Human Computation and Crowdsourcing

Irwin King

Department of Computer Science and Engineering
The Chinese University of Hong Kong, Shatin, N.T., Hong Kong
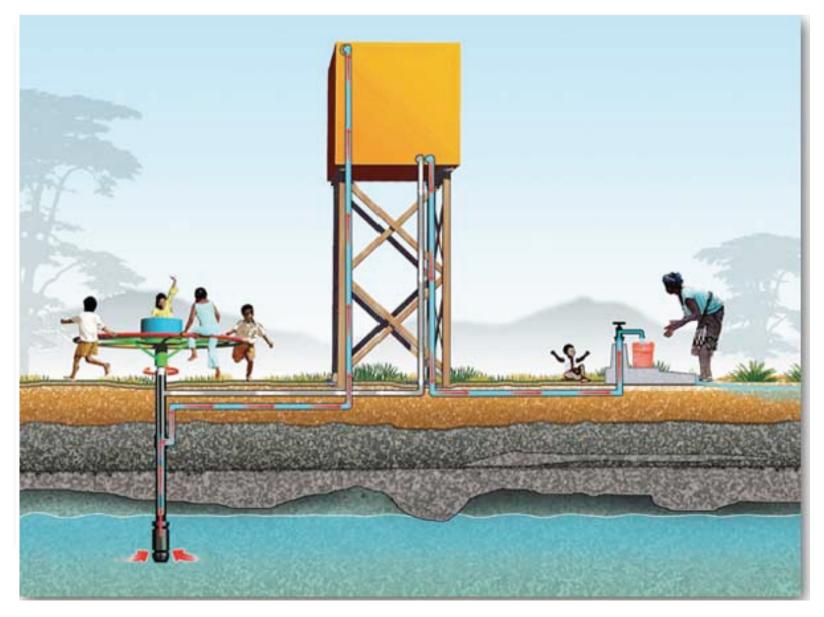
What do you see?

- Human computation is a technique that utilizes humans to carry out computations.

Playing/Having Fun ←❓→ Work/Computation
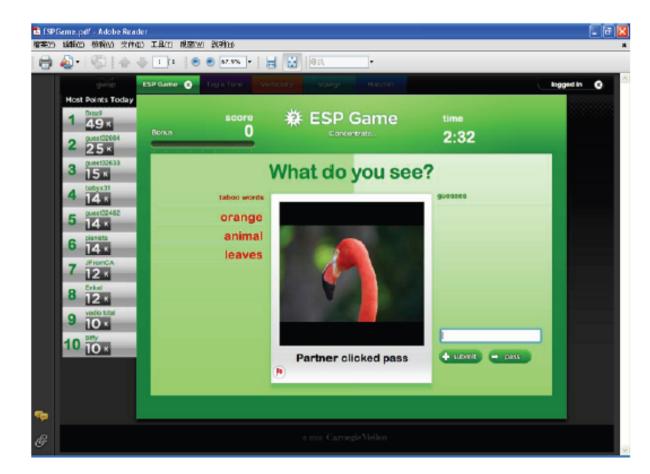
Playing/Having Fun ⬅❓➡ Work/Computation

How to motivate human workers?

- In 2008, an estimated 200 million people play online games every week.

- Games provide an way to channel human effort spent playing games to work on human computational tasks.

- Objectives of games:
  - Because of enjoyment, people play games.
  - As side effect, tasks can be done by players.

Example of Social Games (1)

- To collect text information from images

  - Examples (1): ESP game
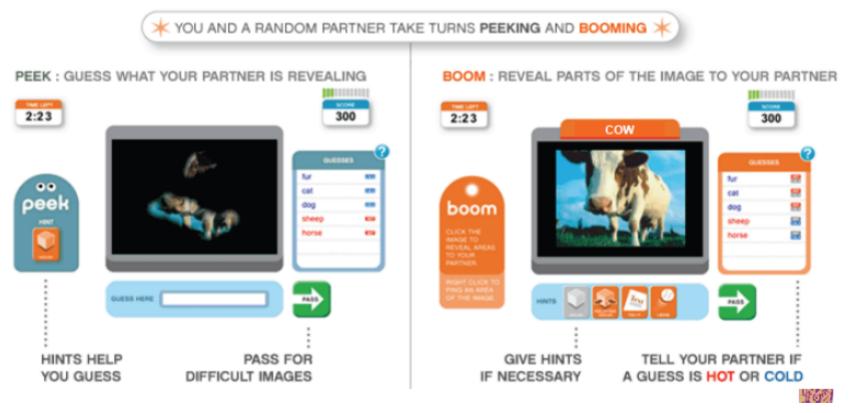
The ESP game for labeling images

