Co-Worker Transparency in a Microtask Marketplace

Peter Kinnaird, Laura Dabbish, Sara Kiesler, Haakon Faste
HCI Institute and Heinz College
Carnegie Mellon University
Pittsburgh, PA 15213
{kinnaird, dabbish, kiesler, hfaste}@cs.cmu.edu

ABSTRACT
Workers in microtask work environments such as Mechanical Turk typically do not know if or how they fit into a workflow. The research question we posed here was whether displaying information about the number of other workers doing the same task would motivate better or poorer work quality. In experiment 1, we varied the information about co-workers presented to the worker and the number of his or her co-workers: "you" or "you alone" are doing a task, or "you" plus 5, 15, or 50 co-workers. We compared these conditions with a no-social information control. In experiment 2, we crossed the number of co-workers (5 vs. 50) with the type of incentive (individual or group). Results show that visual presentations of co-workers changed workers' perceptions of co-workers, and that the more co-workers participants perceived, the lower their work quality. We suggest future work to determine the kinds of co-worker information that will reduce or increase work quality in microtask settings.

Categories and Subject Descriptors
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms
Human Factors

Keywords
Crowdsourcing, group size, co-workers, workflow, CSCW, work motivation, productivity, entitativity

INTRODUCTION

Employers and researchers are increasingly using microtask platforms such as Amazon's Mechanical Turk (MTurk) to accomplish work that can be separated into discrete tasks and distributed to a number of people. However, fully leveraging these systems remains challenging. Work quality is generally low on such platforms, in part because workers have a fleeting relationship with the employer and low pay [3,15,23]. Furthermore, many crowdsourced tasks are routine, repetitive or boring, requiring little creativity or variety, forcing employers to motivate the work entirely through pay [20].

CSCW '13, February 23–27, 2013, San Antonio, Texas, USA.

MTurk amplifies these threats to motivation through design decisions such as automatic task approval, the absence of a strong reputation management system, and impersonal persistent identifiers. Jobs can employ dozens or hundreds of interchangeable, anonymous workers. If workers perceive they are tiny cogs in a large machine, they potentially feel less accountable for their contributions, that their task lacks meaningfulness, and that their effort does not count for much. Lack of identifiability and the interchangeability of workers reduce trust and motivation and increase free riding [6] and social loafing (reduced individual effort that sometimes occurs in group settings) [5,12,21].

Employers using microtask platforms can mitigate the effect of poor worker motivation on the final product of crowdsourced work with workflows featuring high task redundancy and internal checks to redress low worker quality [1,16,18]. MTurk's best practices guide suggests asking multiple workers to complete each posted task (HIT). Some schemes apply complex incentives including threats of nonpayment for poor or incomplete work (e.g.[3,9,20,22]).

Another approach is to use social transparency to increase worker motivation [14]. Social transparency, in this context, means creating or increasing the visibility of the social aspects of an online setting to the people in that setting [11,24]. Almost all crowdsourced tasks present the worker with information only about their own small task even when it is embedded within a larger workflow and may be interdependent with others' work. For example, workers may be asked to search for information. Their search task might be simultaneously assigned to many other workers in order to replicate and check on their quality of work [2,12,13]. Workers typically do not know how many other workers are also working on the same task or how their work may influence or be influenced by other tasks.

Employers on microtask platforms have the option of informing their workers that there are others doing the same task, but we do not understand how this information influences motivation and work quality. In this paper, we describe two experiments in which MTurk workers were

