Collecting Social Dynamics on AMT
or
How to become a fake artist?

Subhransu Maji

Social Choice and Networks

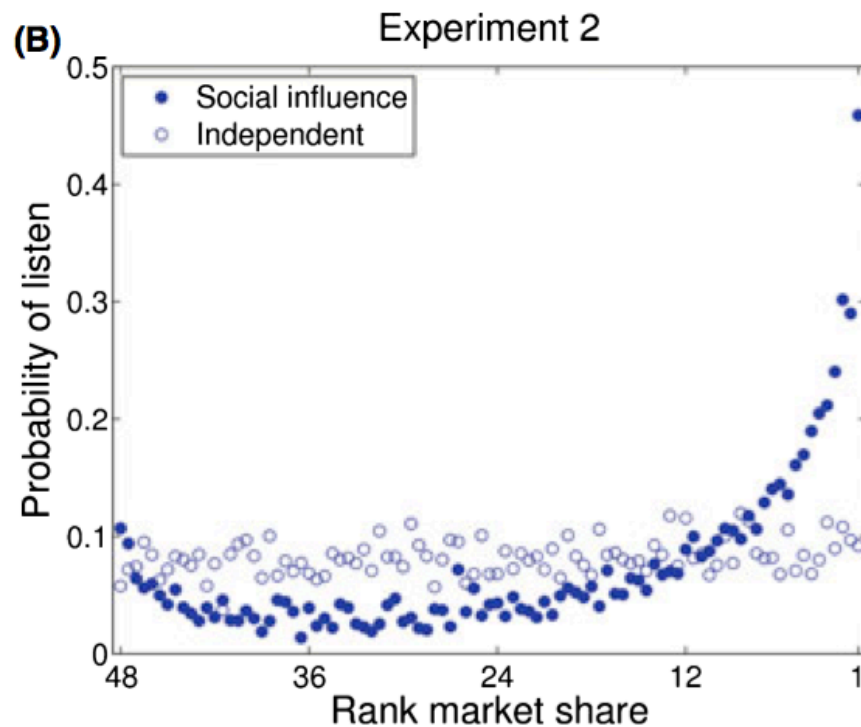Fall 2010

Success and Failure in Cultural Markets

- "winner take all", "rich getting richer"
- Questions:
  - Is harry potter a really good book?
    - Was rejected by 8 publishers before it sold millions..
  - Friends was rejected by FOX before it went to NBC
  - Beatles struggled to get a record deal…
  - Starwars almost never got made…
- Why:
  - Too many things to consider
  - Social acceptance/commonality
  - Difficulty in predicting success
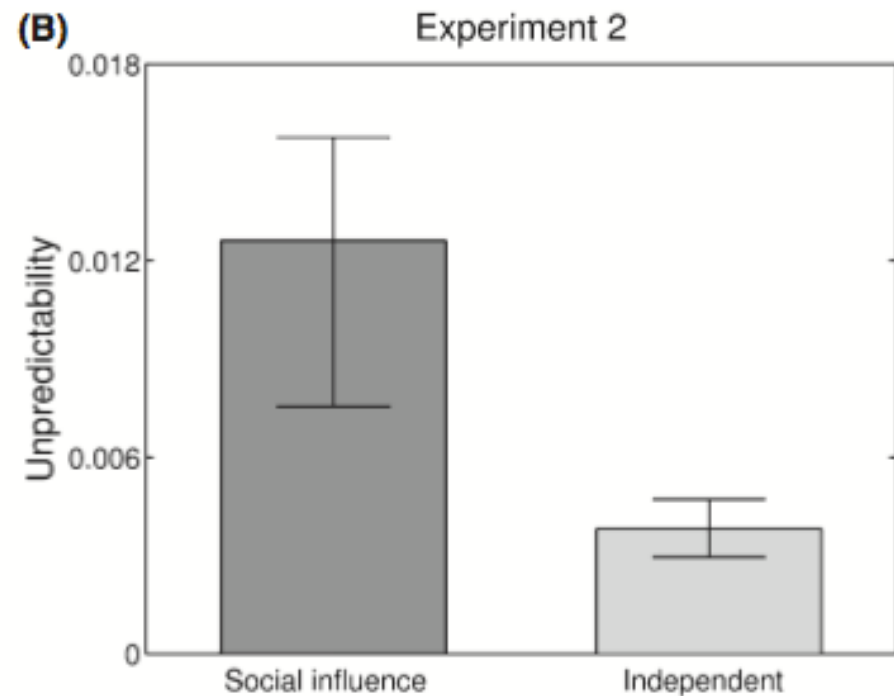
Possible World Model

- Imagine if we could "rewind" time to when Harry Potter was released and "rerun history" a number of times

- If Harry Potter becomes popular in enough of these worlds then it must be a good book

