



DESIGNING
HUMAN STUDIES
(EXPERIMENTS IN GENERAL)

2020.06.01

JUHO KIM



WHAT MAKES GOOD RESEARCH?

- x Research is creation and communication of knowledge that is
 - o Novel
 - o Generalizable
 - o Valuable
 - o Valid



WHAT MAKES GOOD RESEARCH?

X Research is creation and communication of knowledge that is

- Novel
- Generalizable
- Valuable
- Valid

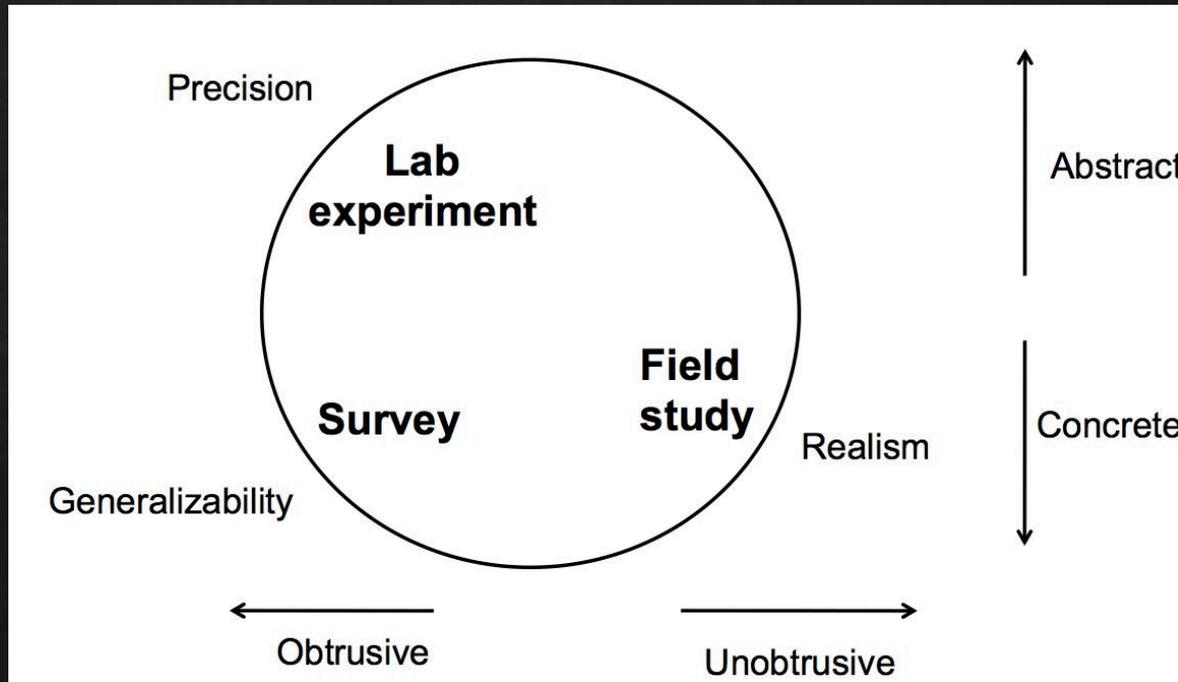
today's focus



WHY RUN EVALUATION WITH HUMANS?

- X More computing technology is designed to directly help humans.
- X Examples
 - HCI: GUI, mobile, input, output, web, social, AR, VR
 - Graphics: video / photo editing tools
 - SE: technology for supporting programming tasks
 - CV: aesthetics, scene detection, video summarization
 - NLP: translation quality evaluation, framing detection
 - Networking/Mobile: sensing technology, quality of service
 - ...

