

# Don't Hide in the Crowd! Increasing Social Transparency Between Peer Workers Improves Crowdsourcing Outcomes

Shih-Wen Huang and Wai-Tat Fu  
University of Illinois at Urbana-Champaign  
Department of Computer Science  
Urbana, IL 61801  
{shuang51,wfu}@illinois.edu

## ABSTRACT

This paper studied how social transparency and different peer-dependent reward schemes (i.e., individual, teamwork, and competition) affect the outcomes of crowdsourcing. The results showed that when social transparency was increased by asking otherwise anonymous workers to share their demographic information (e.g., name, nationality) to the paired worker, they performed significantly better. A more detailed analysis showed that in a teamwork reward scheme, in which the reward of the paired workers depended only on the collective outcomes, increasing social transparency could offset effects of social loafing by making them more accountable to their teammates. In a competition reward scheme, in which workers competed against each other and the reward depended on how much they outperformed their opponent, increasing social transparency could augment effects of social facilitation by providing more incentives for them to outperform their opponent. The results suggested that a careful combination of methods that increase social transparency and different reward schemes can significantly improve crowdsourcing outcomes.

## Author Keywords

Crowdsourcing; human computation; social transparency; social facilitation; social loafing

## ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

## INTRODUCTION

Crowdsourcing has been proven as an effective way to solve various kinds of problems [8]. One of the most notable examples is the ESP game [26], which recruits people to generate image labels while playing an online game. By 2008, this game had recruited 200,000 players and collected more than 50 million labels [27]. In addition, individuals can also

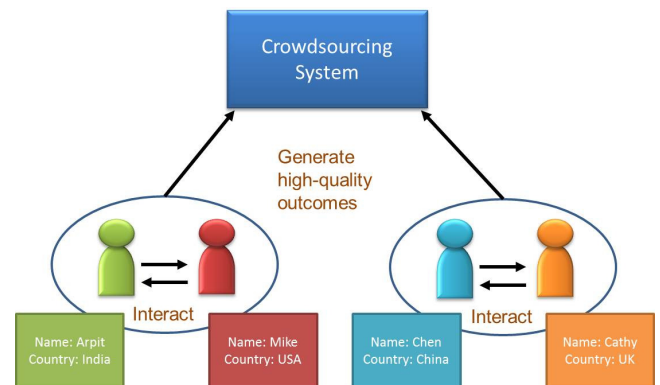


Figure 1. Sharing demographic information between paired workers allows the crowdsourcing system to collect outcomes with higher quality.

easily recruit online workers from Amazon Mechanical Turk (AMT)<sup>1</sup> to solve problems that are difficult for digital computers (e.g., collecting labeled data for natural language processing [23] and relevance evaluation [1]) at a very low cost. These examples have merely begun to demonstrate the potential of crowdsourcing as a social computing technique that can be applied in a wide range of situations.

However, quality control is still one of the biggest issues for crowdsourcing<sup>2</sup> [11]. For example, although the ESP game can successfully recruit players to create a large amount of image labels, a study [22] showed that many of these are low-quality labels that can be generated by robots with little real world knowledge. Moreover, the reliability of workers recruited from AMT is also questionable. People have found that there are many spammers or even robots in AMT, which greatly reduce the quality of the outputs collected from this platform [8]. As a result, designing mechanisms that can ensure the quality of crowdsourcing outcomes is crucial to its success.

<sup>1</sup><https://www.mturk.com/mturk/welcome>

<sup>2</sup>Crowdsourcing and human computation are two highly overlapped fields. Though the foci of these two fields are slightly different, researchers sometimes use these two terms interchangeably. We use the term crowdsourcing throughout because the findings of our study can be applied to diverse applications; however, research in human computation is also highly related to this study. For more discussion about the similarities and differences between crowdsourcing and human computation, please see Quinn *et al.* [21].

### Increasing social transparency as a possible way to enhance quality of collaborative work

The term social transparency has been well studied. For example, the survey in Stuart *et al.* [24] provided a framework that defines various kinds of social transparency (i.e., identity transparency, content transparency, and action transparency) in networked information exchange. The kind of social transparency discussed in this paper is similar to the identity transparency according to their definition. Specifically, Stuart *et al.* [24] suggested that identity transparency (e.g., real name, persistent identity in applications) can make online information more useful and better evaluated. For instance, a study [19] on an online suburban community network showed that using real name not only created accountable online information exchange but also helped promote local commerce in the community.

The idea of increasing social transparency to enhance accountability in collaborations was first introduced by Erickson and Kellog [5]. They argued that in the physical world, people make everyday decisions based on various “social cues.” Some of these social cues allow people to perceive the extent to which their behavior is acceptable for others (e.g., smiles, frowning, etc), and people often learn to adjust their behavior based on the perception of these social cues. In addition, the more that two persons know each other, the more likely people will be influenced by the social cues. In other words, the availability of social cues often increases the perception that a person is accountable for his or her actions. However, these social cues are often missing when people collaborate with each other remotely through a digital system. To address this problem, they implemented a social translucent system that made social cues salient to the collaborators using the system. They showed that a social translucent system can make the users more aware of the presence of others, and as a result, the users can know that they are accountable for their actions.

Kittur *et al.* [15] studied how social transparency help Wikipedia readers adjust their trust toward the information. They designed WikiDashboard [25], a visualization tool that helps users identify the page that was edited by which editors and when the edits were made. Their study [15] showed that manipulating the editor-accountable visualization significantly affected the trust of the readers. Therefore, they argued that their visualization can improve the trustworthiness of Wikipedia articles. They also mentioned it is possible that this social transparency created by their visualization tool can encourage writers to be more responsible [25]. As a result, the quality of the articles also improves.

In addition, research in social identity theory [20] suggests that identity formation process (e.g., information sharing) makes group members more commit to their group.

