



Crowdsourcing: The good, the bad, the ugly

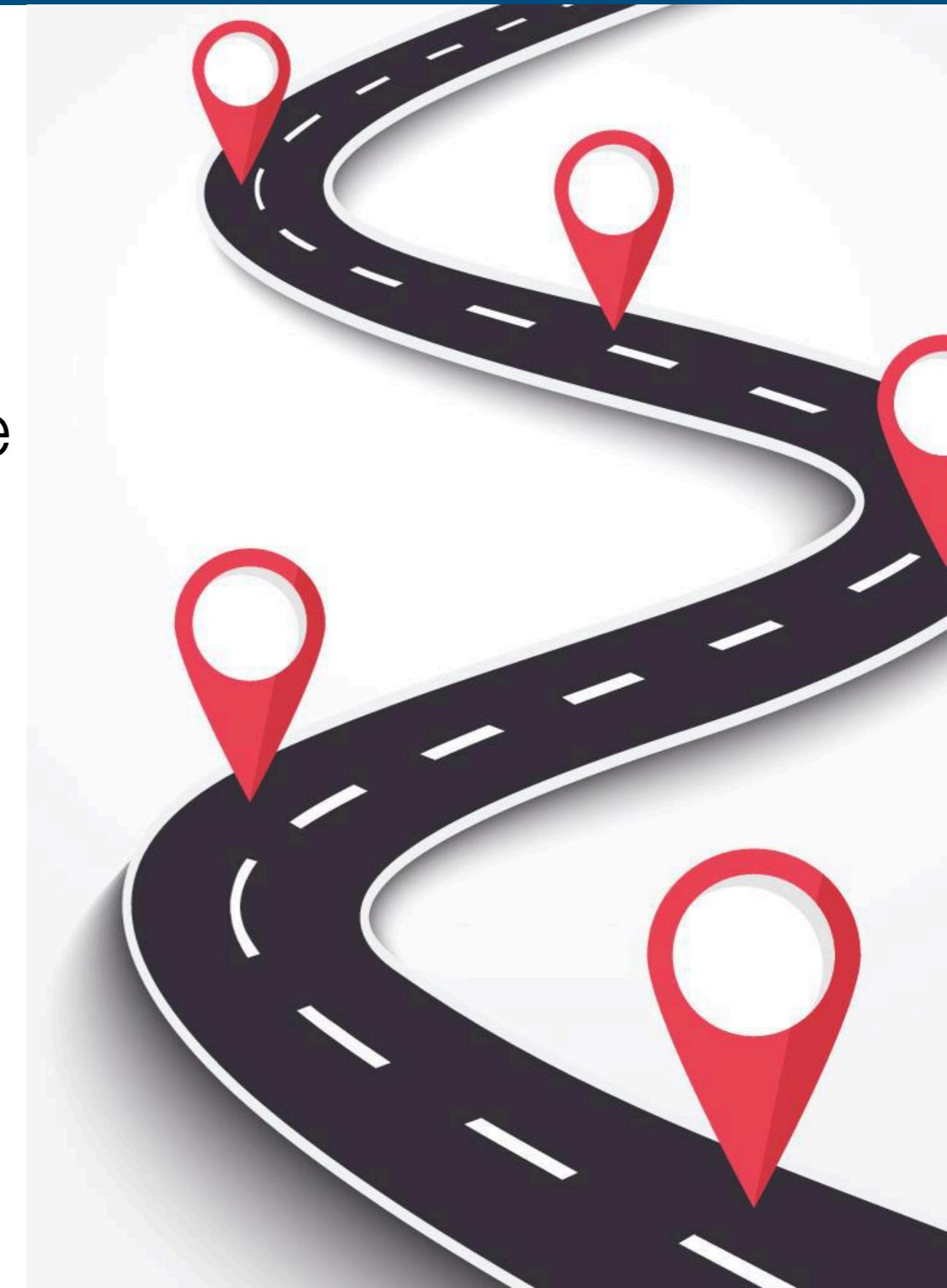
University of Glasgow M&M Talk
2 October 2019

Carolyn Saund



Roadmap

- Crowdsourcing & You
 - Uses
- Designing tasks for online data collection
- Demo





Crowdsourcing You



What Is It?

Crowdsourcing & You



Open Innovation

The crowd provides innovative product ideas or develops problem solutions.

Creative

The crowd participates in tasks of the creative sector.

Crowd Sourcing

Crowdfunding

The crowd finances projects and business ideas of all sorts.

Microtasking

The crowd edits microtasks that result out of major projects.

Crowdtesting

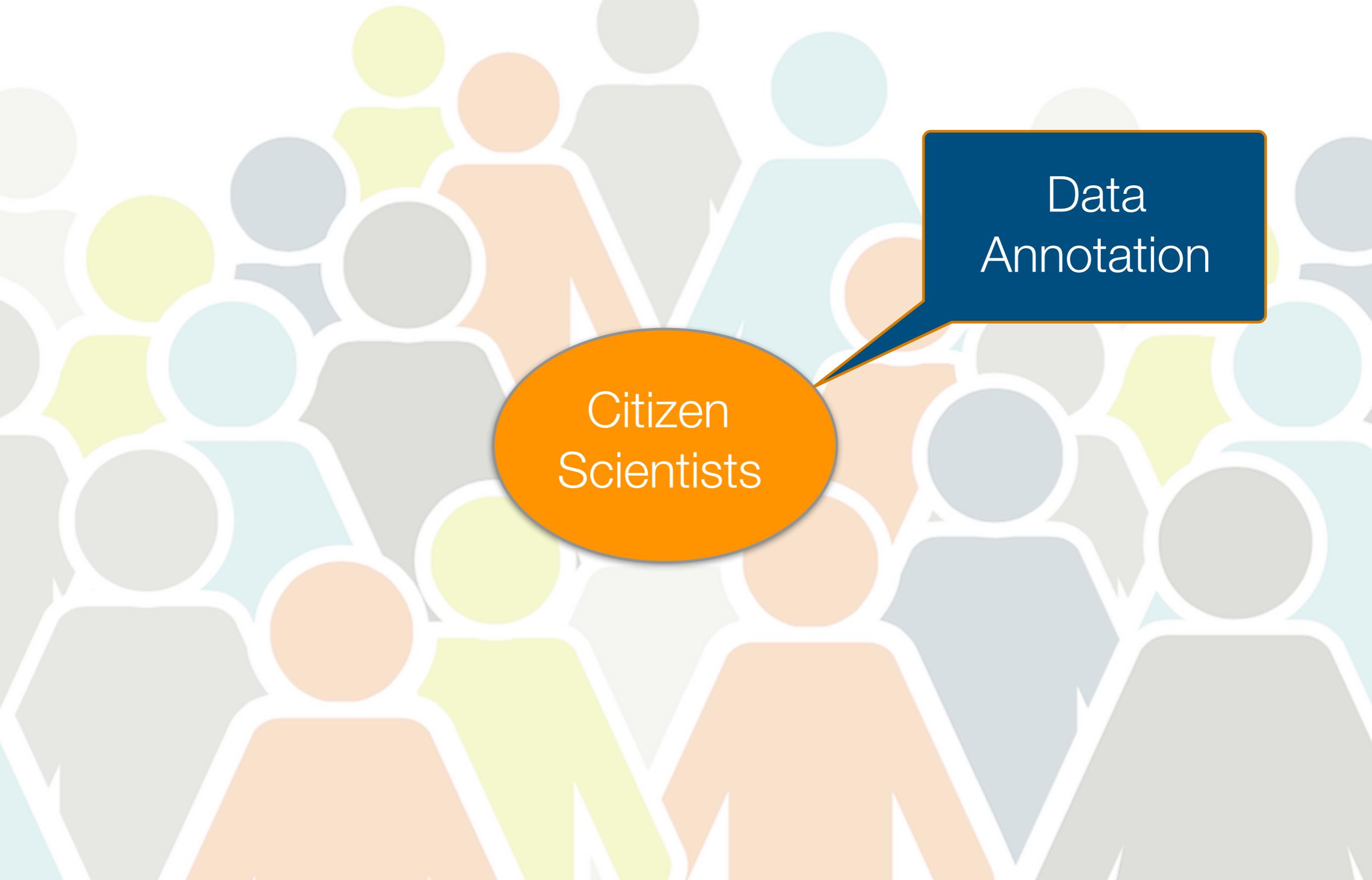
The crowd tests software.