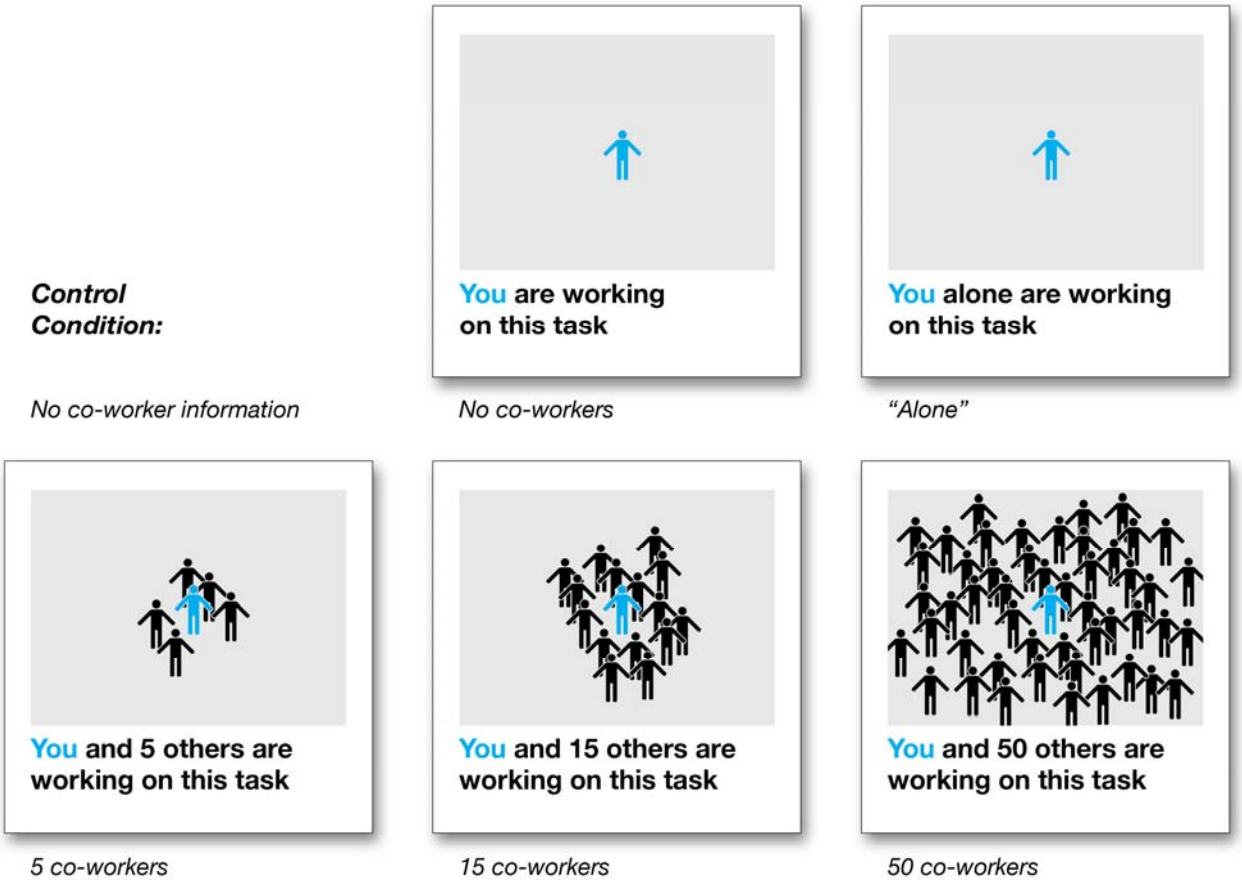

Figure 1. Co-worker information in MTurk Experiment 1.

shown an image displaying whether or not they had co-workers (see Figure 1).

The effect of co-worker information on worker motivation must be carefully tested. On the one hand, if workers learn they have some co-workers, they might become more motivated because they feel like part of a social entity. Common identity engenders loyalty and social feelings among members [10,12]. A worker learning he has co-workers might feel good that he has compatriots, and is sharing their fate. A recent study in which MTurk workers were shown where they stood in a workflow suggests this is a possibility [14]. On the other hand, information about co-workers could backfire if workers receiving this information feel their work is redundant and they are a small part of a large, impersonal process.

How many co-workers?

A related research question is whether employers should inform workers about the number of their co-workers. What if this number is large? We propose that too many co-workers will undermine worker motivation. Research on the group size shows that small groups of fewer than 10 people are more cohesive than larger groups whereas people in larger groups also feel less effective and tend to contribute less [2,8]. In one study of helping conducted in three cities, a large crowd of bystanders was less likely to help someone who dropped coins or pencils than a small group of bystanders, suggesting diffusion of responsibility in larger crowds [19]. Larger numbers of co-workers could also suggest to workers that they are redundant, meaning their work is less valuable to the final product. Studies show that perceived redundancy reduces worker effort and motivation [7,8]

TWO EXPERIMENTS ON CO-WORKER VISIBILITY

We conducted two between-subjects experiments on co-worker visibility in Mechanical Turk. We drew on the literature above to pose the following hypotheses:

***H1**: Workers informed they have a small number of coworkers will perform better quality work than those not informed about the existence of co-workers.*

***H2:** Workers informed they have a small number of coworkers will perform better quality work than those informed they have more coworkers.*

Experiment 1: Method

Participants were paid 25 cents for completing a HIT in MTurk. Workers were recruited between 2 PM EST on April 10, 2012 and 8 AM EST on April 12, 2012. In the HIT, we presented participants with a list of 18 statements made in the past year by prominent politicians in the U.S., evenly split between Democrats and Republicans and displayed in random order. For example, one statement was, "Health care is something that the American people

need, and they need it for everyone, not just a few people who are rich." (Harry Reid, Senate majority leader.) The worker's task was to identify the speaker of each statement by selecting the correct politician from a dropdown box with 11 choices.

