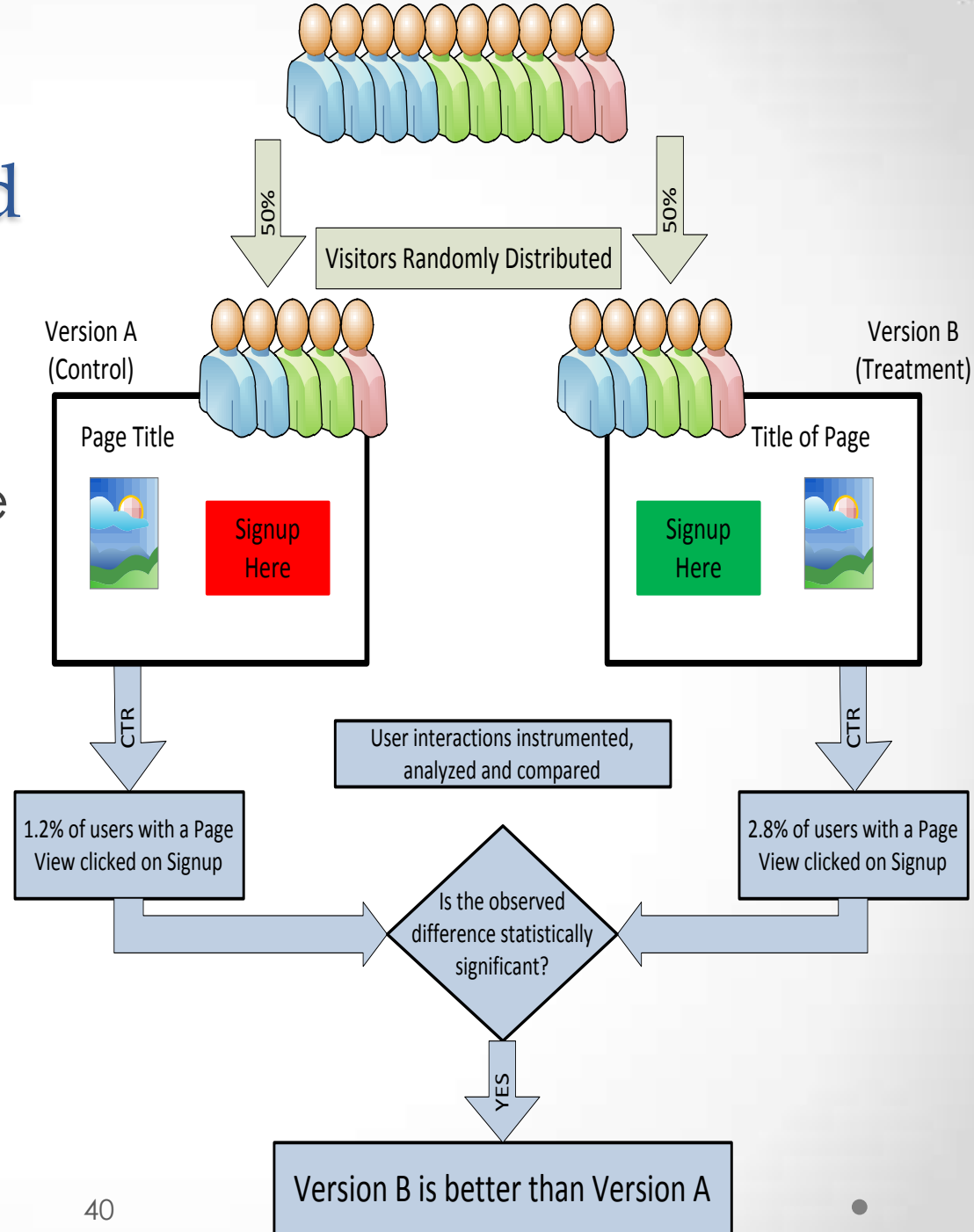
• • •

With Examples

Online Controlled Experimentation

This is an “**A/B**” test
...the simplest example

- A and B are **Variants**
 - A is **Control**
 - B is **Treatment**



Example: Microsoft Xbox Live

Goal: Sign More People up for Gold Subscriptions



<http://www.xbox.com/en-US/live/joinlive.htm>

Which has higher Gold Sign-up...???

A. Control

B. Treatment – up 29.9%

C. Neither

Research reveals Latent Demand

Example: Amazon Shopping Cart Recs

- Amazon.com engineer had the idea of showing recommendations based on cart items [Linden, Apr 2006]
 - Pro: cross-sell more items
 - increase average basket size
 - Con: distract people from checking out
 - reduce conversion
- A marketing senior vice-president was dead set against it.
- Ran an Experiment...



Amazon Shopping Cart Recs

The Scenario

Maria is a new mother. She needs to keep a steady supply of diapers in the house. With the new baby she can't easily get to the store. With the new little one joining the family, money is tighter and she also needs to stay within her budget. And there are so many other things she needs other than diapers to keep her baby happy and healthy.