- Licensed by Google as the Google Image Labeler in 2006

- As of July 2008, 200,000 players had contributed more than 50 million labels for images.
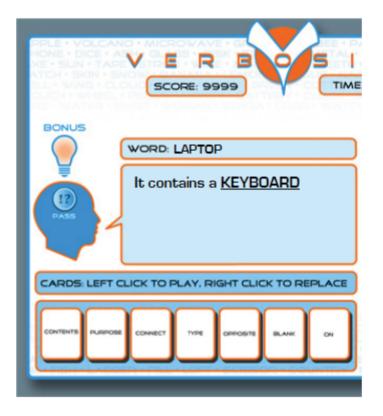
Example of Social Games (2)

- To collect text information for images:

  - Examples (2): Peekaboom

Example of Social Games (3)

- To collect commonsense knowledge:
  - Examples (3): Verbosity

Example of Social Games (4)

- To collect subjective descriptions of sounds and music:

  - Example (4): Tagatune

Example of Social Games (5)

- To learn colleagues' bookmarks in an organizational goal:

  - Example (5): Dogear Game

Example of Social Games (6)

- To tag locations in the real world through gameplay in mobile social games:

  - Example (6): Gopher guessing game

Figure 1. Real world experience, acquiring gophers

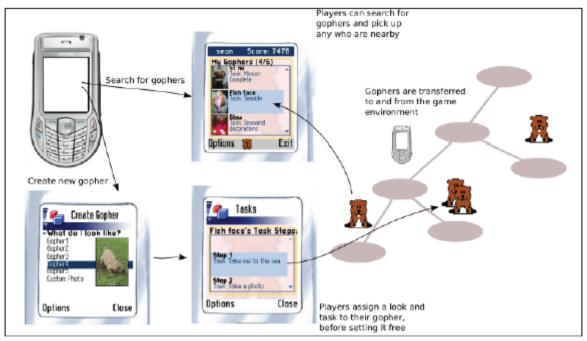Visual feedback can be provided in the form of camera phone images - players photograph their current location and supply this to the gopher. The gopher responds with an image from its history, taken at a spatially nearby location.

Gophers can participate in a word guessing game, based on their real-world location. Players supply semantic descriptions relative to their current whereabouts. They are awarded points depending on the accuracy of their guesses.

Players can provide text information by exchanging some gossip with the gopher - a player supplies textual information to the gopher. The gopher responds with some gossip from it's history, taken at a nearby location.
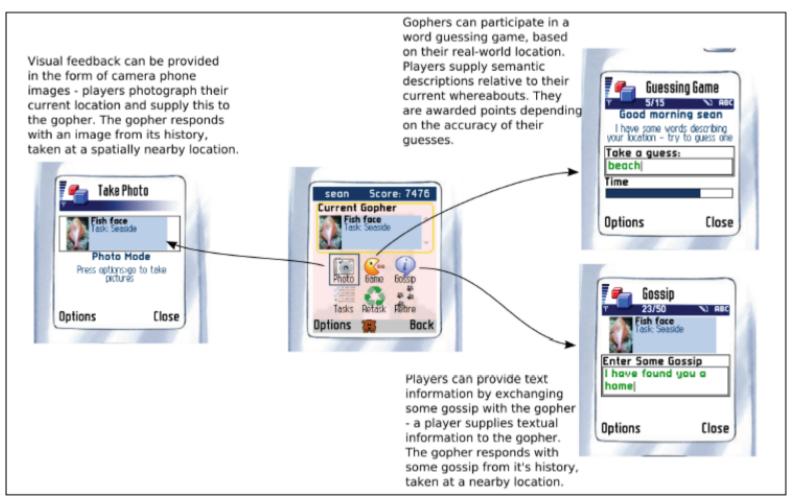
Figure 2. Real world experience, interacting with gophers

Definition of Human Computation

- Human computation is computations carried out by humans where computation is the mapping process of some input representations to some output representations.

- Human computation systems are systems that humans interact with the systems to carry out computations.