We limited participation to the United States to reduce the impact of language understanding on the results. Workers could participate in just one of the experiments. After participants accepted the HIT and completed the minimum required work they advanced to a brief survey. Participants were only required to try to identify the speaker of one of the statements.

Independent variable – co-worker information

Participants in experiment 1 were randomly placed into one of six conditions in a 1 x 6 experimental design as shown in Figure 1. Participants in the control group were shown no information about co-workers. Participants in the no-coworker condition were shown a simple cartoon image of a worker ("you") and a caption explaining they were doing the task. After pretesting, we discovered many participants in the no co-worker condition assumed that they had co-workers, so we added the "Alone" condition to emphasize that they had no co-workers and were doing the task alone. The participants in the other three conditions were shown a cartoon with a caption to emphasize the number of their co-workers. The purpose of the brief text and simple drawings was to make the number of co-workers visually salient (see [4,25].)

Dependent measures

Our primary dependent measure was work quality, accurately identifying the correct speaker as a percent of the total speaker identifications attempted. (This total did not differ across conditions). We were mainly concerned with accuracy because selecting choices randomly would be valueless for an employer. We used a log score to normalize these data but the results are in the same direction without the transformation.

We also measured perceived number of co-workers. On the post-task survey we asked participants how many other Turkers were working on their task. We did this to determine whether coworker information influenced perceptions of co-workers. In the analyses, we truncated the number of perceived co-workers to a range of 0 to 101 because a few participants estimated 1000 or 2000 co-workers, which would have biased the data. We then normalized responses using a log score. This technique is suitable to prevent very high scores from distorting the results.

Experiment 1: Results

One hundred seventy-seven participants completed the HIT, with 28 to 31 per condition. The post-task survey showed that the manipulations differentially affected the perceived number of co-workers (F [5, 171] = 5.7, p < .0001). Control participants, who were not informed they had co-workers, nevertheless estimated they had almost 30 co-workers. This number fell in between the "50 co-workers" and "15 co-workers" conditions. The lowest estimate was in the "You alone" condition; evidently the single word "alone" made it clear to most participants they were not working with others (Figure 2).

Experimental Conditions	N	Mean Number Co-Workers Reported* (*SE*)	Mean Number Items Completed (*SE*) (n.s.)
"You Alone"	25	6.2 [a] (*5.2*)	12.6 (*1.3*)
"You"	30	17.8[abc] (*5.0*)	11.4 (*1.3*)
"You and 5 others"	28	7.6[ab] (*5.2*)	12.7 (*1.3*)
"You and 15 others"	32	21.4[bc] (*4.9*)	11.1 (*1.2*)
"You and 50 others"	31	37.4[d] (*5.0*)	13.8 (*1.2*)
Control – No Information	28	29.4[cd] (*5.2*)	13.9 (*1.3*)

Table 1. Experiment 1 mean number of co-workers reported and items completed by condition.
**Conditions with unique superscripts are significantly different at $p < .05$ based on a Student's t test. There are no differences in the number of items completed across conditions.*

Although the images altered the perceptions of coworkers differentially, it is evident from Figure 2 that some participants ignored or did not believe the coworker information. About 26% of the participants reported a larger or smaller number of perceived co-workers than the intended number. This inattention (or disbelief) was associated with lower work quality and mitigated the direct effect of the coworker information. Thus the direct effect of the manipulations did not significantly change the quality of work. Instead, the perceived number of co-workers (driven by the manipulations) influenced work quality. We checked for selection bias problems with a chi-squared test on whether co-worker information resulted in a higher dropout rate. We discovered no difference in dropout rate across conditions (p = .96) and no difference in accuracy between those who completed the HIT and those who dropped out mid-way through the task (p = .69).