- Can create such markets artificially on the web

Salganik and Watts, 2009

- Experiment with music downloads
- About 3,000 users are asked to listen/rate 48 songs from a website
- 8 possible worlds. Total experiment time = 150 days
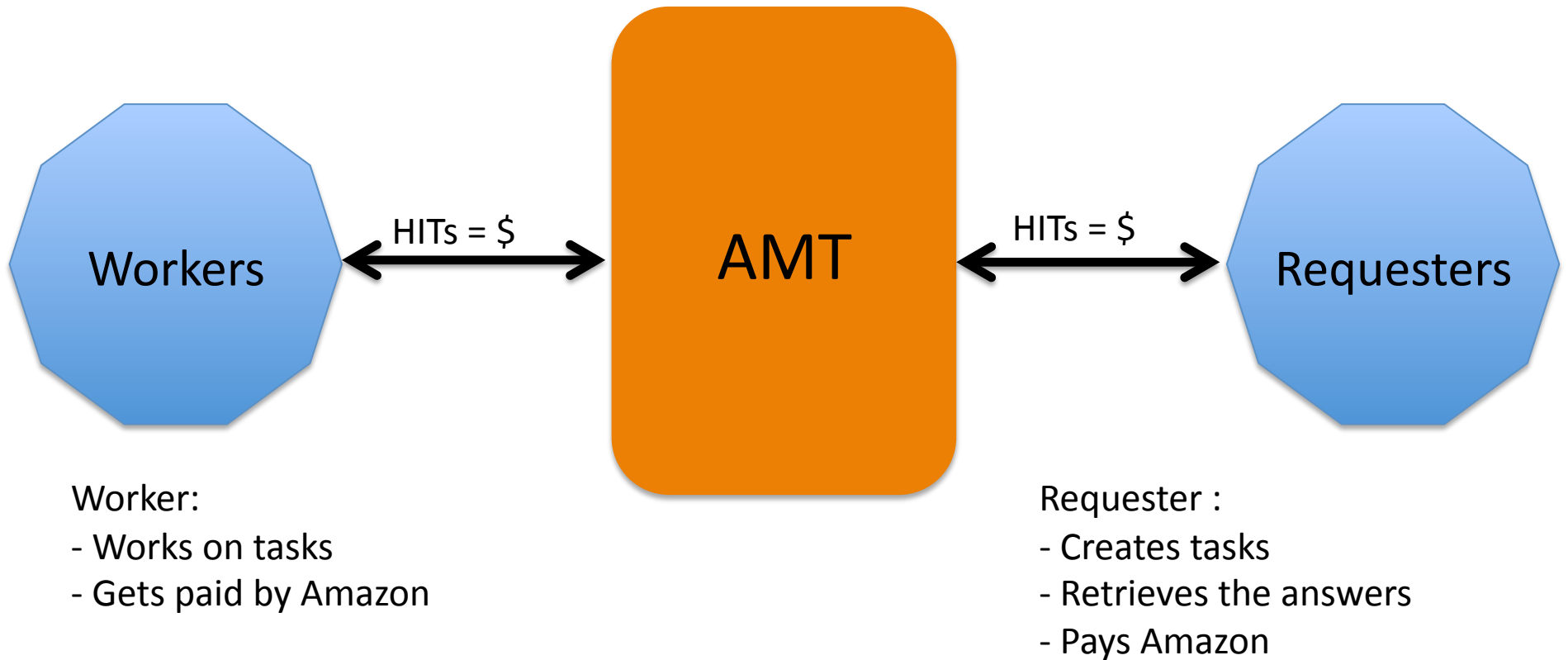

rich-getting-richer dynamics

Success is also difficult to predict

Overview

- Amazon Mechanical Turk's architecture
- Experimental design
  - Rate abstract art paintings w/ side information
- Results and Conclusions

Amazon Mechanical Turk

Workers ←→ HITs = \$ ←→ **AMT** ←→ HITs = \$ ←→ **Requesters**

Worker:
- Works on tasks
- Gets paid by Amazon

Requester :
- Creates tasks
- Retrieves the answers
- Pays Amazon

amazonmechanical turk
beta
Artificial Artificial Intelligence

http://www.mturk.com

HIT : Human Intelligence Task

- "An Amazon Mechanical Turk HIT or Human Intelligence Task is the task you ask a Worker to complete. It may be a task that is inherently difficult for a computer to do." – *AMT Best Practices Guide*

All HITs
1-10 of 2165 Results

Sort by: Reward Amount (most first) GO!   Show all details | Hide all details

Signup and Activate & Report Experience					
Requester: PPD Inc.	HIT Expiration Date:	Oct 16, 2010 (6 days 23 hours)	Reward:	$15.00	
	Time Allotted:	40 minutes	HITs Available:	1	

