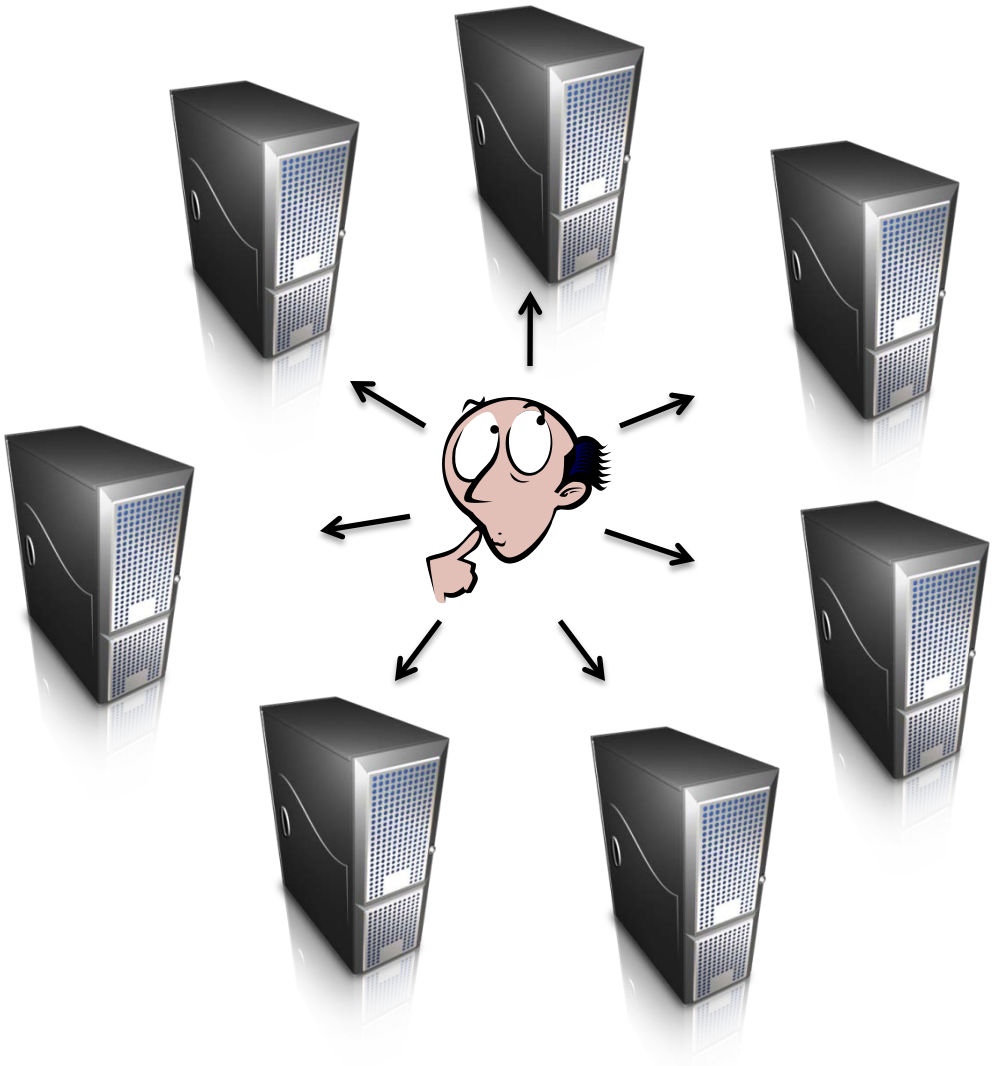
Collaborative Human Computing

Zack Zhu

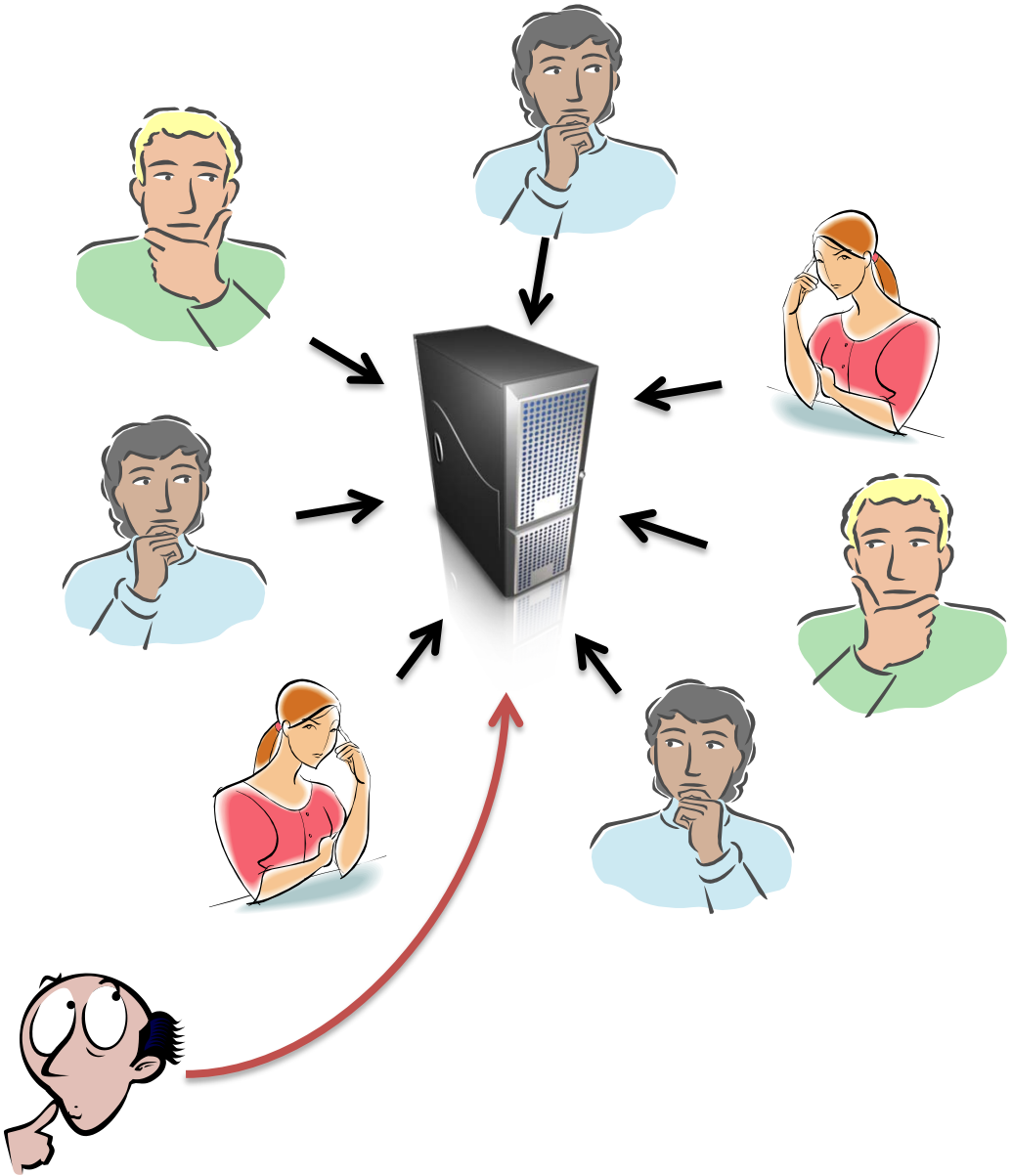
March 31, 2010

Seminar for Distributed Computing

Distributed Computing...



...redefined: Distributed ***Thinking***



“Crowdsourcing”