When participants perceived they had more co-workers, their quality of work declined, as predicted in Hypothesis 2 (F [5, 158] = 5.7, p = .01; Figure 3). This result holds whether or not we control for the number of items completed or condition. This result suggests that, without

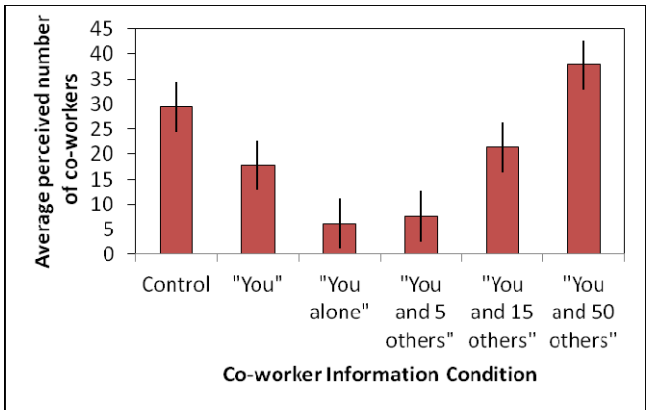

Figure 3. Reported number of co-workers in experimental conditions of Experiment 1.

other interventions creating group identity or other feelings of similarity or common bonds with other workers, simply having more co-workers leads to more social loafing or free riding, and poorer work quality. We found no evidence to support Hypothesis 1, that simply having co-workers would increase work quality.

In summary, our results show that increasing the number of displayed coworkers positively influences perceived coworkers, which in turn reduces work quality.

Experiment 2: Method

We conducted Experiment 2 to investigate whether adding interdependence with other workers would change the negative effects of perceiving co-workers, found in Experiment 1. Workers were again recruited only from the United States and were recruited between 1 PM EST on May 20, 2012 and 10 AM EST on May 21, 2012. The design was a 2 x 2 between-groups experiment, with two levels of co-worker number (5 vs. 50 co-workers, as in Figure 1) and the control no coworker information condition. The second independent variable was individual versus group incentive. In the individual incentive condition, the HIT had a 4-cent bonus, plus an addition 2-cent bonus for a correct answer, and one cent off the bonus for each incorrect answer. In the group incentive condition, participants had some level of outcome interdependence with their co-workers, earning a 2-cent bonus for each correct answer that they and the majority of others answered, and losing 1 cent for each quote they and the majority of others got wrong.

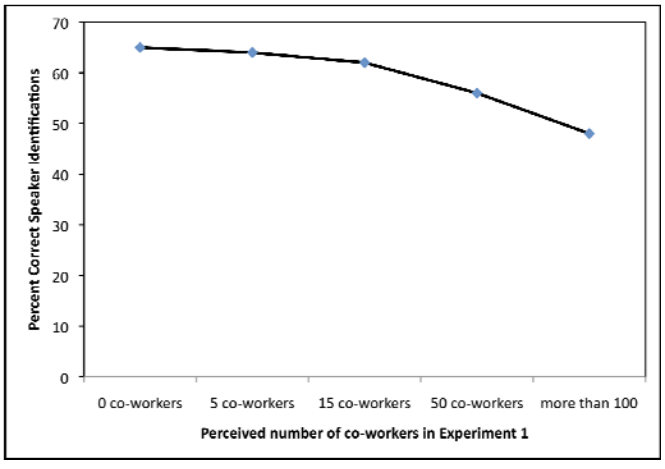

Figure 2. Perceived number of co-workers negatively affects work quality (Experiment 1).

The method was identical to that in Experiment 1 except participants were given 36 statements and 20 speaker options (rather than 18 statements and 11 options as in Experiment 1), increasing the possible total amount of work and reducing the chances of getting a correct answer by guessing.

Experiment 2: Results

There were 186 participants who completed the HIT and the survey. As in Experiment 1, the coworker information changed how participants perceived the work environment. In the control condition, they estimated, on average, they had 20 (SE = 3.95) co-workers; in the "5 other workers" condition, they estimated they had 16.2 (SE = 2.8) co-workers and in the "50 other workers" condition, they estimated they had 44.4 (SE = 2.9) co-workers (F [2, 179] = 26, $p < .0001$). (These averages use the truncated distributions so as not to bias the results with estimates over 100.) We checked for selection bias problems with a chi-squared test on whether either co-worker information or incentive condition resulted in a higher dropout rate. We discovered no difference in dropout rate across conditions ($p = .61$) and no difference in accuracy between those who completed the HIT and those who dropped out mid-way through the task ($p = .13$).