RESEARCH METHODS WITH HUMANS INVOLVED



CONTROLLED EXPERIMENTS

SHOWING THE SUCCESS AND CONVINCING OTHERS

- X “Do you like my system?”
- X “How much do you like my system?”
- X “This is an awesome system. Agree or disagree?”
- X Huge individual variance, “please the experimenter” bias

TOWARD MORE SYSTEMATIC EVALUATION

- X What are the core tasks you're trying to support?
 - Back to earlier stage of research
- X How can you measure the performance of these tasks?
 - Dimensions crucial to your system (e.g., **learnability** – time to completion, **efficiency** – number of steps used, **safety** – accuracy)
- X What's the comparison?
 - Useful to have a baseline (control condition without your magic)

TOWARD MORE SYSTEMATIC EVALUATION

- X Baserates:** How often does Y occur?
 - Requires measuring Y.
- X Correlations:** Do X and Y co-vary?
 - Requires measuring X and Y.
- X Causes:** Does X cause Y?
 - Requires measuring X and Y, and manipulating X.
 - Also requires somehow accounting for the effects of other independent variables (confounds)!

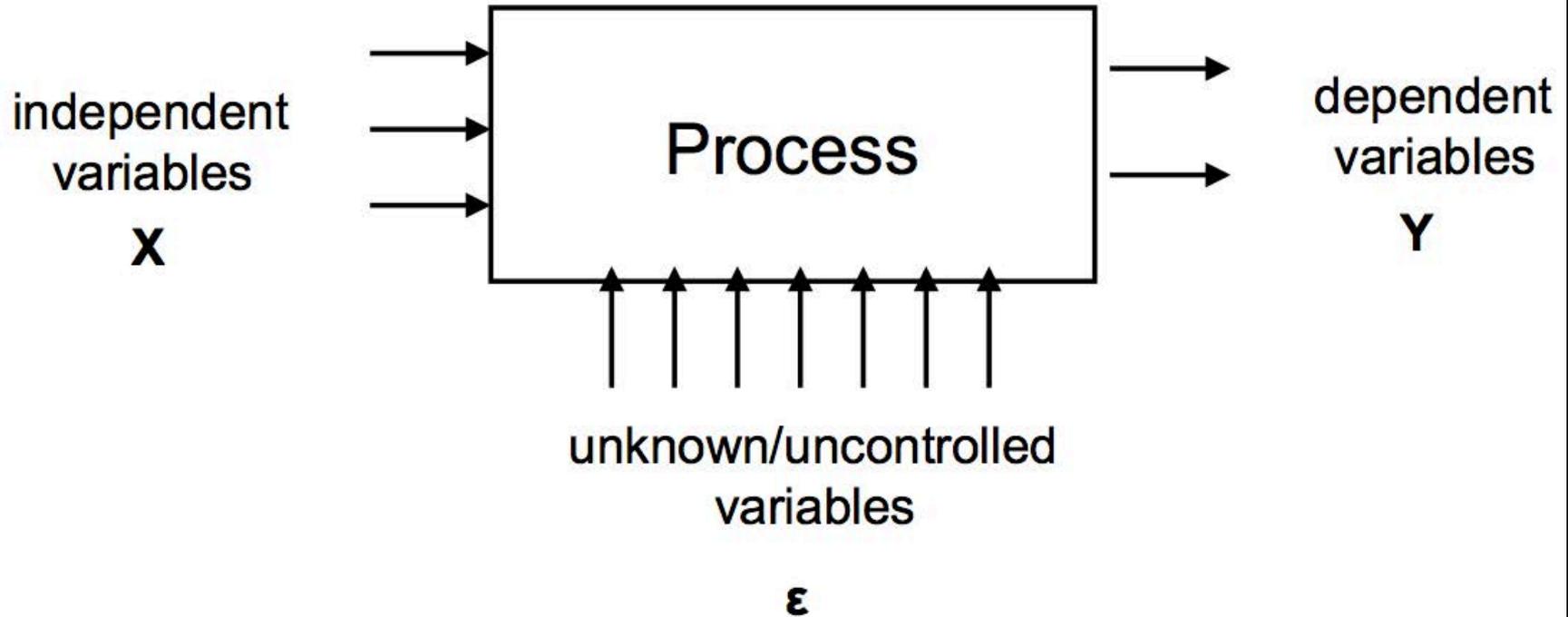
IS X BETTER THAN Y?

- X Most of time, the answer is, “it depends”.
- X On what condition, is X better than Y?
- X Controlled comparison enables causal inference.