+

Human Resource = $$$$$!!

Internet

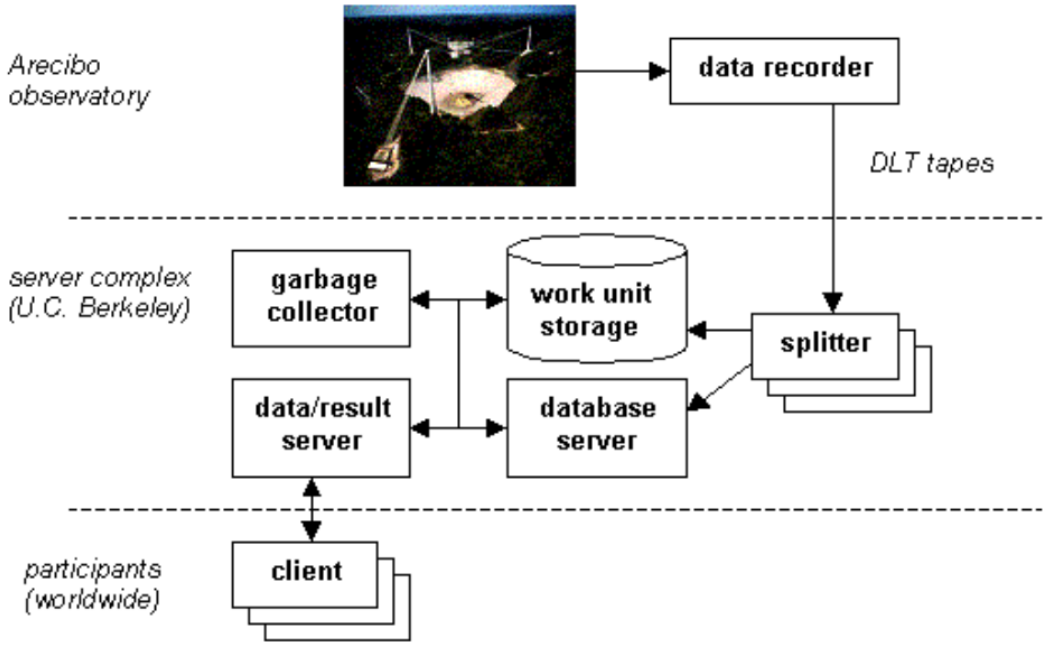
+

Web 2.0

Crowdsourcing

SETI HOME

- Search for Extraterrestrial Intelligence
- Earliest project utilizing the idea (launched in May 1999)
- Voluntary distributed *computing*