Colaborative Knowledge

The crowd provides knowledge.

What Is It?

Crowdsourcing & You



A background of numerous overlapping human silhouettes in various colors (light blue, grey, orange, yellow) creates a sense of a diverse crowd.

Citizen
Scientists

Data
Annotation

What Is It Good For?

Crowdsourcing & You



- Annotation

- <http://pagan.institutedigitalgames.com/howto.php>

The screenshot displays the PAGAN annotation interface. On the left, a vertical sidebar lists project-related buttons: Project Title, My Project, Annotate, Annotation Target, RankTrace, Paste the address, Play Video, and Submit. The main area shows a video feed of a woman and a 3D male head model. Below the video is a line graph labeled "arousal" with an upward-pointing arrow labeled "increase" and a downward-pointing arrow labeled "decrease". To the left of the video, a large text box contains instructions: "Hello!", "By pasting your labelling task URL here, you can quickly get started.", "Please turn on your microphone and camera.", "The video will play automatically.", "Please use the [up] (in red) key to indicate the level of arousal.", "Please wait until the video has loaded completely.", and "label trace". A small "label trace" icon is shown at the bottom of the video frame. On the right, a "TEST MODE" window is open, showing a form to "Create a new project below". The form fields include:

- Project Title: My Project
- Annotation Target: arousal
- Optional Target Description: An optional short description displayed both above and below the target at the beginning of the annotation, use it to help participants understand the labelling task (max 200 characters).
- Annotation Type: RankTrace (radio button selected)
- Project Source:
 - Uploaded Videos (radio button selected)
 - YouTube
 - Subject Upload
 - Subject YouTube
- Choose Files: No file chosen
- How to Load Videos:
 - Randomly (radio button selected)
 - In Sequence
 - Endless Mode
- Number of Annotations a Participant Completes:
 - play all (default)
 - out of 0
- Optional End-Plate Message: An optional message displayed to your participants below the automatic thank you message at the end of the annotation (max 200 characters).
- Optional Google Forms Survey Link: Survey button and not be selected if left empty.
- Autofill Participant ID to Google Forms: id number of item entry (e.g. 1234567890)

At the bottom of the "TEST MODE" window are "submit" and "reset" buttons.

What Is It Good For?

Crowdsourcing & You



- Annotation (& Segmentation)

Table 1. Detection results on 80 images (Pathologists' annotation as GT) GT nuclei = 4436

Annotations	TP	FN	FP	TPR %	PPV %	F-M %
Research Fellow	4109	327	227	92.63 ± 0.8	94.76 ± 0.7	93.68
Automated Method	3735	701	416	84.20 ± 1.1	89.98 ± 1.0	86.99
Contributor Level 1	3814	622	434	85.98 ± 1.1	89.78 ± 1.0	87.84
Contributor Level 2	4016	420	625	90.53 ± 0.9	86.53 ± 1.0	88.49
Contributor Level 3	3787	649	457	85.37 ± 1.1	89.23 ± 0.9	87.26

