

On Mice and Men: The Role of Biology in Crowdsourcing

Panos Ipeirotis

oDesk & New York University

Twitter: @ipeirotis

“A Computer Scientist in a Business School”
<http://behind-the-enemy-lines.com>

A Quick Survey

- How many have received low quality submissions (or spam) on your Mechanical Turk tasks?
 - Spam → quality control
- How many implemented quality control ?
 - Quality control → Blocking bad, rewarding good workers
- How many display a “quality” score to the workers?

Scores may leads to strange interactions

The *“I will just open another account”* spammer:

- *New account, (often) same IP, always same behavior*

The *“angry, has-been-burnt-too-many-times”* worker:

- *“F*** YOU! I am doing everything correctly and you know it! Stop trying to reject me with your stupid ‘scores’!”*

The *overachiever* worker:

- *“What am I doing wrong?? My score is 92% and I want to have 100%”*

National Academy of Sciences Dec 2010 “Frontiers of Science” conference



Your workers
behave like my
mice!

An unexpected connection...

National Academy of Sciences Dec 2010 “Frontiers of Science” conference



Your workers
behave like my
mice!

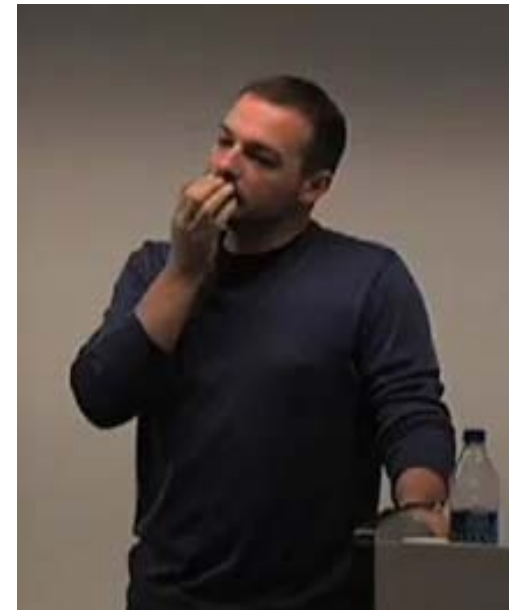
Eh?



An unexpected connection at the NAS “Frontiers of Science” conf.



Your workers want to use
only their **motor skills**,
not their **cognitive skills**



The Biology Fundamentals

- Brain functions are biologically expensive (20% of total energy consumption in humans)
- Motor skills are more energy efficient than cognitive skills (e.g., walking)
- Brain tends to delegate easy tasks to part of the neural system that handles motor skills

An unexpected connection at the NAS “Frontiers of Science” conf.



Your workers want to use
only their **motor skills**,
not their **cognitive skills**

Makes sense



An unexpected connection at the NAS “Frontiers of Science” conf.



And here is how
I train my mice
to behave...



The Mice Experiment



Cognitive
Solve maze
Find pellet



Motor
Push lever three times
Pellet drops

How to Train the Mice?



Confuse motor skills!
Reward cognition!

*I should try this the
moment that I get
back to my room*



Punishing Worker's Motor Skills

- **Punish bad answers** with frustration of motor skills (e.g., add delays between tasks)
 - “Loading image, please wait...”
 - “Image did not load, press here to reload”
 - “404 error. Return the HIT and accept again”

→ Make this **probabilistic** to keep feedback implicit

Misery

[View](#) [Version control](#)

Posted by [danielb](#) on *June 22, 2009 at 10:10am*

Misery is a module designed to make life difficult for certain users.

It can be used:

- As an alternative to banning or deleting users from a community.
- As a means by which to punish members of your website.
- To delight in the suffering of others.

Currently you can force users (via permissions/roles, editing their user account, or using [Troll](#) IP blacklists) to endure the following misery:

- **Delay:** Create a random-length delay, giving the appearance of a slow connection. (by default this happens 40% of the time)
- **White screen:** Present the user with a white-screen. (by default this happens 10% of the time)
- **Wrong page:** Redirect to a random URL in a predefined list. (by default this happens 0% of the time)
- **Random node:** Redirect to a random node accessible by the user. (by default this happens 10% of the time)
- **403 Access Denied:** Present the user with an "Access Denied" error. (by default this happens 10% of the time)
- **404 Not Found:** Present the user with a "Not Found" error. (by default this happens 10% of the time)



Thanks to Rob Miller for the pointer!