Distributed Thinking

+

Crowdsourcing

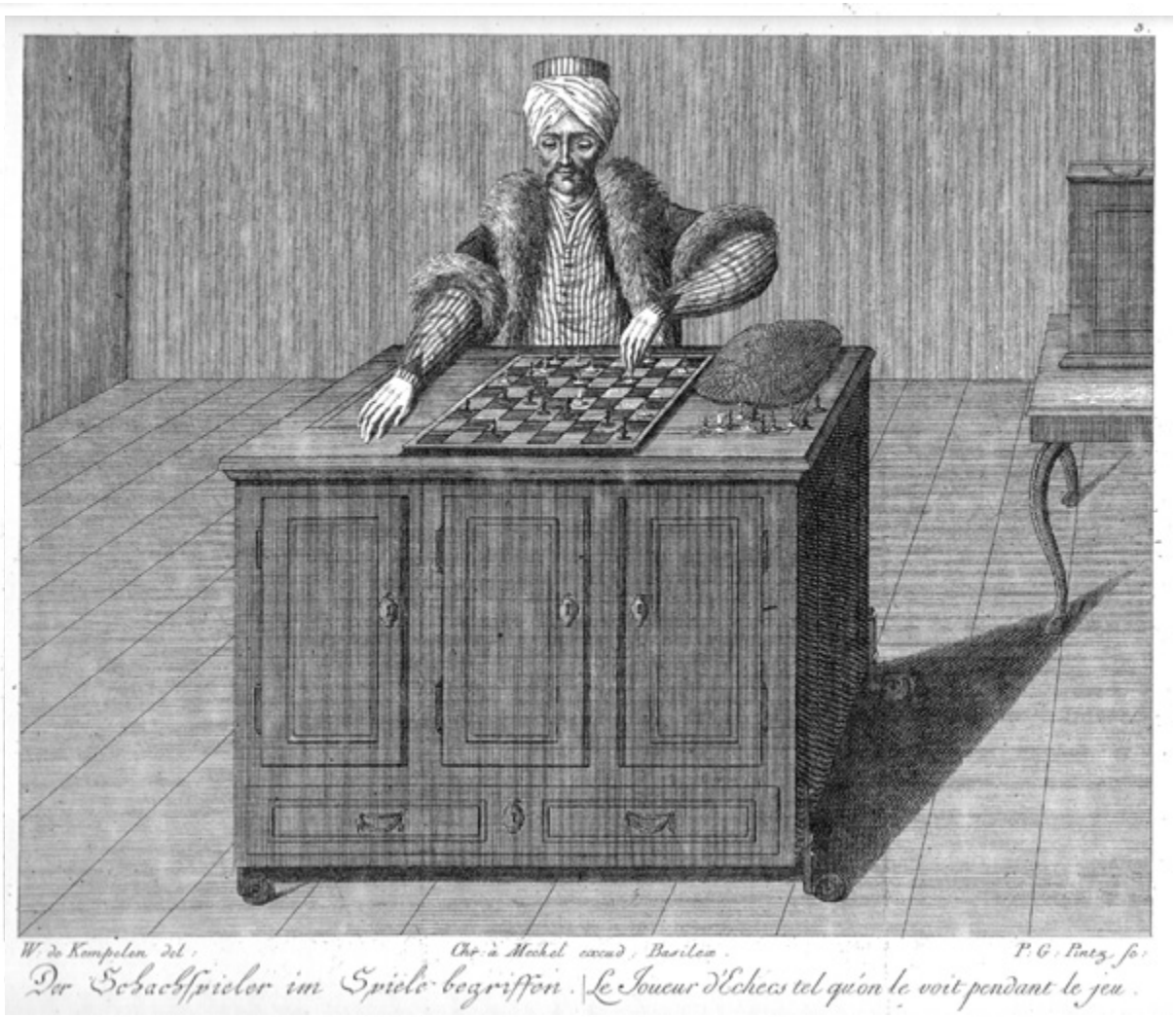
Collaborative Human Computing

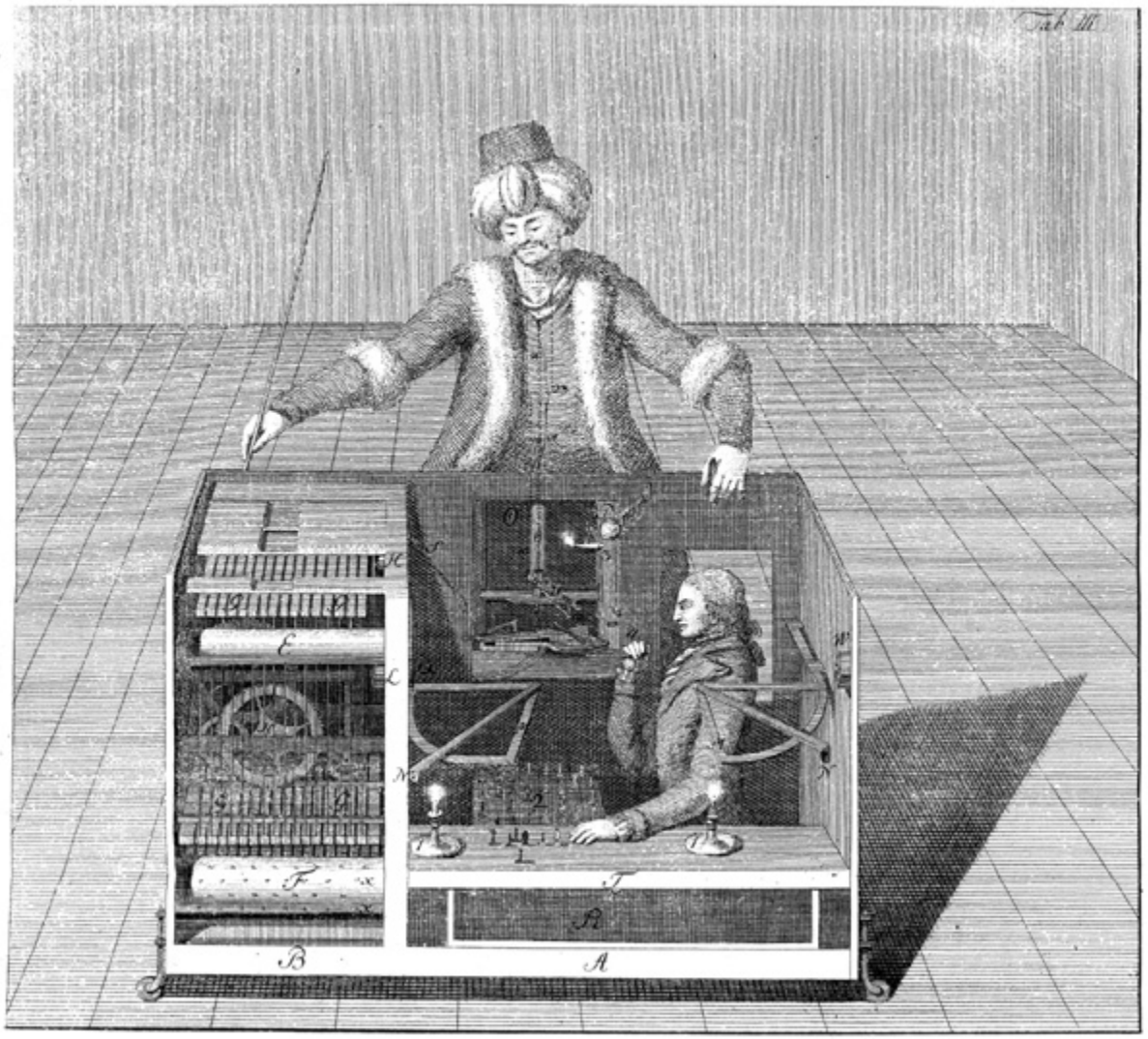
Collaborative Human Computing

INNOCENTIVE®

- Crowdsourced R&D

- Why it works:
  - Solver Diversity
  - Workforce Mentality
  - Vetted Input


W. de Kempelen del.
Chr. à Mechel excud. Basileæ.
P. G. Pintz. sc.