Customers Who Bought *Pampers Baby Dry Diapers Economy Plus Pack*,... Also Bought



Pampers Sensitive Baby Wipes Refills, 192 Count...

★★★★★ (143)

~~\$29.99~~ \$21.18

13 used & new from \$24.25

Add to Cart



Playtex Diaper Genie II Refill (Pack of 3)

★★★★★ (61)

~~\$19.99~~ \$10.40

13 used & new from \$17.30

Add to Cart



Lansinoh 20265 Disposable Nursing Pads, 60-Count Boxes...

★★★★★ (68)

~~\$59.99~~ \$25.95

11 used & new from \$21.99

Add to Cart



Dr. Brown's Natural Flow Level 2 Standard...

★★★★★ (17)

\$3.99

8 used & new from \$2.99

Add to Cart

Active User Feedback

• • •

With Examples

What's a DSAT?

We're not firing local answers when we spell correct

Web Images Videos Shopping News Maps More | MSN Hotmail

bing

Web

Pizza hut redmnd

Web More▼

RELATED SEARCHES

- Pizza Hut Redmond Washington
- Wal-Mart Redmond
- Red Lobster Redmond
- Pizza Hut Kirkland
- Pizza Hut Locator
- Pizza Hut in Seattle
- Pizza Hut Idaho Falls
- Redmond Pizza Delivery

SEARCH HISTORY

Search more to see your

We're including results for **pizza hut redmond**. Do you want results only for **P redmnd**?

ALL RESULTS 1-10 of 91,500 results

Pizza Hut - America's Favorite Pizza Delivered to Your Door

Pizza Hut - home of America's favorite **pizza**, delivering delicious hot **pizza** and WingStreet buffalo wings straight to your door. Order your **Pizza Hut** pizzas, PZones ...

www.pizzahut.com - Cached page

Pizza Hut - Redmond, WA

(Map) · User rating: 2/5 · 7 reviews

(425) 861-9900 · 2560 152nd Ave NE Ste C, **Redmond**, WA 98052 ... "I came awhile back 4 dinner & got the pan crust Pepperoni Lover's, breadsticks & a **Pizza** was ...

That's Better!

Web Images Videos Shopping News Maps More | MSN Hotmail Sign in

bing MS Beta br1011

Web Local More▼

RELATED SEARCHES

- Pizza Hut Redmond Washington
- Wal-Mart Redmond
- Red Lobster Redmond
- Pizza Hut Kirkland
- Pizza Hut Locator
- Pizza Hut in Seattle
- Pizza Hut Idaho Falls
- Redmond Pizza Delivery

SEARCH HISTORY

Search more to see your history

See all

Clear all · Turn off

► NARROW BY DATE

We're including results for **pizza hut redmond**. Do you want results only for **pizza hut redmnd**?

ALL RESULTS 1-10 of 92,900 results · [Advanced](#)

Pizza Delivery, Order Pizza Online, Dinner Deals, Catering - ...
www.pizzahut.com · Official site
Pizza Hut - home of America's favorite **pizza**, delivering delicious hot **pizza** and WingStreet buffalo wings straight to your door. Order your **Pizza Hut** pizzas, pastas, ...



Pizza Hut Bing.com/local
2560 152nd Ave NE Ste C · Redmond · (425) 861-9900
★★★★★ · [Directions](#) · [Website](#) · [Menu](#)
Cuisine: PIZZA
Hours: Monday 10:00 AM - 11:00 PM [more](#)
Price: Under \$20

See more from: [Yelp](#) · [Citysearch](#) · [Urbanspoon](#)

Pizza Hut - Redmond, WA
(Map) · User rating: 2/5 · 6 reviews
(425) 861-9900 · 2560 152nd Ave NE Ste C, Redmond, WA 98052 ... "I came here awhile back 4 dinner & got the pan crust Pepperoni Lover's, breadsticks & a ...
www.yelp.com/biz/pizza-hut-redmond · [Mark as spam](#)

Example: Crowd Source DSATs

Submit

- By Tool, By e-mail
- >10,000 in Jan from Bing alone

Auto Analysis

- Ad
- Related search
- Speller
- Image result
- Algo result



Each DSAT reveals
a Scenario.
Our solutions need
to focus on intent.

Crowd Source User Testing



- On-Demand Scalable “Workforce”
- Pay users to walk through **scenarios** on your site.

Con

- Like other “Interviews”, there is bias
 - uTest may be more biased but more detailed – Testers not Users

Pro

- Cheap and scalable



Example: Crowd Source User Testing



get opinions online or on the go

Learn how to get started. Take the Video Tour. 

