Crowdcomputing and Citizen Science for Large-scale Experiments

Snehalkumar `Neil’ S. Gaikwad, Sohan Dsouza, Oana Vuculescu, Andrew Mao, Iyad Rahwan

Massachusetts Institute of Technology  Microsoft Research New York  Aarhus University

The International Conference on Computational Social Science, 2017, Germany

July 10, 2017
A BRIEF HISTORY OF CROWD COMPUTING

CROWD MECHANICS

INSIGHTS
I

A Brief History of Crowd Computing
In the 19th-20th century
when computers were human
The Mathematical Tables Project

450 human computers

[Grier 2007]
Hidden Figures

Human Computers and IBM 704
source: JPL

Human Computers 1936
source: JPL

Ranger 7, the first successful U.S. mission to the moon
source: JPL
In the 21st century
Rise of the Connected World
Power of the Crowd

crowd as a computational process
Count Craters
Clickworkers & Be a Martian

source: JPL
reCAPTCHA

[Luis von Ahn et al. Science, 2008]
reCAPTCHA

[Luis von Ahn et al. Science, 2008]
Machine Vision
Machine Vision
Games with Purpose

[Von Ahn et al. CHI, 2004]
Games with Purpose

[Von Ahn et al. CHI, 2004]
Games with Purpose

Crowdcomputing and Citizen Science for Large-scale Experiments

[ Von Ahn et al. CHI, 2004 ]
Games with Purpose

[Games with Purpose][Von Ahn et al. CHI, 2004]
Games with Purpose

[Games with Purpose: Von Ahn et al. CHI, 2004]

Crowdcomputing and Citizen Science for Large-scale Experiments
Micro Tasks

low cognitive overload
Metacognition

Creativity

Social Learning

Motivation

Intuitions

Incentives

Collaboration

Empathy
Macro Tasks

high cognitive overload
Thinking at computational and cognitive scale
What can we accomplish if we harness crowd and machine intelligence?
The Internet as our laboratory and a crowd as collaborator
Research and Solve

the biggest unanswered questions
Exploring the life at the cellular level
Understanding RNAs

Why RNA?

Dark matter of Biology

Key to understand life at the cellular level

Manipulate RNAs to help cure Parkinson’s disease
Discover the sequence

Hairpin RNA
Target Shape

RNAs are string of Nucleotide Base \( \in \{ A, C, U, G \} \)

Nucleotide Sequence

<table>
<thead>
<tr>
<th>G</th>
<th>G</th>
<th>A</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>G</td>
<td>U</td>
<td>U</td>
</tr>
<tr>
<td>G</td>
<td>A</td>
<td>U</td>
<td>A</td>
</tr>
<tr>
<td>C</td>
<td>G</td>
<td>A</td>
<td>U</td>
</tr>
<tr>
<td>C</td>
<td>G</td>
<td>C</td>
<td>U</td>
</tr>
<tr>
<td>U</td>
<td>G</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>U</td>
<td>. . .</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The power of games

Every year millions of human hours are spent solving puzzles in games
Games for democratizing science
RNA Nano Engineering
Over 100,000 users from more than 90 Countries

[Lee et al. PNAS, 2014]
Pushing the Boundaries of Quantum Physics
Quantum Control Optima

Search for high dimensional complex challenges
Science@Home

[j.j.w.h. sørensen et al. nature, 2014]
Players

Highest education level graduated

<table>
<thead>
<tr>
<th>Number of years (if any) of post-secondary education in physics</th>
<th>%</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 (I have only studied physics up to/in high-school)</td>
<td>47.54%</td>
<td>29</td>
</tr>
<tr>
<td>one to three</td>
<td>24.59%</td>
<td>15</td>
</tr>
<tr>
<td>three to five</td>
<td>16.39%</td>
<td>10</td>
</tr>
<tr>
<td>more than 5</td>
<td>11.48%</td>
<td>7</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>61</td>
</tr>
</tbody>
</table>

United States
France
Spain
Germany
Romania
Austria
United Kingdom
Russian Federation
Switzerland
Italy
Iceland
Belgium
Netherlands
Sweden
Australia
Finland
Denmark
Democratizing research to increased diversity in the scientific workforce
Scientific research remains the **domain of the privileged few**—
to those blessed with the **socioeconomic** opportunities
The man who knew ∞

