

**Inside the Turk:  
Methodological Concerns and Solutions  
in Mechanical Turk Experimentation**

ACR 2012

# Research on Mechanical Turk

---

- 50% in SPSP and SJDM listservs have used Mturk. Why?
  - Speed (58%) – Hundreds of answers in a few hours
  - Affordability (31%) – Few cents per subject
  - Diversity (31%) – Age 18-80, different education levels, etc.
- Data quality meets standards
  - Buhrmester, Kwang, & Gosling (2011) *Perspectives in Psychological Sciences*
  - Paolacci, Chandler, & Ipeirotis (2010) *Judgment and Decision Making*
  - Horton, Rand, & Zeckhauser (2012) *Experimental Economics*

**MTurk is good. However...**

# Some Open Questions

---

1) Critical Differences (Joseph Goodman)

2) Attention (Emily Peel)

3) Honesty (Dan Goldstein)

4) Non-Naivety (Gabriele Paolacci)

---

# Participant Non-Naivety on MTurk

---

**Gabriele Paolacci**

Erasmus University Rotterdam

*Joint work with:*

**Jesse Chandler** (PRIME Research & Princeton University)

**Pam Mueller** (Princeton University)

---

# Participant Non-Naivety and MTurk

---

- Foreknowledge is often bad
- Two potential causes of non-naivety:
  1. Crosstalk
  2. Repeated participants
- Exacerbated problem on Mturk:
  - Half of us share this subject pool
  - It is not continuously replenished
- Only 4% researchers seem concerned

# An example

---

## A worker on a forum:

It makes me wonder if some of the folks posting surveys are somehow unaware that turkers take hundreds of surveys. If they're unaware, they don't know what they're doing.

If they *are* aware, why do I see so many surveys cutting-and-pasting questions like, as an example, these:

***A bat and a ball cost \$1.10 in total. The bat costs \$1.00 more than the ball. How much does the ball cost?***

***If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?***

***In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?***

These questions can only work to measure what they're supposed to measure the very first time a survey-taker sees them. After that, the cat's out of the bag. Why would a serious researcher put that sort of content into a survey that goes to a pool of professional survey takers? It's a tainted pool. Do they not notice that MTurkers get those questions right way more than the general population? Again, I suspect many of these researchers simply don't care. They're either students, or they're just ripping off whomever gave them the money for the research.

# Mturk Crosstalk

---


- Mturk forums:
  - [turkernation.com](http://turkernation.com)
  - [mturkforums.com](http://mturkforums.com)
  - [turkopticon.com](http://turkopticon.com)
- On these forums turkers:
  - talk about HITs
  - talk about Requesters
  - link to HITs

# Mturk Crosstalk

- Mturk forums:
  - turkernation.com
  - mturkforums.com
  - turkopticon.com
- On these forums turkers:
  - talk about HITs
  - talk about Requesters
  - link to HITs

09-06-2012 01:25 AM #10

**wendi5000** ◦  
Member  
★



Join Date: May 2012  
Gender: ♀  
Location: Houston, TX  
Posts: 259

Not really funny but a good heads-up. A new thing I have seen in a couple of surveys today is special instructions about only copying part of the "code" at the end. For example, the last one I did said this:

*On the next page, you will see a complete code which will consist of three letters, two numbers, and then the letter "x" followed by three additional numbers. In order to receive credit on Mechanical Turk, copy only the three letters and two numbers that precede the "x". If you copy the three numbers that follow the "x", we will assume that you did not follow our directions and you will not be paid.*

*First they came for the Communists, and I didn't speak up, because I wasn't a Communist.  
Then they came for the Jews, and I didn't speak up, because I wasn't a Jew.  
Then they came for the Catholics, and I didn't speak up, because I was a Protestant.  
Then they came for me, and by that time there was no one left to speak up for me.*



# Is crosstalk a problem?

---

# Is crosstalk a problem?

---

Worker survey (N = 320):

- 26% reported knowing someone else who used MTurk personally
- 28% reported reading forums and blogs about MTurk
- 13% reported ever seeing a discussion about the contents of a study

<b>Ranked Topics of conversation</b>
1. How much the HIT pays
2. Requester reputation
3. How long the HIT takes
4. How difficult it is to complete
5. How to successfully complete the HIT
6. How fun the HIT was
7. Purpose of the HIT

