Improving Reactions to Rejection in Crowdsourcing through Self-Reflection

Edixhoven, Tom; Qiu, Sihang; Kuiper, Lucie; Dikken, Olivier; Smitskamp, Gwennan; Gadiraju, Ujwal

Improving Reactions to Rejection in Crowdsourcing Through Self-Reflection

Tom Edixhoven
Delft University of Technology
Delft, Netherlands
T.F.Edixhoven@student.tudelft.nl

Sihang Qiu
Delft University of Technology
Delft, Netherlands
S.Qiu-1@tudelft.nl

Lucie Kuiper
Delft University of Technology
Delft, Netherlands
L.A.Kuiper@student.tudelft.nl

Olivier Dikken
Delft University of Technology
Delft, Netherlands
O.D.F.Dikken@student.tudelft.nl

Gwennan Smitskamp
Delft University of Technology
Delft, Netherlands
G.M.Smitskamp@student.tudelft.nl

Ujwal Gadiraju
Delft University of Technology
Delft, Netherlands
U.K.Gadiraju@tudelft.nl

ABSTRACT

In popular crowdsourcing marketplaces like Amazon Mechanical Turk, crowd workers complete tasks posted by requesters in return for monetary rewards. Task requesters are solely responsible for deciding whether to accept or reject submitted work. Rejecting work can directly affect the monetary reward of corresponding workers, and indirectly influence worker qualifications and their future work opportunities in the marketplace. Unexpected or unwarranted rejections therefore result in negative emotions and reactions among workers. Despite the high prevalence of rejections in crowdsourcing marketplaces, little research has explored ways to mitigate the negative emotional repercussions of rejections on crowd workers. Addressing this important research gap, we investigate whether introducing self-reflection at different stages after task execution can alleviate the emotional toll of rejection decisions on crowd workers. Our work is inspired by prior studies in psychology that have shown that self-reflection on negative personal experiences can positively affect one's emotion. To this end, we carried out an experimental study investigating the impact of explicit self-reflection on the emotions of rejected crowd workers. Results show that allowing workers to self-reflect on their delivered work, especially before receiving a rejection, has a significantly positive impact on their self-reported emotions in terms of valence and dominance. Our findings reveal that introducing a self-reflection stage before workers receive acceptance or rejection decisions on submitted work, can help in positively influencing the emotions of a worker. These findings have important design implications towards fostering a healthier requester-worker relationship and contributing towards the sustainability of the crowdsourcing marketplace.

CCS CONCEPTS

• **Information systems** → **Crowdsourcing**; • **Applied computing** → **Law, social and behavioral sciences**; • **Human-centered computing**;

KEYWORDS

crowdsourcing, microtask, rejection, emotion, mood, self-reflection

1 INTRODUCTION

Popular crowdsourcing marketplaces such as Amazon's Mechanical Turk (MTurk) have attracted an increasing number of people, working as full-time crowd workers [8, 9]. Workers plying their trade in such marketplaces, earn their living by completing human intelligence tasks (HITs) published by requesters [22]. The challenges that workers face on a regular basis due to platform dynamics and power asymmetry with task requesters, have been studied widely and are well-documented [27]. An important result of the power asymmetry is that requesters are solely responsible for deciding whether to accept or reject submitted work. Work rejections are ingrained into the fabric of crowdsourcing marketplaces, and often play a central role in defining worker experiences [13, 32].

Several works in the past have explored and established the prevalence of unreliable worker behavior in microtasking platforms [10, 14, 24, 26]. Rejection in crowd work can be seen as a necessary action to ensure high-quality outcomes, making it possible to provide a signal to the workers who perform sub-optimally and reward the workers who perform well. However, rejecting submitted work can have grave consequences on worker reputation and thereby their future access to work opportunities. Requesters may reject submitted work due to a variety of reasons − certain performance thresholds (e.g. based on worker accuracy), workers failing honeypot/attention-check questions spread throughout the task, or due to post-hoc assessments. Inappropriate rejection however, can also negatively affect task requesters. While the requesters are largely in control of making decisions pertaining to accepting or rejecting submitted work, the workers can also take

action when they do not agree with these decisions, by leveraging external tools [25, 37]. Workers can decide to stop working for certain requesters and share their experiences with other workers via online forums [51, 52]. Communication and exchange of experiences among workers can result in checking unreasonable and unfair task requesters.