These studies have demonstrated that social transparency can effectively improve online collaborative work. Because crowdsourcing also utilizes the power of multiple workers to complete a collective task, it is reasonable for us to hypothesize that increasing social transparency (either between colleagues or between requesters and workers) in crowdsourc-

ing can also enhance outcome quality. To test this hypothesis, we designed a study to test how increasing social transparency would affect the quality of outcomes generated by crowd workers. We are also interested in how social transparency may interact with different reward schemes to influence outcomes. Before we discuss the details of the study, we will first review how different reward schemes can affect the performance of crowd workers below.

### Three possible effects of peer-dependent reward schemes on motivating crowd workers

Micro-tasks in crowdsourcing typically are trivial and tedious to workers. Therefore, one notable reason that crowd workers fail to generate high-quality work is the lack of motivation [21]. Therefore, it is important for requesters to design a crowdsourcing system that can offer enough incentives to workers [16]. The most common incentive used to recruit workers in crowdsourcing is monetary reward. Nevertheless, previous research on financial incentives [18] in crowdsourcing suggested that increasing monetary reward seemed to augment the amount of work produced but not the quality. This finding implies that monetary reward alone might not be enough to motivate workers to generate high-quality work. Interestingly, [10] found that making the reward of the workers codependent on each other can significantly increase the quality of their work. Specifically, the crowdsourcing system created by them used the answers of its workers to evaluate future workers of the system. Every worker in the system knew that his/her answer would be used to evaluate the answer of future workers. Therefore, since the chance that two workers generated the same wrong answer was small, if a worker produced a wrong answer, the workers that were evaluated by this erroneous answer could not earn rewards regardless of whether or not they generated the correct answer. Their study showed that workers put more effort in solving the task to prevent other workers from being penalized unjustly. This study suggests that it is possible to design other peer-dependent reward schemes to enhance the performance of crowd workers.

However, not all peer-dependent reward schemes can motivate workers to perform better. Previous research has shown that poor designs of peer-dependent reward schemes in a collaborative work environment could even lead to worse outcomes [12]. We summarized three possible effects of peer-dependent reward schemes between crowd workers on their motivations below:

- **Social loafing:** Previous studies have shown that when individuals work in a group and only the outcome of the group is evaluated, individual performance becomes significantly worse than when they work alone [12]. For example, an experiment showed that when subjects conducted a rope pulling task, individual effort decreased as the group size increased. Harkins [7] argued that because it is difficult to evaluate each individual in the group when working on a collective task, people feel that they can *hide in the crowd* such that there is a diffusion of responsibility from the individuals to the group. As a result, their effort levels tend to drop. Kerr *et al.* [13] argued that people exert less

effort when working in groups because they may feel that their inputs are not essential to the collective outputs of the group.

Since crowdsourcing is essentially a large collective task, this phenomenon of *social loafing* can decrease the motivation of workers to do well. Therefore, it is important for researchers to study how social loafing can impact the performance of crowd workers and design mechanisms to mitigate its potential negative effects.

- **Altruistic motives:** In contrast to social loafing, the existence of *altruistic motives* is believed to lead to better performance when workers are encouraged to work collaboratively [10]. Altruistic motives refers to the ideas that workers have the tendency to behave so as to benefit their colleagues, even if the behavior entails a negative utility for themselves. Studies [2, 17] have shown that workers will adjust their effort level when the rewards of other workers depend on them. Previous research [10] also showed that when workers in a crowdsourcing system know that their answers will be used as evaluation standards by the system, they work harder to prevent other crowd workers from being unjustly penalized due to their errors.
- **Social facilitation:** The other possible social effect that encourages workers to generate better crowdsourcing outcomes is *social facilitation*. Social facilitation is an effect occurs when that individuals are motivated to perform better when they know they can be compared to others<sup>3</sup>. A study [7] showed that when conducting a brainstorming task, subjects who were paired with another subject and knew their individual work would be evaluated generated significantly more ideas than subjects who worked alone. Harkins [7] suggested that social facilitation can be conceived as the complement of social loafing. That is, when the outcomes of each individual in the group are evaluable, the workers are motivated to outperform other members in the group. As a result, the collective outputs of the entire group are also better. Study [17] also highlighted that worker effort is positively related to the performance of workers who see the work being executed. Similarly, social comparison theory [6] suggests that people have the tendency to compare their own work to others. These results suggest that social facilitation can effectively enhance individual performance in collaborative work.

To summarize, previous research has shown that a set of complex social effects may interact to influence the motivation of the workers. It is not clear, however, how directly manipulating social transparency and reward schemes may make each of these social effects more or less salient, such that a designer of a crowdsourcing system can more reliably predict how these effects can be utilized to improve outcomes. To this end, we design our experiments by focusing on two common reward schemes: a teamwork scheme and a competition

scheme. In a teamwork scheme, in which workers in a group are rewarded based on the collective outcomes of the group, there are two possible effects that may influence the motivation of the workers. First, social loafing can negatively impact the performance of workers because they feel that they can hide in the crowd and rely on their partner's effort. If this is the case, we should find that social loafing is more salient when a worker perceives that their partner is already at a high level of performance. In such situation, the worker may not be motivated to put extra effort in the task as the collective outcome is likely to be good. In contrast, a second factor, altruistic motives, can positively impact the performance of workers in the same reward scheme because a worker may not want their partner to feel that his or her poor performance may impact their collective reward. If this is the case, we should find that altruistic motives may be more salient when social transparency is high, as the perception of the presence of their partner may enhance the effect of altruistic motives. In a competition scheme, in which rewards of workers depend on the differences between the outcomes of the workers and their colleagues, it is possible that social facilitation can improve their performance because they are motivated to outperform each other. If this is the case, we should find that pairing with a worker with a high level of performance should motivate a worker to work even harder.

To investigate how different peer-dependent reward schemes improve or degrade the performance of crowd workers, we created three conditions that differ in the reward schemes. Specifically, the way the reward of the workers in our experiment depend on 1) *only the worker's performance*, 2) *the collective outcomes of paired workers*, or 3) *the difference between the performance of paired workers*.

