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.
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?
Which country does this land border correspond to?

Answer
- United Kingdom
- Spain
- Tunisia
- Kazakhstan
- Australia

How confident are you?
- Clueless
- Somewhat
- Most likely
- Certain

Most recent responses:
1. Spain
2. Spain
3. Tunisia
Assessing crowd performance
Assessing crowd performance

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

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

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Crowd answer</td>
<td></td>
</tr>
<tr>
<td>Open-ended</td>
<td>Median answer</td>
</tr>
<tr>
<td>Discrete</td>
<td>Modal answer</td>
</tr>
</tbody>
</table>
Assessing crowd performance

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

<table>
<thead>
<tr>
<th>Question Type</th>
<th>Crowd Answer</th>
<th>Number of Questions</th>
<th>Crowd Percentile Rank (Question-Level)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open-ended</td>
<td>Median answer</td>
<td>400</td>
<td>66</td>
</tr>
<tr>
<td>Discrete</td>
<td>Modal answer</td>
<td>600</td>
<td>60</td>
</tr>
</tbody>
</table>
Assessing crowd performance

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

<table>
<thead>
<tr>
<th></th>
<th>Crowd answer</th>
<th>Number of questions</th>
<th>Crowd percentile rank (question-level)</th>
<th>Number of domains</th>
<th>Crowd percentile rank (domain-level)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open-ended</td>
<td>Median answer</td>
<td>400</td>
<td>66</td>
<td>20</td>
<td>86</td>
</tr>
<tr>
<td>Discrete</td>
<td>Modal answer</td>
<td>600</td>
<td>60</td>
<td>30</td>
<td>87</td>
</tr>
</tbody>
</table>
Differentiation in expertise drives improvement

![Graph showing the relationship between improvement in crowd percentile rank and average SD of subject performance.](image-url)
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