FULL LENGTH Usability Test on www.tictocwatches.co.uk #215					Take Qu
Requester: TryMyUi Administrator	HIT Expiration Date:	Oct 29, 2010 (2 weeks 5 days)	Reward:	$10.00	
	Time Allotted:	2 hours 43 minutes	HITs Available:	1	

Data Entry. Quick data entry verification.					
Requester: Karen Burnow	HIT Expiration Date:	Oct 10, 2010 (19 hours 44 minutes)	Reward:	$6.00	
	Time Allotted:	60 minutes	HITs Available:	1	

Help test our Credit Report system and make $5! Takes 2 minutes, costs you NOTHING.					
Requester: Laurence Crawford	HIT Expiration Date:	Nov 22, 2010 (6 weeks 2 days)	Reward:	$5.00	
	Time Allotted:	10 minutes	HITs Available:	1	

Health System Recognition Survey					Take Qu
Requester: Deborah L Moore	HIT Expiration Date:	Oct 21, 2010 (1 week 5 days)	Reward:	$4.50	
	Time Allotted:	30 minutes	HITs Available:	1	

Answer a survey about yourself					
Requester: otilia	HIT Expiration Date:	Oct 16, 2010 (6 days 23 hours)	Reward:	$4.00	
	Time Allotted:	1 hour 30 minutes	HITs Available:	1	

UK English Voice Mail Messages					Not Qualified to wo
Requester: Karl Brown	HIT Expiration Date:	Oct 14, 2010 (5 days 4 hours)	Reward:	$4.00	
	Time Allotted:	3 hours	HITs Available:	1	

Experimental Design

Rate these paintings on a scale of 1-5.

Use 1 for bad and 5 for excellent.
Click on "Submit Results" once done.

Instructions:

For each painting tell us (1) how much you like it and (2) your guess of the ratings by a professional. You will be paid based on how close your guess of the professional's ratings is to the true answer. For example, do not simply say all the professional's ratings are excellent. You will not be judged based on what your ratings of these paintings are. The descriptions below each of the paintings may be of help.
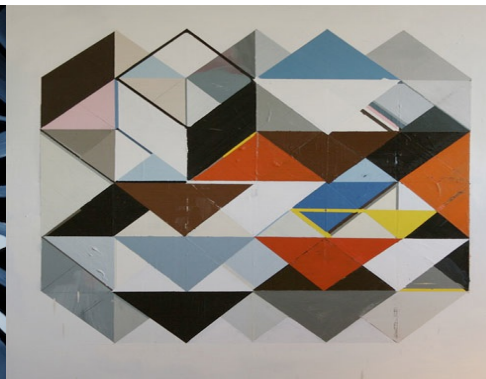
Side information (fake)

your rating : ○1 ○2 ○3 ○4 ○5

professional's rating : ○1 ○2 ○3 ○4 ○5

It is not in public display.

Experimental Design

- Artist
  - It is painted by an amateur
  - It is painted by an average artist
  - It is painted by a well known artist
- Location
  - It is not in public display
  - It is at display at a top art museum
  - It is part of someone's private collection
- Price
  - It is not so expensive
  - It is a moderately expensive painting
  - It is an expensive painting

Results..

- So far 60 different subjects, 360 inputs, 16 hours
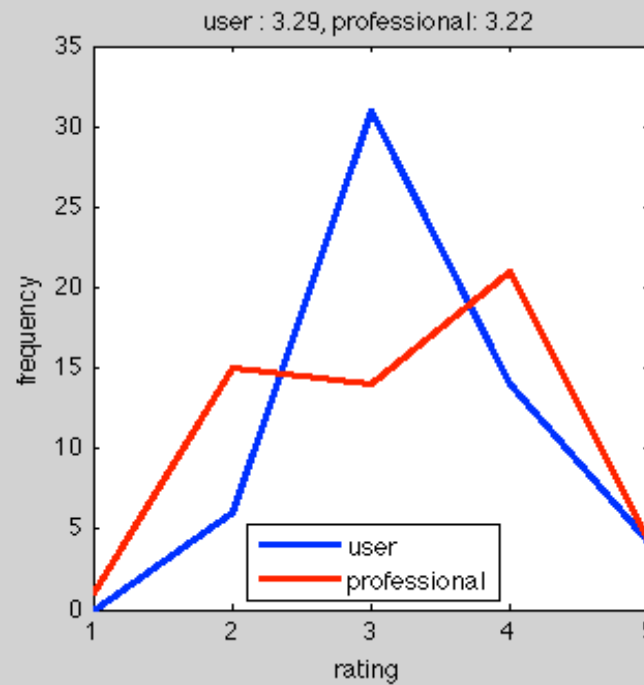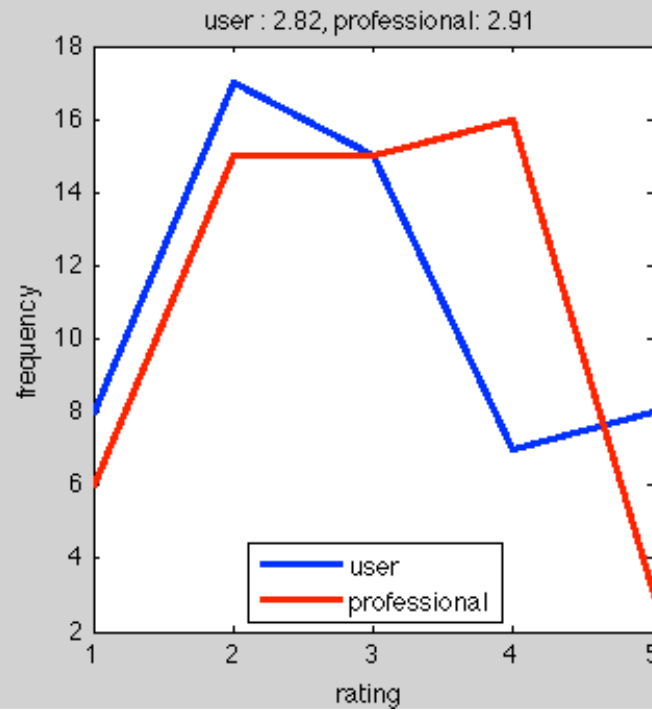- Raw user ratings