## EXPERIMENTAL DESIGN

We used a 3x2 design to see how different peer-dependent reward schemes and social transparency levels affect the performance of crowd workers. 300 subjects were recruited from AMT. We asked the workers to perform tasks individually (**I**), paired with another worker as a team (**T**), and competitively against another paired worker (**C**). Half of the workers worked anonymously (**A**), and the other half of them were asked to report their demographic information (i.e., name, gender, age, and nationality) (**D**). When the workers who revealed their demographic information were paired with another worker, their demographic information was shared with their paired workers. We also examined if there was any interaction effect between social transparency levels and reward schemes.

### Noun counting task

The workers in our experiment were asked to count the words with a particular part-of-speech (noun) in a list of 30 words. Each word in the word list had equal probabilities to be a noun or not a noun. There were two advantages of this task that made it a very good test-bed for us to examine how the factors affected the quality of the outcomes. First, it had a ground truth (objectively correct answer) to serve as the gold standard, making it feasible to measure the quality of an answer by observing the difference between the answer and the

<sup>3</sup>There was a controversy in the literature whether social facilitation includes solely the mere presence of others [28] or should also include comparisons between peers [7]. In this paper, we adopted Harkins' [7] definition of social facilitation, which includes the effect of peer comparison.

gold standard. Moreover, the workers needed to put in a certain level of effort to solve the task, and to a large extent, outcome quality increases with the level of effort. This task is also representative of the tasks used in previous crowdsourcing systems (e.g., [14, 23]).

To evaluate the performance of the workers, we first computed the label accuracy of each collected label. The label accuracy shows how close the label is to the gold standard answer. Therefore, the formula for label accuracy is:

$$\text{Label Accuracy} = 1 - \frac{|N_{GS} - N_{worker}|}{\text{Max}(30 - N_{GS}, N_{GS})}$$

where  $N_{GS}$  is the correct number of nouns in the list and  $N_{worker}$  is the number of nouns in the list reported by the worker. Therefore, the numerator of the second term in the formula is the distance between the label generated by the worker and the gold standard answer, and the denominator is the distance between the worst possible answer and the gold standard answer. (If  $N_{GS} > 15$ , the worst possible answer is 0; otherwise, it is 30.) Therefore, the label accuracy is a number between 1 and 0, where 1 represents that the label is exactly the same as the gold standard, and 0 means that the answer is the worst possible answer. Then, the average label accuracy of each worker was computed by averaging the label accuracy of the ten labels generated by the worker. We used the average label accuracy of the workers to measure their work quality.

#### Variable 1: Peer-dependent reward schemes

In our experiment, we compared the performance of workers under three different peer-dependent reward schemes:

- **Individual (I):** This system was used to imitate traditional crowdsourcing systems, where the bonus of workers only depends on their own performance. Therefore, in this system, worker performed the task individually, and their bonus did not depend on other peer workers.
- **Teamwork (T):** The workers within this reward scheme were paired with another worker as a team. Their bonus was decided by the average of the performance of both workers. In other words, the workers on a team were codependent on each other because only the collective outcomes were evaluated. This allowed us to examine whether social loafing (negative) or altruistic motives (positive) really created an effect on worker performance in a team environment.
- **Competition (C):** This reward scheme aims to test the effect of social facilitation. The workers under competition reward scheme were also paired with another worker. However, unlike the previous condition, the paired workers were opponents. The bonus of a worker was decided by the positive difference between the worker and his/her opponent.<sup>4</sup> Therefore, the workers would have the incentives to outperform their opponents in order to earn more monetary reward.

<sup>4</sup>The worker could not earn bonus if his/her performance was worse than the opponent.

#### Variable 2: Social transparency levels

In our experiment, we also controlled the levels of social transparency of the workers. The two levels of social transparency used in our experiment were the following:

- **Anonymous (A):** This was the lowest level of social transparency; the only information shared by workers and the system was the worker ID in AMT. If the system paired the workers, the paired workers cannot get any information from each other.
- **Demographic information revealed (D):** Under this level of social transparency, the workers were asked to reveal their demographic information, which included their name, gender, age, and nationality. When the workers worked with another worker, this information was also shared with their teammate or opponent.

#### Subject recruitment on Amazon Mechanical Turk

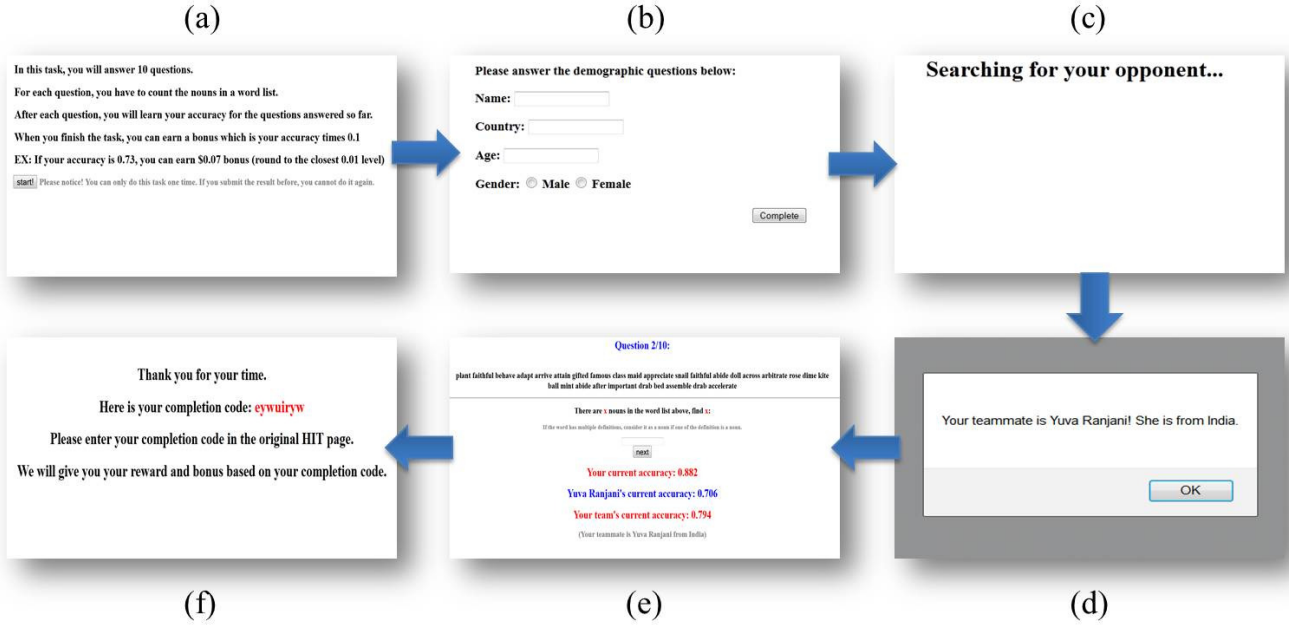
We published our HITs on AMT from 8/25/2012 to 9/4/2012. The title and the description of these HITs were both “Count the nouns in a word list”. We did not require our workers to pass any qualification tests. The price tags for the HITs were \$0.05, which were the rewards the workers could earn regardless of their performance. On the HIT page, we provided a link to route the workers to our experiment website and a text area for the workers to enter an eight-digit completion code. When the workers completed their jobs, our website showed them the total bonus they could earn for their performance and the unique completion code that they would have to enter on the original HIT page on AMT. We then paid the workers both of their rewards (performance independent) and bonuses (performance codependent) based on the completion codes they entered. The system also automatically rejected repeated workers by checking if their AMT worker ID already existed in the database.

