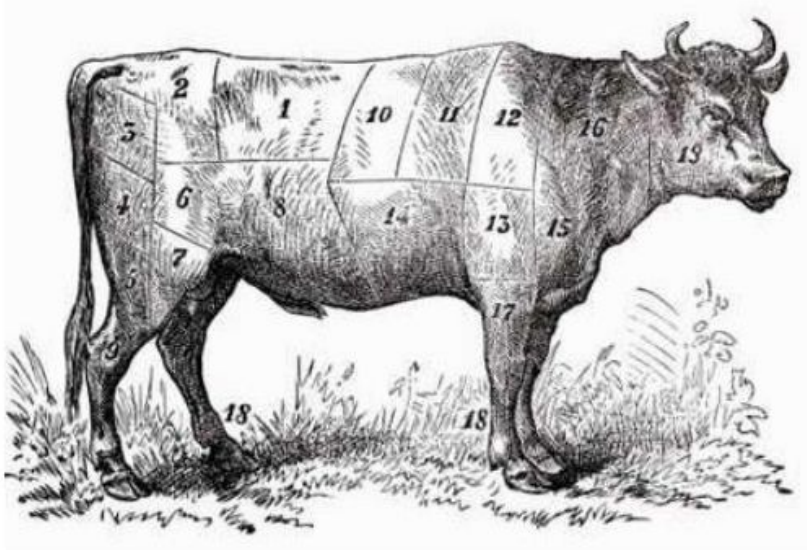
Studying the “Wisdom of Crowds” at scale

