With a Little Help from the Crowd: Receiving Unauthorized Academic Assistance through Online Labor Markets

Christopher G. Harris
Informatics Program
The University of Iowa
Iowa City, Iowa
christopher-harris@uiowa.edu

Padmini Srinivasan
Informatics Program and Computer Science Department
The University of Iowa
Iowa City, Iowa
padmini-srinivasan@uiowa.edu

Abstract—Although a vast majority of crowdsourcing tasks are for ethical purposes, the anonymity and global reach of online labor markets also create a clearinghouse for unethical crowdsourcing tasks. Recent studies show a majority of students have engaged in academic dishonesty using the Internet, and a growing number find this behavior is acceptable. We conduct a study to see if crowdworkers will provide solutions to exams and homework assignments, and knowingly permit these solutions to be used for this purpose. For those who don’t agree, we examine if additional financial incentives can entice them. Our findings indicate most crowdworkers are willing to permit the use of their work; however, for those that are unwilling, additional financial incentives have little effect on altering their decision.

Keywords—ethics, crowdsourcing, incentives, online labor markets, academic dishonesty, Mechanical Turk

I. INTRODUCTION

Work conducted through online labor markets, such as Amazon Mechanical Turk (MTurk), typically involve offering small, discrete tasks in exchange for payments of a few US cents each through an “open call”. The ubiquity of the Internet and ease of sending micropayments across the globe has facilitated the growth of these online labor markets, providing access and opportunities for workers who previously would be unable to participate. A vast majority of these online labor market tasks are conducted to support purposes many would view as ethical, such as annotating images, summarizing text, and providing relevance judgments.

Most online labor markets are designed to maintain mutual anonymity between workers and requestors. While this anonymity benefits the online labor market itself by preventing disintermediation, there are other potential unintended consequences. First, anonymity can lead to quality distortions associated with imperfect information, much like Akerlof’s “lemon market” [2]. Second, it makes long-term studies, which rely on worker continuity, difficult to achieve. Third, the anonymity provides “social distance” between worker and requestor, which has been shown in other studies to be a facilitator of unethical behavior [8, 16]. This anonymity, as well as the ability to send and receive micropayments and the difference in global attitudes on ethics, provides a rich environment for unethical tasks to transpire.

In this paper, we examine the role of unethical behavior in online labor markets within the context of obtaining unauthorized help on academic assignments. Our contributions are as follows. First, we conduct a study to determine the ease at which unethical behavior may occur through online labor markets. Second, we examine if the size of financial incentives affect unethical behavior. Third, we investigate if the amount of assistance provided affects a crowdworker’s willingness to participate in unethical behavior.

This remainder of this paper is organized as follows. In the next section, we discuss the background and motivation. In Section III, we cover related work. Section IV and VI discuss our experimental design and results, respectively. We conclude our work in Section VII.

II. BACKGROUND AND MOTivation

Ethics, defined as “the principles that define the boundary between right and wrong” [27], have been explored since the time of Socrates. A number of studies have demonstrated that ethics can differ across cultural and geographic boundaries (e.g., [3, 19, 24]); what is considered ethical in one culture may be considered unethical in another. These differences present an opportunity for “ethical arbitrage” to occur in online labor markets.

Academic dishonesty includes the behaviors of fraud, plagiarism, falsification, delinquency and unauthorized help. Electronic dishonesty (e-dishonesty) employs online resources to facilitate these five types of unethical behavior, enhanced through the use of Internet services [23]. In this paper, we focus specifically on only one aspect of e-dishonesty: obtaining unauthorized help online using online labor markets.

A number of recent studies have shown that academic dishonesty is on the rise (e.g. [12, 17, 21]). A study by the Center for Academic Integrity (CAI) Assessment Project encompassing nearly 50,000 undergraduates at more than 60 institutions revealed that 43% of students had engaged in some
form of unauthorized collaboration and nearly 22% admitted to cheating on a test or exam [30]. Online courses reportedly have even higher rates of academic dishonesty than traditional courses [1].

Some researchers report that academic dishonesty no longer carries the stigma it once did – less social disapproval coupled with increased competition for university admissions have increased students willingness to cheat [7, 15]. Unfortunately, this unethical behavior is not limited to academic gain – recent studies have shown an association between those engaging in academic dishonesty and those later engaging in similar behavior in the workplace [6, 10]. Advocates of academic dishonesty have even going as far as posting successful cheating techniques on YouTube for others to copy [11].