EXPERIMENT DESIGN 101: REMEMBER SCIENCE CLASS?

- X **Hypothesis:** testable, quantifiable, and measurable
- X **Independent variable:** what is manipulated to test the hypothesis?
 - Different systems, user classes, tasks
- X **Dependent variable:** what is measured?
 - Time, errors, accuracy, # tasks done, satisfaction
 - Statistical tests to accept/reject the hypothesis

$$Y = f(X) + \varepsilon$$

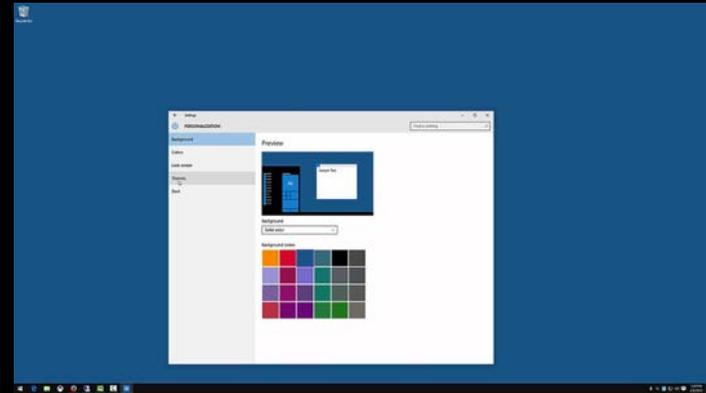


FACTORS TO CONSIDER

- X Users
 - How to sample them? Posting on Ara is sufficient?
- X Implementation
 - Real vs Controlled
- X Tasks
 - Realistic vs Artificial
- X Measurement
 - Capturing dependent variables
- X Ordering
 - Tasks and conditions
- X Hardware
 - What particular devices/machines?

LET'S PRACTICE: MAC VS WINDOWS MENU BAR SPEED COMPARISON

- X Hypothesis?
- X Independent variable?
- X Dependent variable?
- X Users
- X Implementation
- X Tasks
- X Measurement
- X Ordering
- X Hardware



THREATS TO VALIDITY

CONCERNS DRIVING EXPERIMENT DESIGN

- X **Internal validity: precision**
 - Are observed results actually caused by IV?
- X **External validity: generalizability**
 - Can observed results be generalized to the world outside the lab?
- X **Reliability: consistency**
 - Will consistent results be obtained by repeating the experiment?

THREATS TO INTERNAL VALIDITY

- X Ordering effects
 - Always A before B
- X Selection bias
 - CS majors to use my menubar, non-majors to use baseline
- X Experimenter bias
 - “My interface” vs “Some baseline I should beat”

THREATS TO EXTERNAL VALIDITY

- X Population
 - KAIST CS undergrads to evaluate my menubar
 - Selection bias
- X Ecological
 - “In-lab” study of a car dashboard interface
- X Training
 - 1-hour tutorial on the menubar
- X Task
 - Realistic and representative?

THREATS TO RELIABILITY

- X Uncontrolled variation
 - Previous experience
 - User differences
 - Task design
 - Measurement error

POTENTIAL CONCERNS?

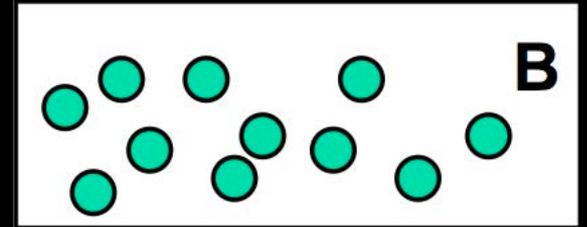
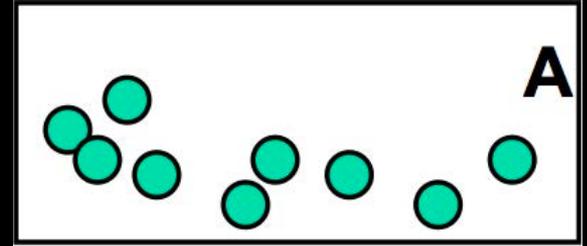
X Hypothesis: *“box A has a different number of balls from box B.”*

X IV? DV?