Presented at HCOMP 2019

Camelia Simoiu

with Chiraag Sumanth, Alok Mysore, Sharad Goel

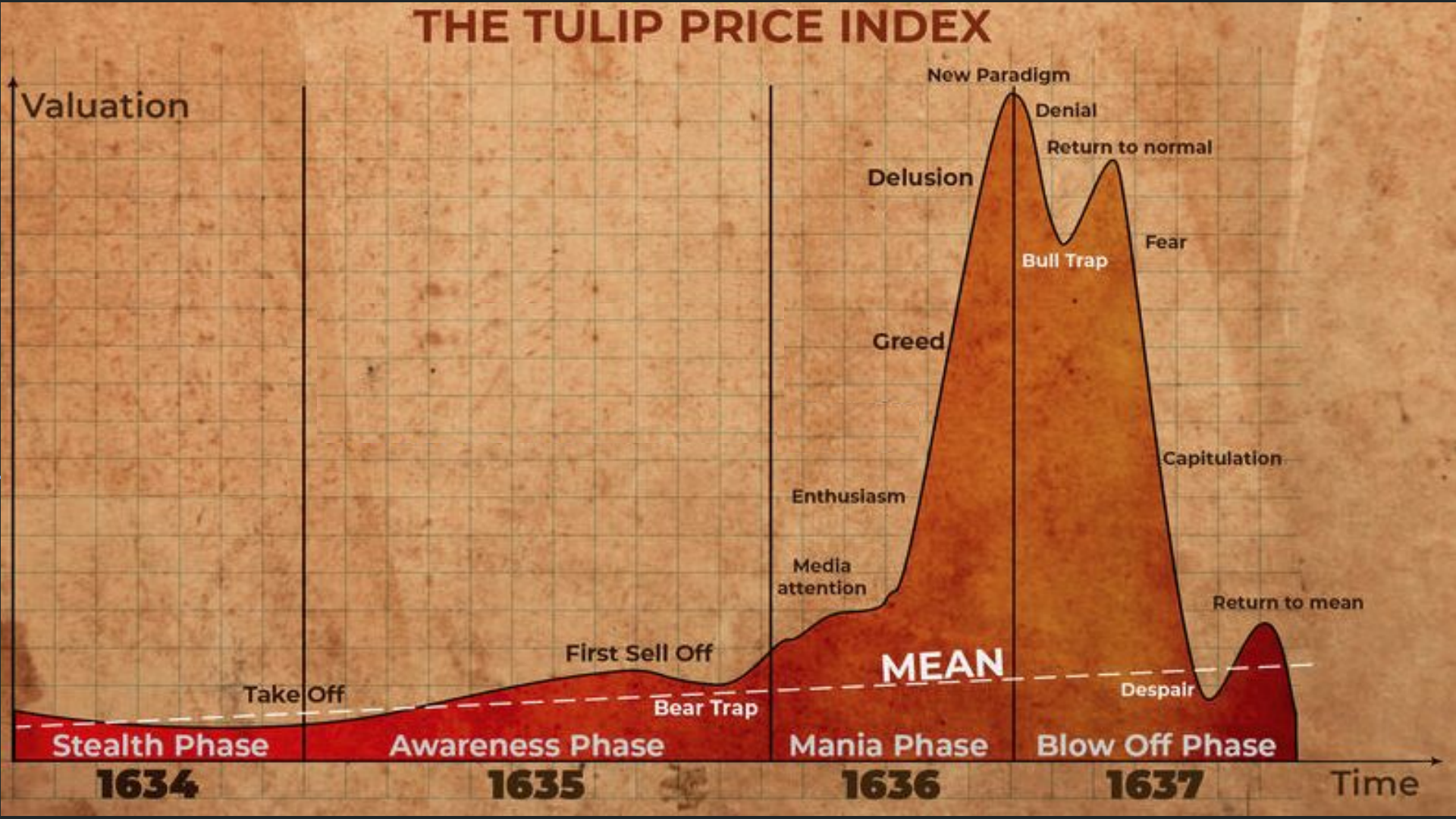


The first “wise” crowd

At a 1906 country fair, 800 people participated in a contest to estimate the weight of an ox.

Francis Galton observed that the median guess, 1,207 pounds, was accurate within 1% of the true weight.

THE TULIP PRICE INDEX



1. Systematic, large-scale study of the “wisdom of the crowds” phenomenon.

2. Examine the effect of social influence on crowd performance.

1. Systematic, large-scale study of the “wisdom of the crowds” phenomenon.

   Consistent evidence for the “wisdom of the crowd” effect at the domain-level.

2. Examine the effect of social influence on crowd performance.

1. Systematic, large-scale study of the “wisdom of the crowds” phenomenon.

   Consistent evidence for the “wisdom of the crowd” effect at the domain-level.

2. Examine the effect of social influence on crowd performance.

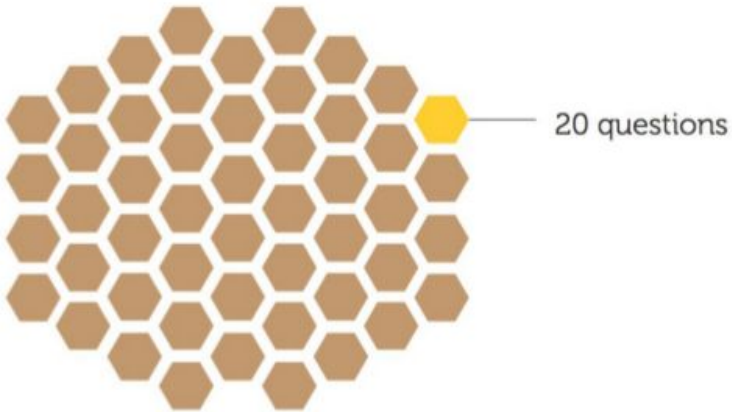
   Disclosing the consensus (e.g., most popular) answer undermines collective judgement.