# Repeated Participants

---

- How likely is it that:
  - any MTurk worker has done a similar HIT before?
  - a worker in my HIT has done something similar before?
- Method:
  - We pooled past HITs from us and other researchers:
    - 132 batches (studies)
    - 16,408 HITs (individual surveys taken)
  - We looked at concentration of working activity
  - We supplemented this with our Worker survey data

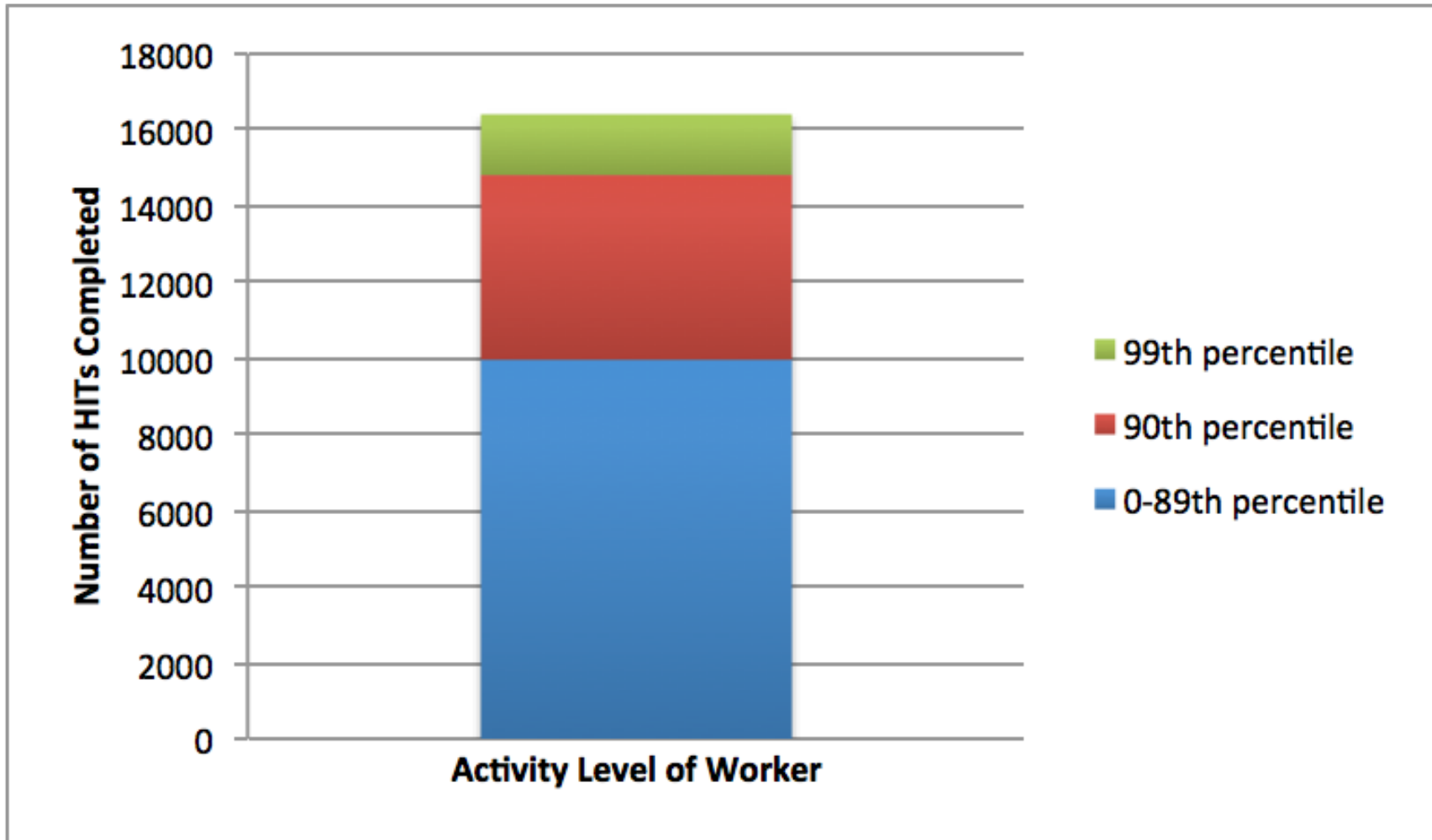
# Worker Activity

---

- 7,498 workers completed 16,408 HITs (avg. worker = 2.24 HITs)

