Workshop on Citizen Science and Crowdsourcing

Amrapali Zaveri, Deniz Iren

Maastricht University
Institute of Data Science

Open Universiteit
www.ou.nl

May 3, 2019

Maastricht, Netherlands

Funded by

Universiteitsfonds Limburg

| SWOL |

Introductions

- Instructors
- Participants

Agenda

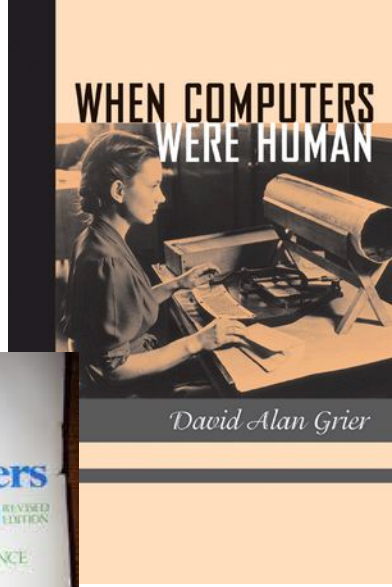
9:15 - 10:00	Introduction to Human Computation
10:00 - 11:00	Fundamentals of microtask crowdsourcing
11:00 - 11:15	Coffee & Tea break
11:15 - 12:30	Hands-on I: designing a microtask on Figure Eight
12:30 - 1:30	Lunch
1:30 - 2:30	Hands-on II: MIA: Crowdsourcing medical image annotation task
2:30 - 3:30	Hands-on III: EMO Annotation
3:30 - 4:00	Coffee & Tea break
4:00 - 4:45	Applications of crowdsourcing, summary and conclusion
4:45 - 5:00	Q&A, Wrap-up and feedback

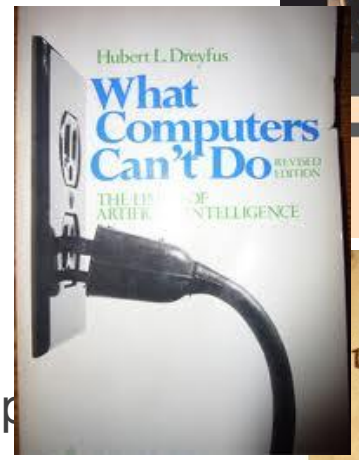
Introduction to Human Computation

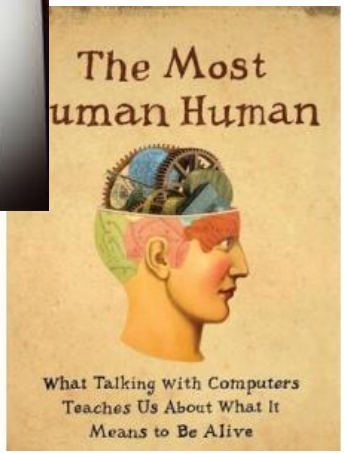
Human Computation

- **Outsourcing tasks that machines find difficult to solve to humans**
  - Difficult not the same as impossible
  - Accuracy, efficiency, cost

- **Historically humans were the first computers**
  - 17th century: Halley‘s commet
  - 19th century: computing factories
  - 20th century: professionalization of human comp







In search for extra terrestial

- In **1996** Berkeley needed powerful computers to download and analyze radio telescope data
- Handful of astronomers and computer scientists proposed a novel revolution: **recruit the public** to donate computer time to the task;
- Volunteers download a screen saver which would kick into gear when the user stopped using his or her machine; when finished automatically send it back
- By 2004, 5.2 million users: 3 million years of computer time; Guiness book of world records - **largest computation in history**

Many can work together to outperform a few!