Is volunteer computing an example of human computation?

- Volunteer computing is the process that allows Internet users to donate their computer resources in a joint effort to solve problems, e.g., SETI.

- No.

- Volunteer computing does not require humans to perform computation themselves.

3 Categories of human computation systems (1)

- Annotation

  – Distinguish between humans and computer programs

  – Image annotation

  – Sound annotation

  – Text annotation

  – Video annotation

  – Web content annotation

3 Categories of
human computation systems (2)

- **Possess knowledge commonsense about the real world**

  – Collect knowledge commonsense about the real world
     knowledge

  – Collect geospatial information

3 Categories of human computation systems (3)

- Indicate human preferences
    - Rank human preferences for images
    - Rank Web search results
    - Identify human intentions behind Web search queries
    - Collect patterns of social interactions
    - Deploy human intelligences

4 Dimensions of
human computation systems (1)

- Motivation - systems have to find their ways to motivate people to participate, because systems require human participation.

Dimension	Value	Example
Motivation	Pay	Amazon Mechanical Turk
	Altruism	Help Find Jim
	Enjoyment	ESP Game [77]
	Reputation	Threadless
	Implicit work	reCAPTCHA [82]

4 Dimensions of human computation systems (2)

- Output quality control – systems have to collect valid and useful output, because people may make mistakes, misunderstand the instructions, or try to cheat.

Dimension	Value			Example
Output quality control	Verification by machine	Algorithmic check		Foldit [19, 20, 18, 42, 26, 41]
		Reputation check		Amazon Mechanical Turk
		Voting		Threadless
		Statistical filtering		Crowdsourceable QoE evaluation framework for multimedia content [14]
	Verification by human	Redundancy		Amazon Mechanical Turk
		Multilevel review		Soylent [8]
		Expert review		Amazon Mechanical Turk
	Verification by human computation system	Defensive game design	Collaborative	ESP Game [77]
			Competitive	Threadless
			Hybrid	Gopher Game [12]
		Ground truth seeding		reCAPTCHA [82]
	Financial incentive			Amazon Mechanical Turk
	None			ChaCha [15]

Wisdom

4 Dimensions of
human computation systems (3)

- Output aggregation - systems have to combine a set of outputs with good quality.

Dimension	Value	Example
Output aggregation	Collection	ESP Game [77]
	Statistical processing of data	Ask500People
	Iterative improvement	MonoTrans [37, 38]
	None	VizWiz [9]

4 Dimensions of
human computation systems (4)

- Worker-task cardinality - The cardinality shows how many workers contribute to how many tasks.

Dimension	Value	Example
Worker-task cardinality	One-to-one	ChaCha [15]
	Few-to-one	VizWiz [9]
	Many-to-one	Help Find Jim