X Internal validity?

X External validity?

X Reliability?



BETWEEN-SUBJECTS & WITHIN-SUBJECTS

X Common user study patterns

X Between-subjects

- Users are divided into two groups.
- One IV condition within each group (group A: cond X, group B: cond Y)
- Results are compared between groups.
- Eliminates variation due to ordering effects.

X Within-subjects

- Each user sees both conditions (in random order).
- Results are compared within each user.
- Eliminates variation due to user differences.

ACTIVITY: ANALYZE A PAPER'S EXPERIMENT

- X In pairs, spend 15 mins to analyze the experiment & answer questions.
- X [Static] vs [Automatic+Dynamic] vs [Customizable+Dynamic] split menus
- X Skim pages 92-93.

bit.ly/i2r-exp1



Move items up and down in top partition

Move items from bottom to top partition

DISCUSSION: ANALYZE A PAPER'S EXPERIMENT

Study setup

- Hypothesis / IV / DV?
- Lab study? Field Study? Survey?
- Between-subjects vs Within-subjects? Ordering?
- Users / Implementation / Tasks / Measurement / Hardware

Internal validity

- Ordering effects / Selection effects / Experimenter bias?

External validity

- Population / Ecological / Training / Task?

Reliability

- User differences / Measurement error / Repetition

ONLINE EXPERIMENTATION

A/B TESTING OBAMA CAMPAIGN

- X 500 A/B tests over 20 months
- X Donation conversion increased by 49%
- X Sign up conversions increased by 161%

CONTROL

OBAMA  BIDEN

Log in | Create account



You could be there
Michael Jordan, Patrick Ewing, Shaquille O'Neal, Carmelo Anthony, and many, many more basketball greats are all stepping up to support President Obama.
You and a guest will meet the President—and shoot hoops with some of the best basketball players alive.
[More players to be announced](#)

Contributor

First name: Last name:

Address:

City: State: ZIP:

Email: Phone number:

Select amount

Other:

Credit card

VISA   

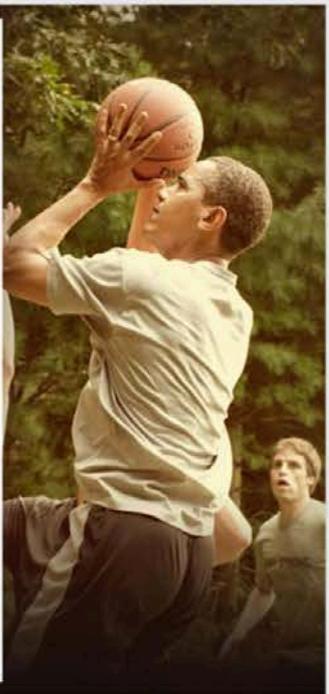
Card number: Expiration: Month Year

Make this a monthly donation to sustain this campaign in the time we have left — Election Day is closer than you think.

Employment

Federal law requires us to use our best efforts to collect and report the name, mailing address, occupation, and employer of individuals whose contributions exceed \$200 in an election cycle.

Employer: Occupation:



“SEQUENTIAL”

OBAMA  BIDEN

Log in | Create account



You could be there
Michael Jordan, Patrick Ewing, Shaquille O'Neal, Carmelo Anthony, and many, many more basketball greats are all stepping up to support President Obama.
You and a guest will meet the President—and shoot hoops with some of the best basketball players alive.
[More players to be announced](#)

How much would you like to donate today?

Select amount

[Home](#) [Sign in](#) [Privacy](#) [Feedback](#)



↑ +5%

<http://kylerush.net>

CONTROL

OBAMA BIDEN

✓ Your donation was successful. Thank you. Secure

Save your payment information for next time.

Save the payment information from this donation into your existing BarackObama.com account so that you don't have to enter it more than once.

You can manage your saved payment information by clicking your name in the top right when you are logged in.

Email:

Enter your BarackObama.com account password:

[Forgot your password?](#)

Link my mobile phone number with my account so that I can donate and receive campaign updates via text message. Message and data rates may apply. My mobile number is: [change this number](#)

Keep me logged in. (If this is a public computer or you are concerned about unauthorized access to this device, you should not enable this setting.)