#### Experimental interface

When the workers visited the experiment website, they were randomly assigned to one of the 6 conditions (**IA**, **TA**, **CA**, **ID**, **TD**, or **CD**). On the introduction page (Figure 2.a), the system told the workers that they had to answer 10 questions about counting the nouns in a word list<sup>5</sup>. It also showed how the bonuses for the workers were decided; the exact words we used for each condition in the introduction were as follows:

- **IA, ID:** When you finish the task, you can earn a bonus which is your accuracy times 0.1. EX: If your accuracy is 0.73, you can earn \$0.07 bonus (rounded to the closest 0.01 level)
- **TA, TD:** You will be paired with another worker in the system as a team. Your bonus will be decided by the average of your accuracy times 0.1. EX: If your teammate’s accuracy is 0.93 and your accuracy is 0.53. The average

<sup>5</sup>We used the same 10 word lists for all workers. One possible concern is that some of the workers released the information to others. However, we found that the correlation between the performance and the arrival time of the worker was very low ( $r = 0.04$ ), which shows that the workers should not get extra information because the average accuracy didn’t go up.



**Figure 2. Experimental interfaces used in our experiment: (a) Introduction page (b) Demographic information collection (c) Asynchronous interaction design (d) Information sharing between workers (e) Noun counting page (f) Completion page**

accuracy of your team is 0.73. Therefore, you and your teammate can get \$0.07 bonus each (rounded to the closest 0.01 level)

- **CA, CD:** Another worker will be chosen to be your opponent. Your bonus will be the positive difference between your accuracy and your opponent's times 0.5. EX: If your opponent's accuracy is 0.79 and yours is 0.93, your bonus will be  $0.14 \times 0.5 = 0.07$  (rounded to the closest 0.01 level). However, if your performance is worse than your opponent, you won't earn the bonus.<sup>6</sup>

After workers read the introduction, those that needed to reveal their demographic information (**ID**, **TD**, and **CD**) were asked to enter their information (Figure 2.b). The anonymous workers (**IA**, **TA**, and **CA**) skipped this step and went to the next step directly.

Then, we used an asynchronous interaction design [9] to make workers in conditions **TA**, **CA**, **TD**, and **CD** feel that they were matched with another worker. Instead of really matching the workers, we provided recorded data from a randomly selected previous worker in the same condition<sup>7</sup>. We adopted this asynchronous interaction design to prevent the workers from gaming the system by starting multiple tasks concurrently in the experiment website, which allowed them to be paired with themselves. This design has been adopted by many GWAPs (e.g., ESP game), and previous research [9]

<sup>6</sup>This design was based on a pilot study that an average worker's accuracy is around 0.8; therefore, it adjusted the amount of bonus to be similar to other conditions.

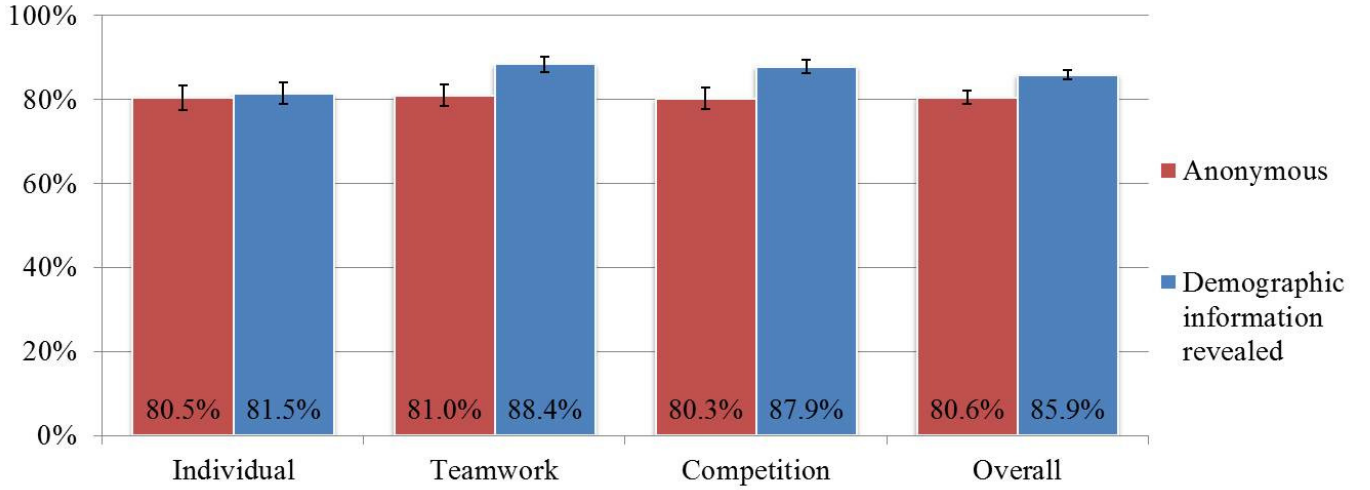
<sup>7</sup>The first worker in the paired conditions received data from randomly selected individual worker

