CSE 440: Introduction to HCI User Interface Design, Prototyping, and Evaluation

Lecture 10: Testing James Fogarty Alex Fiannaca Lauren Milne Saba Kawas Kelsey Munsell

Tuesday/Thursday 12:00 to 1:20

dub design: use: build:

University of Washington



Today

Presentations on Thursday

- BalanceEcotopiaFoodPicMiPhoneSocial ReconnectionTagLinePresentations on Friday Afternoon
 - NeatPoliscopeSchoolViewSitlessSmartClothingTimeout

No Section Friday Morning





Today

For Presentations PowerPoint or PDF Mind Your Time Limits Peer Feedback Forms

Project 3a Due for Tuesday

In-Class Design, Prototype, Test Testing and Patterns





In-Class Design, Prototype, Test

Design and prototype a new touchscreen alarm clock to be deployed in a very high end hotel brand. Your alarm clock should be immediately usable for tired, busy, or just-don't-want-to-be-bothered travelers who will spend zero time learning your interface.

In addition to displaying the current time, your alarm clock should include basic functionality for:

turning the alarm on/off

setting the wake-up time

anything else you think is appropriate

Guests will interact with your alarm using a touch panel.



Task Design is Important

The goal of a test is to figure out how a person interacts with an interface in the wild...

There are two possible explanations for why a test does not find significant problems:

The interface does not have significant problems

The test itself has significant problems





Task Design is Important

Testing is not entirely in the wild

As a part of focusing the test, you often need to give a person a somewhat artificial task

The artificiality of the task may influence how people interact with an interface...

...and thus may influence the outcomes and insights gained through user testing



Bad: Artificial Subgoals

People using the design "in the wild" may not necessarily form these same subgoals

The task should give one top-level goal, a people should form their subgoals while pursuing this

Now you want to choose the type of paper you want to print your **document on. Lets imagine that Bin "B" has the paper you want to** print your paper on, please complete this task.

Now set the darkness of your copies to about 50% dark. After setting the darkness, you decide you want to print 2 sides of copies on two sides of paper. Please complete this task.





Bad: Artificial Ordering

With an artificial ordering of information or subgoals, people might not proceed in this order

The ordering might also be biased towards the layout of the interface, which would conceal any problems with finding the appropriate control

- Enter in 10 copies, with lightness set to 10%.
- Choose 1 sided to 2 sided, use paper source bin A.
- Cover sheet needed, using paper bin B for cover sheet.
- Set stapling feature on and collating on.
- Start printing.





Bad: Changing the Task

The task is to make copies, and this happens to involve entering information in the copier interface

But this task description is an data entry task, "Here is some information. Put it in the interface."

- Make 23 copies
- With collate
- Cover sheets
- Default darkness
- 1 Sided-> 1 Sided





Bad: Giving the Answers

Tells the person what terminology the interface uses, which they might not otherwise know

lighten = contrast, sorted = collated?

You are a teacher and are trying to make 40 copies of a one-sided magazine article that is 10 pages long for your class tomorrow. Due to the large number of copies, you print the article double-sided, in other words 10 page article would be printed on 5 sheets of paper. Due to the high contrast of the article, you must lighten the copy, in other words change the contrast. You then want the copies to be collated and stapled.





Good: Giving Context

Giving realistic context through scenarios can reduce the artificiality of the task

It's your first day in the office, starting a new job. You would like to make some copies of several documents that your boss gave you to browse through. Your colleague in the next cubicle tells you that you need an access code to make copies. The code is 5150. You walk over to the copy machine at the end of the hall and realize that it is not the Xerox copier that you are accustomed too... Make 2 copies of the "Company Annual Report".





Consider: Under-Specified Tasks

Many realistic goals are under-specified, as people have only a general idea what they want

By under-specifying the task, you can elicit realistic confusion and decision-making

You just finished fixing up the old hot rod in the garage and now its time to sell her. Make a couple copies of the pictures you took to **send into the used car sales magazines. It's ok that they're in black** and white but maybe you should lighten them up a bit. Your account billing code is 5150.