Login or Register

Need quick feedback between 2 things on the go? [Get the free iPhone app.](#)

 Available on the App Store

HomeView SwayablesCreate SwayableiPhone AppGetting StartedRegisterLoginFollow Us  

Featured Swayable

Should I get an iPhone 4 or Windows 7 Phone

Apple iPhone4Windows 7 Phone

 VS 

Click HERE to SwayClick HERE to Sway

14 Sways78 views 1 comments

[show results](#) | [see comments](#) | [add comment](#) | [embed code](#) | [FLAG](#)

URL: <http://swy.bz/gzgYSJ> | share   

The Sway Stream

 VS 

my cap OR my hat

10 Sways

11 views 0 comments

Click here to Sway the vote

 VS 

my cap or my hat

Should I get an iPhone 4 or Windows 7 Phone

14 Sways

78 views 1 comments

Click here to Sway the vote

 VS 

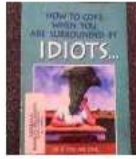
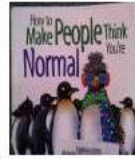
Apple iPhone4 or Windows 7 Phone

idiots book OR how to be normal book

7 Sways

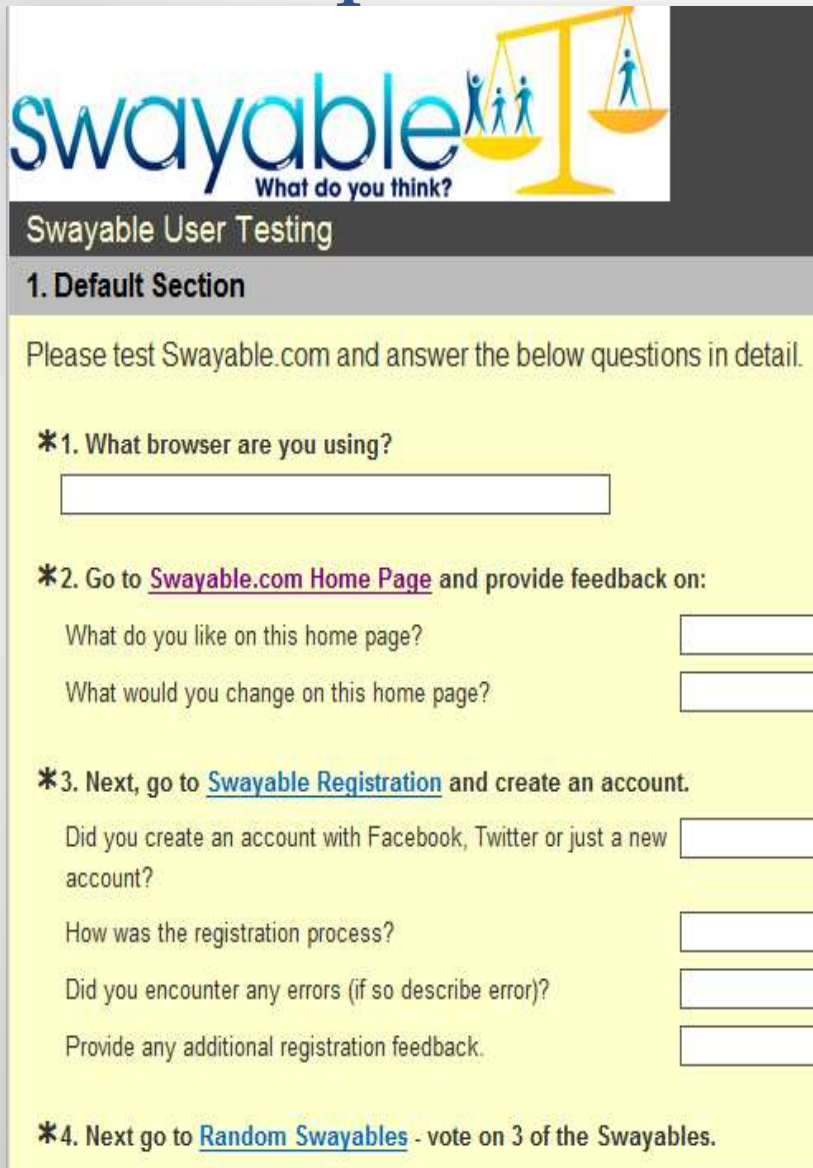
18 views 0 comments

Click here to Sway the vote

 VS 

idiots book or how to be normal book

Example: Crowd Source User Testing



The screenshot shows a survey form titled "Swayable User Testing". At the top left is the Swayable logo with the tagline "What do you think?" and an icon of a balance scale with three figures. The form is divided into sections. Section 1, "Default Section", instructs the user to test Swayable.com and answer questions in detail. It contains four tasks: 1. Identify the browser used. 2. Provide feedback on the Swayable.com Home Page, including what is liked and what would be changed. 3. Create an account via Facebook, Twitter, or a new account, and provide feedback on the registration process, any errors encountered, and additional feedback. 4. Vote on 3 of the Swayables.

swayable
What do you think?

Swayable User Testing

1. Default Section

Please test Swayable.com and answer the below questions in detail.

*1. What browser are you using?

*2. Go to [Swayable.com Home Page](#) and provide feedback on:

What do you like on this home page?

What would you change on this home page?

*3. Next, go to [Swayable Registration](#) and create an account.

