
Does Group Size Affect Problem Solving Performance in Crowds Working on a Hidden Profile Task?

Xiaoyun Huang

University of Maryland
College Park, MD 20740, USA
xyhuang@umd.edu

Yla Tausczik

University of Maryland
College Park, MD 20740, USA
ylatau@umd.edu

Abstract

Individuals and organizations increasingly use online platforms to broadcast difficult problems to crowds. According to the “wisdom of the crowd” because crowds are so large they are able to bring together many diverse experts, effectively pool distributed knowledge, and thus solve challenging problems. In this study we test whether crowds of increasing size, from 4 to 32 members, perform better on a classic psychology problem that requires pooling distributed facts.

Author Keywords

Crowdsourcing; problem solving; group performance; information exchange; hidden profile task.

ACM Classification Keywords

H.5.3 [Group and Organization Interfaces]: Organizational design

Introduction

Crowds discuss problems online in a variety of contexts. Individuals and organizations use online platforms to broadcast difficult problems to crowds. These online platforms specialize in a diversity of problems from R&D (InnoCentive) to medicine (CrowdMed) to data science (Kaggle) (e.g. [1]). Crowds solve user problems by answering questions on Q&A sites, such as Stack Overflow and Cross Validated

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

Copyright held by the owner/author(s).

CHI'18 Extended Abstracts, April 21–26, 2018, Montreal, QC, Canada

ACM 978-1-4503-5621-3/18/04.

<https://doi.org/10.1145/3170427.3188669>

(e.g. [7]). Crowds engage in the same cognitive processes as problem solving when crowds collectively make sense of information in online communities, such as when discussing a news article or understanding treatment options for a chronic health problem (e.g. [3]). Crowdsourcing applications use crowds to make judgements and make predictions about the future (e.g. [4]).

To solve problems crowds engage in collective sensemaking. They share relevant knowledge, discuss and debate this information, and draw a conclusion. “Wisdom of the crowd” supposes that crowds are good at these activities because crowds draw on the knowledge of a large, diverse set of individuals [6].

While researchers hypothesize that crowds are good at solving problems because they pool distributed information from many individuals, a large body of literature on small groups in social psychology has shown that small groups display systematic biases in how they share information and that these biases impede solving some types of problems [8]. In the current study we investigate whether crowds up to a size of 32 can effectively solve a classic psychology problem that stymies small groups and how the size of the crowd affects performance on this problem.

Thirty years of research on the hidden profile task in social psychology has shown that small groups do not solve problems effectively when the problem requires the group to pool distributed knowledge [8]. Small groups exhibit an information exchange bias in which they tend to share and discuss information that everyone in the group is privy too, while neglecting to discuss information that is known by only one person in the group [5]. As a result small groups perform badly on hidden profile tasks. A hidden profile task is a classic problem in psychology in which multiple facts needed to make a decision are distributed among members

of a group. Some facts are given to everyone (or most people) and others are given to only one person. The problem is designed so that the facts given to only one person must be discussed in order to solve the problem correctly.

In the current study we investigate whether crowds are able to solve the hidden profile task correctly. Researchers argue that crowds are effective at pooling distributed knowledge (e.g. [1, 6, 7]). However, crowds may suffer from the same information exchange bias that afflicts small groups. Limited evidence from a meta-analysis that investigates mainly groups of size 3 to 4 and a few larger groups of size 6 and 10, suggests that the information exchange bias associated with the hidden profile task as well as performance on the problem gets *worse* as group size increases [2]. To investigate this issue we conducted an experiment in which we manipulated group size in crowds to answer the following research question:

Research Question 1: How does group size affect solution quality on the hidden profile task?

Theoretically, one of the reasons larger groups are believed to do worse than smaller groups on the hidden profile task is that as group size increases social loafing increases [2]. Social loafing is the tendency of individuals in larger groups not to contribute because they believe others in the group will do the work. If social loafing is high, then group members with critical facts may not share their facts because they do not participate in the discussion. We investigated whether social loafing affected crowds of different sizes by answering the following research question:

Research Question 2: How does group size affect participation in the crowd discussion?