Srinivasa Ramanujan
[1887-1920]
Stanford Crowd Research

[Vaish et al. ACM UIST, 2017]
Crowdcomputing and Citizen Science for Large-scale Experiments

Over 1500 Crowd Researchers
Crowdcomputing and Citizen Science for Large-scale Experiments

Your results get better automatically.
Workers who you like will get first access to your tasks. As you give feedback, your favorite workers will do more and more of your work.
Researching AI ethics and algorithmic biases at scale
Crowdcomputing and Citizen Science for Large-scale Experiments

[Bonnefon et al. Science, 2016]
**Moral Machine**

- Human subject data needed on vast number of combinations of factors
- Subscription crowdsourced experiment services would be prohibitive in cost and/or not as flexible
Moral Machine

- Minimalist and direct
- Abstract intuitive interface
- Cross-platform
- Expressing creativity
- Holding mirror to subject
- Sharing results and scenarios
- Discussion and feedback

What should the self-driving car do?

[www.moralmachine.mit.edu/]

Crowdcomputing and Citizen Science for Large-scale Experiments
CASE

Let’s choose a problem
II
Crowd Mechanics
Crowd Mechanics

DYNAMICS & LOGISTICS

EXPERIMENT DESIGN & DATA SCIENCE

PLATFORM & INFRASTRUCTURE DESIGN

Interactive Case Discussions about the problem you want to solve
Dynamics & Logistics
When to use the crowd?

- Reasonably well-defined problem
- Size, diversity, impact, and time
- Can give the crowd something of value
- Scale and reach over selectivity and acquaintance
Making the Crowd

Recruitment and Retention
Recruitment

• Social Network
  • Reputation
    » [Stanford Crowd Research]

• Volunteer Science / Lab in the wild: Sharing Test-results on Facebook
  » [Science at Home, EteRNA]

• “Let’s Play” and commentary videos by YouTube stars
  » [Moral Machine]
Recruitment

• Press
  • A popular journal (Nature) + the associated press attention
    » [Science at Home]
  • A popular newspaper coverage (e.g. New York Times)
    » [EteRNA, Moral Machine]

• Challenge
  • Build a quantum computer
  • Help invent Medicine
  • Do ethical thing
  • Do research and become a scientist
Crowdcomputing and Citizen Science for Large-scale Experiments

Retention & Incentive Engineering

Motivations: Intrinsic, Extrinsic

- Reputation
- Competition
- Gamification/Fun
- Virtual Market
- Vanity
- Purpose
Bonuses: Incentives and motivation

Periodic bonuses

lead contributor: rewarding bonuses

admin: rewarding bonuses

friends who speak similar language
Organizing the Crowd

Community, Task Distribution, Recognition
Community

• Task Distributions
  • Various paradigm (e.g., Parallel vs Sequential)
  • Milestones and feedback

• Engagement Channels
  • Asynchronous Communication: Forums, Emails, Wikipedia
  • Synchronous Communication: Slack, Hangouts, Chats

• Protocols
  • Inclusive and constructive environment
  • Empathy — Volunteers’ (workers’) and requesters’ relationship
  • Remember that the crowd is comprised of people, not “noisy error-prone computers”

People reading and writing on Slack
Crowdcomputing and Citizen Science for Large-scale Experiments
Crowdcomputing and Citizen Science for Large-scale Experiments

Process
Crowdcomputing and Citizen Science for Large-scale Experiments

eterna
Make Molecules Advance Science

Process

DScore x/100

Crowdcomputing and Citizen Science for Large-scale Experiments
Crowdcomputing and Citizen Science for Large-scale Experiments

Process

DScore x/100
Recognition

- Admins cannot have a full picture of who is doing what
- Most of the voting rules are confusing and can be gamed
- Voluntary projects—
  *how to spot HUBS in a large network and empower them*
Nodes in Credit Distribution

\{ \text{nodes} \}
Links in Credit Distribution

\[ \{ s_1, s_2, s_3, \ldots, s_n \} \]