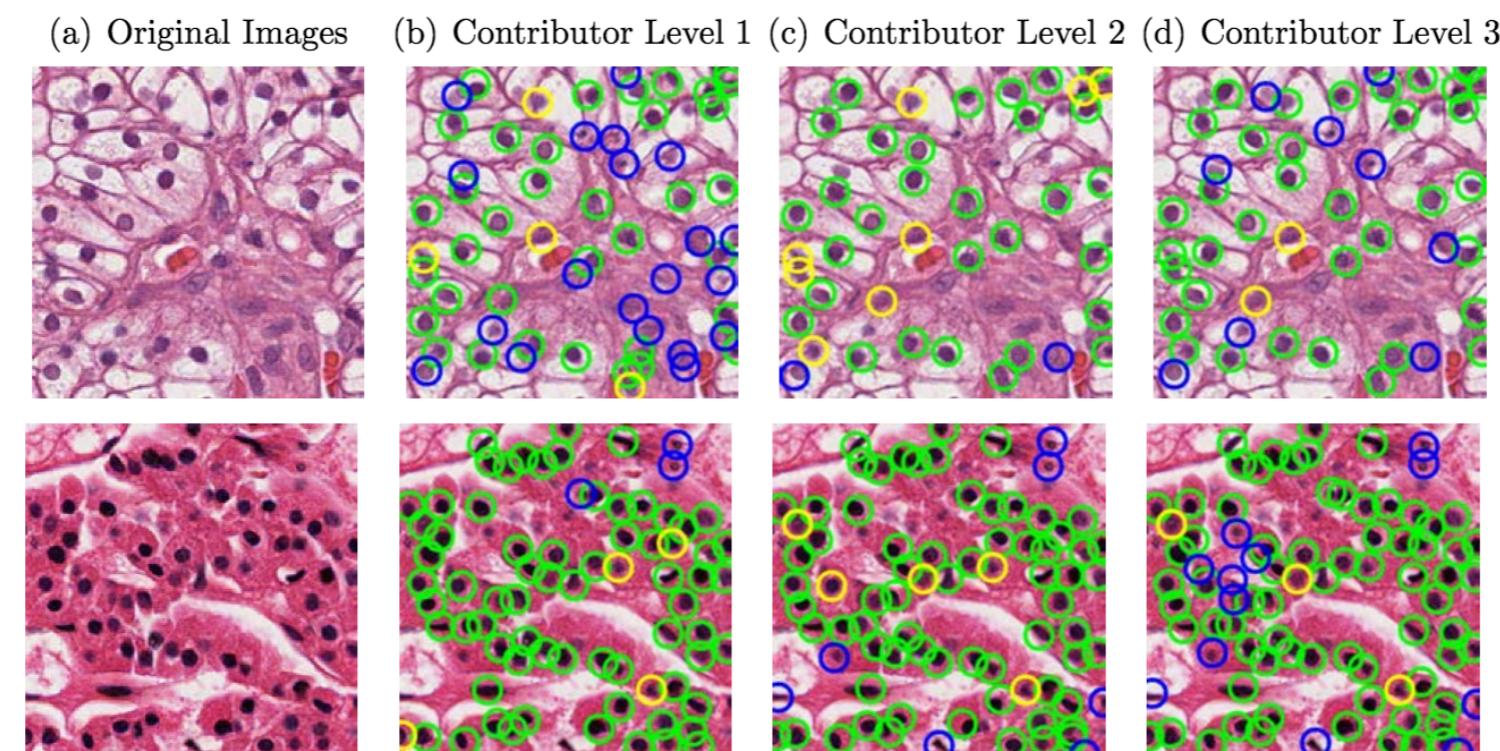


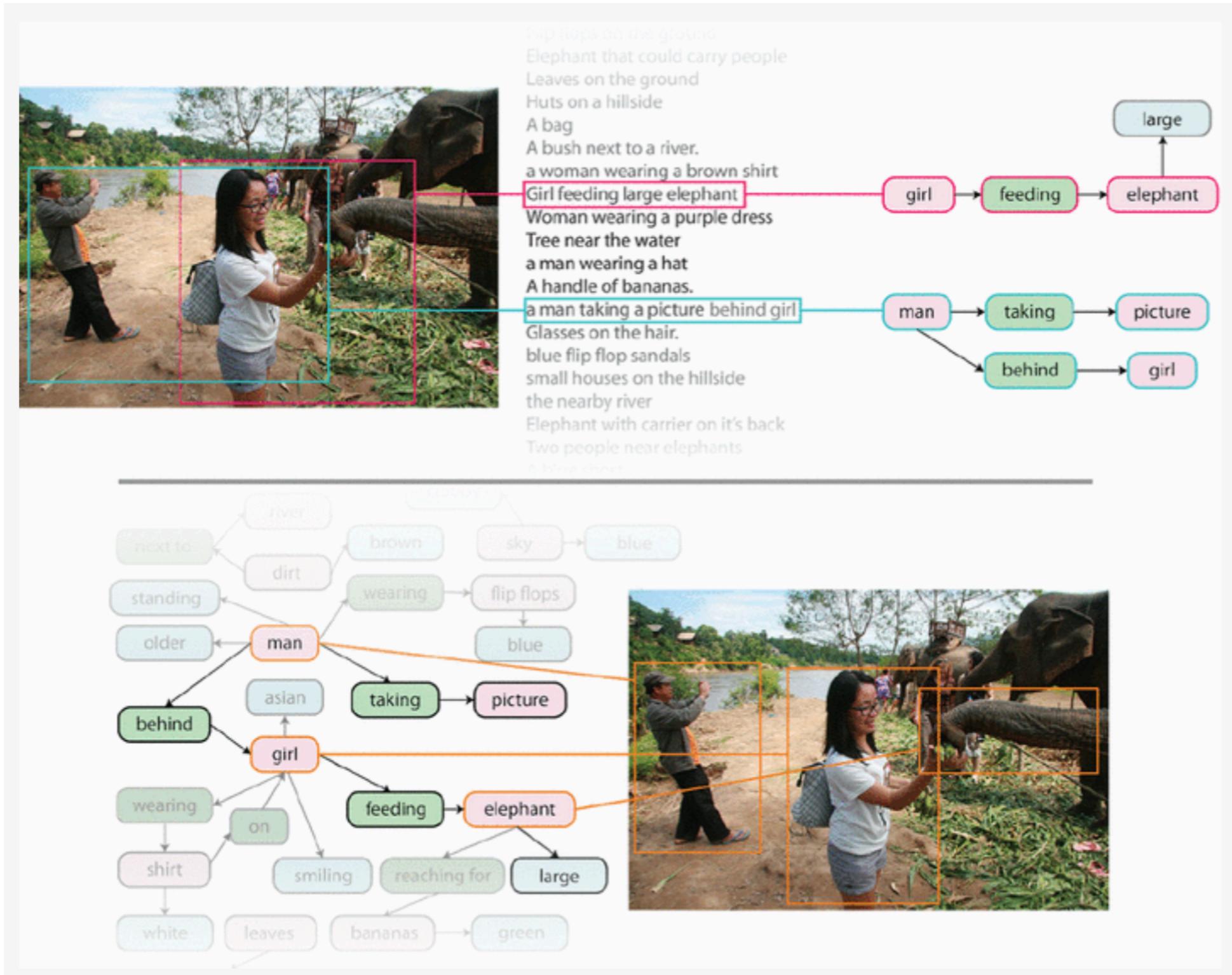
Fig. 1. Examples of nucleus detection results produced by different contributor levels (Green circle indicates TP nuclei, yellow circle indicates FN nuclei and blue circle indicates FP). The automated detected nuclei were used as ground truth.

What Is It Good For?

Crowdsourcing & You

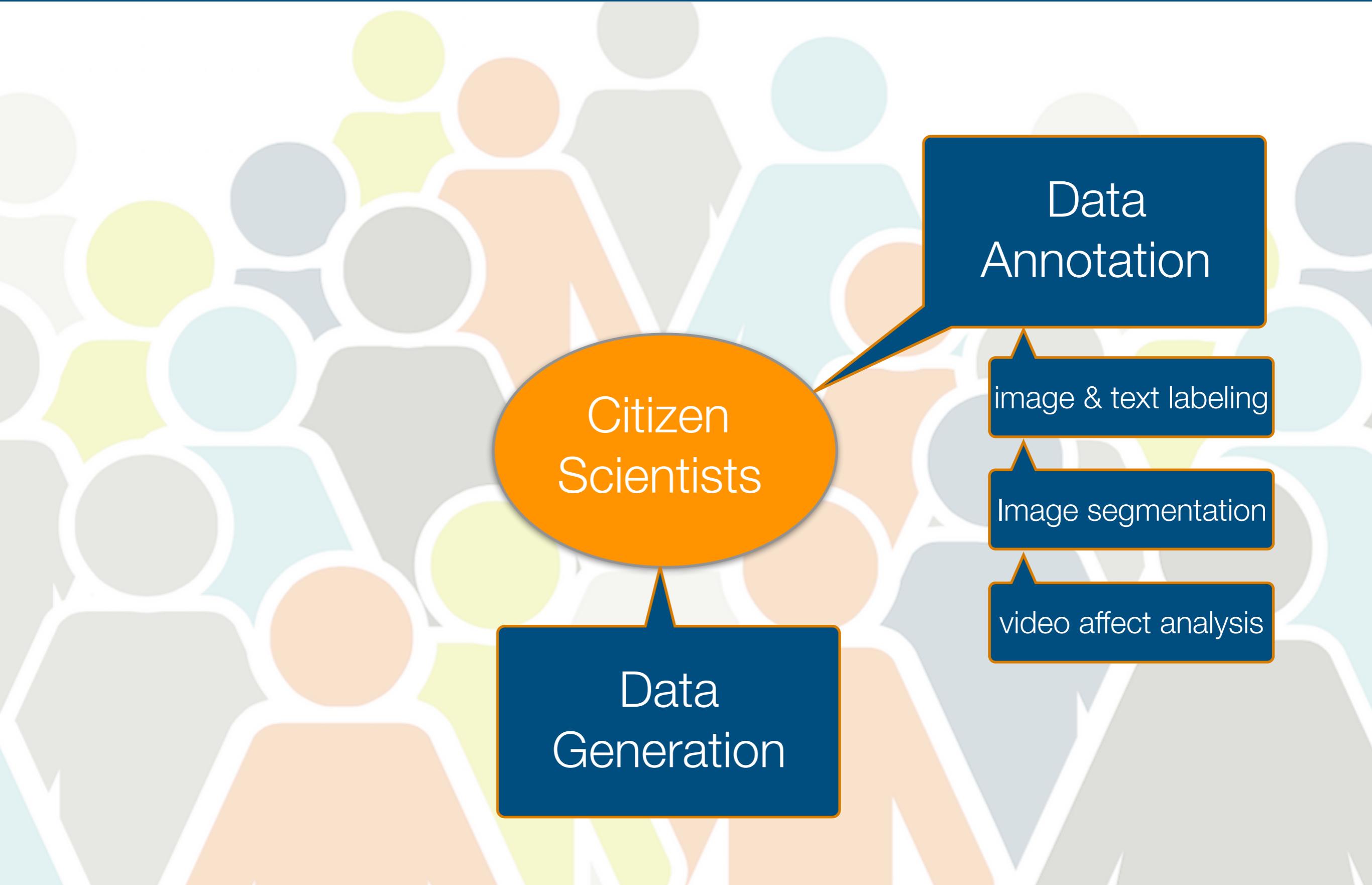


- Annotation (& Labeling)



What Is It?