Task Design Summary

Task design is difficult and important

Poorly designed tasks mask interface failures

If you are not confident in your task descriptions, have others help you "debug" them before testing





Remote Usability Testing

Conferencing-based testing

Use tools like video conferencing, instant messaging, and screencasting to test with a remote participant

Semi-automated remote testing Automatic logging and some analysis of usage

Controlled online A/B experiments

Carefully measure results of showing different versions to different sets of live customers





Move usability testing online

- participants access the "lab" via web
- answer questions & complete tasks in "survey"
- records actions or screens for playback
- can test many people completing many tasks
- Analyze data individually or in aggregate
 - playback individual sessions
 - find general problem areas

if needed, look more closely with traditional methods



Washington





University of Washington

€ N	etRak	er Int	elligen	ce Center 1	our - N	etscape] >
ile	<u>E</u> dit	⊻iew	<u>G</u> o <u>C</u>	ommunicator	<u>H</u> elp	⊘ Send			~	~	./22/5.			
	Back	Eo	award	Beload		earch.	My. Netscape	I Print	Securitu	Shop.	Stor			N
	- Z≱	7	invara	Ticioda	Home	Search	Пессере	1 mix	Jecuny	эпор	otop			
	Per	cent	ages		<u>Total</u>	<u>5</u>	Respo	<u>indents</u>		<u>Detai</u>	<u>ls</u>	Demo	ographics	
		Γ	1 Fi	nd a flat	nanol	monito	rthat cos	te loce	than \$1	200 0	loaso ti	w to	1	
		ž	accon	nplish thi	s task	withou	t using th	ie searc	h functi	ion.	ease a	γio		
			Task								Respor	ise(s)		
			I was	able to c	complet	e the ta	ask			90%				
			I was	not able	to con	nplete th	ne task			10%				
			I thin not si	k that I w ure	vas abl	e to com	plete the	task, bi	ut I`m	0%				
		Ē	<u>Resp</u>	onse Tim	ies									
			Faste	st: 00:0	D:28									
			Media	in: 00:0	0:41									
			Avera	ige: 00:0	0:48.4									
			Slowe	est: 00:0	1:14									
		_											-	
			2. W	hat is th	e price	of the	monitor	you just	found?	,				
			Short	Freefor	m									
			\$1129	9										









Goals

link page elements to actions
identify behavior/navigation patterns
highlight potential problems areas

Interactive graph based on web content

designers can indicate expected paths color code common usability interests

filtering to show only target participants

use zooming for analyzing data at varying granularity





University of Washington





University of Washington



University of Washington

Now available through a variety of services

Loop11UserZoomTryMyUIValidatelyUserlyticsWhatUsersDoUsertesting.comYouEye

Unlikely you need to bake your own Some include mobile testing



http://www.nngroup.com/articles/unmoderated-user-testing-toois.

Controlled A/B Experiments





(this section mostly due Ronny Kohavi)

Controlled A/B Experiments

Example: Amazon Shopping Cart Recommendations

Add an item to your shopping cart

Most sites show the cart

At Amazon, Greg Linden had idea to show recommendations based on cart items



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

Controlled A/B Experiments

Evaluation

Pro: cross-sell more items

Con: distract people from checking out

Highest Paid Person's Opinion: Stop the project

Simple experiment run: Wildly successful



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

Marketplace: Solitaire vs Poker

Experiment run in Windows Marketplace / Game Downloads

Which image has the higher clickthrough? By how much?



Washington

Marketplace: Solitaire vs Poker

Experiment run in Windows Marketplace / Game Downloads Which image has the higher clickthrough? By how much?



Washington

Never Underestimate Solitaire





Never Underestimate Solitaire



University of Washington

Checkout Page

Conversion rate is percentage of visits that include purchase

Docto	r FootCare"	Α	F s	shopping Cart	Docto	r FootCare"	В	Rs	hopping Cart
Home Produc	tts Learn More Tips Te	estimonials FAQ Abo	at Us Contact Us	1-866-211-9793	Home Produc	cts Learn More Tips	Testimonials FAQ Abou	t Us Contact Us	1-866-211-9733
Shop With Co	onfidence Guaranteed (2 ^{e'} 30-d Secured shopping (3 ^{e'} We a	ay, hassle-free Returns ssure your Privacy			Shop With Co	onfidence Guaranteed 대한 30 Secured shopping 대한 W	0-day, hassle-free Returns le assure your Privacy		
A 18805 Secured	Charlemat	Continue Sha	noing Process	d To Checkaut		Checkout		* Proces	d To Checkout
- reasonance	checkout	Contande Sile	pping Proces	to checkode	Item Name	Item Number	Quantity Remove	Unit Price	Subtotal
Item Name	Item Number	Quantity Remove	Unit Price	Subtotal	Trial Kit	FFCS	1 1	\$0.00	\$0.00
Trial Kit	FFCS	1 0	\$0.00	\$0.00				Discount	\$0.00
		Update		Total: \$0.00				Total	\$0.00
							Enter Coupon Code		
		Select Shipping Metho	d Standard (\$5.5	95) 👻			Select Shipping Method	Standard (\$5.95)	· · ·
100% Secured	Checkout	Continue Sho	pping > Procee	ed To Checkout	A 198% Secured	Checkout Recald	culate Continue Shop	ping > Proceed	d To Checkout
Home Prod Cart	ucts Learn More Tips	Testimonials FAQ Ab	out Us Contact Us	Shopping	Home Produ	uts Learn More Tips Te	estimonials EAQ About Vs	<u>Contact Us</u> Shop	ping Cart
Cosyunght @ 20	003 Doctor Foot Care Inc. All B	ights Reverved Privacy Pol	leg l		Copyright @ 2	003 Dotter Foot Care Inc. H	Rights Received, Privaty Poli	EV	



Washington

Which version has a higher conversion rate?