Did you create an account with Facebook, Twitter or just a new account?

How was the registration process?

Did you encounter any errors (if so describe error)?

Provide any additional registration feedback.

*4. Next go to [Random Swayables](#) - vote on 3 of the Swayables.

- Paid 0.90 per feedback
- Errors in workflow
- Functional bugs
- Design feedback
- Dead links
- Removed components to make it quicker/easier

[Swayable, Oct 2010]

End to end testing
with real customers

Prototypes

• • •

With Examples

Prototypes

Focused

- used to test a specific idea or question
- focus on what you don't know

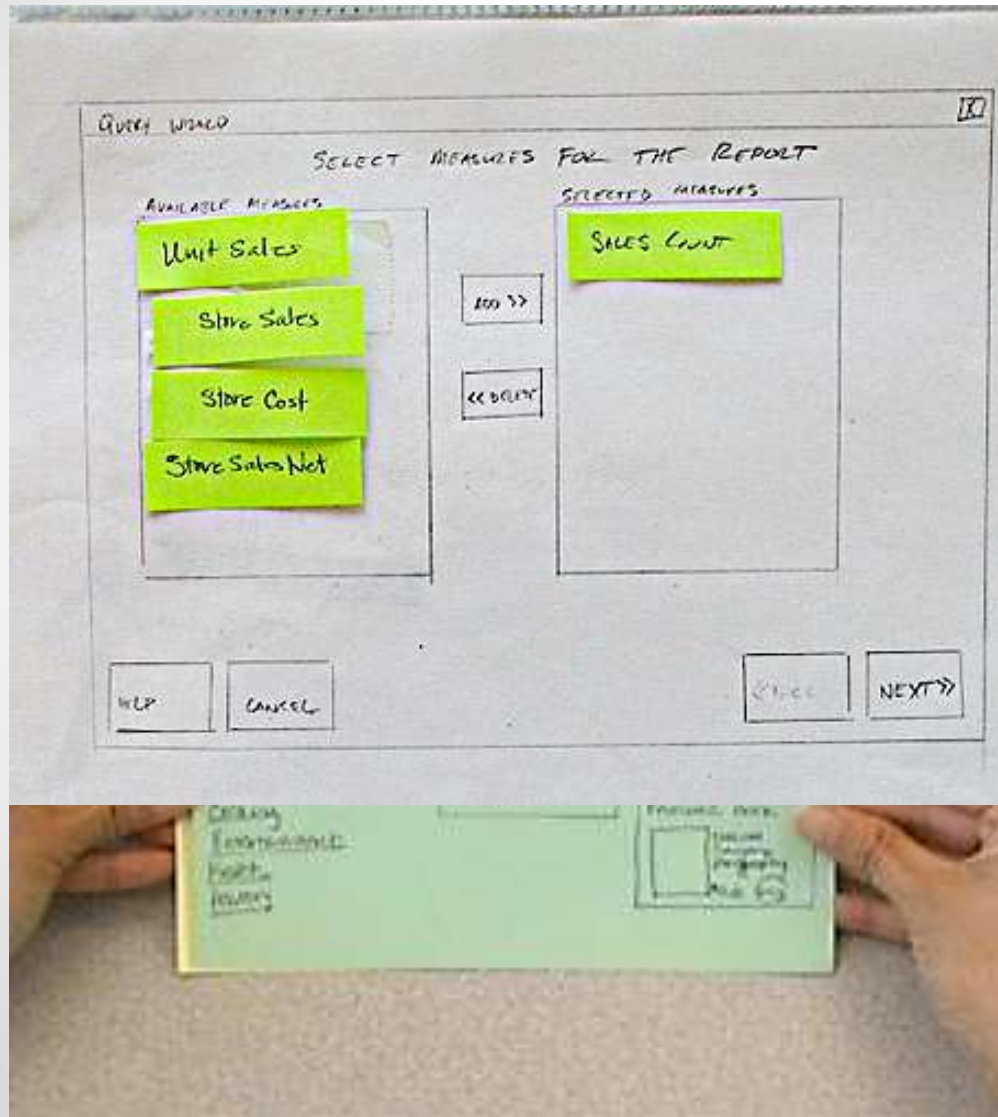
Rapid

- don't waste time on unnecessary detail
- fastest way to get customer feedback early

Disposable

- built to be thrown away
- not robust, not maintainable

Paper Prototypes



Code Prototypes

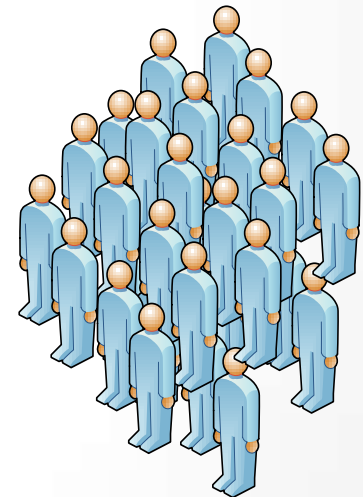
- Launch a new Service – Everyone sees it
- Exposure Control – only some see it