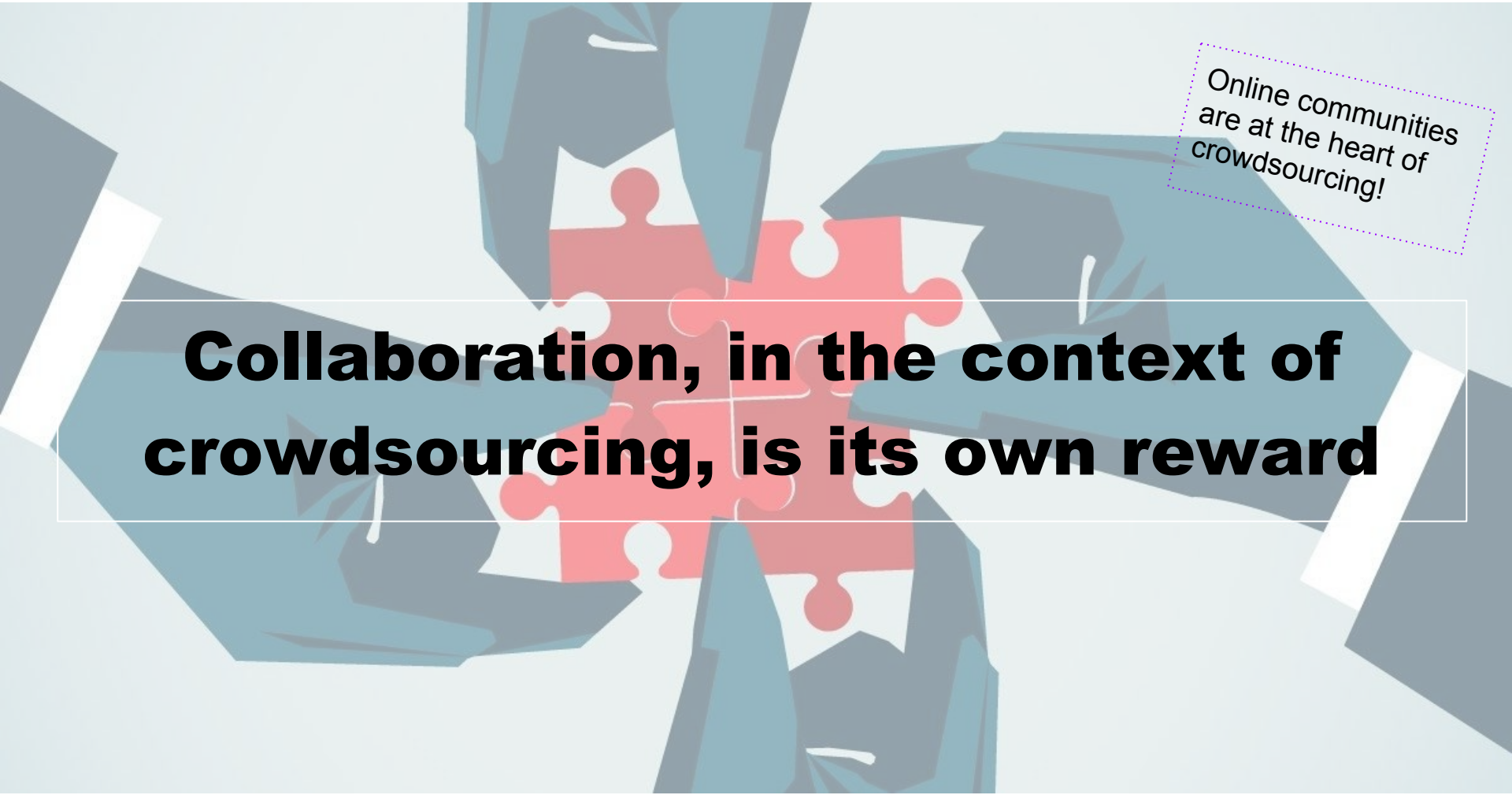
- 3rd largest purveyor of stock images
- 96% of its workforce is comprised of people whose bread is primarily buttered through some other vocation

- Classic case of demand and supply
- Generated 17$ mil in revenue
- Rewards community with store credit every time someone submits a photo, or a friend buys

Unprecedented levels of collaboration

- CS emerged organically- out of uncoordinated actions of thousands of people, who were **doing things that people like to do**
  - Especially in the **companionship** of other people
- **Internet** provided a way for them to pursue their interests

CS capitalizes on the deeply *social nature* of human behavior & on the fact that our interests are more diverse than our business cards



Online communities are at the heart of crowdsourcing!