	IA	ID	TA	TD	CA	CD
Introduction Page	✓	✓	✓	✓	✓	✓
Demographic Info Collection	✗	✓	✗	✓	✗	✓
Asynchronous Interaction Design	✗	✗	✓	✓	✓	✓
Info Sharing between workers	✗	✗	✗	✓	✗	✓
Noun Counting Page	✓	✓	✓	✓	✓	✓
Completion Page	✓	✓	✓	✓	✓	✓

**Figure 3. Summary of experimental interfaces used in the six conditions.**

has shown that this asynchronous interaction design can effectively create the perception that workers are actually working with another worker concurrently. In this design, the workers were redirected to a worker matching page (Figure 2.c). This page paused for a few seconds and indicated that the system was trying to match the worker with another worker (either teammate or opponent). These workers also were redirected to a waiting page after each question to wait for the paired worker to respond. The workers who revealed their demographic information and were paired with another worker (**TD** and **CD**) also received a message that showed the demographic information of their paired worker before the task began (Figure 2.d).

After that, the workers in all conditions would see the interface for performing the noun counting task (Figure 2.e). In



**Figure 4.** The mean of the average label accuracies of the workers under different peer-dependent reward schemes and social transparency levels (with standard error). In general, the workers who revealed their demographic information outperformed the ones who worked anonymously.

	Individual vs. Teamwork	Individual vs. Competition	Teamwork vs. Competition
Anonymous	80.5%(20.6%) vs. 81.0%(17.6%) ( $t(96) = 0.13, p = 0.45$ )	80.5%(20.6%) vs. 80.3%(18.6%) ( $t(97) = 0.06, p = 0.48$ )	81.0%(17.6%) vs. 80.3%(18.6%) ( $t(98) = 0.20, p = 0.42$ )
Demographic info. revealed	81.5%(17.6%) vs. 88.4%(13.0%)* ( $t(90) = 2.25, p < 0.02$ )	81.5%(17.6%) vs. 87.9%(11.7%)* ( $t(86) = 2.15, p < 0.02$ )	88.4%(13.0%) vs. 87.9%(11.7%) ( $t(97) = 0.22, p = 0.41$ )

**Table 1.** The comparison between the mean (standard deviation) of the average label accuracy of workers with different peer-dependent reward schemes with social transparency level fixed and their t-test statistics. The significant results ( $p$ -value  $< 0.05$ ) are marked with \*.

the upper part of the interface, the system showed a word list and an input box for them to provide their answers. The system specifically told the workers to count a word as a noun if one of its definitions is a noun because it is possible for a word to have multiple definitions and only some of them are nouns. This made the gold standard answer of the question clearer to the workers. At the bottom part of the interface, it showed the workers their current label accuracy for the questions they had answered. The workers within the teamwork reward scheme could see their teammate's current label accuracy and the current average label accuracy for their team (**TA** and **TD**). The workers under the competition reward scheme could see their opponent's current label accuracy and the current difference between them (**CA** and **CD**). The workers that were paired with another worker and shared demographic information could also see their colleague's information at the bottom of the interface (**TD** and **CD**).

Finally, after the workers completed 10 questions, they would be redirected to the completion page that showed the completion code (Figure 2.f). The design of the experimental interface of the six conditions is summarized in Figure 3.

## RESULTS

300 subjects (50 subjects per treatment) were recruited from AMT; each of them completed 10 noun counting tasks and generated 10 labels; therefore, 3,000 labels (500 labels per treatment) were collected for our results analysis.

We conducted a 3X2 Two-Way ANOVA to see if there were

significant effects from the peer-dependent reward schemes, social transparency levels, and their interaction. The results showed that social transparency level did significantly affect the performance of workers ( $F(1, 294) = 7.58, p < 0.01$ ). However, the effect of peer-dependent reward schemes was not significant ( $F(2, 294) = 1.41, p = 0.24$ ). Moreover, the effect of the interactions between the two factors was also insignificant ( $F(2, 294) = 1.26, p = 0.28$ ).

### Social transparency effectively improved outcome quality of crowdsourcing

The mean of the average label accuracies for the workers with different social transparency levels and peer-dependent reward schemes are summarized in Figure 4. The results show that the average label accuracy of the workers who revealed their demographic information ( $\mu = 85.9\%$ ,  $SD = 14.5\%$ ) was significantly higher than those who did not ( $\mu = 80.6\%$ ,  $SD = 18.9\%$ ) ( $t(280) = 2.75, p < 0.01$ ). This result supports our hypothesis that increasing the social transparency of workers makes them more accountable. Therefore, the quality of outcomes of crowdsourcing system also became better.

Moreover, although the ANOVA shows that the interaction effect between social transparency levels and peer-dependent reward schemes was not significant, we did find some interesting patterns in the result. Specifically, when the workers did not reveal their demographic information, there were no significant differences between the workers under the three reward schemes, which means that neither teamwork nor



competition motivated the workers to perform better when the workers knew that they were working anonymously.

On the other hand, when the workers revealed their demographic information, the differences between the workers whose rewards were codependent on their peers and those who worked individually became significant (i.e., between **TA** and **TD** and between **CA** and **CD**). The average label accuracies of the workers who worked as a team and competed against each other were both significantly higher than individual workers. (The comparisons and their t-test statistics are summarized in Table 1)

These differences, of course, need to be interpreted cautiously, as the overall interaction effect between social transparency levels and peer-dependent reward schemes was not significant. Nevertheless, we can see that social transparency had a better ability to enhance outcome quality when the workers felt that their information was shared with their colleagues. In contrast, if they felt that their information was only shared with the system (requester), social transparency did not have a significant effect on enhancing the performance of crowd workers.