By Browser



By Location



By Percent
(scale)

Exposure Control

Saves you from having to BUFT if product is a dud

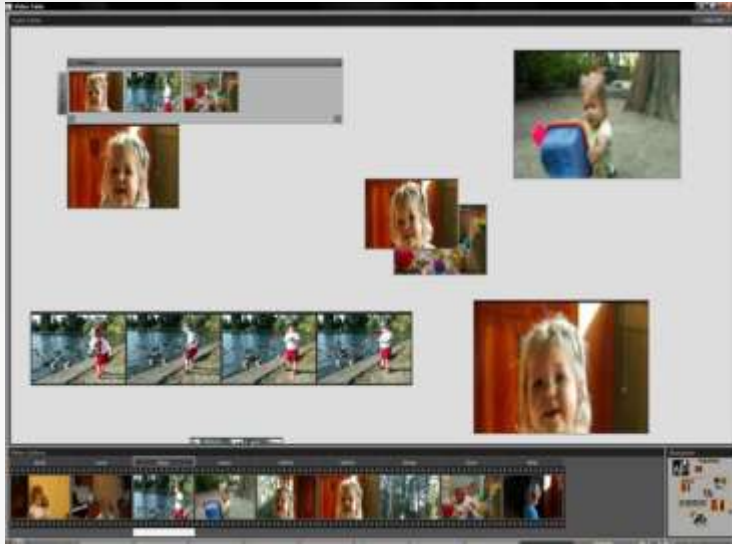


“To have a great idea, have a lot of them”

-- Thomas Edison

Fail Fast – Fail Cheap

Example: Prototype – Video Light Table



Focused

Try out a light-table paradigm for consumer-centric, snippet-based video sharing

Disposable

throw-away code... written as such from the start

Rapid

About 3 days to write

- Written using C# and WPF.
- Expression Blend was used for all visuals, and VS for coding.

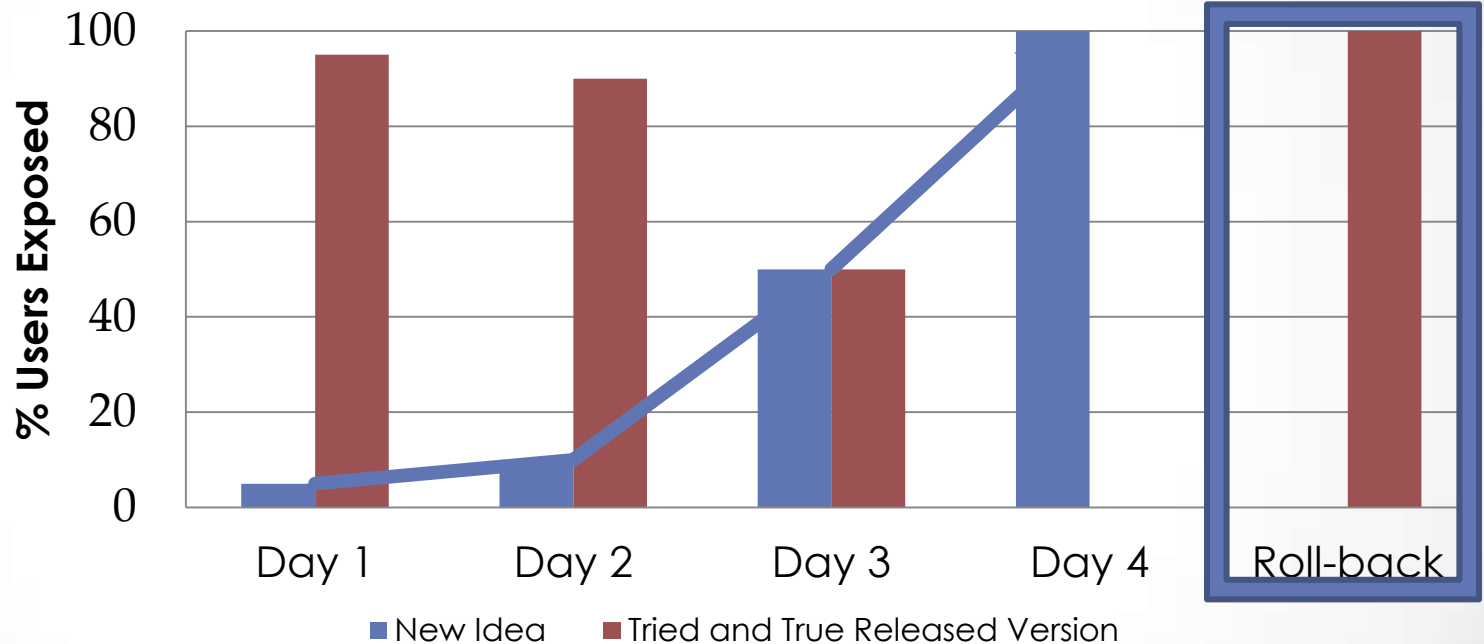
No real video engine backing the prototype

- This was just a UI study
- Just operates on image sequences.