Der Schachspieler im Spiele begriffen. | Le Joueur d'Echecs tel qu'on le voit pendant le jeu.


Tab III
B
A

amazon mechanical turk™
Artificial Artificial Intelligence

Virtual Sweatshop????

Human Intelligence Tasks (HIT)

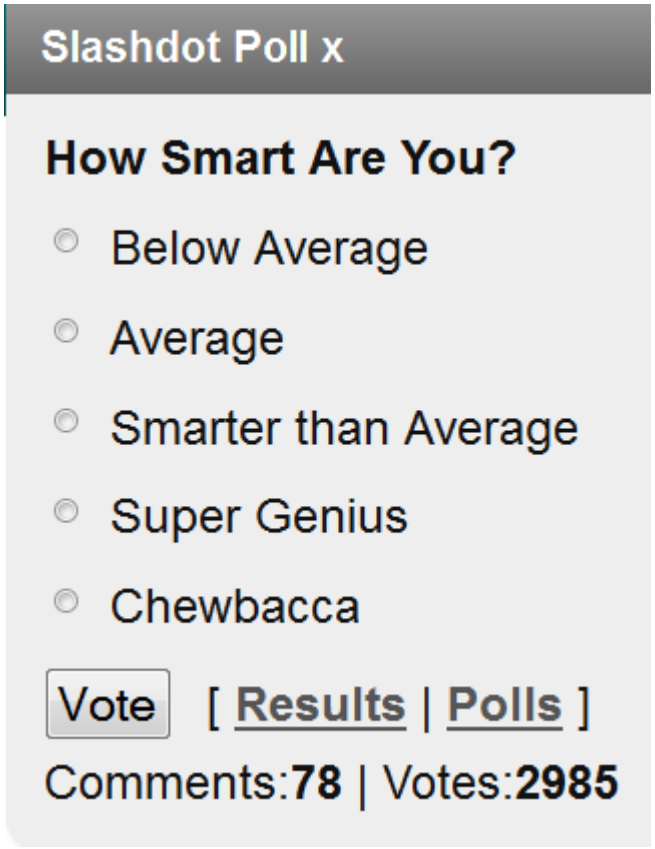
- Relatively trivial for users
- Difficult to automate
- Low payout: $0.01-$5/HIT

For example:

- Image tagging
- Write a review (movies, CDs)
- Rank a series of pictures

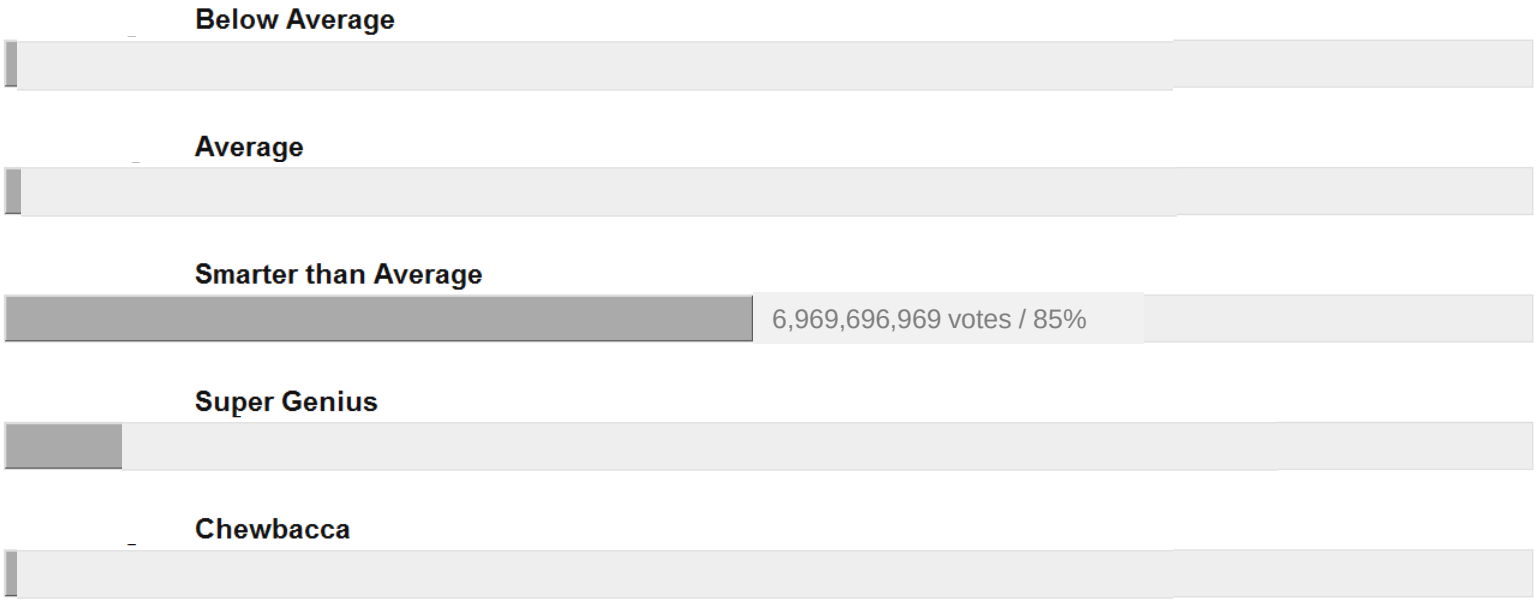
How about harnessing the power of masses for

FREE and **Get Paid?**

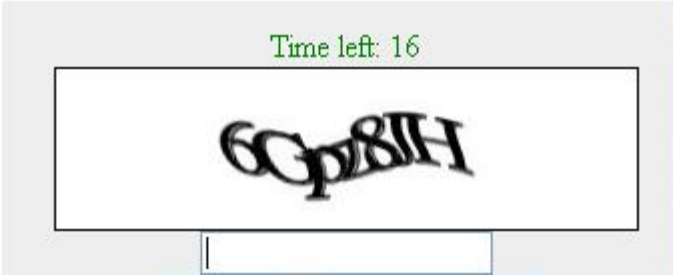

Slashdot Poll x
How Smart Are You?
Below Average
Average
Smarter than Average
Super Genius
Chewbacca
Vote
[Results | Polls]
Comments:78 | Votes:2985

How Smart Are You?

Displaying poll results.



To see the next picture…



Lesson: Give the crowd something they need...

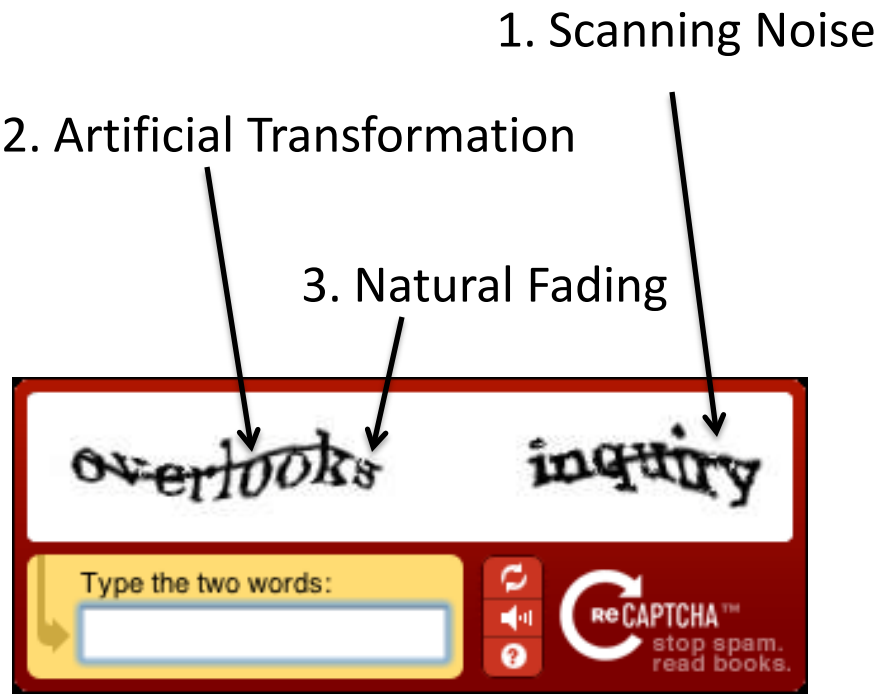
- Initiative to digitize typeset text
  - Today: OCR fails to recognize 20% of scanned text