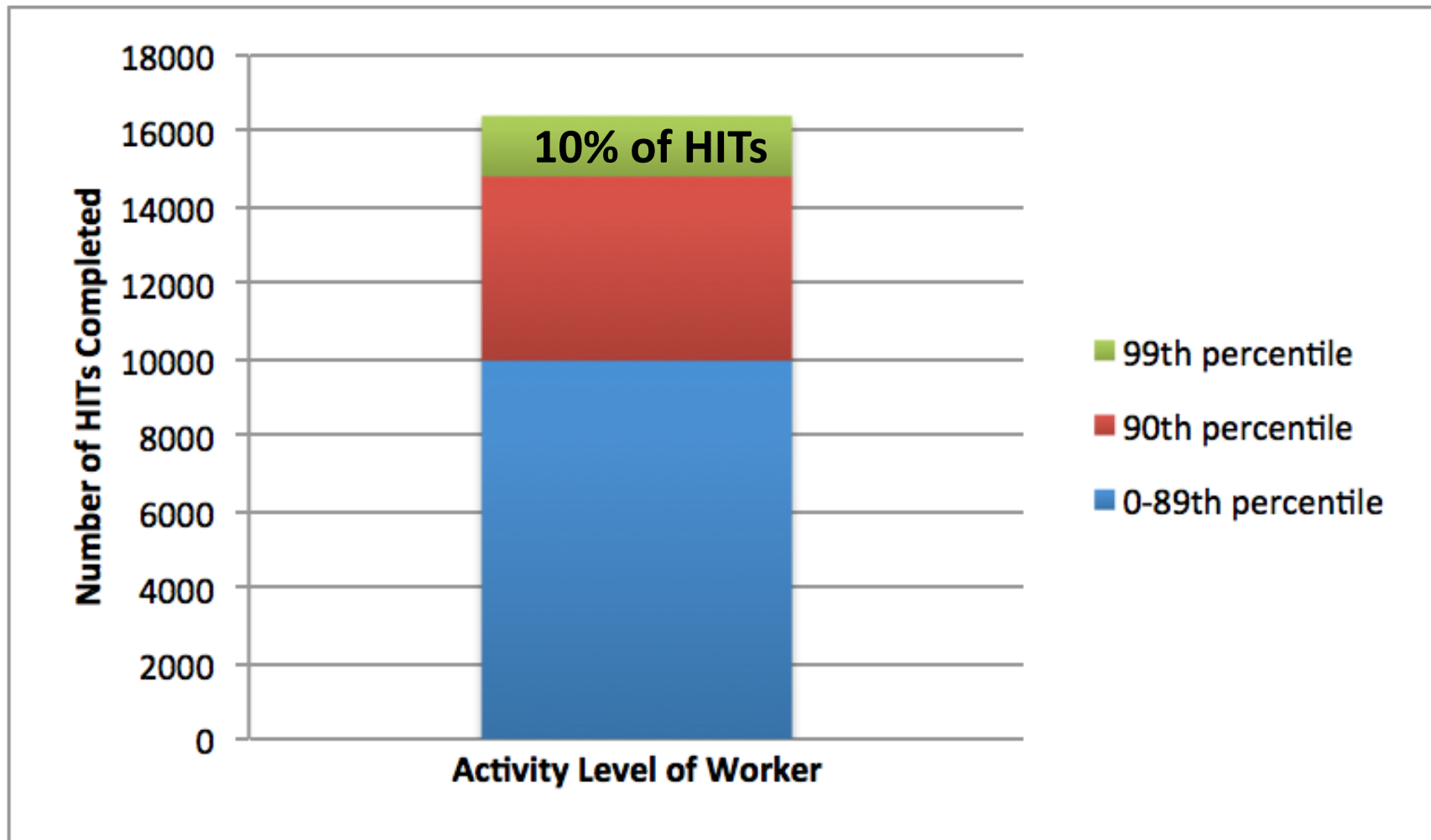
# Worker Activity

- 7,498 workers completed 16,408 HITs (avg. worker = 2.24 HITs)