Minimal informal testing

[Schormann, May 2008]

Ramp-up and Deployment



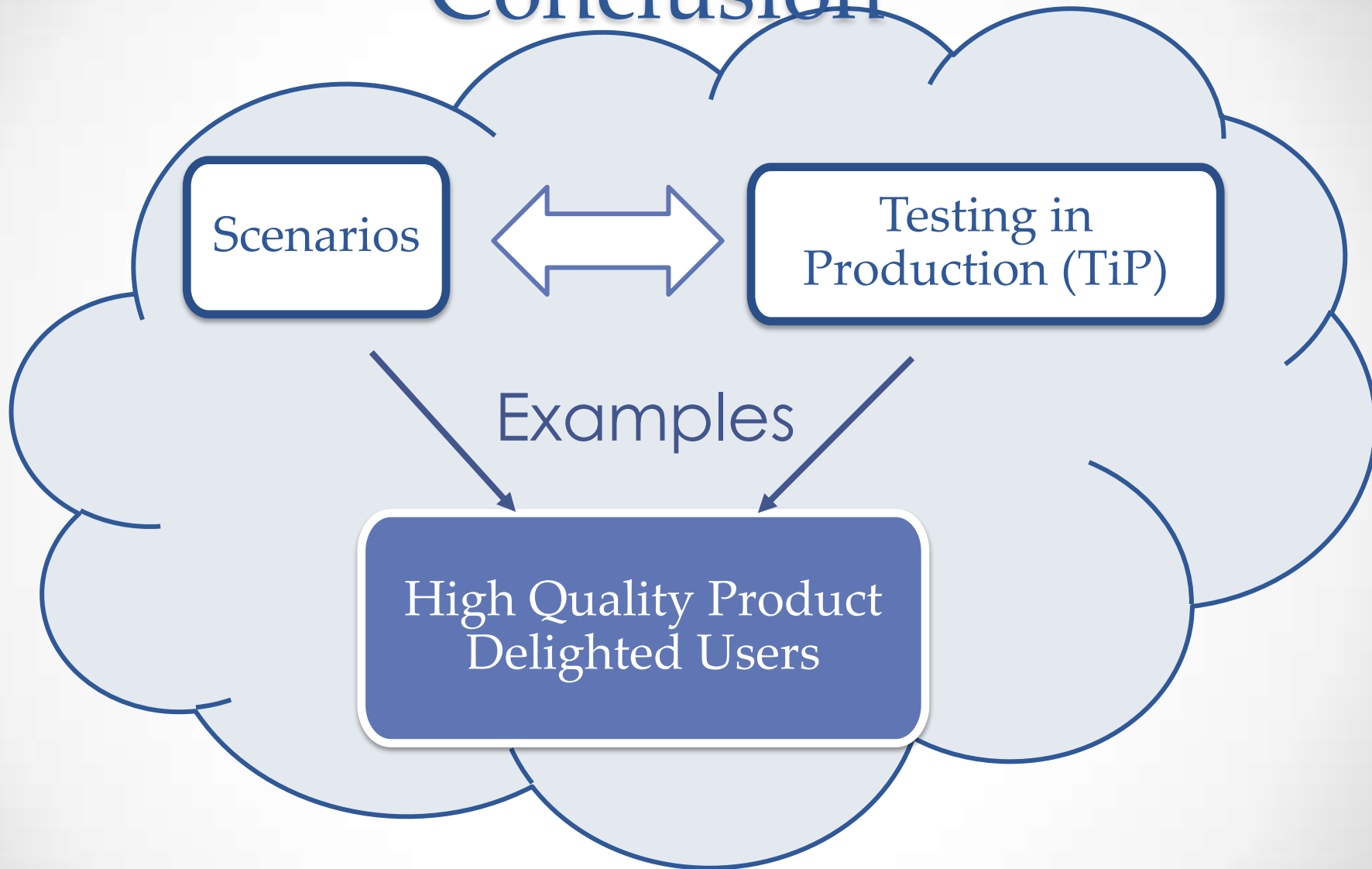
Example: Ramp-up and Deployment: IMVU

“Meet New People in 3-D”

- [v-next is deployed to] a small subset of the machines throwing the code live to its first few customers
- if there has been a statistically significant regression then the revision is automatically rolled back.
- If not, then it gets pushed to 100% of the cluster and monitored in the same way for another five minutes.
- This whole process is simple enough that it's implemented by a handful of shell scripts. [Fitz, Feb 2009]



Conclusion



Resources

...