Despite the fact that rejections are a part and parcel of crowd work, little research has explored ways to improve the impact of rejections on crowd workers. Recent work has shown that rejections can seriously influence workers' emotions, and workers in general are hurt by rejections in crowdsourcing [13]. The authors attempted to present social comparative explanations in order to improve the workers' reaction to rejection, albeit without success. We aim to address this research gap in this work, by drawing inspiration from prior works in psychology. It has been shown that facilitating self-reflection can lead to low depression [17, 40], and positive self-images [23]. Although workers may always be able to self-reflect on their work, we hypothesize that explicitly facilitating their reflection on different facets of their work, through questions regarding a particular task they performed, can potentially improve their reactions to rejection. However, little is currently understood about the point at which self-reflection can be effective in crowdsourcing workflows. Does self-reflection before receiving rejection decisions on submitted work, improve the emotions of workers? Or is self-reflection more effective after receiving rejection decisions? In this paper, we aim to fill this knowledge gap by addressing the following research questions.

RQ1 *Can inducing explicit self-reflection improve the emotional reactions of crowd workers to rejection?*
RQ2 *When is self-reflection effective in reducing the emotional toll of rejections on crowd workers?*

Original Contributions. Inspired by prior works in psychology and education about self-reflection, we propose a new approach towards improving worker reaction towards rejection by allowing for *explicit self-reflection* after a task has been completed. We discuss our approach towards implementing explicit self-reflection into HITs, and propose an experiment to evaluate its effectiveness. The self-reflection component of our experiment is designed to be modular. This allows the component to be widely applicable and easy to plug into existing HITs. We set up three experimental conditions, namely control (no self-reflection), self-reflection before rejection, and self-reflection after rejection. On MTurk, we recruited 460 workers in total, to have at least 50 valid rejected workers for each condition. Results revealed that using self-reflection, especially before rejection, can effectively improve workers' reactions to the rejection, particularly with regard to their happiness. This work has important implications in improving workers' negative reactions, enhancing a healthy requester-worker relationship, and building a harmonious online work environment in crowdsourcing marketplaces.

2 RELATED WORK

2.1 Power Asymmetry and Invisible Labor
A bulk of research in the realm of crowdsourcing over the last decade has paid a substantial amount of attention to improving

quality-related outcomes. Relatively little work however, has focused on understanding the challenges faced by crowd workers and characterizing the hidden factors that shape the quality of their work. Power asymmetry and invisible labor have been shown to be widely prevalent in crowdsourcing marketplaces [18, 25, 31, 38]. Within the dynamics of controlling for quality, task requesters enjoy the most power in crowdsourcing marketplaces. Workers have been observed to invest huge efforts on underpaid tasks [22], struggle to make a minimum wage [46], or even get rejected without being rewarded. Addressing this power imbalance, Irani et al. developed Turkopticon [25], a system that helped workers to evaluate task requesters and check unfair practices. To improve trust dynamics and empower crowd workers, Gaikwad et al. proposed a self-governed crowdsourcing marketplace called Daemo [15]. More recently, Fan et al. designed a novel rewarding mechanism for crowdsourcing marketplaces that enables crowd workers to share risks and earn a fair reward [12]. Recent work has shown that such tools and scripts based on current crowdsourcing marketplaces are effective in addressing worker invisibility issues by promoting both productivity and well-being [48]. Complementing such existing works, in this study we specifically aim to address workers' reactions to rejection of their work, and investigate how we can help improve their emotions in such vulnerable situations.

2.2 Worker Moods, Emotions, and Reactions to Rejection
A prior study in educational psychology showed that people could be more productive if they were happier [41]. In the realm of microtask crowdsourcing, researchers have also found that crowd workers in a pleasant mood tended to produce outcomes of higher quality, exhibited better worker engagement, and perceived less cognitive load [13, 35, 50, 54]. It is useful to note however, that mood and emotion are neither the same, nor entirely independent, as explained in previous works [1, 7, 45]. Although both mood and emotion are valenced affective responses, prior work has elaborately discussed the difference between the two: Firstly, moods last longer than emotions [7, 45]. Secondly, emotions are always targeted towards an event/person/object, while moods are globally diffused [17]. Emotions are more sensitive, representing workers' subjective perception right after the rejection, while moods reflect workers' feelings throughout the whole task. Different tools are needed to assess both. Some of the relevant instruments to measure moods and emotions are: Pick-a-Mood [7], the Self Assessment Manikin [3], the Affective Slider [2], and the Achievement Emotions Questionnaire [33]. In our study, considering the length of a crowdsourced microtask and the suitability of existing instruments, we assess workers' moods and emotions using Pick-a-Mood and the Self-Assessment Manikin respectively.