Overall, the results replicated those of Experiment 1 even more strongly. For those who completed the HIT and survey, there were no direct effects of co-worker information or incentive, but the perceived number of co-

Experimental Conditions	N	Mean Number Co-Workers Reported* (*SE*)	Mean Number Items Completed (*SE*) (n.s.)
Five Co-workers	77	16.2[a] (*2.8*)	28.4 (*1.9*)
Fifty Co-workers	70	44.4[b] (*2.9*)	30.3 (*1.4*)
Control – No Information	39	7.6[a] (*3.9*)	28.8 (*1.3*)

Table 2. Experiment 2 mean number of co-workers reported and items completed by condition.

*****_Conditions with unique superscripts are significantly different at $p < .05$, based on a Student's t test. There are no differences in the number of items completed across conditions._

workers significantly affected work quality. Those who perceived more co-workers did poorer quality work ($F[1,181] = 2.9$, $p = 0.08$). This result held whether or not we controlled for number of items completed. The accuracy rates were 78% correct when no co-workers were reported, 76% when 5 co-workers were reported, 58% when 50 co-workers were reported, and only 38% correct when more than 100 co-workers were reported. Given that the control group who saw no co-worker information may have just been guessing about the number of co-workers, we conducted one further ANOVA excluding their data. In the rest of the conditions, workers who reported larger numbers of co-workers performed poorer quality work ($F[1, 140] = 7.7$, $p = .006$).

Interestingly, outcome interdependence appeared to increase attention to the number of co-workers. Workers in the group incentive condition perceived significantly more co-workers (34) than those in the individual condition (23) ($F[1,137] = 5$, $p = .03$). The group incentive also tended to reduce work quality (60% correct versus 65% in the individual incentive condition) but there was no interaction between incentive and displayed or perceived number of co-workers.

DISCUSSION

Our results suggest that co-worker information in the task environment in microtask marketplaces can significantly influence worker motivation and work quality. This influence is not always positive, as our results indicate. Information about other workers increased the number of others perceived to be doing the same task, which ultimately reduced work quality. We believe that the coworker information we presented heightened workers awareness that they were simply one of many, potentially reducing feelings of task significance and work importance. Workers may have felt they were interchangeable with others and that the outputs of their many co-workers would be combined to account for any mistakes in their work, and thus less accountable for the output of their work. This result contradicts other recent work suggesting workflow information does in some cases have a positive influence on motivation. Future research should examine the boundary conditions of our result.

It is important to note that participants may have gotten cues about coworkers from other areas on MTurk. In both experiments, multiple HITs from the same requestor were available at the same time. It is possible that the number of HITs available could suggest a particular number of coworkers to a given participant. On the other hand, the prominent image of coworkers within the HIT coupled with the common practice of including slightly different tasks in a single group of HITs suggests it is unlikely the number of HITs influenced perceived coworkers. In addition, by describing other workers as coworkers on a task" we intended to decouple the manipulation of coworker information from the number of HITs available in the group.

The strong correlation between the coworker information in the condition and participants' perceived number of coworkers suggests our manipulation was successful. It may be that under certain conditions larger numbers of co-workers can have a positive influence on motivation and work quality and it is important to identify those conditions. In particular the nature of the attachment among co-workers (bonds or common identity) and the nature of interdependence may change the influence of perceived group size. For example, if workers felt more strongly connected to their co-workers, especially if they felt accountable to one another, through a common social identity such as an organizational affiliation, as suggested by Kraut and Resnick [17], co-workers could increase their motivation according to collective effort theory. In addition, if their work output was transmitted to a larger rather than a smaller set of co-workers, meaning those workers were dependent on them, this consequence could enhance perceived task significance.

One implication of our result is that task designers should make it clear to workers when they are posting individual tasks. Knowing that one is working alone improved work quality in our studies.

Our work suggests that we must be careful in presenting workflow information in microtask marketplaces. In some cases this information can reduce motivation and work quality. Further research is needed to identify the social and task structures that impact how individuals perceive social transparency.

ACKNOWLEDGEMENTS

This research is supported by the Center for the Future of Work at Carnegie Mellon University's Heinz College and a National Science Foundation grant, CNS-1040801.