Experiment design

Experiment design

1,000 questions

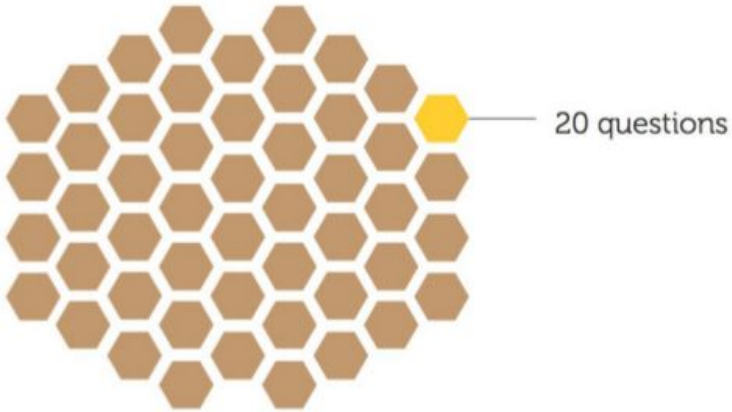
- 50 domains
- 20 questions per domain



Experiment design

1,000 questions

- 50 domains
- 20 questions per domain



4 types of media

4 social conditions

1. Most recent
2. Most confident
3. Consensus
   a. Most popular (discrete)
   b. Median (open-ended)
4. Control

Received over 100 responses per question and social condition, and 500,000 responses in total.

The Wisdom of Crowds Challenge

Let's start, Worker Bee!

How many calories does 1 apple contain?

24

Answer

How confident are you?

Clueless

Certain

Next Question

Which country does this land border correspond to?

Most recent responses

i. Spain
ii. Spain
iii. Tunisia

20

Answer

United Kingdom | Spain | Tunisia | Kazakhstan

Australia

How confident are you?

Clueless — Certain

Next Question

Assessing crowd performance

Assessing crowd performance

Rank the crowd accuracy against the accuracy of subjects in the control group.

vs.

Assessing crowd performance

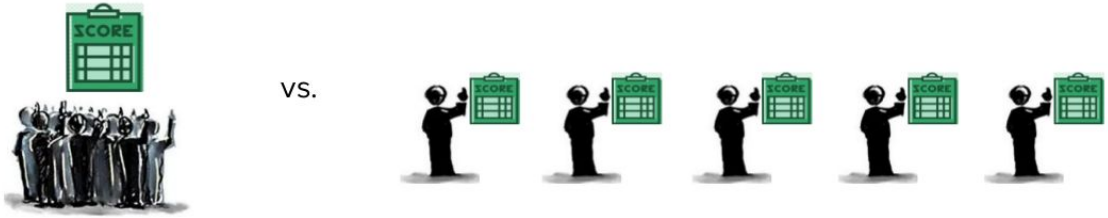
Rank the crowd accuracy against the accuracy of subjects in the control group.

vs.

	Crowd answer
Open-ended	Median answer
Discrete	Modal answer

Assessing crowd performance

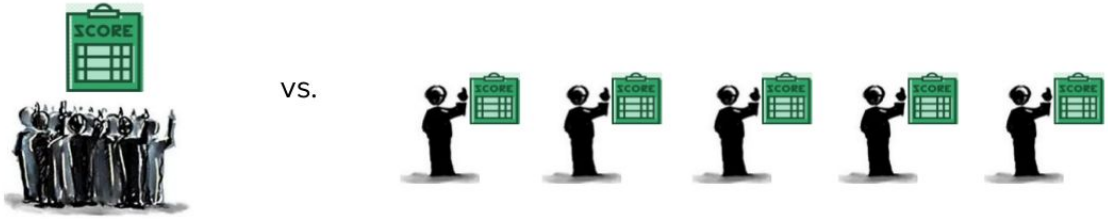
Rank the crowd accuracy against the accuracy of subjects in the control group.