Collaboration, in the context of crowdsourcing, is its own reward

Incentives

- desire to create something from which the larger community would benefit
- sheer joy of practicing a craft at which they excel
- pleasure in cultivating talent and from passing on what they’ve learnt


CROWDSOURCING is
OUTSOURCING on
STEROIDS

The Daily CROWDSOURCE
#1 site for crowdsourcing news

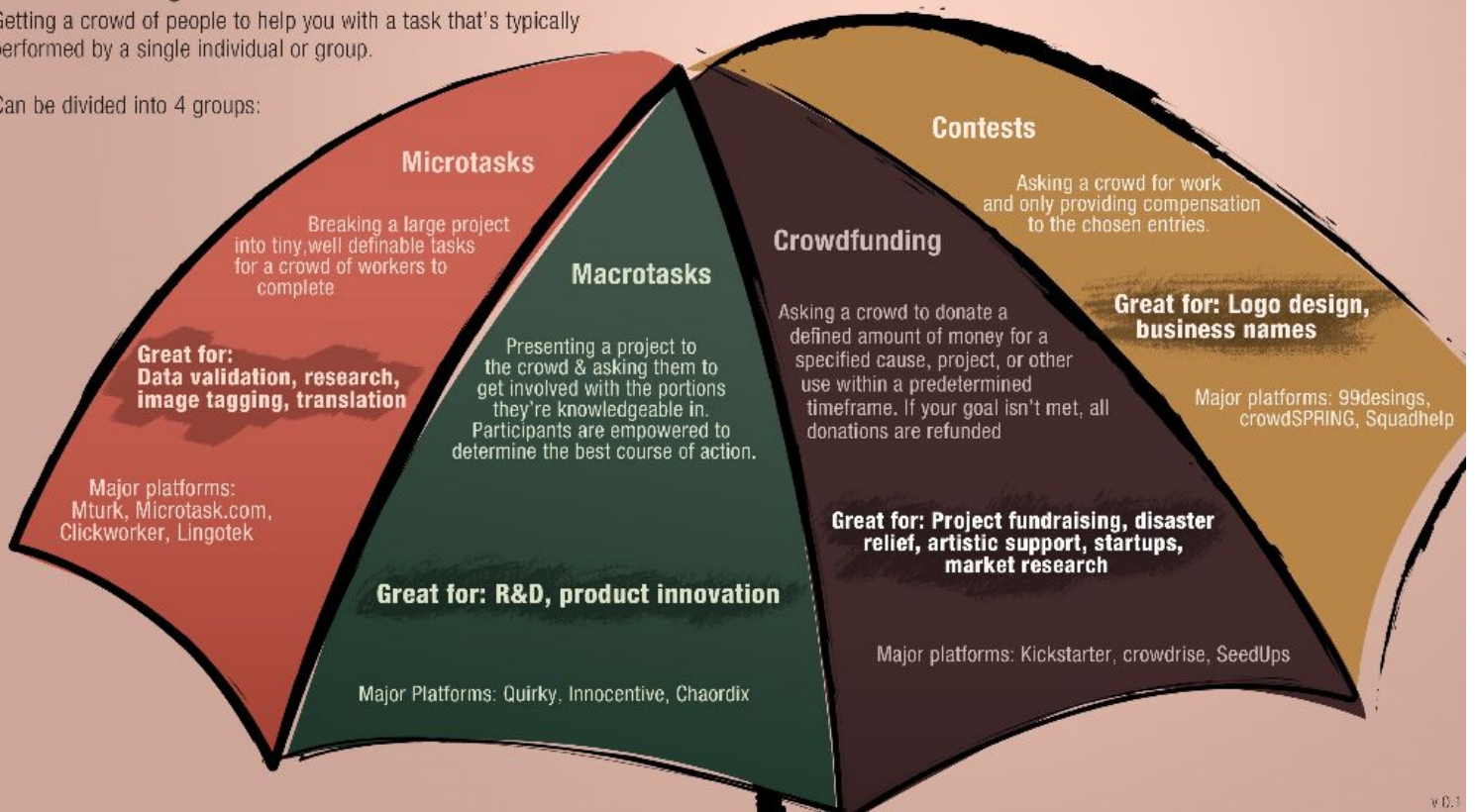
Crowdsourcing's proposed taxonomy. Learn more & voice your ideas at **http://dailycrowdsource.com/taxonomy**

Umbrella of Crowdsourcing

Crowdsourcing

Getting a crowd of people to help you with a task that's typically performed by a single individual or group.

Can be divided into 4 groups:



v C.1

Crowdsourcing
Crowdsourcing Examples

Which projects have YOU come across?