#### Effects of different peer-dependent reward schemes

The overall performance of the workers suggests that the two different peer-dependent reward schemes (i.e., teamwork and competition) used in our experiment had no significant difference on enhancing worker performance. They both worked well when the workers revealed their demographic information and did not have significant effects when workers worked anonymously.

To gain more insight about how these two reward schemes differed, we measured how workers reacted to feedback with different accuracy levels. If social loafing created a negative effect on workers in teamwork conditions, then one would expect that when workers received high-accuracy feedback from their teammates, their effort levels should decrease because they already could earn a high bonus by relying on the performance of their teammates. In contrast, workers who received high-accuracy feedback from their opponents in the competition conditions should perform better because social facilitation suggests that workers were motivated to outperform their colleagues.

To test if these hypotheses are true, we divided the feedback into two levels (high, low) using the median label accuracy of all feedback.<sup>8</sup> We then calculated the average label accuracy of the next question based on the feedback accuracy level of the previous question. We then performed a three-way ANOVA on accuracies with label accuracy level (low vs high), reward schemes (teamwork vs competition), and social transparency (anonymous vs sharing demographic information) as independent variables. Results showed that the three-way interaction was significant ( $F(1, 1782) = 6.831, p < 0.01$ ). As shown in Figure 5 & 6, the three-way interaction was caused by the differential effects of social transparency

on how workers performed in each reward scheme when received with a low or high accuracy feedback. We will perform separate two-way ANOVAs on these to understand this three-way interaction below.

#### Social loafing made workers who received high-accuracy feedback from teammate perform worse

A 2X2 two-Way ANOVA (social transparency level and feedback accuracy level) for the workers who cooperated with the paired worker (teamwork) showed that social transparency significantly affected the performance of the workers ( $F(1, 896) = 34.23, p < 0.01$ ). Moreover, the effect of the accuracy level of feedback from teammates was also significant ( $F(1, 896) = 6.43, p < 0.02$ ). However, the effect of the interaction of these two factors was not significant ( $F(1, 896) = 0.15, p = 0.70$ ).

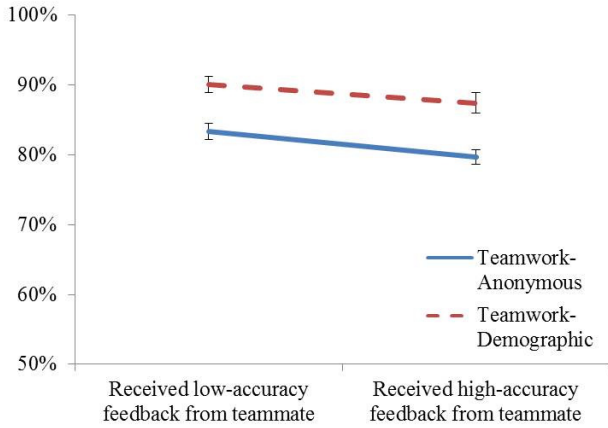
This result has two important implications. First, it indicates that social loafing negatively affected worker performance in teamwork conditions. That is, team members were less diligent when they found that they could earn a higher bonus because their teammates already performed very well. Second, social transparency between team members can enhance the performance of the workers regardless of their teammates' performance. This shows that sharing demographic information between team members increased social transparency, which possibly increased the motivation of the workers through altruistic motives to put more effort into the task. As a result, simply asking workers to share demographic information before the tasks was enough to motivate workers in teamwork conditions to perform better than workers in individual conditions.

#### Social transparency enhanced social facilitation in competition conditions

A 2X2 Two-Way ANOVA (social transparency level and feedback accuracy level) for the workers who competed against their paired worker (competition) showed that the effect of social transparency was significant ( $F(1, 896) = 20.92, p < 0.01$ ). However, the accuracy level of the feedback from opponents did not significantly affect the average label accuracy of workers ( $F(1, 896) = 0.12, p = 0.91$ ). Finally, the effect of the interaction between these two factors was significant ( $F(1, 896) = 17.19, p < 0.01$ ). Simple effects analysis showed that the difference between low and high accuracy of feedback was significant in both the anonymous condition ( $t(444) = 3.04, p < 0.001$ ) and in the high social transparency condition ( $t(425) = 2.09, p < 0.05$ ).

This interesting result shows that when workers competed against each other anonymously, performance was better when they received a low accuracy feedback. However, when the social transparency was high, their performance was significantly higher when they received a high accuracy feedback. This suggested that social transparency had a significant interaction with the level of accuracy of the feedback provided by their opponents in the competition reward scheme. It appears that social transparency was key to the social facilitation in the competition conditions. It is possible that when the workers did not share their information with

<sup>8</sup>Across all conditions, 88.5%.



**Figure 5.** When workers who worked in the teamwork condition received high-accuracy feedback from teammates, they performed significantly worse, which validates the negative effect of social loafing on crowd workers.

each other, they in general were less motivated than when social transparency was high. As a result, they were more likely to work harder when they received a low accuracy feedback than when they received a high accuracy feedback. On the other hand, our analysis showed that when the workers received high-accuracy feedback from their opponent, they were motivated to perform even better than their opponents. This led to better overall performance for workers in the high social transparency competition conditions.