	Crowd answer	Number of questions	Crowd percentile rank (question-level)
Open-ended	Median answer	400	66
Discrete	Modal answer	600	60

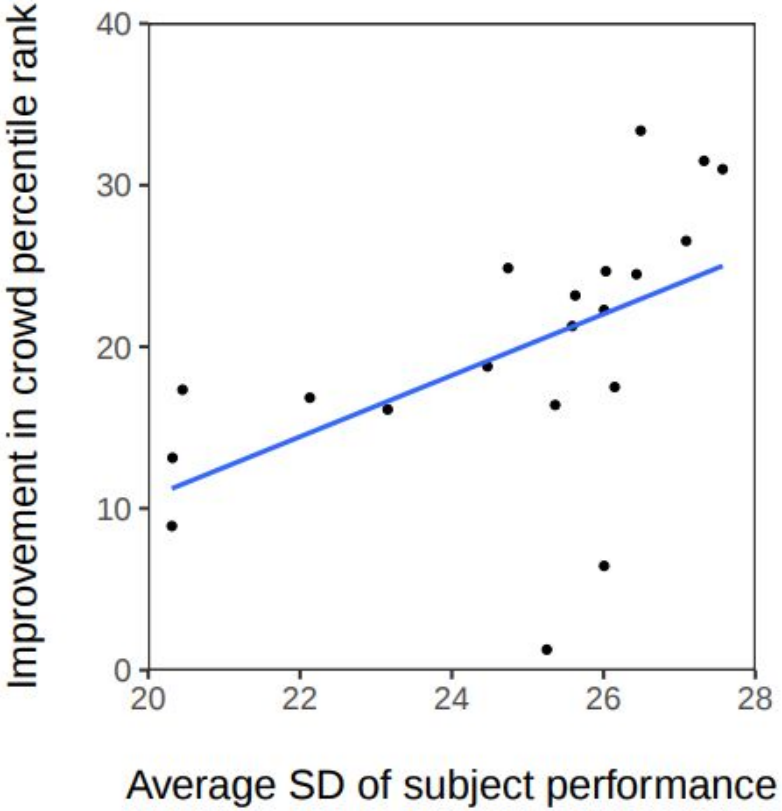
Assessing crowd performance

Rank the crowd accuracy against the accuracy of subjects in the control group.

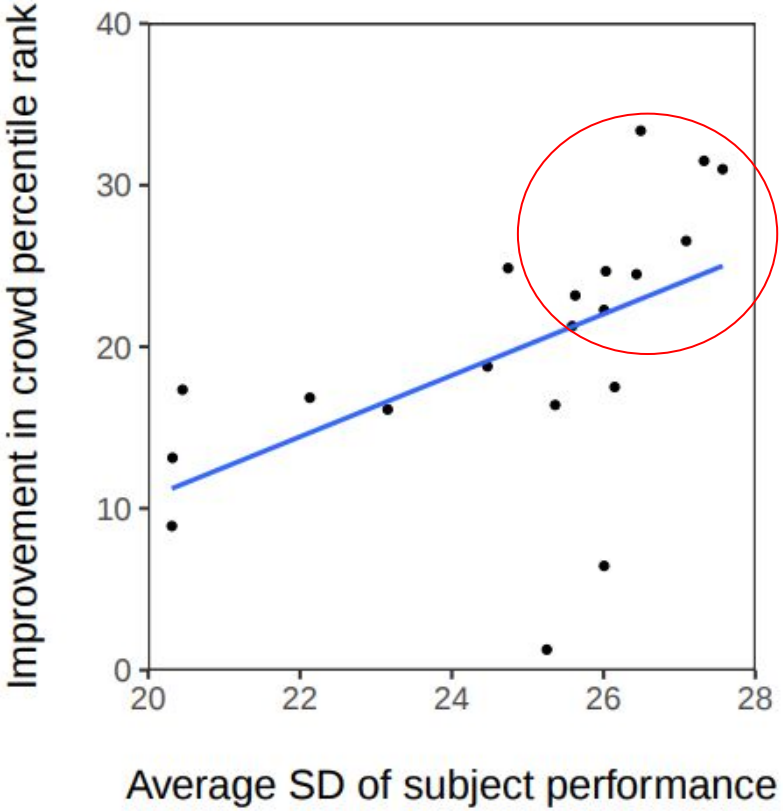


	Crowd answer	Number of questions	Crowd percentile rank (question-level)	Number of domains	Crowd percentile rank (domain-level)
Open-ended	Median answer	400	66	20	86
Discrete	Modal answer	600	60	30	87