More Information

- seth.eliot@microsoft.com
- Seth's Blog: http://bit.ly/seth_qa

References

Google Talk, June 2007

Google: Seattle Conference on Scalability: Lessons In Building Scalable Systems, Reza Behforooz

<http://video.google.com/videoplay?docid=6202268628085731280>

Kohavi, et al; 2009

R, Kohavi, T. Crook, R. Longbotham; Online Experimentation at Microsoft

<http://exp-platform.com/expMicrosoft.aspx>

Anderson, July 2006

Chris Anderson, The Long Tail

<http://www.longtail.com/about.html>

Unpingco, Feb 2011

Edward Unpingco; Bug Miner; Internal Microsoft Presentation, Bing Quality Day Feb 11, 2011

Linden, Apr 2006

Greg Linden's Blog: <http://glinden.blogspot.com/2006/04/early-amazon-shopping-cart.html>

Swayable, Oct 2010

Lindsey Harper's Blog: 5 Lessons Learned: User Testing with Amazon's Mechanical Turk

<http://harperlindsey.com/2010/10/28/5-lessons-learned-user-testing-with-amazon%E2%80%99s-mechanical-turk/>

Fitz, Feb 2009

Timothy Fitz, IMVU, Continuous Deployment at IMVU: Doing the impossible fifty times a day,

<http://timothyfitz.wordpress.com/2009/02/10/continuous-deployment-at-imvu-doing-the-impossible-fifty-times-a-day/>

Schormann, May 2008

Christian Schormann, Video Table: A Concept Study , <http://electricbeach.org/?p=79>

With Special Thanks to

- The Microsoft Engineering Excellence Team: Alex Blanton, Austina De Bonte, Michael Corning, Kent Sullivan
- Scott Marks for the Automated Failure Analysis data for Cosmos

END

Session 801
Testing in Production,
Your Key to Engaging Customers

Seth Eliot

Thank you for attending this session.
Please fill out an evaluation form.

More Examples

...

Example: Amazon ordering pipeline

- Amazon's ordering pipeline (checkout) systems were migrated to a new platform.
- Team had tested and was going to launch.
- Quality advocates asked for a limited user test using Exposure Control.
- Five Launches and Five Experiments until A=B (showed no difference.)
- The cost had it launched initially to the 100% users could have easily been in the millions of dollars of lost orders.



TiP is Not New (cont.)

Event

Link Print

Test store - USA

1600 Lind Ave (11.8 miles)
Renton, WA 98057
United States

Get Directions: To Here - From Here

Events

Name: Dungeons & Dragons WPN: D&D WPN-2010-02-20-Renton
Format: D&D WPN
Phone Number: 1 (555) 555-5555
Coordinator: tom.ko@wizards.com
Date: 2/20/2010

Other Events by this Location

Name: D&D Encounters: Undermountain
Format: D&D WPN
Phone Number: 1 (425) 687-2135
Coordinator: brian.larabee@wizards.com

Test store - USA

1600 Lind Ave (11.8 miles)
Renton, WA 98057
Get Directions: To Here - From Here

Add: Immunization

Type of immunization *

- Select -
- meningococcal vaccine, NOS
- mumps virus vaccine
- no vaccine administered
- parainfluenza-3 virus vaccine
- pertussis vaccine
- plague vaccine
- pneumococcal conjugate vaccine, polyvalent
- pneumococcal polysaccharide vaccine
- pneumococcal vaccine, NOS
- poliovirus vaccine, inactivated
- poliovirus vaccine, live, oral
- poliovirus vaccine, NOS
- Q fever vaccine
- rabies immune globulin
- rabies vaccine, for intradermal injection
- rabies vaccine, for intramuscular injection
- rabies vaccine, NOS
- RESERVED - do not use
- respiratory syncytial virus immune globulin, intravenous
- respiratory syncytial virus monoclonal antibody (palivizumab), intramuscular

But leveraging it as a legitimate Test Methodology may be new.
Later... many examples how we do this right

Example: Man vs. Machine

Goal: Increase Clicks on Page per User

HOPS = Headline Optimization System



Which has higher page clicks per user...???