Methods

Participants

We recruited 836 participants from Amazon's Mechanical Turk (AMT). 53.1% of participants were female; they ranged in age from 18 to 78 ($M = 34.7$, $SD=10.6$). Most participants came from the United States (86.2%) and India (11.2%). Participants were randomly assigned to a group; in total there were 55 groups¹

Design

The main objective of this study was to understand how crowds of different sizes would perform on a hidden profile task. We experimentally manipulated group size, assigning individuals to groups of size 4, 8, 16, or 32. This was a between subjects design. We choose 4 as the smallest group size because it is a typical group size used in prior studies of traditional small groups [2]. We choose 32 as the largest group size because it is comparable to group sizes found in the wild, such as those on Q&A platforms and in crowdsourcing applications.

Task

We designed our hidden profile task based on tasks used in prior studies, choosing elements of prior studies which were most suitable to work on a crowdsourcing platform like AMT. Prior studies using the hidden profile task have used a number of different tasks, such as solving a murder mystery, choosing the best job candidate, or selecting the right information technology. We designed our task as a job candidate selection task. Participants were asked to identify the best candidate for a job from a set of four candidates.

We developed a set of facts to describe the strengths and weaknesses of the four candidates. We focused on soft

skills, since many of these are important regardless of the specifics of the job. To ensure that the soft skills we used for the task were seen as equally important, we conducted a pre-experiment survey on AMT in which we had individuals rate the importance of a list of soft skills. We then selected two sets of 5 soft skills that were rated as being important and which had similar scores to each other. We developed a set of materials which consisted of 16 facts, each fact reported a strength (e.g. "is good at thinking creatively"), a weakness (e.g. "is a bad team player"), or a generally neutral statement (e.g. "is average at learning new skills") about a candidate (see Table 1). To ensure the exact choice of materials didn't bias the results we varied the skills used for the facts.

In hidden profile tasks information is distributed among group members. In a typical small group study with 3-4 members some facts are given to everyone, while other facts are given to only one member. Facts given to everyone function as common knowledge, while facts given to only one member function as rare knowledge. In prior work information is treated differently depending on how many people have access to it (i.e. is it common or rare). We generalized this to a variety of sized groups by operationalizing common facts as those that are known by a majority of group members and rare facts as those that are known by a minority of group members. More specifically in this study we created 8 common facts which were distributed to 75% of members in each group and 8 rare facts which were distributed to 25% of members in each group (Table 2).

Each participant was given a set of 8 facts, which we call an information profile. Each information profile consisted of 6 common facts and 2 rare facts. Information profiles were designed so that regardless of the size of the group (e.g. 4 or 32) each person in the group had a different information

Soft Skills	Candidates			
	A	B	C	D
learn new skills	+	-	0	0
team player	+			+
listening skills	-	-	-	+
creative thinking	0	0	+	+
work ethic			+	+

Table 1: Example distribution of soft skills for four job candidates.

Fact Type	Group Size				
	4	8	16	32	%
Common	3	6	12	24	75%
Rare	1	2	4	8	25%

Table 2: The number of group members given each type of fact for the four group size conditions.

¹We excluded groups that did not fill up because there were too few active AMT workers at that time.

Profile	Common Facts	Rare Facts	Preference
1	$A_1+, B_1-, B_2-, B_0, C_1+, C_2+$	C_0, D_1+	C
2	$A_1+, A_2+, B_1-, B_2-, C_2+, D_0$	A_0, D_2+	A
3	$A_2+, B_1-, B_0, C_1+, C_2+, D_0$	$A-, D_3+$	C
4	$A_1+, A_2+, B_2-, B_0, C_1+, D_0$	$C-, D_4+$	A

Ideal Candidate: D

Table 3: The distribution of common and rare facts among group members in groups of size 4. Larger groups had an analogous distribution with a unique information profile for each group member.

profile. These information profiles were specially crafted so that on their own they would favor a non-ideal candidate. Prior work shows that participants are more likely to share information when there is conflict in the group, so we designed these information profiles so that half the participants would favor one non-ideal candidate while the other half would favor another non-ideal candidate (Table 3).