Differentiation in expertise drives improvement

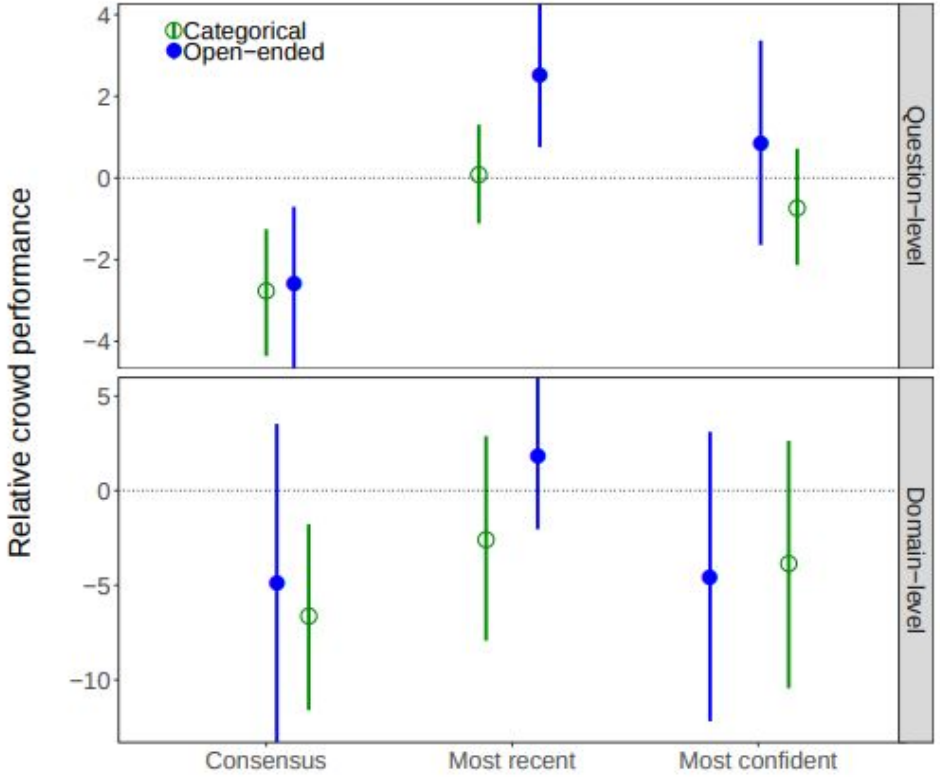


Differentiation in expertise drives improvement