Example from Bryan Eisenberg's article on clickz.com

Checkout Page

University of

Washington

Conversion rate is percentage of visits that include purchase

Doctor FootCare"	Α	民 Shopping Cart	Docto	r FootCare	В	R sh	opping Cart
Home Products Learn More Tips To	estimonials FAQ About	Us Contact Us 1-866-211-9733	Home Produ	cts Learn More Tips	Testimonials FAQ	About Us Contact Us 1	-866-211-9733
Shop With Confidence	lay, hassie-free Returns Issure your Privacy		Shop With Co of Satisfaction of 100% Safe,	onfidence Guaranteed Col Secured shopping Col	30-day, hassle-free Returns We assure your Privacy		
A 18855 Served Charlent	Continue Show	Dennard To Chadraut	a 100% Secured	Checkout		* Process	To Checkout
	Continue Shop	and Proceed to checkode	Them Name	Item Number	Quantity Remo	ave Unit Price	Subsocial
Item Name Item Number	Quantity Remove	Unit Price Subtotal	Trial Kit	FFCS	1 1	\$0.00	50.00
Trial Kit FFCS	1 🗇	\$0.00 \$0.00				Discount	\$0.00 10.00
	abaara				Enter Coupon Code		
	Select Shipping Method	Standard (\$5.95)			Select Shipping Meth	od Standard (\$5.95)	¥
100% Secured Checkout	Continue Shop	sing > Proceed To Checkout	A 196% Secured	d theckout Rec.	alculate Continue 1	Shopping > Proceed	To Checkout
Home Products Learn Nore Tips Cart	Testimonials FAQ Abos	ut Us Contact Us Shopping	Home Prody	acts Learn More Tips	Testimonials EAQ Abou	it.Us Contact.Us Shopp	ing Cart
Copyright 2 2009 stortor First Care Inc. All P	lights Reserved Privacy Policy		Copyright @ 2	003 Dotter Foot Care Inc.	All Rights Reserved. Privaty	Policy	

Which version has a higher conversion rate?

Example from Bryan Eisenberg's article on clickz.com

Checkout Page

Conversion rate is percentage of visits that include purchase

Doctor FootCare"	Α		R Shopping Cart	Doctor	FootCare	В	R si	hopping Cart
Home Products Learn More Tips	Testimonials FAQ	About Us Contact	Us 1-866-211-9733	Home Product	br Learn More Tips	Testimonials FAQ	About Us Contact Us	1-866-211-9733
Shop With Confidence	10-day, hassle-free Returns Ve assure your Privacy			Shop With Co	nfidence suaranteed (sf ecured shopping (sf	30-day, hassle-free Returns We assure your Privacy		
A 18855 Second Charlout	Castinu	Shanning 20	mound To Chashaut		heckout		+ Proces	ad To Checkout
	Continue	e snopping	roceed to checkode	Them Name	Item Number	Quantity Rem	ove Unit Price	Subcotal
Item Name Item Number Trial Kit FECS	Quantity Ren	nove Unit Price	Subtotal	Trial Kit	FFCS	1 1	50.00 Discount	\$0.00 \$0.00
	Update		Total: \$0.00			Enter Coupon Code		
a 100% Secured Checkout	Select Shipping M	ethod Standard	roceed To Checkout	a 186% Secured	Checkout Rec	Select Shipping Met	sod Standard (\$5.95) Shopping > Proceed	d To Checkout
Itome Products Learn Nore Tipe Cart	Testimoniala FAQ	About Us Conta	ct.Us Shopping	Home Produc	35 Learn More Fips 03 Doctor Foot Care Inc.	Testimonials EAQ Abou	nt Us Contact Us Shop	pina Cart



Which version has a higher conversion rate?

Coupon Code decreases by factor of 10

Example from Bryan Elsenberg's article on clickz.com

Office Online Feedback

Please let us know if this content w	as helpful.
Rate this content: ☆☆☆☆☆	
Tell us why you rated the content this wa	ay (optional):
Remaining characters: 650	Submit

Feedback A puts everything together, whereas feedback B is two-stage: question follows rating.