By design, based on the facts in the information profile alone, no individual would favor the ideal candidate who had the most strengths. The strengths of the ideal candidate were instead given as rare facts which needed to be pooled in order to identify the ideal candidate.

Procedure

Participants were recruited with a HIT placed on AMT. After accepting the HIT and being given informed consent, participants proceeded to the main task. Participants were placed in an active group, which had randomly assigned to be of a certain size depending on the condition. Within a group each participant was randomly assigned to a unique information profile. Participants were not told that they had different information profiles than other group members.

Participants assigned to the same group were directed to the same discussion forum. In each forum, the participant could see the general instructions, a private list of facts that made up his/her own information profile, and a threaded list of messages between group members. In the forum participants could post messages and read messages sent by others. The discussion forum allowed threaded comments and voting on comments, similar to discussion forums on other platforms (e.g. Reddit, HackerNews, Coursera). Comments were displayed in the order they were received, with child comments nested below parent comments. Facts had to be shared through discussion, no other tools were provided for sharing information. While working on the task participants could enter and exit the forum as they pleased, all group messages were displayed regardless of whether an individual was present. Once, a participant decided that he/she wanted to complete the task he/she could exit the discussion forum and proceed to the questionnaire.

The questionnaire asked participants to report their final answers to the main task, group satisfaction, and general demographics (e.g. age, sex). Participants were paid \$1.50 for the main task. To incentivize group work participants were also paid a bonus up to \$1.00. A bonus was given if an individual submitted the correct solution; it was given in proportion to the number of other group members who provided the correct solution. Participants were only allowed to participate in the task once.

Results & Discussion

RQ1: How does group size affect solution quality on the hidden profile task?

While researchers hypothesize that crowds are good at pooling distributed knowledge, prior work in psychology on the hidden profile task suggests that groups are bad at pooling distributed knowledge and that groups get worse

Comparison	<i>p</i>
4 8	0.001
4 16	0.485
4 32	0.139
8 16	0.037
8 32	0.194
16 32	0.862

Table 4: Tukey post-hoc tests for pairwise comparisons between the four conditions on performance.

as they increase in size. To address this issue, we tested performance on a hidden profile task, which can only be solved by pooling knowledge with different sized crowds on AMT.

For each crowd we calculated the proportion of group members who reported the correct solution to the problem. We then performed one-way ANOVA to test whether group size affected performance. We found a significant relationship between group size and performance (Tables 5,6; Figure 1). Pairwise comparisons showed that groups of size 8 performed significantly better than groups of size 4 and size 16 (Table 4). We observed a trend in which groups of size 16 and 32 performed better than groups of size 4. However, these differences were not statistically significant.

In conclusion, we find a non-linear relationship between group size and performance on the hidden profile task, in which medium sized groups performed the best. However, even in medium sized groups on average only 40% of group members reported the correct solution. These results suggest that crowds also have difficulty with the hidden profile task, presumably because they don't pool knowledge effectively and that this difficulty is not alleviated by increasing group size.

RQ2: How does group size affect participation in the crowd discussion?