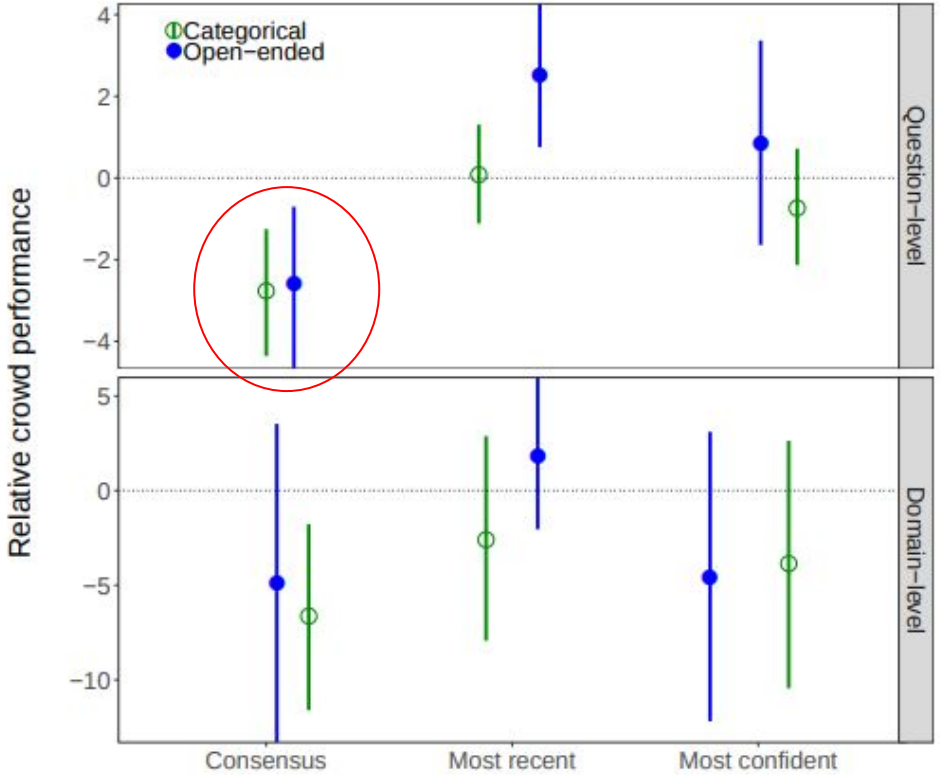


Effects of social influence

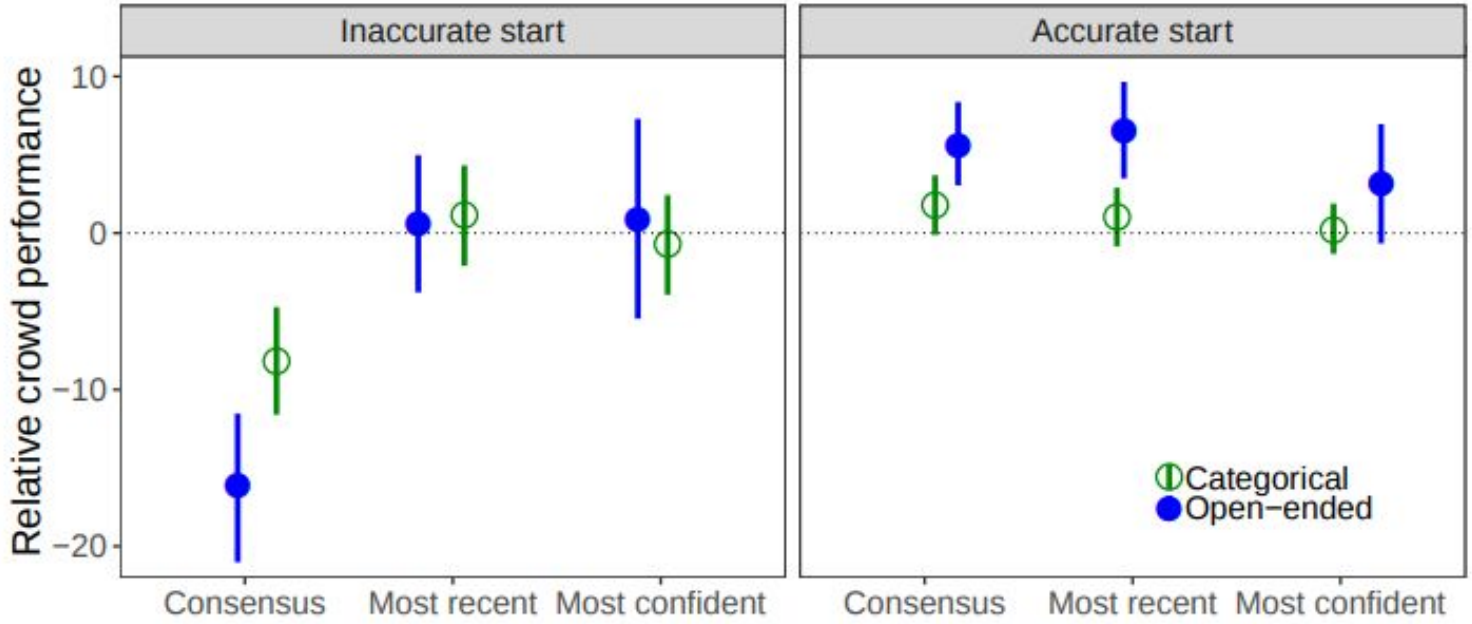
Consensus cues decrease crowd performance