Feedback A just has 5 stars, whereas B annotates the stars with "Not helpful" to "Very helpful" and makes them brighter.

helpful	
Click to rate: 3 out of 5	stars
1	
How helpful was this information?	
How helpful was this information? Click a star.	
How helpful was this information? Click a star. Not Yery helpful 수도 사망 helpful	
How helpful was this information? Click a star. Not helpful A A A A A A A A A A A A A A A A A A A	2
How helpful was this information? Click a star. Not helpful 소교 소교 Very helpful Why did you rate the information this way	2
How helpful was this information? Click a star. Not helpful Very helpful Why did you rate the information this way	12

Which one has a higher response rate? By how much?

B





Office Online Feedback

Please let us know if this content was	s helpful.
Rate this content: ☆☆☆☆☆	
Tell us why you rated the content this way	(optional):
	~
	~
Remaining characters: 650	Submit

Feedback A puts everything together, whereas feedback B is two-stage: question follows rating.

Feedback A just has 5 stars, whereas B annotates the stars with "Not helpful" to "Very helpful" and makes them brighter.

University of Washington

Not helpful the	Very helpful	
Cid	k to rate: 3 out of 5 stars	
How helpful was this Click a star.	s information?	
How helpful was thi Click a star. Not helpful	very helpful	
How helpful was thi Click a star. Not helpful 여호호호호 Why did you rate the ir	very helpful	
How helpful was thi Click a star. Not helpful Why did you rate the in	is information?	
How helpful was thi Click a star. Not helpful 여호호호호 Why did you rate the in	is information?	

Which one has a higher response rate? By how much?

B

B gets more than double response rate.



Another Feedback Variant

Call this variant C. Like B, also two stage.

Which one has a higher response rate, B or C?



Another Feedback Variant

Call this variant C. Like B, also two stage.

Which one has a higher response rate, B or C?



MSN US Home Page

Proposal: New Offers module below Shopping

Shopping

- · Lancôme: Free deluxe compact w/ purchase
- Special promotions at your favorite stores
- * Warm fall fashion styles are here
- · Save on top brand digital cameras
- Free shipping on furniture for every room

Advertisements



- A smart way to buy a diamond
- * Wal-Mart: Back-to-school
- Our editor picks budget electronics
- · Get fit & save money: Sports sale

Control

Shopping

- · Lancôme: Free deluxe compact w/ purchase
- Special promotions at your favorite stores
- · Warm fall fashion styles are here
- · Save on top brand digital cameras
- . Free shipping on furniture for every room

Advertisements



- A smart way to buy a diamond
- Wal-Mart: Back-to-school
- Our editor picks budget electronics
- · Get fit & save money: Sports sale

Offers



Search GM Certified With our 117-Point Inspection GM Certified means no worries



Online University

Earn degree from a top school 100% Online, Get Free Info!

\$200k Loan, Get Low Rates

Secure Financing and Increase Cash Flow, Click Here Now!

Treatment



Ran A/B test for 12 days on 5% of MSN US visitors





Ran A/B test for 12 days on 5% of MSN US visitors

Clickthrough:

Page views per person-day:



Ran A/B test for 12 days on 5% of MSN US visitors

Clickthrough: decreased 0.49%

Page views per person-day: decreased 0.35%



Ran A/B test for 12 days on 5% of MSN US visitors

Clickthrough: decreased 0.49%

Page views per person-day: decreased 0.35%

Value of click from home page: X cents

Net = Expected Revenue – Value Per Click * Direct lost clicks – Value Per Click * Lost Due to Decreased Views



University of Washington

Ran A/B test for 12 days on 5% of MSN US visitors

Clickthrough: decreased 0.49%

Page views per person-day: decreased 0.35%

Value of click from home page: X cents

Net = Expected Revenue – Value Per Click * Direct lost clicks – Value Per Click * Lost Due to Decreased Views

Net was negative (in millions of dollars),offers module did not launch

Data Driven Methods Not Just Online



University of Washington

Data Driven Methods Not Just Online



Washington

Data Driven Methods Not Just Online



University of Washington



Limitations

Drives hill-climbing, but not overall design A design may be better, but is it good?

Impossible for new designs to compete

Can be difficult to scale to many features How about we step through a larger example





CSE 440: Introduction to HCI User Interface Design, Prototyping, and Evaluation

Lecture 10: Testing James Fogarty Alex Fiannaca Lauren Milne Saba Kawas Kelsey Munsell

Tuesday/Thursday 12:00 to 1:20



University of Washington