Rewarding (?) Cognitive Effort

- **Reward good answers** by rewarding the cognitive part of the brain
 - Introduce variety
 - Introduce novelty
 - Give new tasks fast
 - Show score improvements faster (but not the opposite)
 - Show optimistic score estimates

Experiments

- Web page classification
- Image tagging
- Email & URL collection

Experimental Summary (I)

- Spammer workers quickly abandon
 - No need to display scores, or ban
 - Low quality submissions from ~60% to ~3%
 - Half-life of low-quality from 100+ HITs to less than 5
- Good workers unaffected
 - No significant effect on participation of workers with good performance
 - Lifetime of participants unaffected
 - Reduction in response time (*after* removing the “intervention delays”; that was puzzling)

Experimental Summary (II)

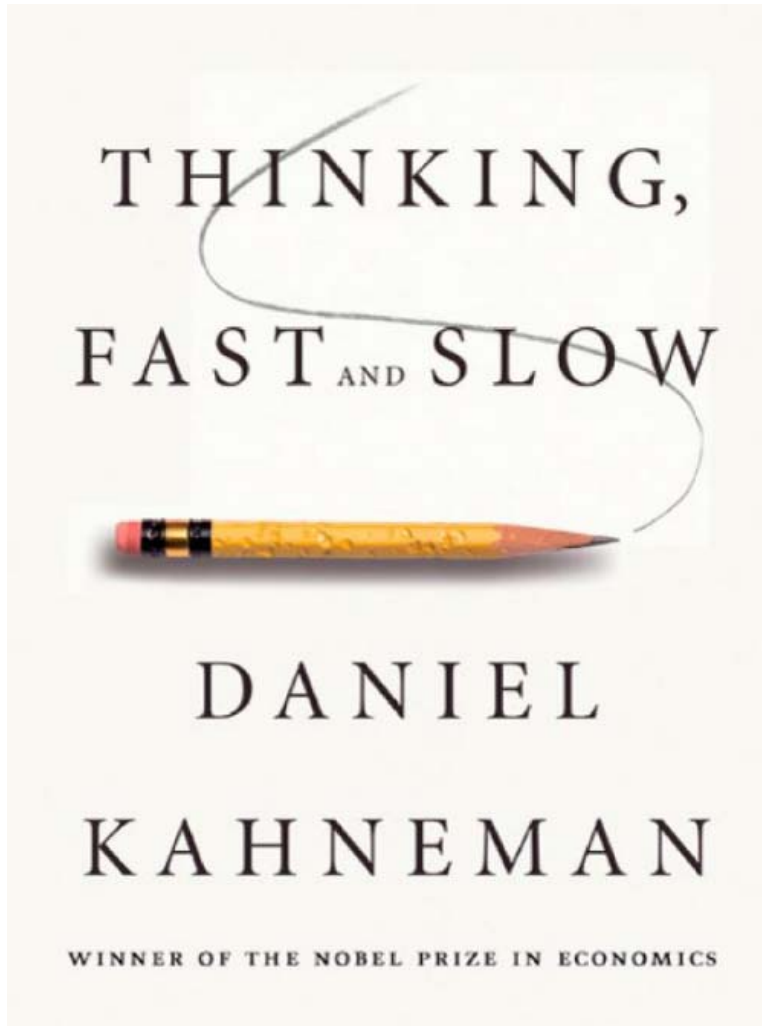
- Remember, scheme was for *training* the mice...
- 15%-20% of the spammers start submitting good work!

????

Two key questions


- Why response time was slower for some good workers?
- Why some low quality workers start working well?

????