Crowdsourcing & You



Citizen
Scientists

Data
Generation

Data
Annotation

image & text labeling

Image segmentation

video affect analysis

What Is It Good For?

Crowdsourcing & You



- Story & Natural Language Generation
 - Natural Text
 - Natural Interactions
 - Motivation/Goal analysis
 - Hugely creative
- Model Analysis

Prompt: You (Allen) are trying to enter the house of a celebrity, Christopher. Christopher's body guard, Beth, is guarding the house. Please use the following chat window to act out this scenario.

Welcome, user Exit Chat

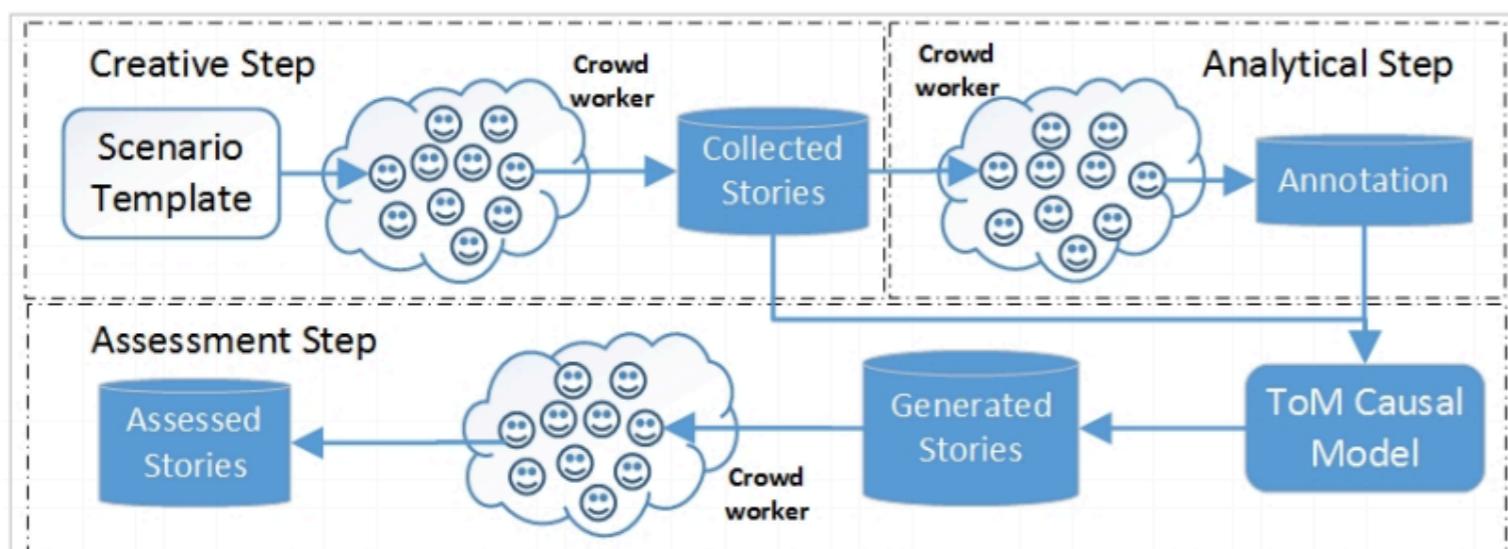
You (Allen): Hello, may I come in?

Them (Beth): No, you are not allowed.

You (Allen): I have a special delivery for Christopher.

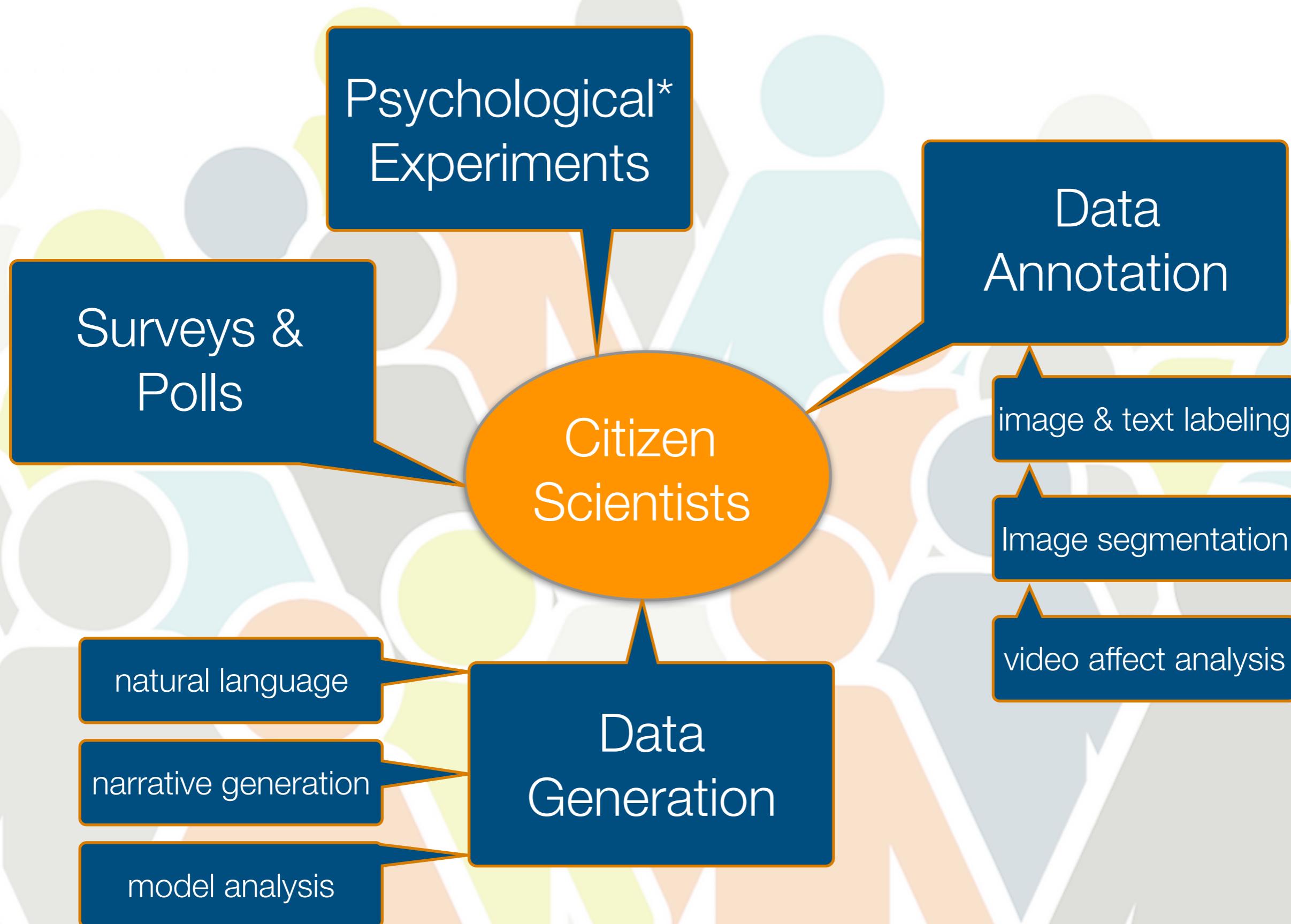
Them (Beth): He didn't tell me about that. You cannot come in.