- How?
  1. Scanned page
  2. Decipher with 2 independent OCR programs
  3. List suspicious words (no consensus)
  4. Distort and send out as reCaptcha

6. Enter unrecognized word into database (consensus established between ***n*** people)

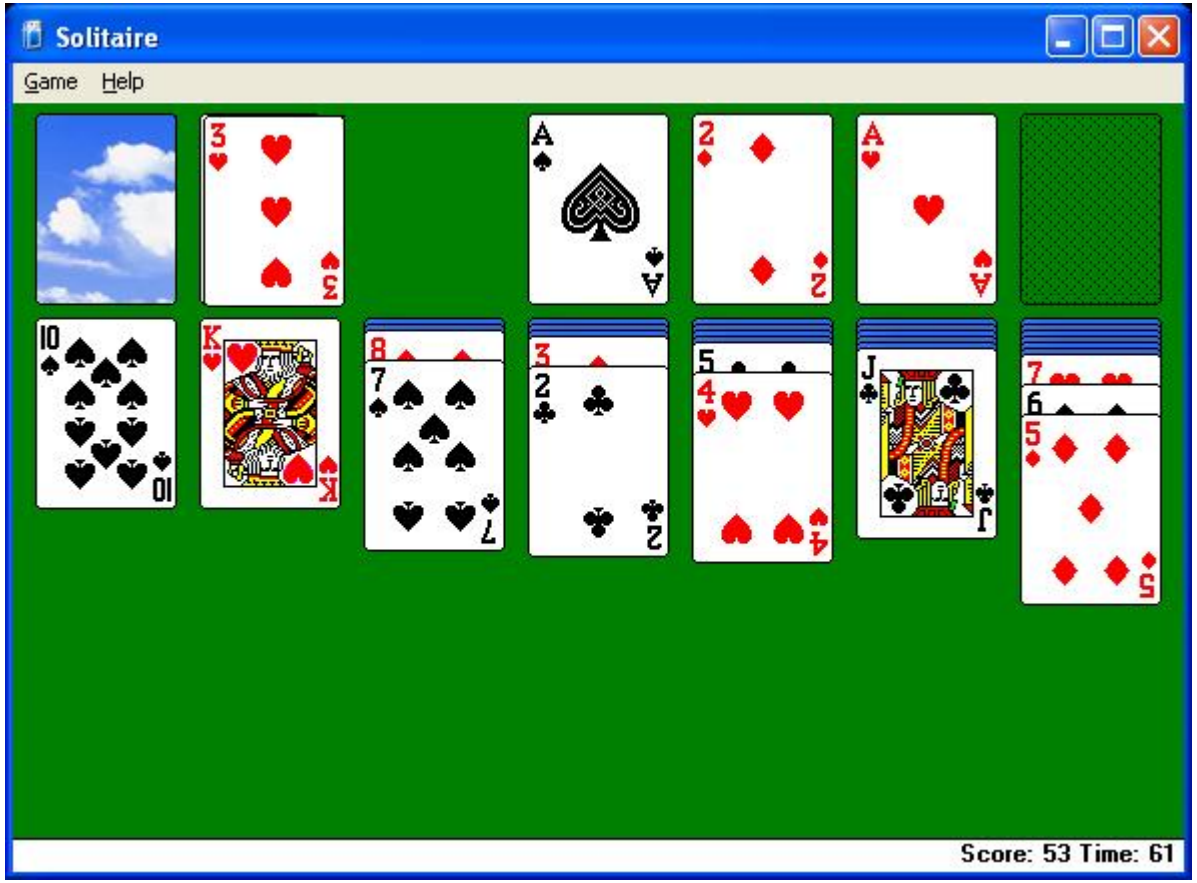
Is it secure?

- More secure than conventional Captchas
  - Anti-captcha algorithms
  - 100% Successful in failing anti-captcha algorithms
  - Computer-generated Captcha 90% successful



Is it successful?