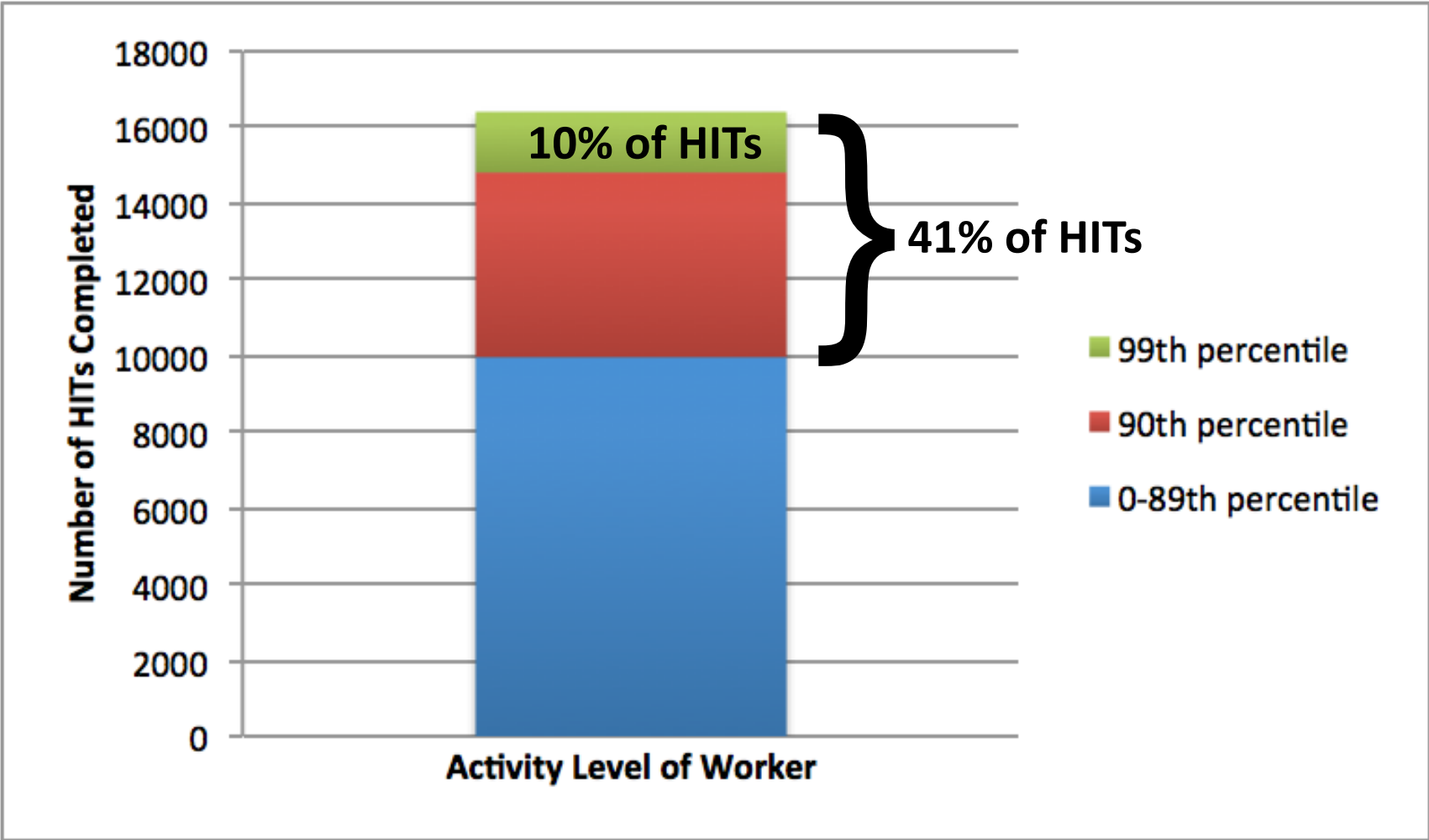
# Superturkers

- 7,498 workers completed 16,408 HITs (avg. worker = 2.24 HITs)