What do you say next? Send



What Is It?

Crowdsourcing & You



What Is It Good For?

Crowdsourcing & You



- Surveys and Polling
 - Many tools have their own crowdsourcing integration

SoGoSurvey

ZOHO

surveygizmo.

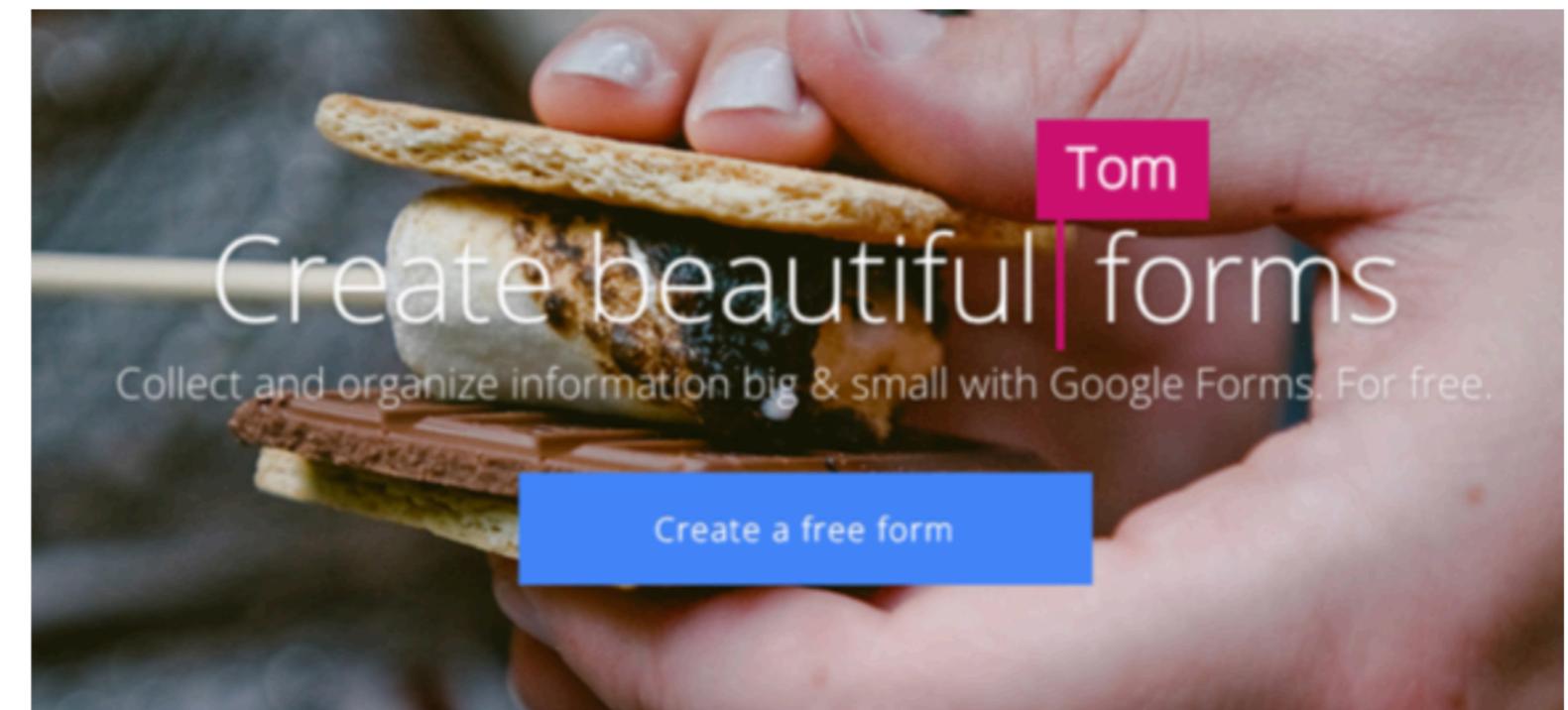
Typeform | Ask awesomely

 **SurveyMonkey™**



 **surveyplanet™**

 **Client Heartbeat**



The Good

Crowdsourcing & You



- Anyone and Everyone... ish
- Huge audiences, super quickly (150-200 / 1-2hrs)
- Much more diverse sample
- Flexible as to what you want

Worker requirements

Require that Workers be Masters to do your tasks (Who are Mechanical Turk Masters?)

Yes No

-- Select --

System Qualifications

HIT Approval Rate (%) for all Requesters' HITs

✓ Location

Number of HITs Approved

Premium Qualifications

Age 18-25

Age 25-30

Age 30-35

Age 35-45

Age 45-55

Age 55 or older

Blogger

Borrower - Auto Loans

Borrower - Business Loan

Borrower - Credit Cards

Borrower - Home Mortgage

Borrower - Personal Loan

Borrower - Student Loan

Car Owner

Current Residence - Owned

Current Residence - Rented

Daily Internet Usage - 1 to 4 hours

Daily Internet Usage - 5 to 7 hours

Daily Internet Usage - 7+ hours

Employment Industry - Banking & Financial Services

Employment Industry - Education

Employment Industry - Food & Beverage

Employment Industry - Government & Non-Profit

Employment Industry - Healthcare

Employment Industry - Manufacturing

A, LAND ISLANDS (AX)
AFGHANISTAN (AF)
ALBANIA (AL)
ALGERIA (DZ)
AMERICAN SAMOA (AS)

Exercise - Every Day
Exercise - Four Plus Times a Week
Exercise - Not at All
Exercise - Once a Week
Exercise - Two to Three Times a Week
Facebook Account Holder
Financial Asset Owned - Certificate of Deposit (CD)
Financial Asset Owned - Checking Account
Financial Asset Owned - Common Share Stocks
Financial Asset Owned - Exchange-Traded Fund (ETF)
Financial Asset Owned - Money Market Account
Financial Asset Owned - Mutual Funds
Financial Asset Owned - Real Estate Investment Trusts (REITs)
Financial Asset Owned - Savings Account
Financial Asset Owned - Stock Options
Financial Asset Owned - U.S. Treasury Bills/Government Debt
Gender - Female
Gender - Male
Google Account Holder
Handedness - Left
Handedness - Right
Household Income - \$100,000 or more
Household Income - \$25,000 - \$49,999
Household Income - \$50,000 - \$74,999
Household Income - \$75,000 - \$99,999
Household Income - Less than \$25,000
Insurance Policyholder - Auto Insurance
Insurance Policyholder - Healthcare Insurance
Insurance Policyholder - Home Owners Insurance
Insurance Policyholder - Life Insurance
Insurance Policyholder - Renters Insurance
Job Function - Accounting & Finance
Job Function - Administrative
Job Function - Arts & Design
Job Function - Education & Training
Job Function - Engineering
Job Function - Information Technology
Job Function - Management
Job Function - Marketing, Sales & Business Development
Job Function - Operations
Language Fluency (Basic) - Brazilian Portuguese