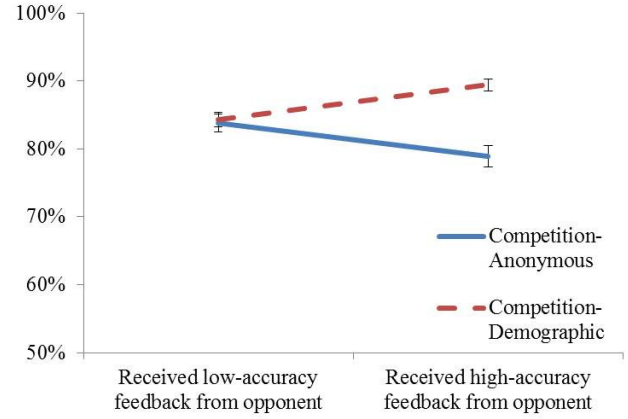
## DISCUSSION

### Social transparency makes social interactions between crowd workers more realistic

Our analysis shows that revealing demographic information can increase social transparency, thereby motivating workers to put more effort into the task when interacting with their teammates. Moreover, in a competition reward scheme, workers who share their information also have more incentives to outperform each other. This suggests that social transparency can make social interactions between crowd workers more engaging. Therefore, it is possible that the strategy of increasing social transparency can be applied to other reward schemes that hinge on social interactions between crowd workers. For example, this strategy can be applied to the external assessment scheme proposed by Dow *et al.* [3]. Asking crowd workers who served as assessors to provide their own personal information is possible to make them provide more responsible feedback and assessment and therefore became better shepherds for other workers. Furthermore, workers in peer consistency evaluation [?] can also work harder if they know that their work will be used as evaluation standards with their personal information attached. We believe that the findings of our experiment could be applied to various mechanism designs that involve social interaction between workers.

### Better utilizing the power of altruistic motives

When workers collaborated with a teammate, the workers with a high-accuracy teammate performed worse than those



**Figure 6.** Workers who shared their demographic information performed significantly better when they received high-accuracy feedback from opponents. This shows that social transparency helped enhance the effect of social facilitation.

with a low-accuracy partner regardless of whether they revealed their demographic information or not. This shows that the effect of social loafing was stronger than the effect of altruistic motives when the bonuses depend on the collective outcomes. However, this does not mean that altruistic motives cannot improve worker performance. In future, mechanisms that better utilize altruistic motives and prevent the system from suffering social loafing could be designed. For instance, the system could make workers feel that their performance can affect the bonus of other workers but does not let them benefit from the good performance of others (e.g., the altruistic motives condition used in [10]). This would allow us to separate the effects of social loafing and altruistic motives so that the system could better utilize the power of altruistic motives.

### How does the original motivation influence the effect of social transparency?

The results of our experiments show that social transparency between paired workers made them perform significantly better. However, a limitation to our study is that we didn't vary the level of financial incentive or the task design. Therefore, it is not clear if the result of our experiment can be applied to diverse settings. Especially, the task used in our experiment offered only a small amount of payment and allowed workers to guess the answer without putting much effort. This could make our experiment attract some spammers which skewed the results. As a result, it is possible that the mechanism of reducing anonymity used in our experiment only works for workers with low motivation.

Though this is an important limitation to our study, given the random assignment, if there really was such effect, it should have influenced all the conditions<sup>9</sup>. Since we have found some

<sup>9</sup>We tested if there was a selection effect by calculating the drop-out rates (the percentages of participants dropped after they saw the instruction). It shows that there was no significant difference between conditions.



significant differences between conditions, the effects should not affect the comparisons presented in this paper.

In the future, it is valuable to vary the settings of our experiment to see if workers with different original motivation react to the mechanism of reducing anonymity.

### Privacy issues and utilizing social translucence

The results of our experiment showed that social transparency between paired workers (i.e., sharing workers demographic information) significantly improved outcome quality. However, increasing social transparency between crowd workers also raised the concern for privacy issues [4]. Some online labor markets do not allow requesters to gather personal information from workers in order to protect the workers. For example, AMT prohibits requesters to collect personal identifiable information<sup>10</sup> (e.g., email) from workers. Reducing the anonymity of crowd workers might even make some of them unwilling to work on HITs. This could decrease the efficiency, which is one of the biggest advantages of crowdsourcing.

One possible way to address this issue is to utilize the idea of social translucence as mentioned in Erickson *et al.* [4], in which they designed graphical interfaces that presented only the social cues of the colleagues without sharing their identifiable information. This could help requesters ensure outcome quality by strengthening the social connections between workers and preventing the workers from feeling that their privacy was invaded. We believe this “virtual workspace” idea can be useful to crowdsourcing researchers and practitioners and is worth further investigation in the future.

### CONCLUSIONS AND FUTURE WORK

This paper found that increasing social transparency between workers effectively enhanced the quality of crowdsourcing outcomes. That is, when workers were asked to provide their demographic information (i.e., name, nationality, gender, and age) to their teammates or opponents, they performed significantly better than those who worked anonymously. Since there are many possible social interactions between the workers in a crowdsourcing system, we believe our findings present a novel way to enhance outcome quality of crowdsourcing systems.

Moreover, we found that social loafing created a negative effect on the performance of team members. When the workers received high-accuracy feedback from their teammates, they put less effort in their work because they realized that they could already reap a bonus based on the performance of their teammate. However, social transparency was able to make the workers more accountable to their teammates, so the workers performed significantly better when they revealed demographic information to their teammates.

In addition, our results suggest that social transparency also made the effect of social facilitation significant. When two anonymous workers competed against each other, the workers were not motivated to perform better. In contrast, when the paired workers shared their information, the workers had

more incentive to outperform their opponents. As a result, when they received high-accuracy feedback from their opponent, the worker put more effort in performing better. This made the workers within the competition reward scheme perform significantly better than those who work individually.

These results suggest that designing various peer-dependent reward schemes can successfully motivate workers to generate outcomes of higher quality. Since previous research [18] has proven that increasing financial incentive does not always motivate workers to perform better, the reward schemes that utilize the existence of peer workers can be both more useful and more cost-effective.

In future studies, we would like to tease out the effects of self-identifiability and peer-identifiability. The information sharing conditions share both the information of the worker and his/her peer. This makes it difficult to know if the better performance is because of the worker shared the information or learned the information from the peer. Therefore, separating the effects will help us gain more insight on how social transparency affect the motivation of crowds. In addition, we would like to examine whether strategically selecting the paired workers can further enhance the performance of crowd workers. Previous research [26] has shown that it is not necessary to really pair two online workers to create social interaction. Using interface design (i.e., worker searching page and worker waiting page) alone can make workers feel that other workers are working with them concurrently. This design allows the system to actively select the opponents or teammates for the workers. For example, since our results have shown that high-accuracy feedback from opponents can motivate the workers in competition conditions perform better, we can strategically select one of the high-accuracy workers as an opponent to the new worker. In addition, we can also deliberately select one of the workers with low-accuracy to be the teammate of the new worker in teamwork mode to reduce the effect of social loafing. Moreover, it is also possible to vary the level of financial incentive to see how it affects the performance of workers with different social transparency levels. The higher financial incentive can make the workers more accountable for their actions, but it also can reduce the intrinsic value of the workers and makes the effect of social loafing even stronger. To understand the effect, a systematic study that involves different levels of financial incentive can be conducted to tease apart the complex interactions of these effects.