WIKIPEDIA

The Free Encyclopedia

Citizen Science

WHAT IS OUTSOURCED

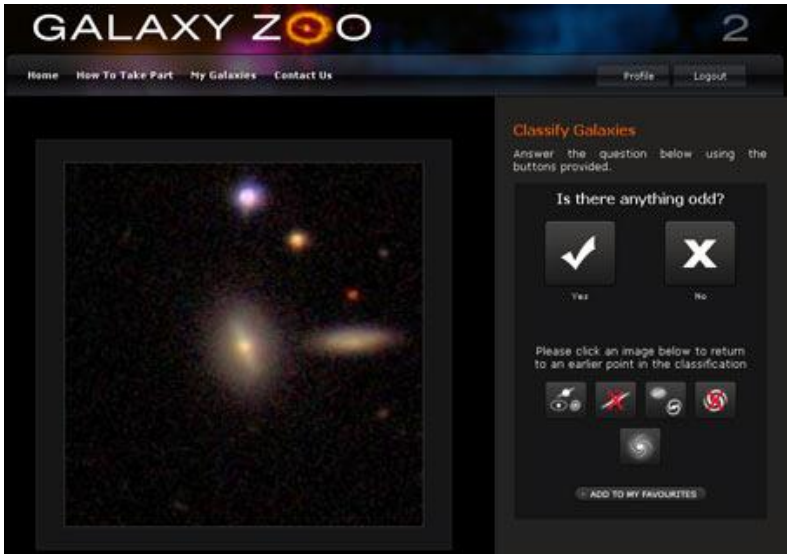
- Object recognition, labeling, categorization in media content

WHO IS THE CROWD

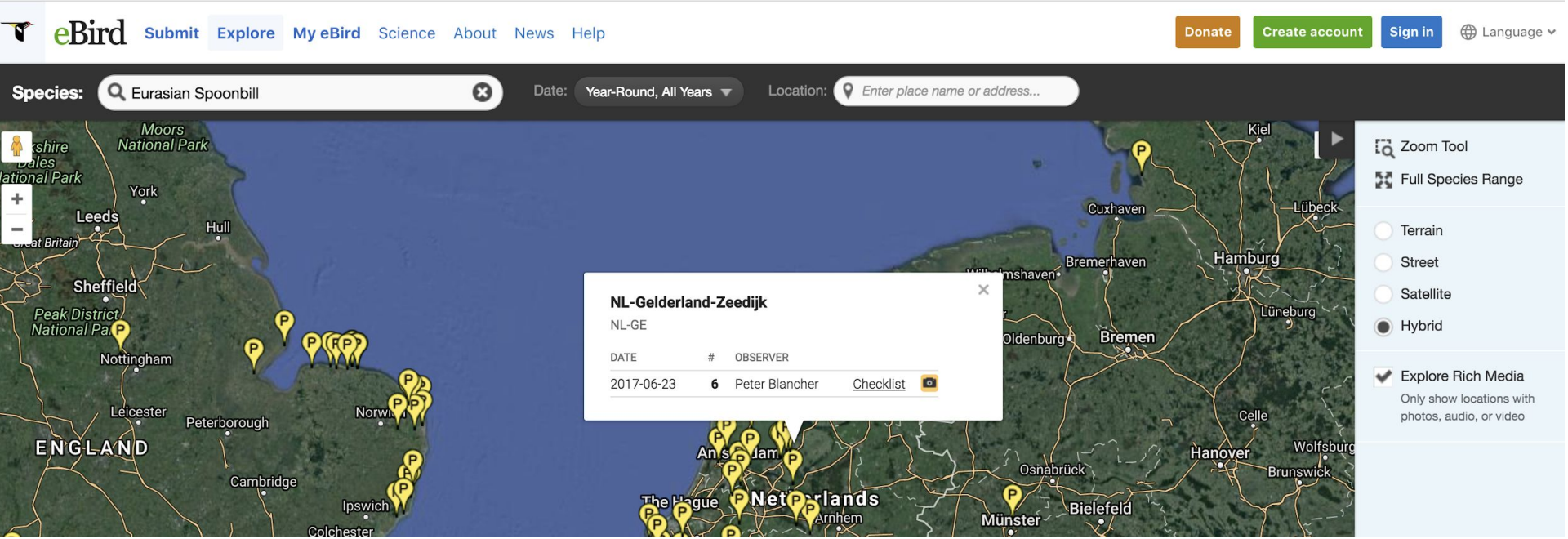
- Anyone

HOW IS THE TASK OUTSOURCED

- Highly parallelizable tasks
- Every item is handled by multiple annotators
- Every annotator provides an answer
- Consolidated answers solve scientific problems



https://ebird.org



- 1,000 participated in the 1st bird count in 1996
- In 2007 - 80,000 people participated
- Unprecedented look at the distribution and migratory patterns of a multitude of species
- Far fewer professionals

"There's simply no other way to have this type of coverage"

It's not about money!
It's about credit &
acknowledgement!