Taxonomy of human computation systems

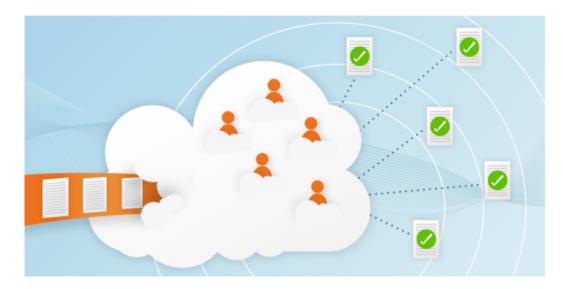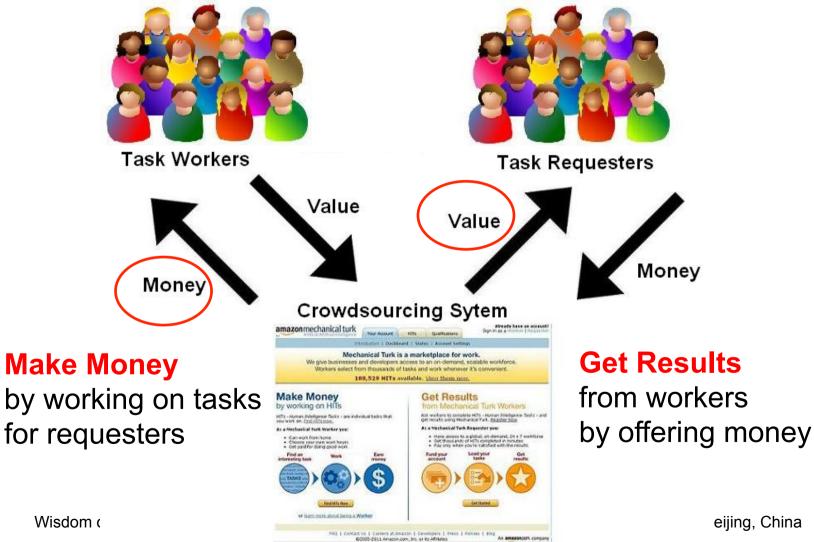| Dimension	Value			Example
Motivation	Pay			Amazon Mechanical Turk
	Altruism			Help Find Jim
	Enjoyment			ESP Game [77]
	Reputation			Threadless
	Implicit work			reCAPTCHA [82]
Output quality control	Verification by machine	Algorithmic check		Foldit [19, 20, 18, 42, 26, 41]
		Reputation check		Amazon Mechanical Turk
		Voting		Threadless
		Statistical filtering		Crowdsourceable QoE evaluation framework for multimedia content [14]
	Verification by human	Redundancy		Amazon Mechanical Turk
		Multilevel review		Soylent [8]
		Expert review		Amazon Mechanical Turk
	Verification by human computation system	Defensive game design	Collaborative	ESP Game [77]
			Competitive	Threadless
			Hybrid	Gopher Game [12]
		Ground truth seeding		reCAPTCHA [82]
	Financial incentive			Amazon Mechanical Turk
	None			ChaCha [15]
Output aggregation	Collection			ESP Game [77]
	Statistical processing of data			Ask500People
	Iterative improvement			MonoTrans [37, 38]
	None			VizWiz [9]
Worker-task cardinality	One-to-one			ChaCha [15]
	Few-to-one			VizWiz [9]
	Many-to-one			Help Find Jim

Crowdsourcing

- Crowdsourcing is an idea of outsourcing a task that is traditionally performed by an employee to a large group of people in the form of an open call.

- Objectives:
  - to reduce a company's production costs
  - to make more efficient use of labor and resources

How to seek for human workers?

- Crowdsourcing = Crowd + sourcing
- Crowdsourcing aims to outsource tasks to the crowd in the form of an open call.

Amazon Mechanical Turk (AMT)

Task Workers

Task Requesters

Value

Value

Money

Money

Crowdsourcing Sytem

Make Money
by working on tasks
for requesters

Get Results
from workers
by offering money

Wisdom ⟨   eijing, China   26

AMT – Statistics

Collection Period	01/2009 to 04/2010 (16 months)
Number of registered workers	over 400,000
Number of tasks	6,701,406
Number of requesters	9,436
Total value of the posted tasks	US$ 529,259

Data are publicly available - http://www.mturk-tracker.com

Categorization of crowdsourcing systems

- Voting Systems
  - Find out the answer that the majority selected

- Information Sharing Systems
  - Share user-generated meta-data

- Creative Systems
  - Cannot replace the role of human in creativity

- Games
  - Produce useful metadata as a by-product

Game With a Purpose (GWAP)

- To effectively collect information from players through a social game, we have to achieve two important factors:

  1. Maintain the enjoyment of players in the game
     - game design
  2. Guarantee the quality of collected information
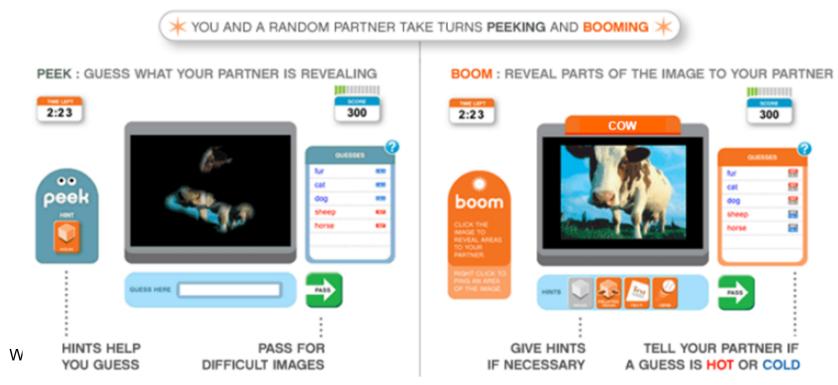   - game mechanism