Many crowdworkers might be unaware that the task they have agreed to participate in is unethical. This may be related to how online labor markets allow the division of a single, larger task into several autonomous components. For example, as Harris pointed out in [13], the scope of a task can be hidden in online labor markets; i.e., asking for assistance with a single question of a 100-question homework assignment may appear ethically acceptable to many, but if 100 workers are each tasked a single question, this behavior may cross the line from ‘ethical’ to ‘unethical’. In our research, we examine if providing a larger portion of an assignment affects the willingness to agree to its use.

III. RELATED WORK

As Gallant points out in [9], academic dishonesty has been a unfortunate aspect of education for centuries. With regard to obtaining unauthorized help online, Lancaster and Clark were the first to define contract cheating, a concept that illustrates how students use websites such as RentACoder.com (now vWorker.com) to have computer science coding assignments in completed but other unattributed authors. Their definition not only includes online labor markets and the purchase of work from custom-essay writing companies, but also the use of “homework help” discussion boards and websites, where students can post questions and receive answers from others. In [4], these authors also investigated contract cheating using internet auctions. In [5], they provide a discussion about “liveware” approaches, such as having other students perform their homework assignments, as a form of cheating. Contract cheating is not exclusively tied to computer science assignments – Walker et. al. illustrate how it spans different levels and different disciplines in [28].

Some others have explored unethical behavior outside of academics. In [29], Wang et. al. described using the crowd for unethical behavior, such as creating spam. In [25], Ott et. al. demonstrated how the crowd can be used to submit deceptive opinion spam, such as false reviews, on websites such as TripAdvisor. In [14], Harris and Srinivasan conducted a study using online labor markets to solicit crowdworkers to compromise targeted email accounts, concluding that unethical behavior can be performed by the crowd for very little compensation. In [26], Suri et. al. looked at how the honestly the crowd would report the roll of a fair die, offering different payment amounts for different results. They conclude that although few of the workers cheated a lot, many cheated a little.

Although many have reported how unauthorized help and unethical behavior can occur, or made some observations on its use, we found no experiments in the literature that examine the propensity for using the crowd to assist with unethical tasks, consider financial incentives, or examine the effects of amount of financial assistance impacts crowdworker behavior.

IV. EXPERIMENTAL DESIGN

In this study, our intention is to answer the following research questions.

1. Is there a sufficient number of crowdworkers willing to consent for use their contributions for unethical behavior? If so, what is the percentage willing to do so?
2. For those crowdworkers not willing to provide consent, will financial incentives change their mind?
3. Are crowdworkers more willing to provide assistance for a homework assignment, compared with help on an exam (which is arguably a bigger ethical violation)?
4. If the unethical task requires greater time and effort to complete, will crowdworkers be more inclined to agree to its use in exchange for compensation?
5. Are crowdworkers more willing to provide assistance with a partial unethical contribution (answering a single question) as compared with an entire unethical contribution (answering the complete homework assignment or exam)?

To answer these, we conducted two separate experiments using MTurk between April 2 and May 4, 2012. We created a website for this experiment called Homework Assist1. In the MTurk task, participants were told the website was a broker for academic assistance much like sources discussed by Lancaster and Clark. Our experiments were conducted as a deception study and we received full IRB approval in advance.

A. Initial Experiment – Answering a Single Question

In the first experiment, participants provide an answer to a single question and were paid $0.25 for their time and effort. We listed 200 tasks on MTurk consisting of a single question from one of two problem types: short essay or probability.

Short-essay question types solicited responses of a few paragraphs in the subjects of history, journalism, or economics. They would be short-essay questions that would be asked in an introductory level college course.

Probability question types solicited answers to problems that may be found in an introductory college level probability course. These questions required not only a numerical answer, but an explanation of the steps used to solve the problem.

We divided workers into two groups of equal size, containing 200 participants each – a control group and a treatment group. If a worker initially accepted the task, but did not contribute (i.e., go beyond the first screen), we did not

1 http://www.hwassist.org
count them as participants in our study and relisted the task. For participants to receive compensation, we required participants to provide a randomly-generated token into MTurk.