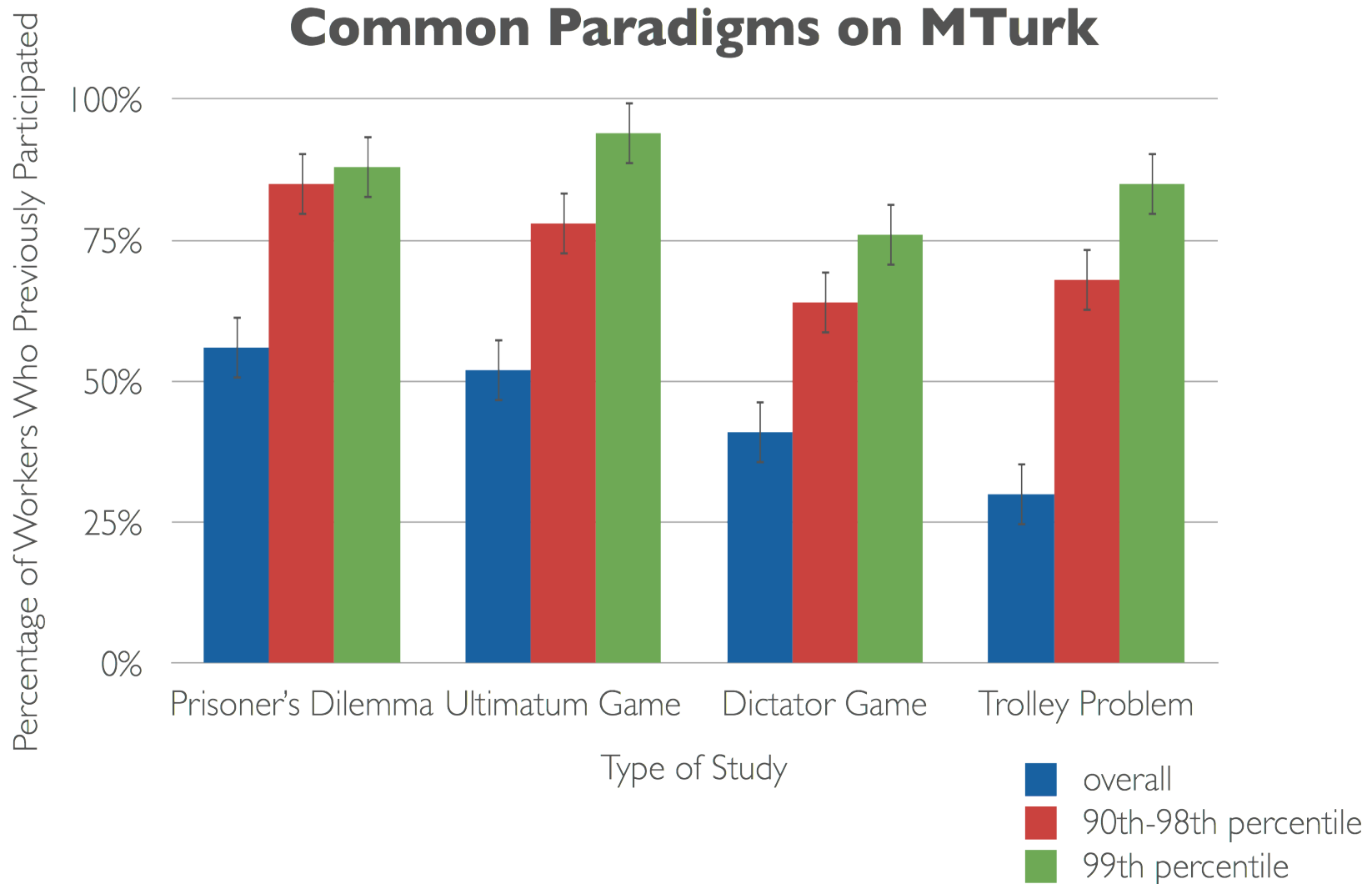


# Superturkers

- 7,498 workers completed 16,408 HITs (avg. worker = 2.24 HITs)



# Classics should better be avoided





# Preventing Repeated Participants

---



- Block previous workers
- “Please don’t do this survey if you’ve done it already”.
- Direct turkers to same page, change the redirect
- “Have you done anything similar before?”
- Qualifications (duplicates won’t see your HIT)
- Use our Qualtrics/MTurk script



# Conclusions

---

- Don't assume that your participants are naïve
- Monitor Mturk forums
- Avoid classic paradigms, manipulations, measures, etc.
- Always save Worker IDs

**<http://experimentalturk.wordpress.com>**