### ACKNOWLEDGEMENTS

Many thanks to the anonymous reviewers for their valuable comments for us to improve the earlier draft of this work. This paper is also benefited from the discussion with Pei-Fen Tu, Jessie Chin, and Vera Liao.

### REFERENCES

1. Alonso, O., Rose, D. E., and Stewart, B. Crowdsourcing for relevance evaluation. *SIGIR Forum* 42, 2 (Nov. 2008), 9–15.
2. Bandiera, O., Barankay, I., and Rasul, I. Social preferences and the response to incentives: Evidence

<sup>10</sup><https://requester.mturk.com/help/faq>

- from personnel data. *The Quarterly Journal of Economics* 120, 3 (2005), 917–962.
3. Dow, S., Kulkarni, A., Klemmer, S., and Hartmann, B. Shepherding the crowd yields better work. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work, CSCW '12*, ACM (2012), 1013–1022.
4. Erickson, T., Halverson, C., Kellogg, W. A., Laff, M., and Wolf, T. Social translucence: designing social infrastructures that make collective activity visible. *Commun. ACM* 45, 4 (Apr. 2002), 40–44.
5. Erickson, T., and Kellogg, W. A. Social translucence: an approach to designing systems that support social processes. *ACM Trans. Comput.-Hum. Interact.* 7, 1 (Mar. 2000), 59–83.
6. Festinger, L. A theory of social comparison processes. *Human relations* 7, 2 (1954), 117–140.
7. Harkins, S. G. Social loafing and social facilitation. *Journal of Experimental Social Psychology* 23, 1 (1987), 1 – 18.
8. Howe, J. The rise of crowdsourcing. *Wired Magazine* (06 2006).
9. Huang, S.-W., and Fu, W.-T. Systematic analysis of output agreement games: Effects of gaming environment, social interaction, and feedback. In *Proceedings of HCOMP12: The 4th Workshop on Human Computation* (2012).
10. Huang, S.-W., and Fu, W.-T. Enhancing reliability using peer consistency evaluation in human computation. In *Proceedings of the ACM 2013 conference on Computer Supported Cooperative Work, CSCW '13*, ACM (2013).
11. Ipeirotis, P. G., Provost, F., and Wang, J. Quality management on amazon mechanical turk. In *Proceedings of the ACM SIGKDD Workshop on Human Computation, HCOMP '10*, ACM (2010), 64–67.
12. Karau, S., and Williams, K. Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology* 65, 4 (1993), 681–706.
13. Kerr, N. L. Motivation losses in small groups: A social dilemma analysis. *Journal of Personality and Social Psychology* 45 (1983), 819–828.
14. Kittur, A., Smus, B., Khamkar, S., and Kraut, R. E. Crowdforge: crowdsourcing complex work. In *Proceedings of the 24th annual ACM symposium on User interface software and technology, UIST '11*, ACM (2011), 43–52.
15. Kittur, A., Suh, B., and Chi, E. H. Can you ever trust a wiki?: impacting perceived trustworthiness in wikipedia. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work, CSCW '08*, ACM (2008), 477–480.
16. Law, E., and Von Ahn, L. *Human Computation. Synthesis Lectures on Artificial Intelligence and Machine Learning*. Morgan & Claypool Publishers, 2011.
17. Mas, A., and Moretti, E. Peers at work. Working Paper 12508, National Bureau of Economic Research, September 2006.
18. Mason, W., and Watts, D. J. Financial incentives and the "performance of crowds". *SIGKDD Explor. Newsl.* 11, 2 (May 2010), 100–108.
19. Millen, D. R., and Patterson, J. F. Identity disclosure and the creation of social capital. In *CHI '03 extended abstracts on Human factors in computing systems, CHI EA '03*, ACM (2003), 720–721.
20. Postmes, T., Spears, R., Lee, A., and Novak, R. Individuality and social influence in groups: inductive and deductive routes to group identity. *Journal of personality and social psychology* 89, 5 (2005), 747.
21. Quinn, A. J., and Bederson, B. B. Human computation: a survey and taxonomy of a growing field. In *Proceedings of the 2011 annual conference on Human factors in computing systems, CHI '11*, ACM (2011), 1403–1412.
22. Robertson, S., Vojnovic, M., and Weber, I. Rethinking the esp game. In *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems, CHI EA '09*, ACM (2009), 3937–3942.
23. Snow, R., O'Connor, B., Jurafsky, D., and Ng, A. Y. Cheap and fast—but is it good?: evaluating non-expert annotations for natural language tasks. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP '08*, Association for Computational Linguistics (Stroudsburg, PA, USA, 2008), 254–263.
24. Stuart, H. C., Dabbish, L., Kiesler, S., Kinnaird, P., and Kang, R. Social transparency in networked information exchange: a theoretical framework. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work, CSCW '12*, ACM (2012), 451–460.
25. Suh, B., Chi, E. H., Kittur, A., and Pendleton, B. A. Lifting the veil: improving accountability and social transparency in wikipedia with wikidashboard. In *Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems, CHI '08* (2008).
26. von Ahn, L., and Dabbish, L. Labeling images with a computer game. In *Proceedings of the SIGCHI conference on Human factors in computing systems, CHI '04*, ACM (2004), 319–326.
27. von Ahn, L., and Dabbish, L. Designing games with a purpose. *Commun. ACM* 51 (Aug. 2008), 58–67.
28. Zajonc, R. Social facilitation. *Science* 17 (1965), 65.