| 2.81 | 3.09 | 3.15 | 3.27 | 3.29 | 3.51 |

Increasing user rating →

(Also same as the order of professional ratings)

user : 2.82, professional: 2.91

user : 3.29, professional: 3.22

Distributions of ratings
Means are similar

Effect of side information

bias = professional's rating − user's rating

-0.40(0.98) : It is painted by an amateur

-0.15(1.17) : It is painted by an average artist.

+0.31(1.14) : It is painted by a well known artist.

+0.16(1.09) : It is not in public display.

+0.00(1.09) : It is at display at a top art museum.

-0.08(0.98) : It is part of someone's private collection.

-0.23(1.10) : It is not so expensive.

+0.20(1.07) : It is a moderately expensive painting.

+0.06(1.16) : It is an expensive painting.

Effect of side information

- **+0.31(1.14) : It is painted by a well known artist.**
- Stratified by painting :

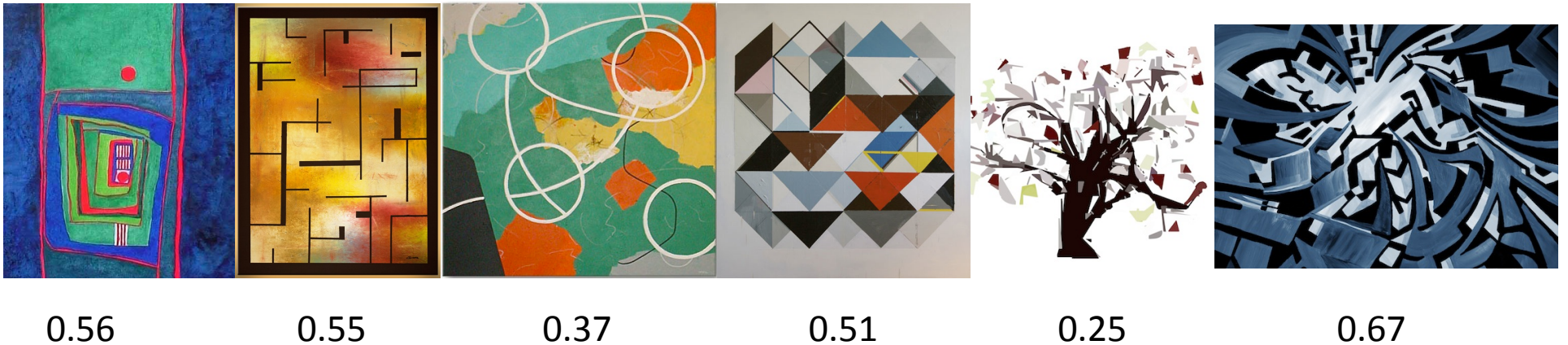| 0.17(0.98) | 0.17(0.98) | 0.00(2.16) | 0.00(0.82) | 0.50(1.38) | 0.56(1.13) |

More effect for the extremes?

Correlated Results?

- Pearson correlation coefficients between user and professional ratings…

- Seems like side information affects the user ratings!

| 0.56 | 0.55 | 0.37 | 0.51 | 0.25 | 0.67 |

More effect for the extremes?

Conclusions

- Proof of concept for social experiments on Amazon Mechanical Turk
- Time spend :
  - 1 hour to create the interface
  - 16 hours to collect user input (at the time of creating slides)
  - 2 hours for slides/matlab code to process data
- About 60 unique users/360 ratings..
- Expenses : 3$