[← SAVE PAYMENT INFORMATION](#)

[Maybe next time](#)

OBAMA BIDEN

[Privacy Policy](#) [Terms of Service](#)

[FIND FOR BY OBAMA FOR AMERICA](#)

© 2011–2012 Obama for America. All Rights Reserved.

SEGUE COPY

OBAMA BIDEN

✓ Your donation was successful. Thank you. Secure

Now, save your payment information.

Save the payment information from this donation into your existing BarackObama.com account so that you don't have to enter it more than once.

You can manage your saved payment information by clicking your name in the top right when you are logged in.

Email:

Enter your BarackObama.com account password:

[Forgot your password?](#)

Link my mobile phone number with my account so that I can donate and receive campaign updates via text message. Message and data rates may apply. My mobile number is: [change this number](#)

Keep me logged in. (If this is a public computer or you are concerned about unauthorized access to this device, you should not enable this setting.)

[← SAVE PAYMENT INFORMATION](#)

[Maybe next time](#)

OBAMA BIDEN

[Privacy Policy](#) [Terms of Service](#)

[FIND FOR BY OBAMA FOR AMERICA](#)

© 2011–2012 Obama for America. All Rights Reserved.

↑ +21%

CONTROL

OBAMA  BIDEN

DINNER WITH BARACK

Your chance to meet the President

[GET STARTED](#)



No purchase, payment, or contribution necessary to enter or win. Contributing will not improve chances of winning. Void where prohibited. Entries must be received by September 20, 2012. You may enter by contributing to Obama Victory Fund 2012 here or click here to enter without contributing. Three winners will each receive the following prize package: round-trip tickets for winner from within the 48 U.S. States, DC, or Puerto Rico to a destination to be determined by the Sponsor; hotel accommodations; and dinner with President Obama on a date to be determined by the Sponsor (approximate retail value of all prizes \$4,800). Odds of winning depend on number of entries received. Promotion open only to U.S. citizens, or lawful permanent U.S. residents who are legal residents of 50 United States, District of Columbia and Puerto Rico and 18 or older (or age of majority under applicable law). Promotion subject to Official Rules. Official rules and additional restrictions on eligible. Sponsor: Obama for America, 130 E. Randolph St., Chicago, IL 60601.

OBAMA  BIDEN

[Privacy Policy](#) [Terms of Service](#)

Contributions or gifts to Obama Victory Fund 2012 are not tax deductible.

FINN FOR BY OBAMA VICTORY FUND 2012, A JOINT FUNDRAISING COMMITTEE AUTHORIZED BY OBAMA FOR AMERICA, THE DEMOCRATIC NATIONAL COMMITTEE, AND THE STATE DEMOCRATIC PARTIES IN THE FOLLOWING STATES: CO, FL, IA, IL, IN, MI, NC, OH, PA, VA, AND WI.

© 2011-2012 Obama for America. All Rights Reserved.

IMAGE VARIATION

OBAMA  BIDEN

DINNER WITH BARACK

Your chance to meet the President

[GET STARTED](#)



No purchase, payment, or contribution necessary to enter or win. Contributing will not improve chances of winning. Void where prohibited. Entries must be received by September 20, 2012. You may enter by contributing to Obama Victory Fund 2012 here or click here to enter without contributing. Three winners will each receive the following prize package: round-trip tickets for winner from within the 48 U.S. States, DC, or Puerto Rico to a destination to be determined by the Sponsor; hotel accommodations; and dinner with President Obama on a date to be determined by the Sponsor (approximate retail value of all prizes \$4,800). Odds of winning depend on number of entries received. Promotion open only to U.S. citizens, or lawful permanent U.S. residents who are legal residents of 50 United States, District of Columbia and Puerto Rico and 18 or older (or age of majority under applicable law). Promotion subject to Official Rules. Official rules and additional restrictions on eligible. Sponsor: Obama for America, 130 E. Randolph St., Chicago, IL 60601.