The problem type was listed in the task description, and participants selected either a short essay question (history, journalism, or economics) or a probability question prior to beginning work on the task. Examples of short essay questions are:

- How has the image of flight attendants changed in airline advertisements over the past half century?
- How has muckraking changed in the last century? Include a discussion of the pioneers of yellow journalism in your answer.
- Describe how the public infrastructure of the resource-rich country of Papua New Guinea has hampered its ability to develop its economy.

Examples of probability questions are:

- In country X, 65% of people wear a seatbelt while driving. If three people are chosen at random while driving, what is the probability that all three of them are wearing a seatbelt?
- Three cards are chosen at random from a deck without replacement. What is the probability of getting a queen, a seven and an ace (of any suit)?
- A class consists of 55% men and 45% women. Of the men, 35% wear glasses, while 45% of the women wear glasses. If a student is chosen at random and is found to wear glasses, what is the probability that the student is a man?

The control group was provided with a unique link on the hwassist.org website. Participants were asked to provide a solution to a single question in a text box (Figure 2 provides an example) and press ‘submit’. They were then taken to a second screen, Screen 2, (see Figure 3) and given information about the intended purpose of their work – either for a homework assignment or an exam – and were asked if their answer could be used for this purpose. Once they provided their choice, they were taken to a Debriefing Statement (not shown). No incentives were offered to the control group.

The treatment group was provided a different link to the same website. After supplying an answer to a question and pressing ‘submit’, participants were taken to Screen 2 (Figure 3). On Screen 2 participants were told the intended use for their answer and asked if they allowed its use for this purpose. If they did not agree, they were taken to Screen 3 (Figure 4). On Screen 3, they were provided another request for permission, but offered an additional incentive that doubled their original expected payment to $0.50 ($0.25 payment and $0.25 bonus). If participants still refused the offer, they were taken to Screen 4 (Figure 5) and provided yet another incentive, quadrupling their original $0.25 payment to $1.00. If participants gave approval in Screens 2 or 3, and regardless of the Screen 4 decision, they were immediately taken to the Debriefing Screen. The flow for the treatment group is given in Figure 1.

In our study, the ability to go back to a previous screen was disabled, though participants could quit the task at any time. In addition, participants had no advance notice of any additional incentives if they refused to allow the use of their solution on any one screen. For all participants, we tracked the IP address for all participants, logged screen information and elapsed time taken on each screen. Since this was a deception study, workers were only allowed to participate once. We monitored IP logs, MTurk results, and the MTurk worker forums to evaluate information communicated about our task to other workers.
B. Second Experiment – Multiple Assignment

Our second experiment was conducted to examine if a greater investment of time and effort or a submission that represented a larger portion of an exam or homework assignment (multiple assignments) would affect the decision to approve its use. We had 400 participants (200 control, 200 treatment) for this experiment. Participants selected from two tasks: either four short essay questions (economics, journalism, or history) or four probability questions. They were required to provide answers that fully explained their response, requiring a significant investment of time. This multiple assignment task was listed on MTurk for $1.00 for time and effort. On Screen 2, participants were told the purpose was for use as an entire homework assignment or an entire exam. The control group went directly to the Debriefing Statement as in the single assignment experiment. For the treatment group, the flow is identical to that used in the single assignment experiment (Figure 1), except Screen 3 provided a possible bonus of $1.00, for a total payment of $2.00 and Screen 4 provided a possible bonus of $3.00, for a total payment of $4.00.

V. RESULTS AND DISCUSSION

Across all experiments we had 1302 participants visit Screen 1; however, only 800 participants went beyond the first screen (400 control, 400 treatment). This conversion rate was consistent across all assignment types and treatment groups. The 502 participants who chose not to participate likely did so due to lack of interest (the task was too hard or required too much effort for the compensation) or because of technical issues. Tables 1 and 2 provide the counts for single and multiple assignment tasks, respectively, broken out by design type (short essay or probability). Tables 3 and 4 provide counts for single and multiple assignment tasks, respectively, broken out by problem type (homework or exam).