Online Purchase - Clothing & Shoes
Online Purchase - Electronics & Computers
Online Purchase - Groceries & Food
Online Purchase - Handmade Products
Online Purchase - Health & Beauty
Online Purchase - Home & Garden
Online Purchase - Jewelry
Online Purchase - Movies
Online Purchase - Music
Online Purchase - Sports & Outdoor Equipment
Online Purchase - Toys
Online Purchase - Videogames
Parenthood Status
Pinterest Account Holder
Primary Internet Device - Desktop
Primary Internet Device - Laptop
Primary Internet Device - Smartphone or Tablet
Primary Mobile Device - Android
Primary Mobile Device - iPhone
Primary News Source - Online News (News Websites, News Apps)
Primary News Source - Podcasts
Primary News Source - Print (Newspapers & Periodicals)
Primary News Source - Radio (AM/FM, Internet, Satellite)
Primary News Source - Social Media
Primary News Source - TV (Late Night Comedy, Other)
Primary News Source - TV (Local/Cable News Broadcast)
Primary News Source - Word of Mouth
Reddit Account Holder
Single Family Home Resident
Smoker
Tablet Owner
Tumblr Account Holder
Twitter Account Holder
US Bachelor's Degree
US Graduate Degree
US High School Graduate
US Political Affiliation - Conservative
US Political Affiliation - Liberal
Vacation Frequency - Every Few Years
Vacation Frequency - Every Month
Vacation Frequency - Every Quarter
Vacation Frequency - Every Year
Vacation Frequency - Never
Voted in 2012 US Presidential Election
Voted in 2016 US Presidential Election
YouTube Account Holder

The Good

Crowdsourcing & You



- Anyone and Everyone... ish
 - Huge audiences, super quickly (150-200 / 1-2hrs)
 - Much more diverse than standard psychology-student sample
 - Flexible as to who can participate
- Incredibly low-effort
- Cheap

The Bad

Crowdsourcing & You



- Cannot* handle precision
 - Perceptive inputs and timing measurements are impossible (or very very sketchy)
 - *However, some relative-timing experiments are shown to work well

- Stroop

Yin et al. 2014

- Word recognition

Cooke et al. 2011

orange blue green red

blue purple green red

orange blue red green

purple orange red blue

green red blue purple

The Bad

Crowdsourcing & You



- Cannot* handle precision
 - Perceptive inputs and timing measurements are impossible (or very very sketchy)
 - *However, some relative-timing experiments are shown to work well
- Must be **very** simple
 - No one will read instructions
- Motivated for time, not quality
- Subjects aren't exactly naive...

The Ugly

Crowdsourcing & You



- Ethics...
- Fair pay
- Worker treatment

Stop using MTurk for research



Ekaterina Damer [Follow](#)

Jul 31 · 7 min read

The Good, The Bad and the Ugly: Why Crowdsourcing Needs Ethics

September 2013

DOI: 10.1109/CGC.2013.89

Conference: Cloud and Green Computing (CGC), 2013 Third International Conference on

Florian Alexander Schmidt

Schmidt 2013

NEXT ECONOMY

The Internet Is Enabling a New Kind of Poorly Paid Hell

For some Americans, sub-minimum-wage online tasks are the only work available.

ALANA SEMUELS JAN 23, 2018

MIRANDA KATZ BACKCHANNEL 08.23.2017 06:55 AM

Amazon's Turker Crowd Has Had Enough

Workers on Amazon Mechanical Turk have long felt ignored and underpaid. A new platform hopes to give them a better home—and could change the future of crowd work.

Amazon's Mechanical Turk workers protest: 'I am a human being, not an algorithm'

A Christmas email campaign is asking Amazon's CEO Jeff Bezos to improve terms for workers providing cheap digital labour

Is Crowdsourcing Bad for Workers?

"Cloud labor" sites like Crowdflower may be the face of the future—but their wages are stuck in a sweatshop past.

By Michelle Chen

The Ugly

Crowdsourcing & You



- Data integrity
 - “up to 25% fraudulent or suspicious responses”
Smith et al. 2016
 - Bots
 - “Expert survey takers”



Designing Tasks to Mitigate Badness

part art, part science

The Obvious(?)

Designing Tasks



- Include ethics as a landing page
- Give participants a way to contact you
 - When they contact you, respond to them
- Protect participant data
 - Redirect to external site for data collection
- **Pay them if they completed the work**

Confirm you want to do this survey

Please confirm that you want to participate in this survey.

I understand the conditions of this study

Click this button to start survey

Important data protection information

When you start, this survey will store your answers and browser information on the [PsyToolkit server](#). The responsibility for this survey rests entirely with the researcher(s) listed above. [Click here if you do not want to participate now.](#)

Ensuring Data Quality

Designing Tasks



- Bot-Proof the Tasks!
 - Surveys should require unique codes
 - “Honey pot” questions
 - If someone answers for a broken URL → spammer
 - Redirect to external sites and generate unique survey codes to paste back in your tool



Please paste the code below in your mturk tab to receive credit for this HIT.

Make sure you write down the following code in the yellow area:

4NNfgPpv

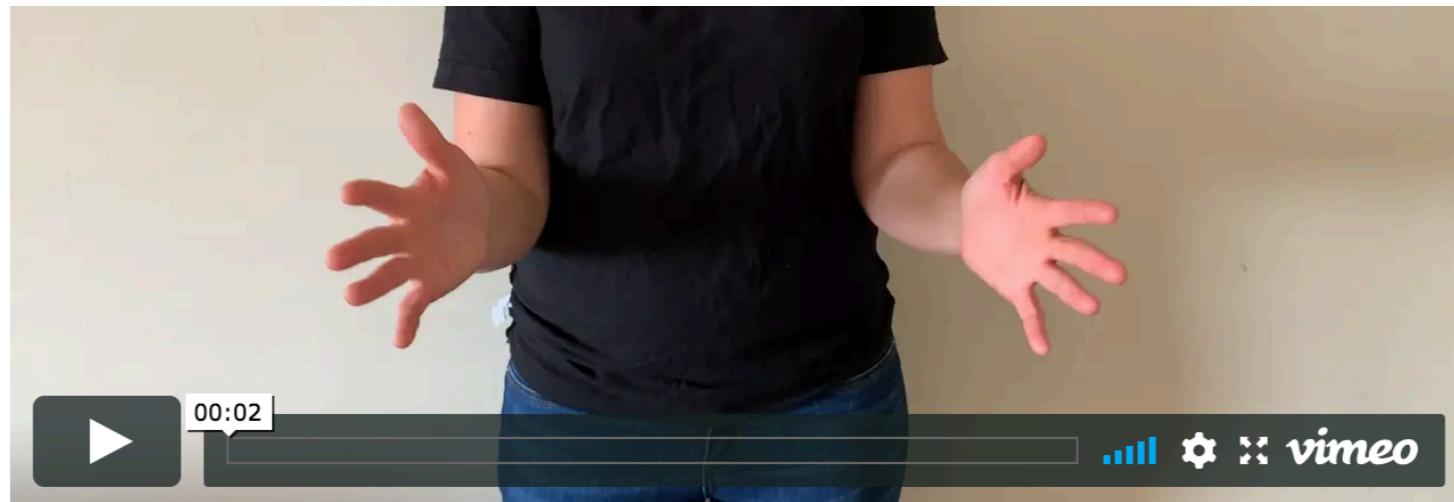
[You are now done with this survey: Now click here to leave this website.](#)

Ensuring Data Quality

Designing Tasks



- Check for Attentiveness
 - Add a vital MC question to check for attentiveness
 - Or better yet, free-response (with word minimum)



If the video link does not work, please go to this link: <https://vimeo.com/334666369>

The speaker is talking about a group of people. Please write what you think they are saying about this group. **Please write at least 10 words.**

What are they saying about the group?

- Optimize to Slow Down ("You will not be able to move on from this page for 30 seconds")

Making it Attractive

Designing Tasks



I'll tell you what I think
of this task.