OBAMA  BIDEN

[Privacy Policy](#) [Terms of Service](#)

Contributions or gifts to Obama Victory Fund 2012 are not tax deductible.

FINN FOR BY OBAMA VICTORY FUND 2012, A JOINT FUNDRAISING COMMITTEE AUTHORIZED BY OBAMA FOR AMERICA, THE DEMOCRATIC NATIONAL COMMITTEE, AND THE STATE DEMOCRATIC PARTIES IN THE FOLLOWING STATES: CO, FL, IA, IL, IN, MI, NC, OH, PA, VA, AND WI.

© 2011-2012 Obama for America. All Rights Reserved.

 +19%

bing

WEB IMAGES VIDEOS MAPS NEWS MORE

amazon

ALSO TRY: Walmart · Target · Netflix

202,000,000 RESULTS Any time ▾

Ad related to amazon

[Amazon.com® Official Site - Huge Selection and Amazing Prices.](#)
[www.Amazon.com](#) - Official Site
 Free Shipping on Orders Over \$25

Amazon Prime	Prime Instant Video
Join Amazon Student	Rent Textbooks
Download eTextbooks	Kindle Store

[Amazon.com: Online Shopping for Electronics, Apparel, Computers ...](#)
[www.amazon.com](#) ▾ Official site
 Customer service 866-216-1072
 Online shopping from the earth's biggest selection of books, magazines, music, DVDs, videos, electronics, computers, software, apparel & accessories, shoes, jewelry ...

Books	Music
Shop online for millions of new and used books	Browse best sellers, new releases, deals, vinyl records, MP3s, and ...

bing

WEB IMAGES VIDEOS MAPS NEWS MORE

amazon

ALSO TRY: Walmart · Target · Netflix

202,000,000 RESULTS Any time ▾

Ad related to amazon

[Amazon.com® Official Site - Huge Selection and Amazing Prices.](#)
[www.Amazon.com](#) - Official Site
 Free Shipping on Orders Over \$25

Amazon Prime	Prime Instant Video
Join Amazon Student	Rent Textbooks
Download eTextbooks	Kindle Store

[Amazon.com: Online Shopping for Electronics, Apparel, Computers ...](#)
[www.amazon.com](#) ▾ Official site
 Customer service 866-216-1072
 Online shopping from the earth's biggest selection of books, magazines, music, DVDs, videos, electronics, computers, software, apparel & accessories, shoes, jewelry ...

Books	Music
Shop online for millions of new and used books	Browse best sellers, new releases, deals, vinyl records, MP3s, and ...

More successful at completing tasks,
 time-to-success shorter, +\$10M annually

WHY USE ONLINE CROWDS VS LAB PARTICIPANTS?

- x Reduced cost
- x Faster completion
- x Ease of recruiting
- x Scalability
- x Diversity
- x Lack of control
- x Lack of expertise
- x Lack of motivation
- x Distraction
- x Quality control

Experimental control vs Ecological validity

crowdsourcing

noun | crowd·sourc·ing | \ˈkraʊd-sɔr-sɪŋ\

Definition of CROWDSOURCING

Popularity: Bottom 50% of words

: the practice of obtaining needed services, ideas, or content by soliciting contributions from a large group of people and especially from the online community rather than from traditional employees or suppliers

scale

outcome

undefined crowd
open call

task

Mechanical Turk is a marketplace for work.

We give businesses and developers access to an on-demand, scalable workforce.
Workers select from thousands of tasks and work whenever it's convenient.

910,392 HITs available. [View them now.](#)

Make Money by working on HITs

HITs - *Human Intelligence Tasks* - are individual tasks that you work on. [Find HITs now.](#)

As a Mechanical Turk Worker you:

- Can work from home
- Choose your own work hours
- Get paid for doing good work



or [learn more about being a Worker](#)

Get Results from Mechanical Turk Workers

Ask workers to complete HITs - *Human Intelligence Tasks* - and get results using Mechanical Turk. [Get Started.](#)

As a Mechanical Turk Requester you:

- Have access to a global, on-demand, 24 x 7 workforce
- Get thousands of HITs completed in minutes
- Pay only when you're satisfied with the results