If you can read **you can help.**

I want to help

https://mark2cure.org/

https://www.innocentive.com/

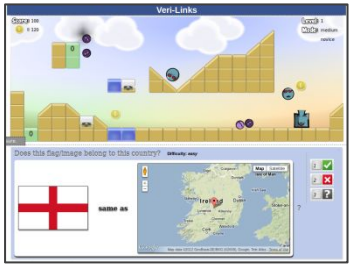
GAMES WITH A PURPOSE (GWAP)

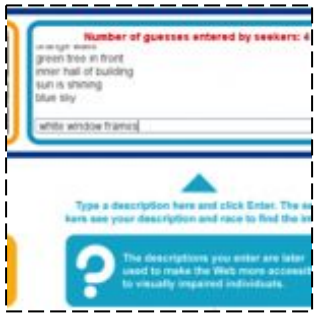
Human computation disguised as casual games

Tasks are divided into parallelizable atomic units (challenges) solved (consensually) by players

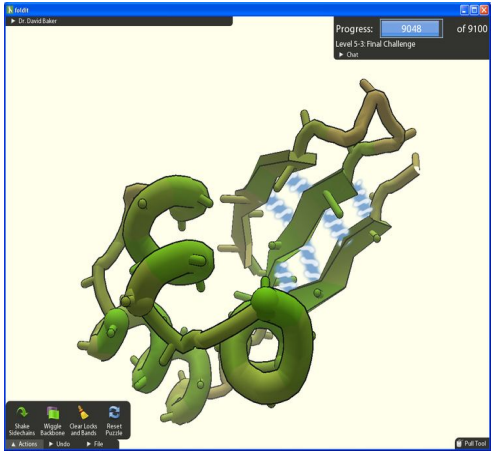
Game models

- Single vs. multi-player
- Selection agreement vs. input agreement vs. inversion-problem games





EXPLICIT VS. IMPLICIT CONTRIBUTION - AFFECTS MOTIVATION AND ENGAGEMENT



Users aware of how their input contributes to the achievement of application's goal (and identify themselves with it)

vs.

Tasks are hidden behind the application narratives. Engagement ensured through other incentives.

Microtask Crowdsourcing

MICROTASKS - WHAT & WHY?

TIME 

- Highly **parallelizable** tasks
- Work is broken down into smaller — ‘micro’ — pieces that can be solved independently

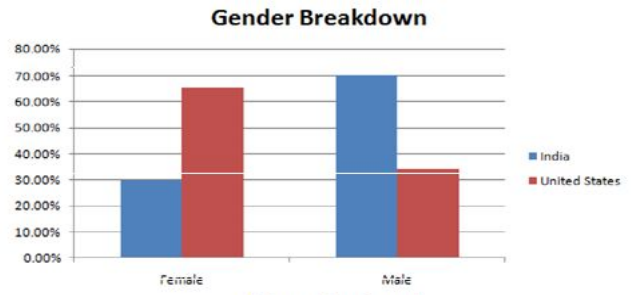
MONEY $ $ $

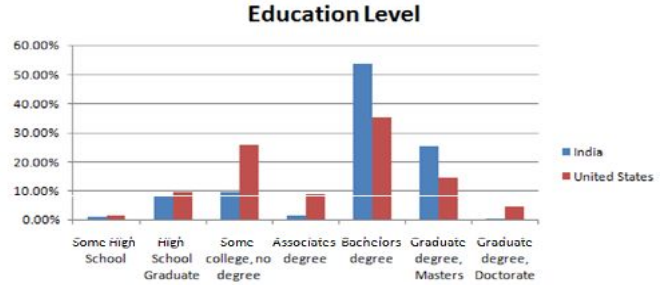
- Tasks based on **human skills** not easily replicable by machines
- Non-expert workers can perform the tasks with a minimal payment

Consolidated answers solve scientific problems !!