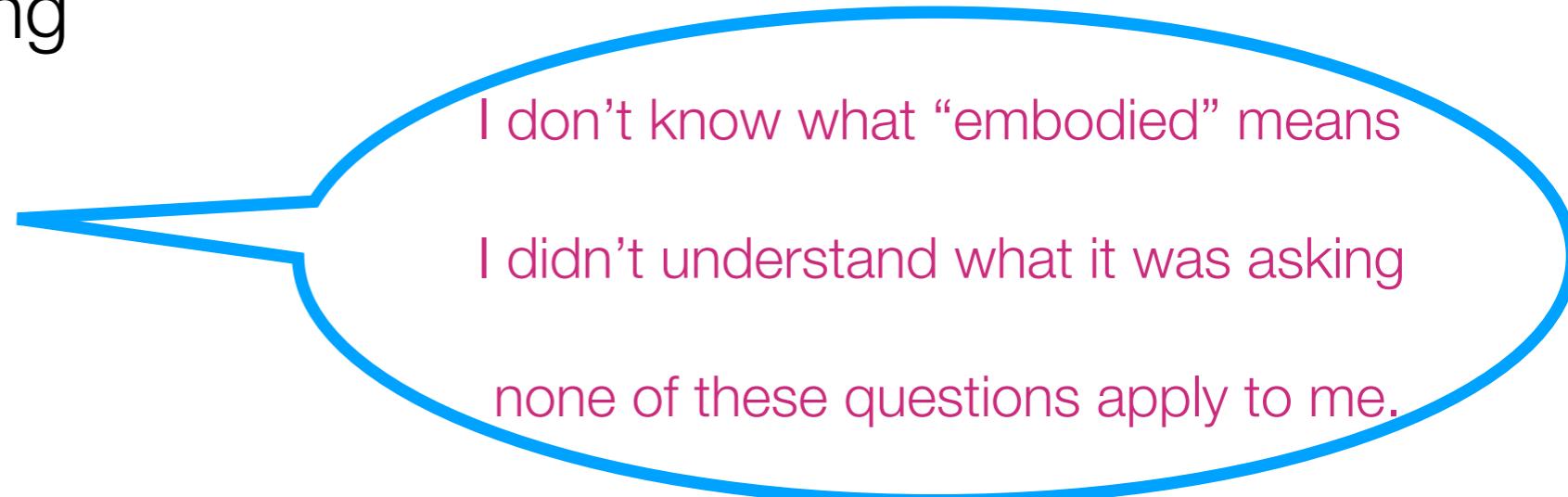
User ID daf439111-e450-11e9-a79d-186590d8f22
aka "daf"

Making it Attractive

Designing Tasks



- Instructions are for chumps and nobody reads them
 - Show examples instead
 - Be **very** clear
- Always ask for feedback in an open-ended input box
 - People are good at knowing what they don't like
 - They can point out flaws we, as experimenters, have a hard time seeing



Making it Attractive

Designing Tasks



- Make the meta-task interesting!
 - Add tags
 - I like fun! oh boy fast money! I'm a scientist!
 - *research, fun, 1 minute, academic, survey, watch video, participate, opinion*
 - I have opinions!
 - Describe payment too!!

Making it Attractive

Designing Tasks



- Pay your workers fairly
 - Don't refuse to pay them if they complete the task
 - Instead, offer bonuses for good behavior
 - Literally no recompense for jerks
 - Too little \$\$ == no interest
 - Too much \$\$ == spammers
 - Block a spammer if you find one





Two Quick Examples

For Example...

Designing Tasks



wait this is a rip off.
Help us describe How-To Videos! Earn \$2.50 bonus for every 25 videos entered!

Watch a how-to video, and write a keyword-friendly synopsis describing the video.

how much do I have to do then...?

1. Click on the link to watch the Film & Theater how-to video => [332492 Get a 35mm film look with a depth of field adaptor](#)
2. Write a description of the video linked in 4 or more sentences.
3. Be detailed in your description. Describe how the procedure is done.
4. Description should be at least 100 words.
5. Description should be fewer than 2000 characters.
6. Use the character and word counters below to help you stay within the limits.
7. You must complete 25 video descriptions in order to earn the \$2.50 bonus. Bonuses are distributed after HITs have been completed. The more HITs completed and approved, the more you will earn.
8. It is not necessary to repeat the headline in your entry. It will NOT count toward your word count.
9. Do NOT describe the following: the format, where the video comes from, or how long the video is. This information is IRRELEVANT.
10. Do NOT describe the video in the following manner: "She turns around to face the camera. Then she faces left." Follow the examples below.

I don't... not do it?

Current Word Count: 0 Current Character Count: 0 / 2000

am I supposed to fill this in?

Criteria for REJECTION:

1. Entries with obvious and multiple spelling or grammatical errors will be rejected.
2. Entries with fewer than 100 words will be automatically rejected.
3. Text copied from the web or other places will be rejected. Multiple plagiarized answers will lead to being BLOCKED. You may use a quotation, but the majority of your content must be ORIGINAL.
4. Incomplete and blank answers will be rejected. Multiple blank answers will result in being blocked.
5. Tasks submitted without descriptions will be rejected.
6. Tasks submitted with inaccurate descriptions will be rejected as well.
7. Do NOT add any personal opinions. Entries with personal opinions or reviews will be automatically REJECTED.
8. If you notify us that a link is broken, we appreciate it but will not be able to accept the submission. The notification will result in rejection.
9. Entries that transcribe the video will be REJECTED.

Am I even going to get paid???

Do I have to answer all of this???

What am I even doing?

there are no pretty pictures :(

For Example...

Designing Tasks



Task

Please evaluate the relevance of the following document for the query **milton keynes**.

The screenshot shows a Bing search results page. The search bar contains 'milton keynes'. The results include a 'MILTON KEYNES' section with links to 'Milton Keynes Map' and 'Milton Keynes Restaurants'. Below this is the main search results area with 'ALL RESULTS' and '1-20 of 7,020,000 results - Advanced'. The top result is a link to 'Milton Keynes - Wikipedia, the free encyclopedia'. A snippet from the Wikipedia page describes Milton Keynes as a large town in Buckinghamshire, south east of England, about 45 miles (72 km) north-west of London. It also mentions it is the capital of... Below the snippet are links to 'History', 'Urban design', 'Culture', and 'Education'. To the right of the results are 'Sponsored sites' for 'Milton Keynes Hotels' with a link to 'www.expedia.com'.

Please rate the above document according to its relevance to **milton keynes** as follows. Note that the task is about how relevant to the topic the document is.

- Relevant.** A relevant document for the topic.
- Not relevant.** The document is not good because it doesn't contain any relevant information.

<http://www.gianlucadmartini.net/crowdsourcing/udine/3-setup.pdf>

- I know what to do
- I know what you're asking
- I know how I'm supposed to respond
- Sweet, I can do this.