<u>Scene Attribute Labeling</u>		View a HIT in this group
Requester: James Hays Lab	HIT Expiration Date: Jun 19, 2011 (6 days 19 hours)	Reward: \$0.03
	Time Allotted: 15 minutes	HITs Available: 5274
<u>NRS V2 FR Electronics Top-Fri Jun 10 20:44:54 PDT 2011</u>		View a HIT in this group
Requester: Amazon Requester Inc.	HIT Expiration Date: Jul 10, 2011 (4 weeks)	Reward: \$0.01
	Time Allotted: 60 minutes	HITs Available: 5160
<u>NRS V2 DE Electronics Bottom-Fri Jun 10 22:51:39 PDT 2011</u>		View a HIT in this group
Requester: Amazon Requester Inc.	HIT Expiration Date: Jul 11, 2011 (4 weeks)	Reward: \$0.01
	Time Allotted: 60 minutes	HITs Available: 3952
<u>Image Annotation</u>		View a HIT in this group
Requester: Liyue Zhao	HIT Expiration Date: Jun 14, 2011 (1 day 22 hours)	Reward: \$0.01
	Time Allotted: 10 minutes	HITs Available: 3642
<u>Copy text from scanned image (Handwritten)</u>		View a HIT in this group
Requester: HAVI Digital	HIT Expiration Date: Jun 25, 2011 (1 week 6 days)	Reward: \$0.18
	Time Allotted: 2 hours	HITs Available: 3170
<u>NRS V2 IT Electronics Top-Fri Jun 10 23:30:07 PDT 2011</u>		View a HIT in this group
Requester: Amazon Requester Inc.	HIT Expiration Date: Jul 10, 2011 (4 weeks)	Reward: \$0.01
	Time Allotted: 60 minutes	HITs Available: 3062
<u>Copy text from scanned image (Caption)</u>		View a HIT in this group
Requester: HAVI Digital	HIT Expiration Date: Jun 25, 2011 (1 week 6 days)	Reward: \$0.20
	Time Allotted: 2 hours	HITs Available: 2768
<u>Create listings for consumer products</u>		View a HIT in this group
Requester: Buyer's Best Friend	HIT Expiration Date: Jun 19, 2011 (6 days 23 hours)	Reward: \$0.05
	Time Allotted: 60 minutes	HITs Available: 2759
<u>Moderate Internet Images for Violence, Hate, and Sexual Content (contains mature content) Job 4</u>		View a HIT in this group
Requester: Dolores Labs	HIT Expiration Date: Jun 19, 2011 (6 days 23 hours)	Reward: \$0.04
	Time Allotted: 60 minutes	HITs Available: 2038



Provide 3 tags for the image (Add 3 tags) [Show hit instructions](#)

Image 1 of 5 [Send feedback](#)

Preview mode please accept the hit



Please enter 3 tags, confirm with enter:

Want to work on this HIT?

[Accept HIT](#)

Want to see other HITs?

[Skip HIT](#)

CROWDSOURCING FOR EXPERIMENTATION & DATA COLLECTION

x Experimentation

- Less WEIRD
- Large scale
- Programmable

x Data Collection

- Large scale datasets
- ML, CV, NLP
- Low cost

FACEBOOK EMOTION STUDY

“We show, via a massive (N = 689,003) experiment on Facebook, that emotional states can be transferred to others via emotional contagion, leading people to experience the same emotions without their awareness.”

Kramer et al., PNAS vol. 111 no. 24

- X A week in 2012
- X Modified News Feed to show more/less emotional posts
- X Seeing more negative posts → more negative posts
- X Seeing more positive posts → more positive posts
- X Seeing less emotional posts → less emotional posts

MAKING ONLINE EXPERIMENTS WORK

- X Instruction comprehension
- X Environment differences
- X Qualification task
- X Outlier removal
- X Implementation issues
- X Ethical concerns

RESOURCES

- X CS374's Experiment Design reading
 - Much of today's content is from this material.
 - <https://www.kixlab.org/courses/cs374-spring-2018/classes/20-Experiment-Design/>