Scores \( s \)

\[ \{ 10, 20, 00, 40, 00, \ldots, 30 \} \]
Network of Credit Distribution

\{ 10, 20, 00, 40, 00, \ldots, 30 \}^n

\sum_{i=1}^{n} s_i = 100

Crowdcomputing and Citizen Science for Large-scale Experiments
Author ranks via page rank
CASE
The Problem and Crowd Dynamics

Recruitment Size, Diversity, Goal/Task

Recognition, Retention, Motivations

Community & Organization
Experiments and Data Science
Data collection

• How to account for Noise
  • Crowd participants are heterogeneous
    » Long tail of contributions and commitment
    » Cultural biases and language barriers
    » Diverse range of expertise

• Data
  • One optimal or “best” solution [Quantum Moves]
  • Hypothesis to discover molecules [EteRNA]
  • Aggregation of data from many individuals [Moral Machine]
  • Scientific interactions [Crowd Research]
Data collection

• Barrier to Entry
  • Registration/forms
  • Extra tutorials, milestones
  • Unfamiliar tools or technologies

• Quality Control
  • Organic—crowd filters out bad designs or submissions
  • Capture—Logs, IP, and device IDs (time to action, cursor movement, intermediate activity) to detect and either block or later filter out unthinking click-throughs, spam, or sabotage
Moral Machine - data expansion

• Internationalization (10 languages)

• Addition of post-quiz user survey

• Expansion of experiment scope with classic scenarios
Moral Machine - data collection

- 3.2M Users
- 30M Decisions
- 350k Full Surveys
“For my first attack, […]
moving all the pieces into a single pile.

[…] However, it seems all you did was lock the pieces together and ban my IP address.”
Sabotage activity logs

“Which led me to my second attack, using a VPN and a neighbors wireless for some new IP addresses [...] select all the pieces and place them on top of each other, but this got old soon.”
Sabotage activity logs

“Which led me to my second attack, using a VPN and a neighbors wireless for some new IP addresses

[...]
select all the pieces and place them on top of each other, but this got old soon.”

“So I decided to get a bit more sneaky

[...]
I selected a number of pieces, enough to make solving the puzzle difficult and not so much that people would immediately notice,
[...]
moved them off the top of the virtual table”
Building vs policing

Both attackers and responders move fast

Users slowed down after attack

dot = 1 user, Red = attacker
Analysis

• Plan out data analysis beforehand

• Think about how you will collect your data. What data do you need
  • “Let’s collect ALL the data (because we don’t know what we will need later)” is a bad strategy

• How will you deal with attrition and dropout, partially observed data?
  • Is attrition correlated with experimental treatments?
Analysis

• What analyses will you run if you had your data?
  • Do a dry run before going and collecting all the data

• The best way to find out what people are thinking/doing: ask them!
  • Qualitative complements Quantitative
Activity Scheduling [time, day]

Crowd Research
Community Formation

equipment & programming

UX research

friends

friends who speak similar languages

social science
Characteristics of leaders
CASE

The Problem and Data Science

Barriers to Entry & Task Logistics

Data Collection

Measurements, Hypotheses, Analysis
Platform and Infrastructure Design
No Infrastructure
Game Mechanics & Engineering

Iterative Development—Prototype, pilot quickly, and fail fast
Web Engineering

[www.moralmachine.mit.edu/]
Moral Machine - Scaling Up

• Demo run and press-only release
• Initial launch at small scale on cloud server
• Optimization and caching for growing audience
• Shift to scalable pay-by-usage services
CASE

The Problem and Platform Design

Design Thinking & Human Factors

Logistics
(e.g., Open Sourced/ Closed Sourced)

System Engineering
III

Insights
At the end of crowd projects we typically show a nice bundled-up presentation, which looks like everything went smoothly
It didn’t!
Murphy’s law

think of as much as you can beforehand,
but don’t be surprised if something goes wrong
Here are some of our stories showing mistakes and recovery
Launch of Long-run Prisoner's Dilemma Experiment
Long-run PD experiment

- Incentives: financial, via MTurk
- Retention: combination of social and financial incentives
- Would people come back each day?
  - We didn’t know
Our experiment

Round 1

Round 10

Anonymous partners

Cooperate

Defect
Our experiment

Game

Round 1

Round 10

anonymous partners

Cooperate

Defect
Random rematching across games
Random rematching across games
Random rematching across games
Random rematching across games
One experiment session – 20 games
One experiment session – 20 games