Deploying Strategically

Designing Tasks



- Pre-Test Your Survey!
 - 10-15 as a test to see:
 - how long it takes
 - all the links work
 - it's not crazy confusing
 - Change and iterate as needed

Deploying Strategically

Designing Tasks



- Deploy in small batches
 - 100-200 *max*
 - Algorithm pushes old stuff down
- Between 11am-5pm in your desired timezone
 - Weekdays, or weekends!
- In general, prepare to Iterate

Assessing Quality

Designing Tasks



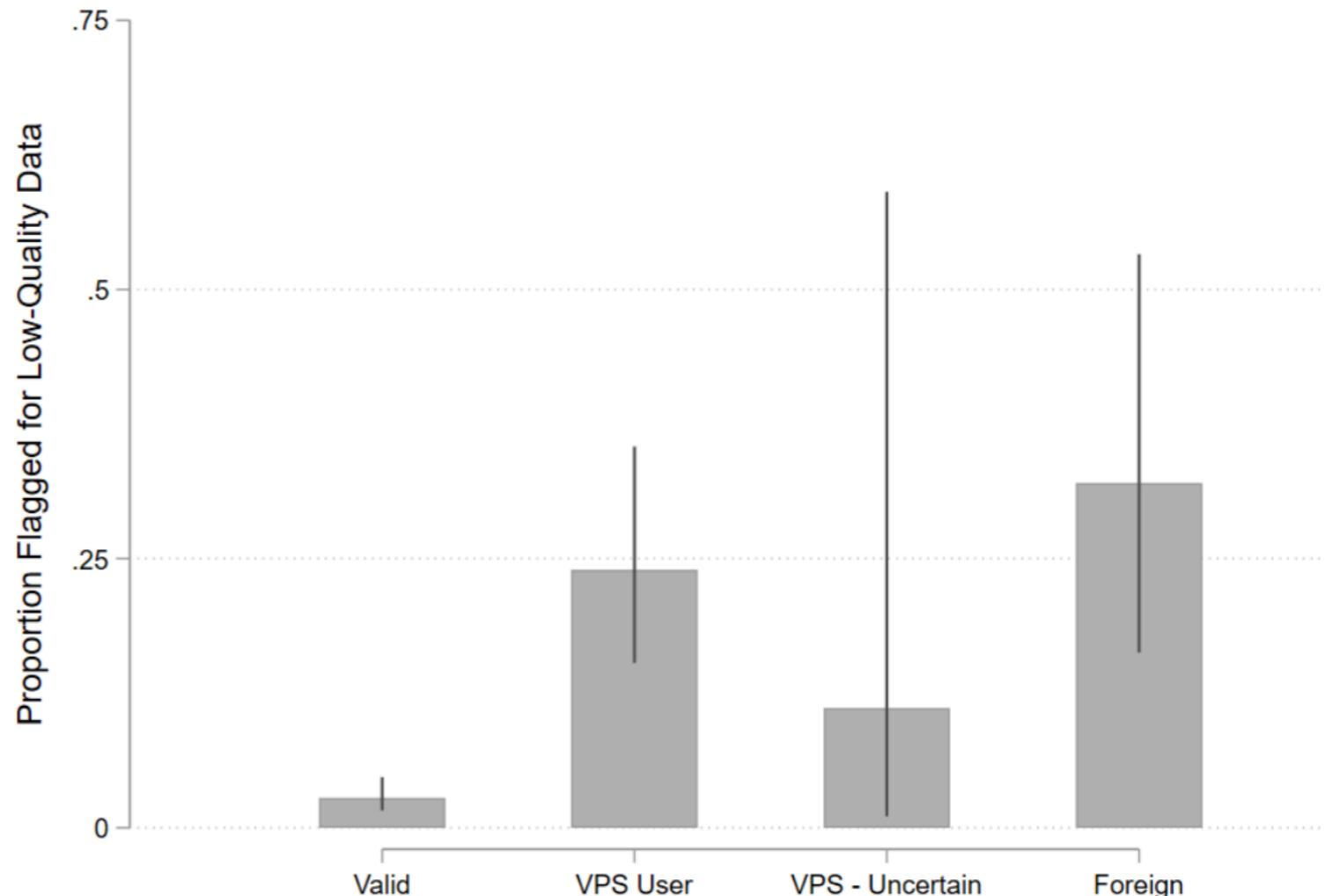
- Agreement
 - Inter-participant agreement percentages
 - Cohen's kappa (2 raters)
 - Leiss' kappa (any number of raters)
 - Majority votes

Assessing Quality

Designing Tasks



- Detect abnormal click patterns
- Trace IP addresses
 - Duplicates
 - Spoofs
- Actually read responses*
 - *and publish your inclusion criteria!



Kennedy et al. 2018

Possible Tools

Designing Tasks



Tool	Pros	Cons
Amazon's Mechanical Turk (mTurk, AMT)	<ul style="list-style-type: none">• Huge worker base• Many fancy features• Lots of worker metadata• Very straightforward survey integration	<ul style="list-style-type: none">• Pay only in Amazon gift cards (internationally)• Easy to accidentally be evil, or even illegal
Prolific	<ul style="list-style-type: none">• Research-focused• Automatically GDPR Compliant	<ul style="list-style-type: none">• Small worker base• New company... fewer features & more bugs
CrowdFlower	<ul style="list-style-type: none">• Wide global participant pool• Wide variety of pre-built tasks	<ul style="list-style-type: none">• Kind of confusing interface• Questionable business practices???
Many Many others	Design focus, specialty work focus, ML data generation focus, specificity out the wazoo	Often small, specialized work-bases.

Other Great Resources

Designing Tasks



- How To Use Crowdsourcing Effectively: Guidelines and Examples (Elena Simperl)
 - <https://www.liberquarterly.eu/articles/10.18352/lq.9948/>
- Guide To Running MTurk Experiments (Martin Hebart)
 - www.gianlucademartini.net/crowdsourcing/udine/3-setup.pdf
- How To Setup a Crowdsourcing Task (Gianluca Demartini)
 - www.gianlucademartini.net/crowdsourcing/udine/3-setup.pdf



Demo



References

- Cooke, M., Barker, J., Lecumberri, M. L. G., & Wasilewski, K. (2011). Crowdsourcing for word recognition in noise. In *Twelfth Annual Conference of the International Speech Communication Association*.
- Feng, D., Carstensdottir, E., Carnicke, S. M., El-Nasr, M. S., & Marsella, S. (2016, November). An active analysis and crowd sourced approach to social training. In *International Conference on Interactive Digital Storytelling* (pp. 156-167). Springer, Cham.
- Irshad, H., Montaser-Kouhsari, L., Waltz, G., Bucur, O., Nowak, J. A., Dong, F., ... & Beck, A. H. (2014). Crowdsourcing image annotation for nucleus detection and segmentation in computational pathology: evaluating experts, automated methods, and the crowd. In *Pacific symposium on biocomputing Co-chairs* (pp. 294-305).
- Kennedy, R., Clifford, S., Burleigh, T., Waggoner, P., & Jewell, R. (2018). The shape of and solutions to the MTurk quality crisis. *Available at SSRN*.
- Krishna, R., Zhu, Y., Groth, O., Johnson, J., Hata, K., Kravitz, J., ... & Bernstein, M. S. (2017). Visual genome: Connecting language and vision using crowdsourced dense image annotations. *International Journal of Computer Vision*, 123(1), 32-73.Chicago
- Melhart, D., Liapis, A., & Yannakakis, G. N. (2019). PAGAN: Video Affect Annotation Made Easy. *arXiv preprint arXiv:1907.01008*.
- Schmidt, F. A. (2013, September). The good, the bad and the ugly: Why crowdsourcing needs ethics. In *2013 International Conference on Cloud and Green Computing* (pp. 531-535). IEEE.
- Sheehan, K. B. (2018). Crowdsourcing research: data collection with Amazon's Mechanical Turk. *Communication Monographs*, 85(1), 140-156.
- Smith, S. M., Roster, C. A., Golden, L. L., & Albaum, G. S. (2016). A multi-group analysis of online survey respondent data quality: Comparing a regular USA consumer panel to MTurk samples. *Journal of Business Research*, 69(8), 3139-3148.
- Yin, M., Chen, Y., & Sun, Y. A. (2014, September). Monetary interventions in crowdsourcing task switching. In *Second AAAI Conference on Human Computation and Crowdsourcing*.



Questions



carolyn.saund@gmail.com



@CarolynBot



csaund