LARGE, BUT NOT ALWAYS DIVERSE CROWD

amazon mechanical turk
Artificial Artificial Intelligence





Country of residence

- United States: 46.80%
- India: 34.00%
- Miscellaneous: 19.20%

CrowdFlower

figure eight

Business Data

Collect data on businesses at massive scale

Content Moderation and Curation

Quickly find both good and bad user generated content

Ranked

Boost conversions with better search results

Content Generation

Improve your search engine ranking with quality content

Custom solutions

We help businesses of all sizes automate really big custom projects

Customer and Lead Data Enhancement

Increase sales by knowing more about your customers

Sentiment and Opinion Analysis

Know exactly what people are saying about you

Categorize

Categorize products, businesses, videos, events, & more

Surveys

Find and interact with highly-qualified digital consumers

Builder

Advanced user? Developer? Build your own crowdsourcing projects

COMPLEX WORKFLOWS CANNOT ALWAYS BE DIRECTLY IMPLEMENTED

WHAT IS OUTSOURCED

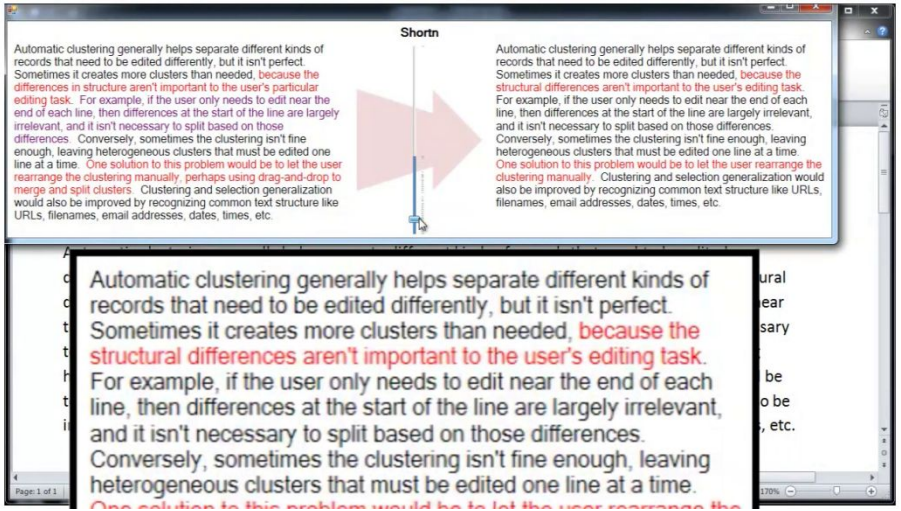
- Text shortening, proof-reading, open editing

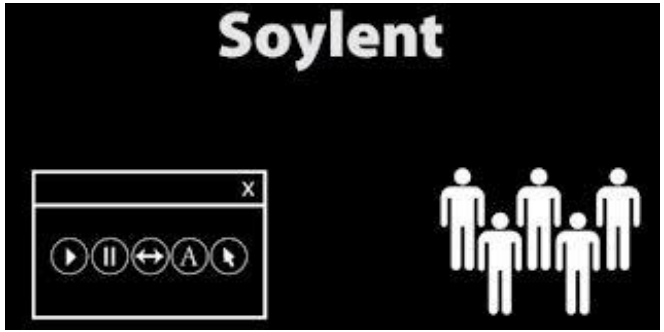
WHO IS THE CROWD

- MTurk

HOW IS THE TASK OUTSOURCED

- Text divided into paragraphs
- Select-fix-verify pattern
- Multiple workers in each step





http://www.youtube.com/watch?v=n_miZqsPwsc

MEASURING PERFORMANCE CAN BE CHALLENGING

WHO AND HOW

- Redundancy
- Excluding spam and obviously wrong answers
- Voting and ratings by the crowd
- Assessment by the requester
- Where does the ground truth come from and is it needed
    - Note: improving recall of algorithms