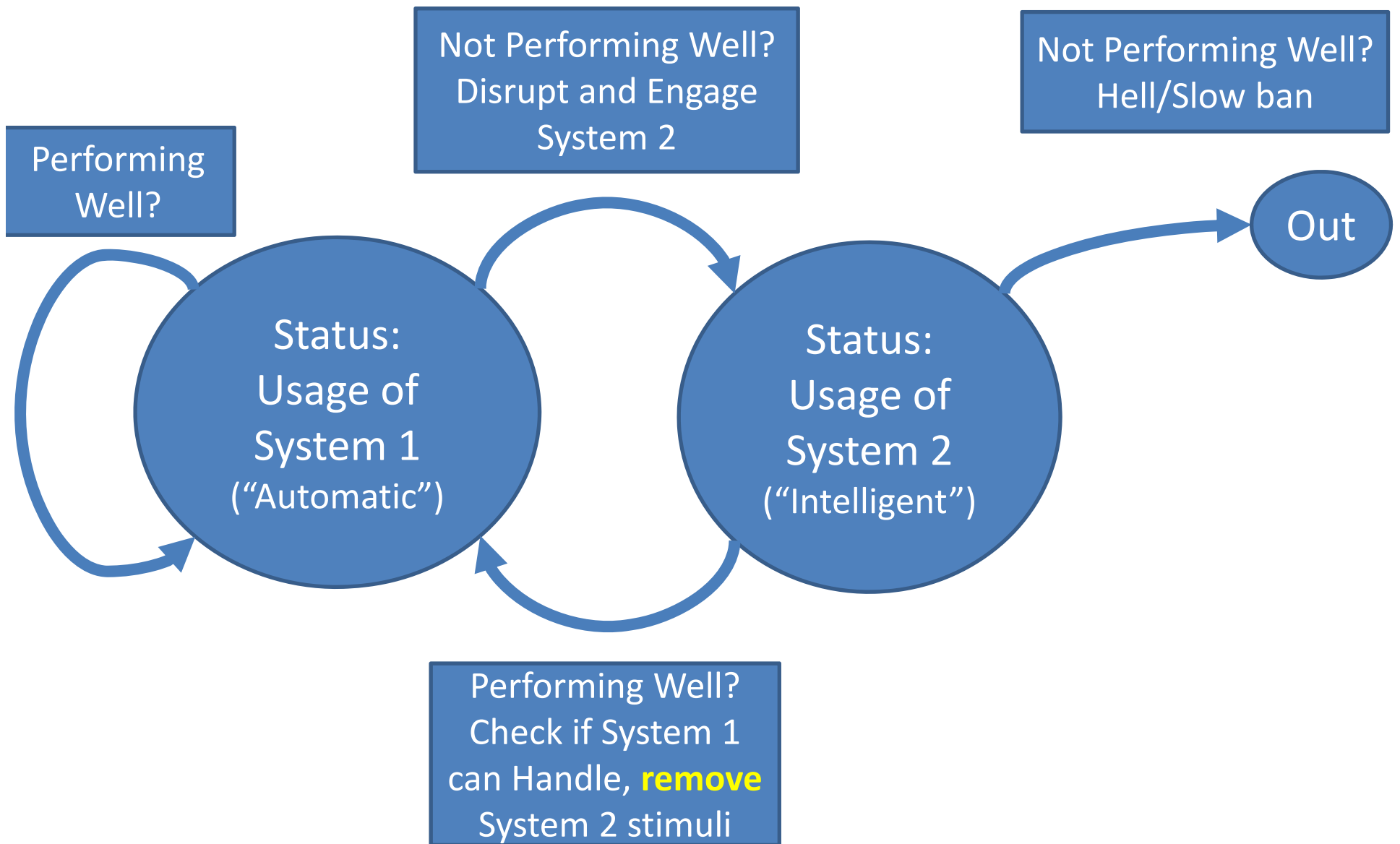
- System 1:
“Automatic” actions
- System 2:
“Intelligent” actions

System 1 Tasks

- Detect that one object is more distant than another.
- Orient to the source of a sudden sound.
- Complete the phrase “bread and...”
- Make a “disgust face” when shown a horrible picture.
- Detect hostility in a voice.
- Answer to $2 + 2 = ?$
- Read words on large billboards.
- Drive a car on an empty road.
- Find a strong move in chess (if you are a chess master) 
- Understand simple sentences.

System 2 Tasks

- Focus attention on the clowns in the circus.
- Look for a woman with white hair.
- Count the occurrences of the letter *a* in a page of text.
- Compare two washing machines for overall value.
- Check the validity of a complex logical argument.



How to Measure Performance?

- **Use gold testing and redundancy** (details in “Managing Crowdsourced Workers” working paper, by Wang et al.)
- **Use behavioral signals** (see “Instrumenting the Crowd” by Rzeszotarski and Kittur, UIST 2011)

Current Experimental Results (web page classification)

- Spammer workers quickly abandon (as before)
- Borderline workers underperforming in System 1 mode perform fine in System 2
- **Some** good workers operating in System 2 (instead of System 1) improve performance (through added novelty, etc.) but are slower
- Good workers operating in System 1 operate as before, no reduction in speed

Thanks!

Q & A?