- A. Control - Editor Selected
- B. Treatment – HOPS +2.8%**
- C. Neither

Research reveals
how to meet
customer needs

Example: Microsoft Xbox Marketplace

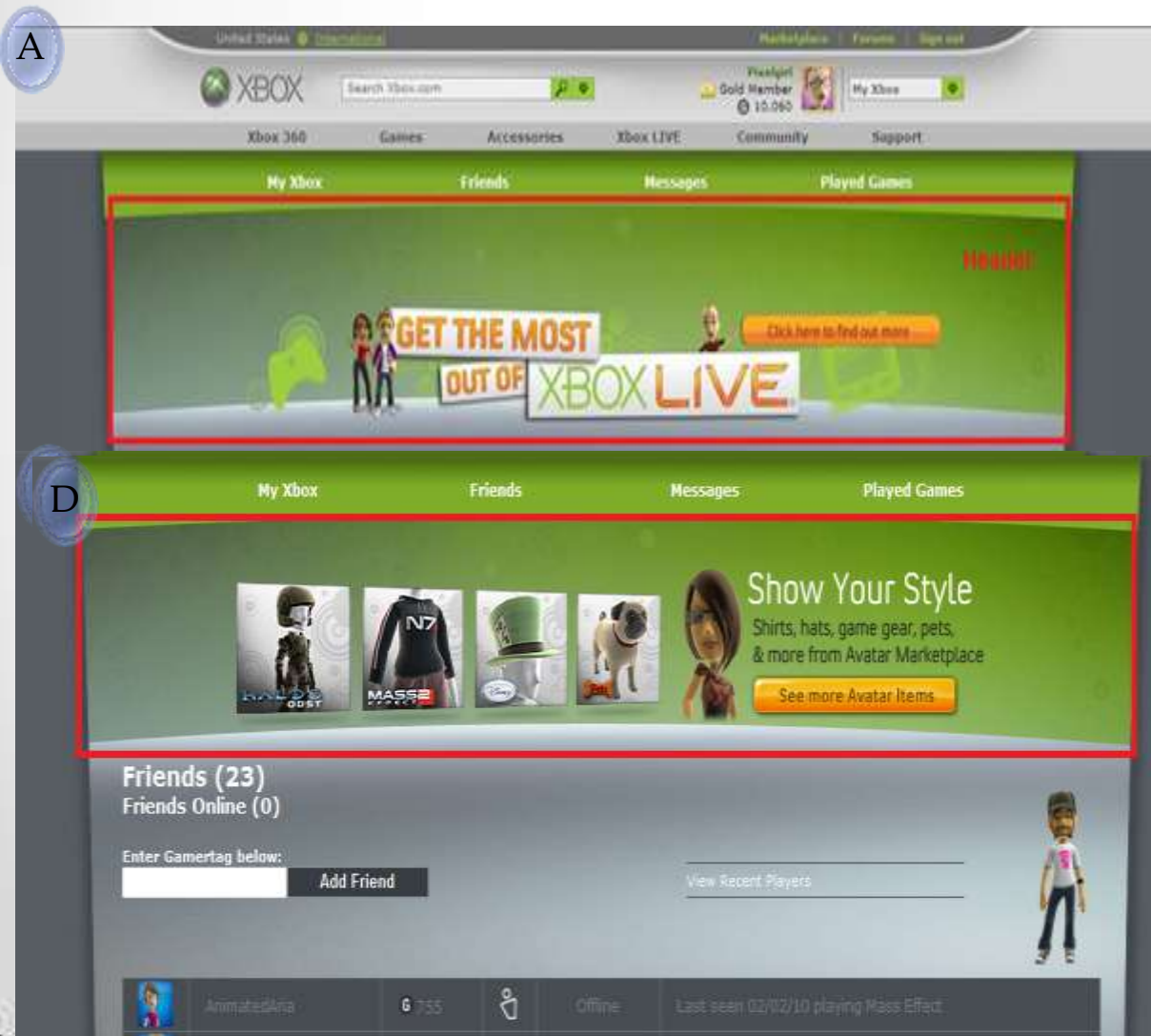
Goal: Increase Total Points Spent per User

<http://marketplace.xbox.com/en-US>

Which has higher Points Spent...???

- A. Control
- B. T1: Game Add-Ons
- C. T2: Game Demo
- D. T3: Avatar Gear
- E. None**

Promoted content up,
but at expense of others



Example: Microsoft Store

Goal: Increase Average Revenue per User

<http://store.microsoft.com/home.aspx>

Which increased revenue...?

A. Control

B. Treatment – up 3.3%

C. Neither

A

Shopping Cart

 Xbox 360 Elite Bundle \$399.99

Quantity: 1

[Remove](#) Delivery Options: Ship It

Shipping Method

☒ UPS Ground - \$0.00 (3-7 business days)

☐ UPS 2nd Day - \$5.00 (2 business days)

☐ UPS Next Day - \$12.00

Subtotal: \$399.99
Shipping: \$0.00
Tax: \$38.00

Estimated Total: \$437.99

Enter Zip to calculate Tax: 98052

[E-mail Cart to a Friend](#)

B

Shopping Cart

 Xbox 360 Elite Bundle \$399.99

Quantity: 1

[Remove](#) Delivery Options: Ship It

Shipping Method

☒ UPS Ground - \$0.00 (3-7 business days)

☐ UPS 2nd Day - \$5.00 (2 business days)

☐ UPS Next Day - \$12.00

Subtotal: \$399.99
Shipping: \$0.00

[E-mail Cart to a Friend](#)

Example: Google Talk

- Use an “Experimentation Framework”
- Limit launch to
 - Explicit People
 - Just Googlers
 - Percent of all users
- Not just features, but it could be a new caching scheme



[Google Talk, June 2007 @ 20:35]

Example: Google Talk Shadowing

- Google Talk Server provides Presence Status
 - Billions of packets per day
- Orkut integration
 - Started fetching presence without showing anything in UI for weeks before launch
 - Ramp-up slowly from 1% of Orkut PVs
- GMail chat integration:
 - Users logged in/out: used this data to trigger presence status changes w/o showing anything on the UI



[Google Talk, June 2007 @ 9:00]