WHEN

- Real-time constraints in games
- Near-real-time microtasks,

HOW ARE THE RESULTS VALIDATED

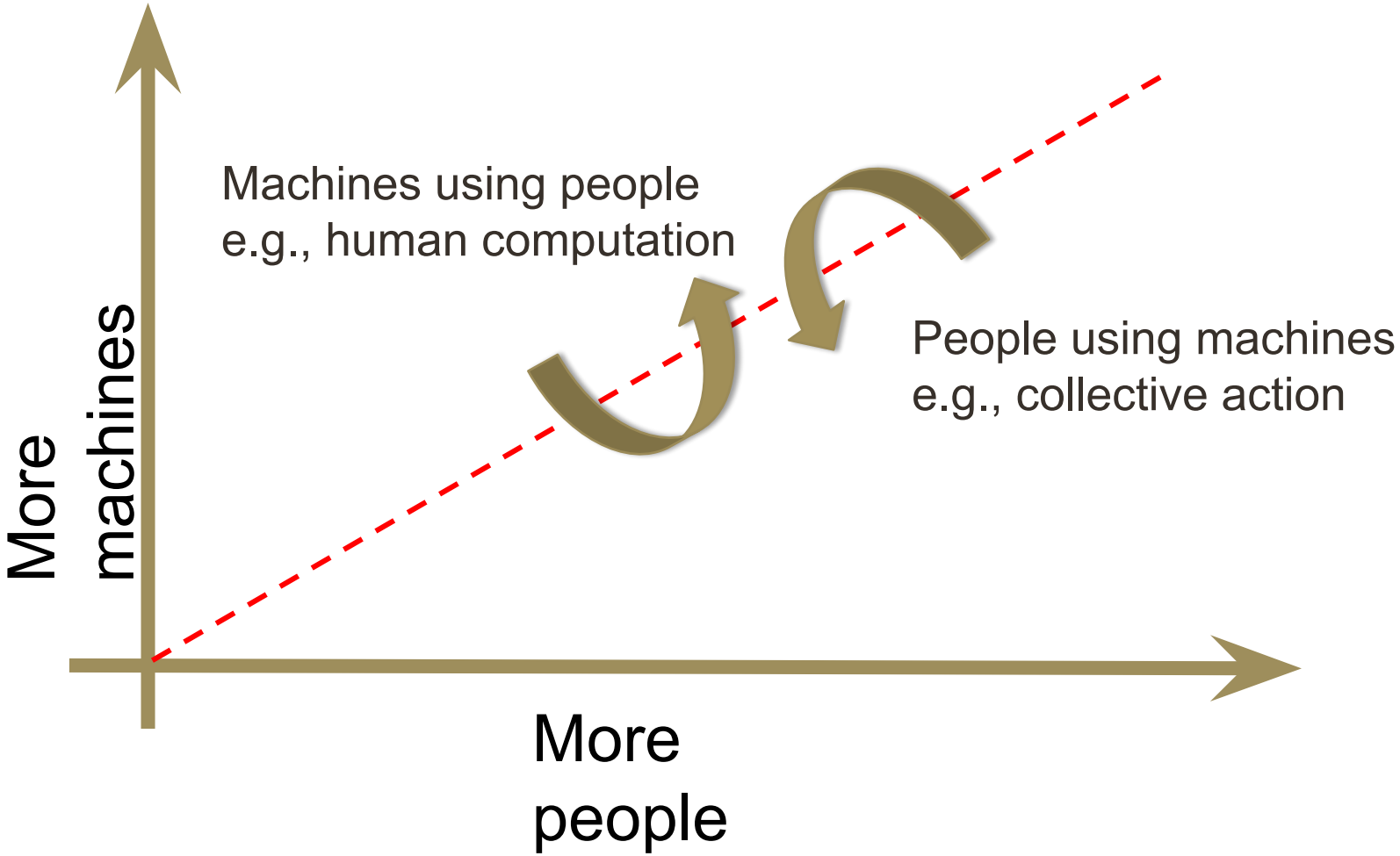
- Solutions space closed vs. open
- Performance measurements/ground truth
- Statistical techniques employed to predict accurate solutions
  - May take into account confidence values of algorithmically generated solutions