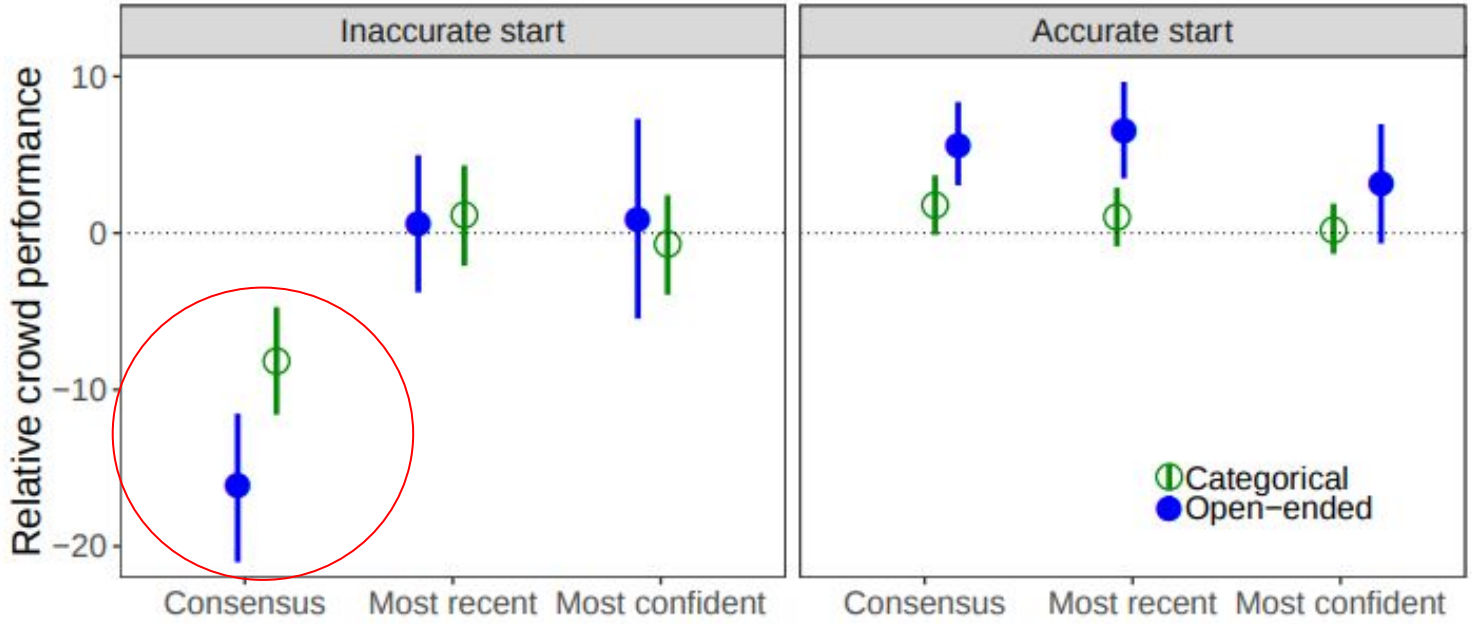
Consensus cues decrease crowd performance



Consensus crowd can’t recover from poor initial judgement



Consensus crowd can’t recover from poor initial judgement



Conclusions

Disclosing the consensus (e.g., most popular) answer undermines collective judgement.

The finding has important implications for the design of information aggregation mechanisms.

Questions?

Camelia Simoiu
csimoiu@stanford.edu

Data + Code
github.com/stanford-policylab/wisdom-of-crowds