50 pairs

Game 1

Game 20
Crowdcomputing and Citizen Science for Large-scale Experiments

Aug 4, 2015 - Day 1

Aug 31, 2015 - Day 20
Logistics

- 20 days
- 20 games per day
- 10 rounds per game
- 113 players to start, 94 completed after attrition
- 375,000 decisions: ~20 times longer than previous experiments
Launch of long-run PD experiment

- **Social incentives**
  - Expectation of month-long experiment
  - Daily reminders of commitment

- **Financial incentives**
  - Players kicked out for missing more than two days
  - Daily cash payments proportional to payoffs
  - Lump sum bonus for completion
Comparison of Dropouts to Main Population
Crowdcomputing and Citizen Science for Large-scale Experiments

Past Prisoner's Dilemma Experiments

Number of Participants

- Never: 48
- 1-4: 35
- 5-9: 15
- 10-49: 10
- 50-99: 5
- 100+: 1
Strategic Behavior in the Stanford Crowd Research
Crowd Dynamics
Strategic behavior
Crowdcomputing and Citizen Science for Large-scale Experiments

Friends

Fairness via page rank
Alice Remote Control
Crowdcomputing and Citizen Science for Large-scale Experiments
Challenges

- **Unprecedented experimental setup leads to unprecedented problems**
  - Social science investigations were dependent on having a team online simultaneously
  - Recruitment, team allocation, waiting room?
  - Data backlog
  - Gamification matters
Discussion
Crowd = Richer Research at Scale

The Internet as a lab

Size, Scale
- More Samples
- Large Social Interactions

Complexity, Realism
- Realistic vs Abstract
- More precise instrumentation

Duration, Participation
- Longer period of time
- Fewer resource constraints
Crowd is synergistic
Make use of the long tail

Scientists, Experts

Contribution Level

% of Participants

Lab experiments

Crowd

Crowdcomputing and Citizen Science for Large-scale Experiments
Tradeoffs in increasing size

![Graph showing tradeoffs between sample size and measurement variance between lab and crowd settings.](image-url)
Global Ecosystem

- **Crowdcomputing and Citizen Science for Large-scale Experiments**
- **Education, Partnership**
- **Fun, Engagement**
- **Scientist**
- **Crowd**
- **Data, Insights**
- **Research Ideas, Feedback**
Non-traditional impact
Democratize research
Stanford Crowd Research
Science at the global scale

EteRNA, Science Home—Quantum Moves
Rethink about AI and society

Moral Machine
Crowdcomputing and Citizen Science for Large-scale Experiments

Snehalkumar `Neil’ S. Gaikwad, Sohan Dsouza, Oana Vuculescu, Andrew Mao, Iyad Rahwan

Massachusetts Institute of Technology * Microsoft Research New York † Aarhus University ‡

The International Conference on Computational Social Science 2017, Germany

July 10, 2017
Daemo is live!

send an email to gaikwad@mit.edu or daemo@stanford.edu
Apply for the Adventurous Science Fellowship ($150,000)

https://www.media.mit.edu/scifellows/
References & Recommended Readings

» [Antin et al. CHI, 2012] Social desirability bias & self-reports of motivation: a study of amazon mechanical turk in the US & India
» [Bernstein et al. UIST, 2010] Soylent: A Word Processor with a Crowd Inside
» [Bernstein et al. UIST, 2011] Crowds in two seconds: Enabling realtime crowd-powered interfaces
» [Bigham et al. UIST, 2010] VizWiz: nearly real-time answers to visual questions
» [Bigham et al. ASSETS, 2011] The design of human-powered access technology
» [Boudreau et al. HBR, 2013] Using the crowd as an innovation partner
» [Boudreau et al. Management science, 2011] Incentives and problem uncertainty in innovation contests: An empirical analysis
» [Cheng et al. CHI, 2015] Measuring crowdsourcing effort with error-time curves
» [Cooper et al. Nature, 2010] Predicting protein structures with a multiplayer online game
» [Dow et al. CSCW, 2012] Shepherding the crowd yields better work
» [Horton et al. ACM EC, 2010] The labor economics of paid crowdsourcing
» [Horton et al. Experimental Economics, 2010] The online laboratory: Conducting experiments in a real labor market
» [Gaikwad et al. UIST, 2016] Boomerang: Rebounding the Consequences of Reputation Feedback on Crowdsourcing Platforms
References & Recommended Readings