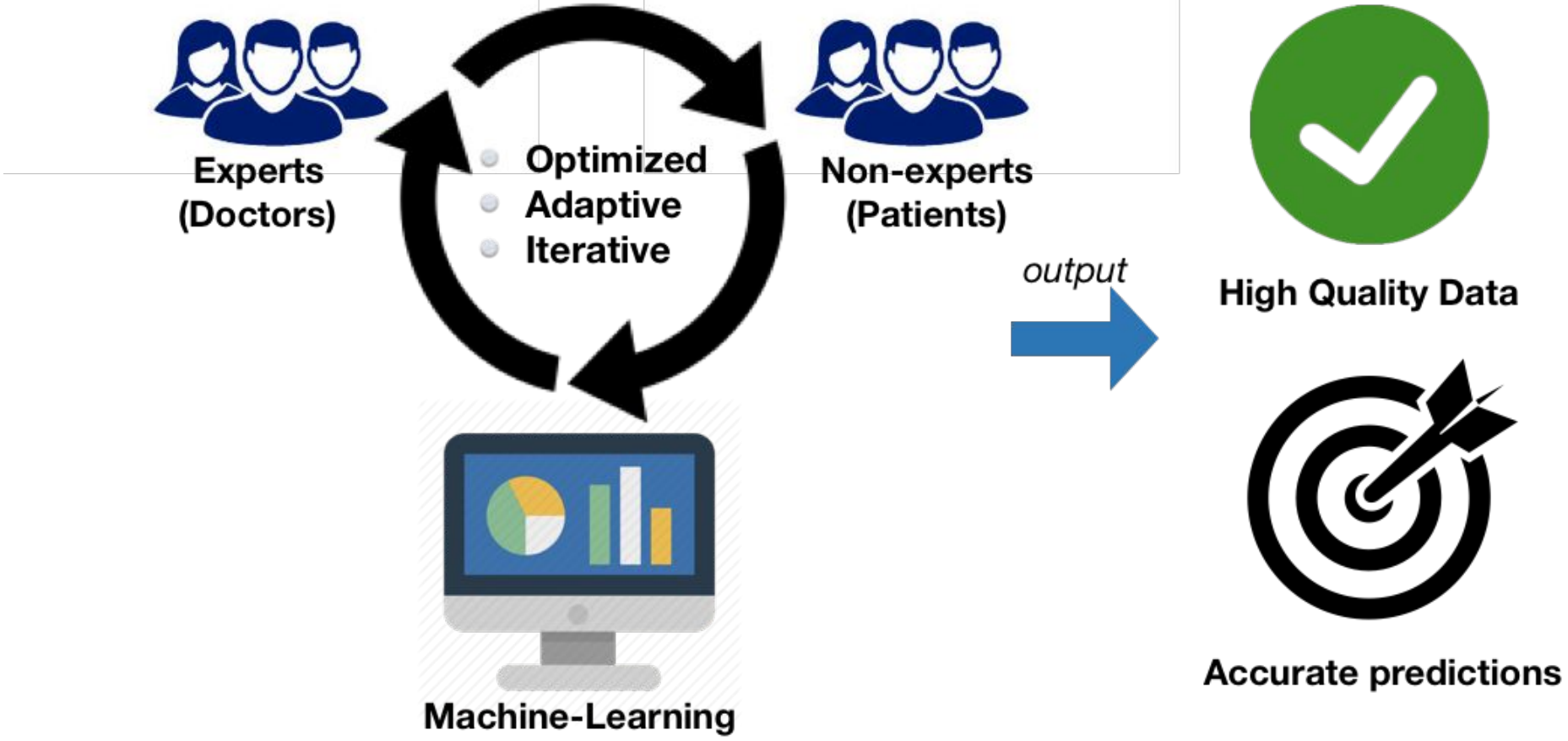
HOW CAN THE PROCESS BE OPTIMIZED

- Incentives and motivators
- Assigning tasks to people based on their skills and performance (as opposed to random assignments
- *Symbiotic combinations of human- and machine-driven computation, including combinations of different forms of crowdsourcing*

HYBRID WORKFLOWS

Combining Humans & Machines


Machines using people
e.g., human computation
People using machines
e.g., collective action
More machines
More people


Experts
(Doctors)
Optimized
Adaptive
Iterative
Non-experts
(Patients)
output
High Quality Data
Accurate predictions
Machine-Learning

Next Session:

Fundamentals of Microtask Crowdsourcing

Questions?

Acknowledgements

Slides adapted from the tutorial “Microtask Crowdsourcing to Solve Semantic Web Problems” by Gianluca Demartini, Elena Simperl, and Maribel Acosta at ISWC 2013.
Source: https://github.com/maribelacosta/crowdsourcing-tutorial