One reason that groups might perform worse on the hidden profile task as they increase in size is that larger groups may suffer more from social loafing. To investigate the effect of social loafing on the different sized groups we calculated the percentage of group members who communicated using the online forum. A one-way ANOVA showed that there was no significant relationship between group size and the rate of participation (Tables 5, 6). Regardless of condition around 90% of group members sent at least one

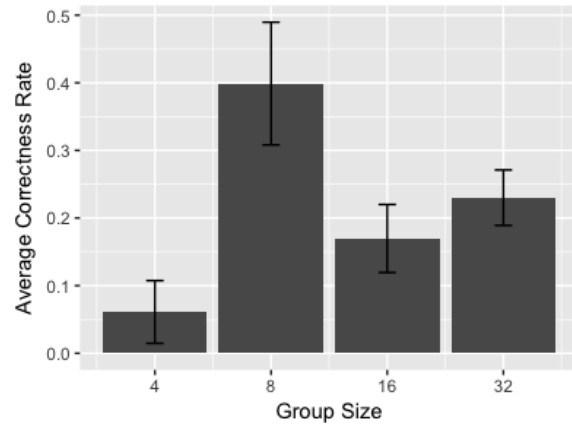


Figure 1: Average performance (SE) across the four experimental conditions.

Outcome	Group Size	Mean	SD
Correctness Rate	4	0.06	0.18
	8	0.40	0.30
	16	0.17	0.19
	32	0.23	0.15
Participation Rate	4	0.90	0.16
	8	0.92	0.08
	16	0.91	0.08
	32	0.88	0.05

Table 5: Descriptive statistics.

Outcome	ANOVA
Correctness Rate	$F(3,51) = 5.81, p < 0.002, \eta^2 = 0.25$
Participation Rate	$F(3,51) = 0.34, p = 0.80, \eta^2 = 0.02$

Table 6: Analysis of Variance for the two outcome measures.

message in the forum. We find no evidence of social loafing in larger groups. In fact, these results suggest groups of all sizes have high rates of participation. Further research is needed to understand why we observed differences in performance across groups of different sizes.

Conclusion

We find that crowds like small groups also struggle to solve a hidden profile task. Presumably crowds like small groups have difficulty pooling distributed knowledge during this task. These findings are in contrast to researchers' arguments that crowds can solve problems well because they can pool knowledge from a large, diverse set of individuals (e.g. [7]). One explanation for these conflicting accounts, is that they may reference two very different types of problems. In the hidden profile task, distributed knowledge is not shared because it does not support initial beliefs about the solution, whereas typically we assume crowds share distributed knowledge before they form beliefs about the solutions. More research is needed to understand how crowds engage in pooling of knowledge, collective sensemaking, and problem solving.

Acknowledgements

This work was funded by the National Science Foundation (CRII:1657308).

REFERENCES

1. Lars Bo Jeppesen and Karim R Lakhani. 2010. Marginality and problem-solving effectiveness in broadcast search. *Organization Science* 21, 5 (2010), 1016–1033.
2. Li Lu, Y Connie Yuan, and Poppy Lauretta McLeod. 2012. Twenty-five years of hidden profiles in group decision making: A meta-analysis. *Personality and Social Psychology Review* 16, 1 (2012), 54–75.
3. Lena Mamykina, Drashko Nakikj, and Noemie Elhadad. 2015. Collective sensemaking in online health forums. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 3217–3226.
4. Barbara Mellers, Lyle Ungar, Jonathan Baron, Jaime Ramos, Burcu Gurcay, Katrina Fincher, Sydney E Scott, Don Moore, Pavel Atanasov, Samuel A Swift, and others. 2014. Psychological strategies for winning a geopolitical forecasting tournament. *Psychological Science* 25, 5 (2014), 1106–1115.
5. Garold Stasser and William Titus. 1985. Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of Personality and Social Psychology* 48, 6 (1985), 1467.
6. James Surowiecki. 2005. *The Wisdom of Crowds*. Anchor.
7. Yla R Tausczik, Ping Wang, and Joohee Choi. 2017. Which Size Matters? Effects of Crowd Size on Solution Quality in Big Data Q&A Communities. In *ICWSM*. 260–269.
8. Gwen M Wittenbaum, Andrea B Hollingshead, and Isabel C Botero. 2004. From cooperative to motivated information sharing in groups: Moving beyond the hidden profile paradigm. *Communication Monographs* 71, 3 (2004), 286–310.