- Accuracy of 99.1%
  - Human: 99%
  - Standard OCR: 83.5%
- 440 Million words deciphered in the 1st year (~17,600 books)
- 35 Million words/day (March, 2009)



9 BILLION human-hours/year

gwap

ESP Game
Concentrate...
How to Play
1 You and a partner see the same image.
2 Each of you must guess what words your partner is typing.
make a ne
Tree
Got it, Let's Play!
View Top Scores

gwap

Image Tagging

- Is it fun?
  - 15 million agreements (tags) from 75,000 players
  - 200,000 **regular** players
  - Many people play >20 hours a week
  - Playing streaks of >15 hours

- Why?
  - Sense of connection with your partner

- Bush
- President
- Man
- **Yuck**

"...the two of you are bringing your minds together in ways lovers would envy."

Single Player Version?

- Record moves of players with time stamps
- Play pre-recorded moves
- ESN Game
  - Moves recorded (Player A):

  (0:02) goddess; (0:03) ziyi

  (0:04) thoughtful; (0:08) hot

Taboo Words	Time	Player 1	Bot (Player A)
Woman	0:01	***ziyi***	
Beautiful	0:02	asian	goddess
Gorgeous	0:03	model	***ziyi***

…0 Player?

<u>Moves recorded</u>

Bot 1: (0:02) goddess; (0:04) face; (0:08) hot ***(0:14) flowers***

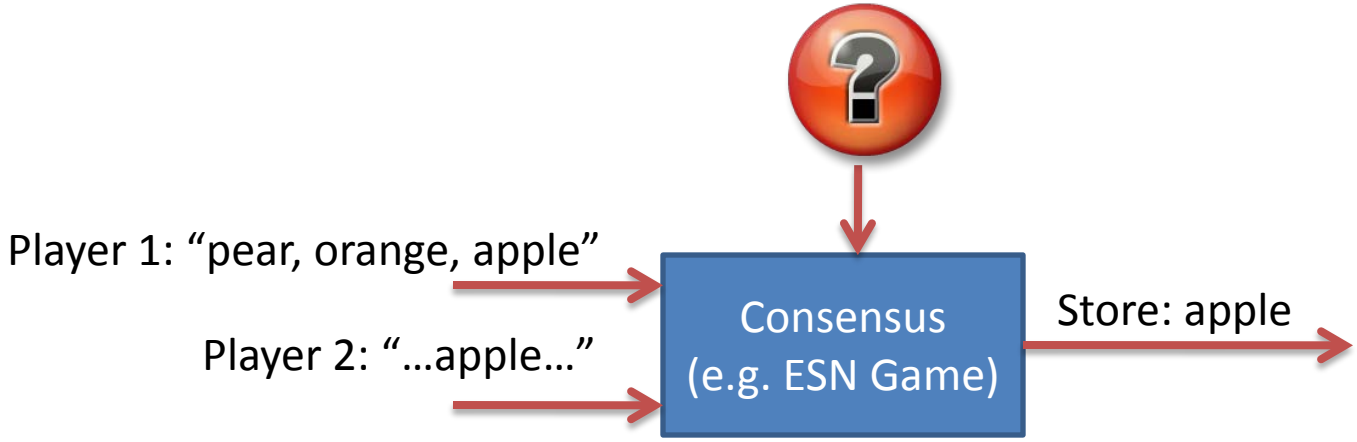
Bot 2**:** ***(0:01) flowers***; (0:02) model; (0:03) asian; (0:09) girl

Generalization

- Game <-> algorithm: Input-Output

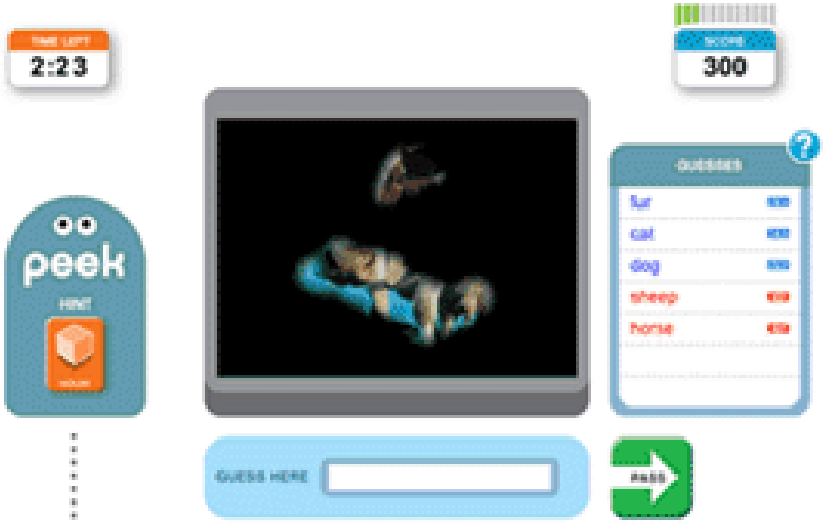
- Symmetric/Parallel:

  n player completing the same task



Peekaboom

PEEK : GUESS WHAT YOUR PARTNER IS REVEALING



HINTS HELP YOU GUESS

PASS FOR DIFFICULT IMAGES

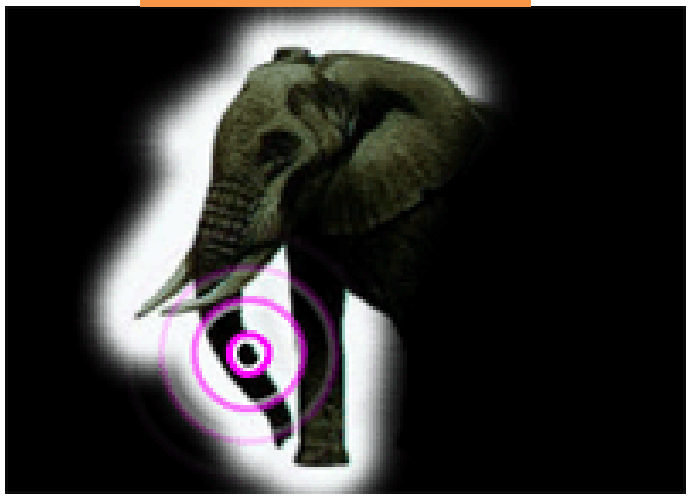
BOOM : REVEAL PARTS OF THE IMAGE TO YOUR PARTNER

GIVE HINTS IF NECESSARY

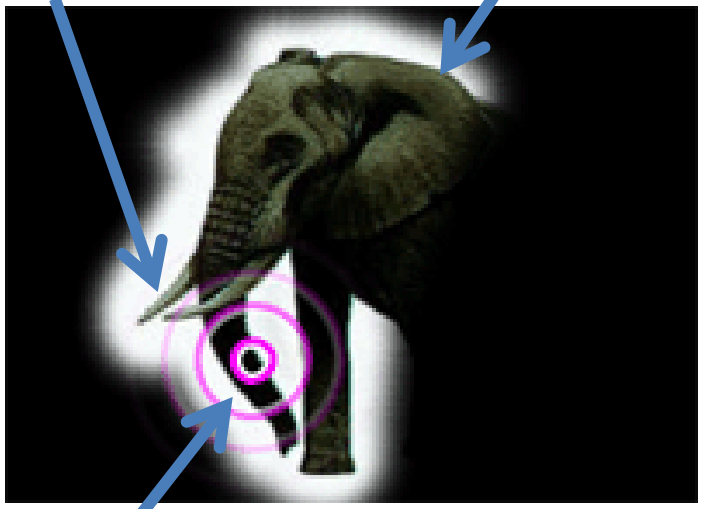
TELL YOUR PARTNER IF A GUESS IS **HOT** OR **COLD**

Peekaboom

Trunk/Tusk/Ear



User-Created Pings



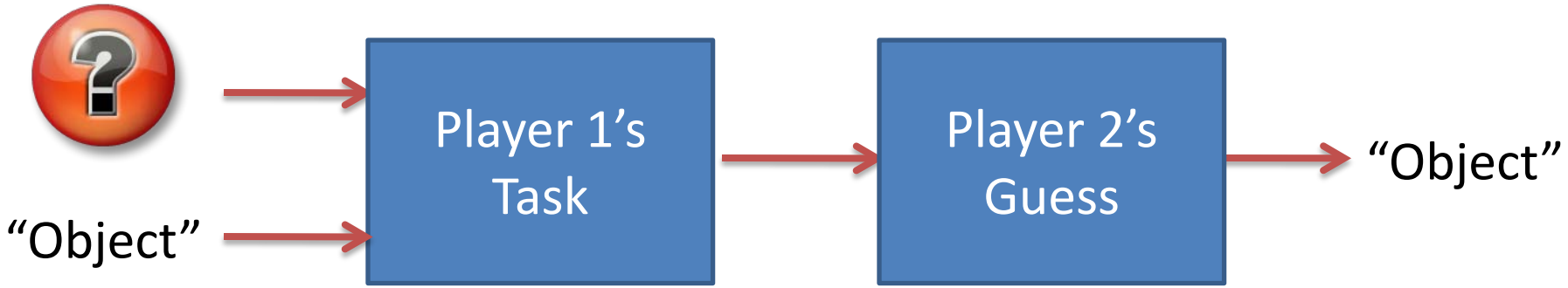
Peekaboom

Hints:

Generalization

Asymmetric/Sequential:

Player 1's output fed to Player 2's input



Security Measures

Pretty standard …

- Player queue
- IP Check (location proximity)

Security Measures

More interesting…

- Test image/behaviour matching
- Aggregated consensus
- reCaptcha the gwap games?

References

- L. von Ahn, M. Blum (2006). Peekaboom: A game for locating objects in images. In *ACM CHI.*
- L. von Ahn, B. Maurer, C. McMillen, D. Abraham, and M. Blum. "reCAPTCHA: Human-Based Character Recognition via Web Security Measures." *Science,* September 2008.
- J. Howe. "The Rise of Crowd Surfing"*, Wired*, June 2006.
- D. P. Anderson , J. Cobb , E. Korpela , M. Lebofsky , D. Werthimer, "SETI@home: an experiment in public-resource computing," *Communications of the ACM,* v.45 n.11, p.56-61, November 2002 .
- *gwap, http://www.gwap.com*
- *Amazon Mechanical Turk, https://www.mturk.com/mturk/welcome*
- *Google Tech Talk, http://www.cs.cmu.edu/~biglou/*

Discussion

- Net productivity?
- Declining popularity with time, repackagable?
- …your input?