» [Ghosh et al. ACM EC, 2011] Who moderates the moderators?: crowdsourcing abuse detection in user-generated content
» [Gomes et al. NIPS, 2011] Crowdclustering
» [Gupta et al. CHI, 2012] mclerk: enabling mobile crowdsourcing in developing regions
» [Ho et al. ICML, 2013] Adaptive task assignment for crowdsourced classification
» [Ho et al. WWW, 2015] Incentivizing high quality crowdwork
» [Ipeirotis et al. ACM SIGKDD, 2010] Quality management on amazon mechanical turk
» [Ipeirotis et al. XRDS, 2010] Analyzing the amazon mechanical turk marketplace
» [Kamar et al. AAMS, 2012] Incentives for truthful reporting in crowdsourcing
» [Kim et al. CSCW, 2017] Mechanical novel: Crowdsourcing complex work through reflection and revision
» [Kittur et al. CHI, 2008] Crowdsourcing user studies with mechanical turk
» [Kittur et al. CSCW, 2008] Harnessing the wisdom of crowds in wikipedia: quality through coordination
» [Kittur et al. UIST, 2011] Crowdforge: Crowdsourcing complex work
» [Kittur et al. CSCW, 2013] The future of crowd work
» [Kraut et al. MIT Press, 2011] Building successful online communities: Evidence-based social design
References & Recommended Readings

» [Krogh et al. Research Policy, 2003] Community, joining, and specialization in open source software innovation: a case study
» [Kulkarni et al. TOCHI, 2013] Peer and self assessment in massive online classes
» [McInnis et al. CHI, 2016] Taking a hit: Designing around rejection, mistrust, risk, & workers’ experiences in amazon mechanical turk
» [Nebeling et al. UIST, 2015] WearWrite: Orchestrating the Crowd to Complete Complex Tasks from Wearables (…)
» [Lasecki et al. UIST, 2012] Real-time captioning by groups of non-experts
» [Lasecki et al. CSCW, 2015] Information extraction and manipulation threats in crowd-powered systems
» [Lee et al. PNAS, 2014] RNA design rules from a massive open laboratory
» [Luther et al. ACM CSCW, 2015] Structuring, aggregating, and evaluating crowdsourced design critique
» [Malone et al. ACM CSCW, 1990] What is coordination theory and how can it help design cooperative work systems?
References & Recommended Readings

» [Mao et al. AAAI HCOMP, 2013] Volunteering versus work for pay: Incentives and tradeoffs in crowdsourcing
» [Mason et al. Behavior research methods, 2012] Conducting behavioral research on amazon’s mechanical turk
» [Martin et al. CSCW, 2014] Being a Turker
» [Parameswaran et al. VLDB, 2011] Human-assisted graph search: it’s okay to ask questions
» [Rutherford et al. PNAS, 2012] Limits of social mobilization
» [Rzeszotarski et al. UIST, 2011] Instrumenting the crowd: using implicit behavioral measures to predict task performance
References & Recommended Readings

» [Salehi et al. CHI, 2015] We are dynamo: Overcoming stalling and friction in collective action for crowd workers
» [Shah et al. HRI, 2011] Improved human-robot team performance using chaski, a human-inspired plan execution system
» [Shaw et al. CSCW, 2011] Designing incentives for inexpert human raters
» [Singer et al. WWW, 2013] Pricing mechanisms for crowdsourcing markets
» [Stephane et al. Upwork, 2015] Hire fast & build things: How to recruit and manage a top-notch team of distributed engineers
» [Vaish et al. CHI, 2014] Twitch crowdsourcing: crowd contributions in short bursts of time
» [Vaish et al. UIST, 2017] Crowd Research: Open and Scalable University Laboratories
» [Valentine et al. CHI, 2017] Flash Organizations: Crowdsourcing Complex Work by Structuring Crowds As Organizations
» [Von Ahn et al. CHI, 2004] Labeling images with a computer game