Output-agreement games

- All players are given the same input.
- Assume no communication between the players.
- Winning condition: Same outputs
- E.g. ESP game for collecting image labels
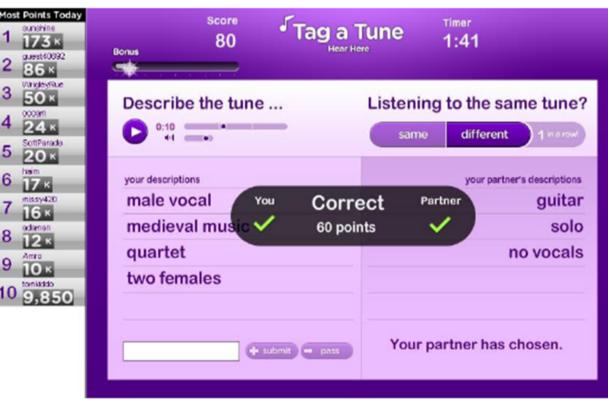
Inversion-problem games

- The first player has access to the whole problem and gives hints to the second player.

- Winning condition: the second player can guess the secret (we assume that the hints given by the first player are correct.)

- E.g. Peekaboom for locating objects in images

YOU AND A RANDOM PARTNER TAKE TURNS **PEEKING** AND **BOOMING**

PEEK : GUESS WHAT YOUR PARTNER IS REVEALING

BOOM : REVEAL PARTS OF THE IMAGE TO YOUR PARTNER

HINTS HELP YOU GUESS — PASS FOR DIFFICULT IMAGES

GIVE HINTS IF NECESSARY — TELL YOUR PARTNER IF A GUESS IS **HOT** OR **COLD**

Input-agreement games

- All players are given inputs separately.

- Each player is asked to describe his/her input.

- Assume no communication between the players.

- Winning condition: All players can guess whether their inputs is the same or different correctly.

- E.g. TagATune for labeling music

Wisdom of the

Output-optimization games

- All players are given the same input, and they can see each other's outputs.

- All players can communicate with each other using their outputs.

- Winning condition: All players agree on each other's output.
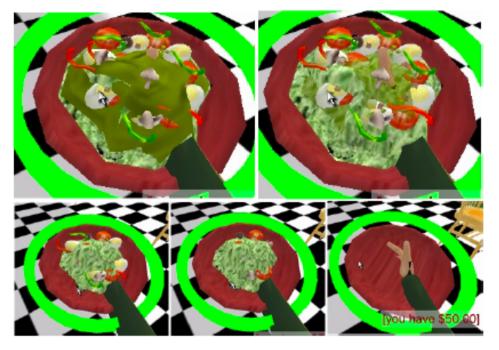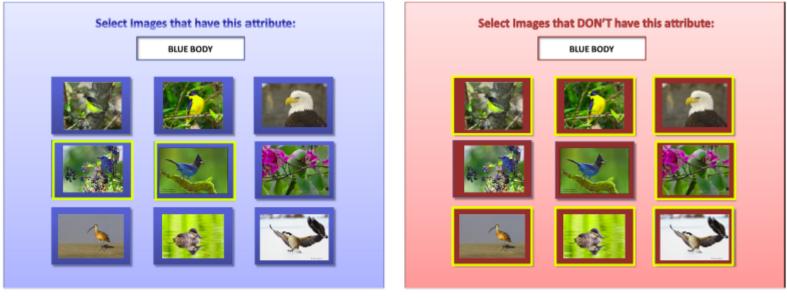
- E.g. Restaurant Game for designing salad

Wisdom of the

Figure 2. Clockwise from top left, bites of a salad disappear using an algorithmically designed munching order

Complementary-agreement games

- Players alternate between the "positive" player and "negative player".
- Given a set of choices, the "positive" player is asked to choose correct answers, while the "negative player" is asked to choose incorrect answers.
- Winning Condition: Players select right answers of right role
- E.g. Polarity for identifying correct answers

(a) "Positive Player"   (b) "Negative Player"

Summary

- 3 categories of human computation systems
- 4 dimensions of human computation systems
- Crowdsourcing systems
  - Output-agreement games
  - Inversion-problem games
  - Input-agreement games
  - Output-optimization games
  - Complementary-agreement games

Q&A

"No, crowdsourcing is not an option!"

"I want everybody to do your own homework. No crowdsourcing on this!"