### Table 1. Counts by Screen, Examining Single Assignment Types for Short Essay and Probability Problem Designs

<table>
<thead>
<tr>
<th></th>
<th>Treatment</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Short Essay</td>
<td>Prob</td>
</tr>
<tr>
<td>Arrived at website</td>
<td>156</td>
<td>151</td>
</tr>
<tr>
<td>'Submit' @ Screen 1</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>'Agree' @ Screen 2</td>
<td>72</td>
<td>74</td>
</tr>
<tr>
<td>'Agree' @ Screen 3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>'Agree' @ Screen 4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Reject All Offers</td>
<td>24</td>
<td>23</td>
</tr>
</tbody>
</table>

### Table 2. Counts by Screen, Examining Multiple Assignment Types for Short Essay and Probability Problem Designs

<table>
<thead>
<tr>
<th></th>
<th>Treatment</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Short Essay</td>
<td>Prob</td>
</tr>
<tr>
<td>Arrived at website</td>
<td>180</td>
<td>166</td>
</tr>
<tr>
<td>'Submit' @ Screen 1</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>'Agree' @ Screen 2</td>
<td>85</td>
<td>88</td>
</tr>
<tr>
<td>'Agree' @ Screen 3</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>'Agree' @ Screen 4</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Reject All Offers</td>
<td>12</td>
<td>6</td>
</tr>
</tbody>
</table>

### Table 3. Counts at Each Screen, Examining Single Assignment Types for Homework and Exam Problem Types

<table>
<thead>
<tr>
<th></th>
<th>Treatment</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HW</td>
<td>Exam</td>
</tr>
<tr>
<td>Arrived at website</td>
<td>156</td>
<td>151</td>
</tr>
<tr>
<td>'Submit' @ Screen 1</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>'Agree' @ Screen 2</td>
<td>72</td>
<td>72</td>
</tr>
<tr>
<td>'Agree' @ Screen 3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>'Agree' @ Screen 4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Reject All Offers</td>
<td>24</td>
<td>25</td>
</tr>
</tbody>
</table>

### Table 4. Counts at Each Screen, Examining Multiple Assignment Types for Homework and Exam Problem Types

<table>
<thead>
<tr>
<th></th>
<th>Treatment</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HW</td>
<td>Exam</td>
</tr>
<tr>
<td>Arrived at website</td>
<td>180</td>
<td>166</td>
</tr>
<tr>
<td>'Submit' @ Screen 1</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>'Agree' @ Screen 2</td>
<td>85</td>
<td>88</td>
</tr>
<tr>
<td>'Agree' @ Screen 3</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>'Agree' @ Screen 4</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Reject All Offers</td>
<td>11</td>
<td>5</td>
</tr>
</tbody>
</table>
The percentage of participants allowing permission on Screen 2 was higher than what we had anticipated. The most suitable comparison we found in the literature is the 43% admitting to “unauthorized collaboration” in the recent CAI study [30]. Using this as our baseline, we found a significant increase in “unauthorized collaboration” using the crowdsourcing workers in our study as compared with this CAI-determined percentage across four factors using a $\chi^2$ goodness-of-fit test: group type (treatment or control), assignment type (single or multiple assignment), and problem type (homework or exam) and problem design (probability question or single essay), (p ≤ 0.001 in all cases). Examining differences between groups, we found there was no significant difference between the acceptance rates of the initial agreement on Screen 2 across group type, assignment type, problem type and problem design, using a $\chi^2$ goodness-of-fit test (p = 0.36, p = 0.08, p = 0.33, p = 0.29, respectively).

None of the four factors we studied provided a better environment for unauthorized collaboration over any other. However, when comparing to what was reported in the CAI study (M_{CAI} = 0.43), the percentage agreeing to unauthorized collaboration in online labor markets (M_{OLM} = 0.79) is considerably higher than the CAI study indicated. Although we are comparing a different population subset than the CAI study, our investigation indicates our participants are more likely to participate in unauthorized collaboration by nearly a factor of two-to-one, which shows the growth potential of using online markets to facilitate unauthorized academic assistance.

Next we examine the effects of additional incentives (agreement at Screens 3 and 4). We evaluate the incentive conversion rate – the conditional probability of approving use in Screen 3 or 4, given they did not approve in Screen 2 ($\sim S_2$). 

$$P(\text{Incentive} | \sim S_2) = \frac{P(\text{Incentive} \land \sim S_2)}{P(\sim S_2)}$$

We found a significant difference in the incentive conversion rate for assignment type (single or multiple assignment), using a $\chi^2$ goodness of fit test (p = 0.003); however, we did not find a significant difference for either problem type or problem design (p = 0.017, p = 0.024 respectively). Participants that answered the four assignment initially did not agree in Screen 2 may have found a “fair price” for their time and effort invested in Screens 3 or 4. As described by Mason and Watts in [22], this may be due to an anchoring effect – the perception that their effort was worth more than our initial offer. Another explanation may reflect loss aversion, an aspect of prospect theory first defined by Kahneman and Tversky [18]. Loss aversion describes a greater preference for avoiding losses (the compensation given up by refusing the offer) relative to seeking gains (maintaining one’s principles in an anonymous setting). For some participants, the larger potential loss may trump the gain from maintaining one’s ethics in an anonymous setting.