REFERENCES

1. Bernstein, M.S., Little, G., Miller, R.C., et al. Soylent: a word processor with a crowd inside. *Proceedings of the 23nd annual ACM symposium on User interface software and technology*, ACM (2010), 313–322.

2. Brewer, M.B. and Kramer, R.M. Choice behavior in social dilemmas: Effects of social identity, group size, and decision framing. *Journal of Personality and Social Psychology 50*, 3 (1986), 543–549.

3. Downs, J.S., Holbrook, M.B., Sheng, S., and Cranor, L.F. Are Your Participants Gaming the System ? Screening Mechanical Turk Workers. *Science*, (2010), 2399–2402.

4. Fekete, J.-D., Wijk, J.J., Stasko, J.T., and North, C. Information Visualization. In A. Kerren, J.T. Stasko,

J.-D. Fekete and C. North, eds., Springer-Verlag, Berlin, Heidelberg, 2008, 1–18.

5. George, J.M. Extrinsic and Intrinsic Origins of Perceived Social Loafing in Organizations. *The Academy of Management Journal 35*, 1 (1992), 191–202.

6. Granovetter, M.S. The Impact of Social Structure on Economic Outcomes. *Journal of Economic Perspectives 19*, 1 (2005), 33–50.

7. Hackman, J.R. and Oldham, G.R. Motivation through the design of work: test of a theory. *Organizational Behavior and Human Performance 16*, 2 (1976), 250–279.

8. Huberman, B.A., Romero, D.M., and Wu, F. Crowdsourcing, attention and productivity . *Journal of Information Science 35* , 6 (2009), 758–765.

9. Ipeirotis, P.G. and Paritosh, P.K. Managing crowdsourced human computation: a tutorial. *Proceedings of the 20th international conference companion on World wide web*, ACM (2011), 287–288.

10. Jans, L., Postmes, T., and Van Der Zee, K.I. The induction of shared identity: The positive role of individual distinctiveness for groups. *Personality social psychology bulletin 37*, 8 (2011), 1130–1141.

11. Jones, G.R. Task Visibility, Free Riding, and Shirking: Explaining the Effect of Structure and Technology on Employee Behavior. *The Academy of Management Review 9*, 4 (1984), 684–695.

12. Karau, S.J. and Williams, K.D. Social Loafing : A Meta-Analytic Review and Theoretical Integration. *Journal of Personality and Social Psychology 65*, 4 (1993), 681–706.

13. Kerr, N.L. Illusions of efficacy: The effects of group-size on perceived efficacy in social dilemmas. *Journal of Experimental Social Psychology 25*, 4 (1989), 287–313.

14. Kinnaird, P., Dabbish, L., and Kiesler, S. The Impact of a Transparent Workflow in Mechanical Turk. *GROUP*, ACM (2012).

15. Kittur, A., Chi, E.H., and Suh, B. Crowdsourcing user studies with Mechanical Turk. *Proceeding of the twentysixth annual CHI conference on Human factors in computing systems CHI 08 08*, April 5-10 (2008), 453.

16. Kittur, A., Smus, B., and Kraut, R.E. CrowdForge : Crowdsourcing Complex Work. *Human-Computer Interaction*, (2011), 1801–1806.

17. Kraut, R.E., Resnick, P., Kiesler, S., et al. *Building Successful Online Communities: Evidence-Based Social Design*. MIT, Boston, MA, 2012.

18. Kulkarni, A., Can, M., and Hartmann, B. Collaboratively crowdsourcing workflows with turkomatic. *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, ACM (2012), 1003–1012.

19. Latane, B. and Dabbs Jr., J.M. Sex, Group Size and Helping in Three Cities. *Sociometry 38*, 2 (1975), 180–194.

20. Mason, W. and Watts, D.J. Financial incentives and the "performance of crowds."*ACM SIGKDD Explorations Newsletter 11*, 2 (2010), 100.

21. Roland E. Kidwell, J. and Bennett, N. Employee Propensity to Withhold Effort: A Conceptual Model to Intersect Three Avenues of Research. *The Academy of Management Review 18*, 3 (1993), 429–456.

22. Shaw, A.D., Horton, J.J., and Chen, D.L. Designing Incentives for Inexpert Human Raters. *CSCW*, (2011), 1–8.

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. *Proceedings of the Conference on Empirical Methods in Natural Language Processing 254*, October (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. *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, ACM (2012), 451–460.

25. Ware, C. *Information visualization: perception for design*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2000.