Last, we examine the difference in the time taken to evaluate the incentive offers in Screens 3 and 4, as compared to the time taken on Screen 2. We infer that more time spent on one of these screens implies more indecision; in other words, we would expect the increase in time would be correlated with the increase in incentive. Performing an analysis of variance, we found a significant difference in the time taken on the three screens [F(2,335) = 76.73, p ≤ 0.001]. A Bonferroni post-hoc evaluation showed a significant difference between the two assignments on all three screens (Screen 2, M = 15.36 sec, SD = 4.16; Screen 3, M = 26.70 sec, SD = 4.97, Screen 4, M = 38.85 sec, SD = 5.04). The correlation between time taken and incentive was slightly positive for all participants (R = 0.31, N = 337, R^2 = 0.096). The strongest correlation occurred for the assignment type factor for multiple assignments (R = 0.52, N = 184, R^2 = 0.27).

We also examined the IP address to see if location played a part in the decisions made by participants. Using the IP address technique only examines the physical location, not the nationality, of the participants. Of the 800 IP addresses recorded, we were able to determine the location for 787 of them. Workers in 17 countries participated in our study, with the five largest contingents being 46% in India, 22% in the United States, 7% each in the Philippines and Canada, and 5% in Bangladesh. We also looked at the subset of 648 participants that allowed the use of their solutions (we could determine location for 642 of them). The five largest contingents were 41% in India, 15% in the United States, 5% from the Philippines, and 4% each from Canada and Bangladesh.

VI. CONCLUSION

A greater number of students rely on the Internet to obtain assistance with their coursework than ever before. Recent surveys find that this increased reliance, coupled with an increase in online course offerings at universities, has fostered a rise in student cheating incidents. Online labor markets have maintained rapid growth since their introduction, primarily because they provide a clearinghouse for requestors to have tasks performed quickly and inexpensively and with reasonable quality. These advantages create a significant potential for unauthorized assistance on homework assignments and exams for students. We conducted a study on ethics in online labor markets to evaluate if the crowd could be used to provide assistance.

We found that 79.0% of crowdworkers agreed to provide their work for assistance on exams or homework assignments without additional incentive; this increased to 81.4% when additional incentives were offered. Although the percentage agreeing to allow use was higher than expected, the actual effects of the additional incentives we offered were relatively small; in most cases, most crowdworkers who had ethical concerns were not swayed by these incentives. One exception was those crowdworkers answering multiple questions; it may be the amount of compensation that would be refused more...
than offset the gain of maintaining ethics in an anonymous setting, as loss aversion, a part of prospect theory suggests.

Crowdworkers allowed use of their solutions for a homework assignment at the same rate as they did for an exam (which is arguably a larger breach of ethics). Likewise we found no difference between participants who chose write a short essay and those who chose to solve a probability problem.

We found that crowdworkers spent more time to make a decision for the use of their work as the value of the incentives presented to them increased. This was particularly true for work on multiple assignments since more compensation was potentially at risk.

After seeing how quick and easy it is to obtain assistance through online labor markets, what can be done to help prevent this unethical behavior? LoShiavo and Shatz [20] point to a few suggestions. Immediate (i.e., perceived social distance) of the instructor is one key factor that influences compliance with honor codes in courses. Lower immediacy brings lower rates of cheating. Also, students who signed an honor code were about 30% less likely to cheat than those who did not. Therefore, given the ease of obtaining unauthorized help, as illustrated in our experiments, the best answer may be to focus on greater involvement by instructors and implementing and enforcing an honor code – two ideas that have been in vogue long before online labor markets were in existence.

There are several directions in which we could extend our work. We did not evaluate the correctness of the answers supplied; however, we did reject those answers that were poorly supported or were clearly off target. In future work, we would like to perform additional evaluations of the quality of the answers supplied relative to the factors we considered. Additionally, we plan to investigate if an “auction format” for unauthorized assistance work might have an effect on incentives and the likelihood of cheating.

REFERENCES