Gadiraju et al. [13] conducted experiments to measure the emotional impact of rejections on workers and explored methods to reduce the emotional toll of rejections. Workers in pleasant moods were found to exhibit relatively better task performance. Results revealed that rejected workers exhibited the least joyful emotions, in-line with previous research [49, 53]. Apart from crowdsourcing, research on improving reactions to rejections has been explored in the field of job applications [16, 39]. In contrast to the findings

of [13], applicants reacted more positively on receiving explanatory feedback along with job application rejections, and different explanations influenced the perception of fairness of the rejections. The goal of our work is to understand how the reactions to rejections can be improved to enhance the overall experience of crowd workers. Note that our work was limited to exploring rejections that are warranted (i.e., when workers fail to meet a reasonable accuracy threshold). We ensured that our study complied with the institutional ethics regulations for human-subjects research.

2.3 The Role of Self-Reflection

Prior work in behavioral psychology has shown a correlation between a high self-reflection score and a low depression score [17, 40]. Positive self-images have been reported after activities of self reflection [23]. Others have shown that the influence of self-reflection can vary depending on the cultural background of individuals [43]. To the best of our knowledge, there are no scientifically formulated instruments to trigger or measure self-reflection of workers in crowdsourcing tasks. Most self-refection research is focused on evaluating one's own thoughts, feelings, and behavior in a participant's life as a whole [4, 36, 42]. Self-reflection processes have been widely applied in the fields of education and learning, to improve learning effects, self-efficacy, and achievement [6, 20]. van Velzen proposed a questionnaire to assess students' use of self-reflective thinking [44]. Similarly, in the field of education, Pintrich et al. [34] designed a set of questionnaires to evaluate the effects of motivated strategies in learning, which also includes an aspect of self-reflection. Moreover, research in sport has leveraged self-reflection, and designed corresponding questionnaires [5, 21, 29]. Inspired by these prior works, to help workers self-reflect on their task performance, we designed a questionnaire based on self-reflection in the field of education.

3 STUDY DESIGN

3.1 Facilitating Self-Reflection

We design the self-reflection step to be modular, so that it can be plugged into HITs with a minimum development overhead. As an essential part of the study, to facilitate self-reflection, workers are asked to respond to a series of 'Yes/No', and Likert-type questions, as shown in Table 1. The questions were designed based on previous work exploring self-reflection, to allow workers to reflect on the HITs they just completed, presenting them an outlet to channel their emotions. These questions are presented in the listing below.

Workers are asked to respond to the self-reflection questions either before or after completing the tasks, depending on the experimental conditions (described below). The questions are designed to help find the root cause of success or failure, as seen by the worker. Therefore, through reflection it is possible for workers to blame-shift, to criticize their own performance, justify mistakes, express emotions such as frustration and pride, and to comment on task fairness, appropriateness, and meaningfulness.

3.2 Measures and Workflow

Figure 1 presents an overview of the workflow in our study. Similar to prior work that explored ways to mitigate negative emotions due to rejections in microtask marketplaces [13], we measured

Table 1: Self-reflection questions.

(1)	**I understood the task.** □ Yes □ No
(2)	**I spent an appropriate amount of time on the task.** □ Yes □ No
(3)	**I worked to the best of my abilities on the task.** 5-point Likert scale – *1: Disagree* to *5: Agree*
(4)	**I have the abilities needed to succeed in this task.** 5-point Likert scale – *1: Disagree* to *5: Agree*
(5)	**I have completed similar tasks before.** □ Yes □ No
(6)	**I understand why I was rejected.** 5-point Likert scale – *1: Disagree* to *5: Agree*
(7)	**The task was well explained.** 5-point Likert scale – *1: Disagree* to *5: Agree*
(8)	**The task was of a high difficulty.** □ Yes □ No
(9)	**The task was achievable within the available time.** □ Yes □ No
(10)	**The task acceptance/rejection criteria are reasonable.** 5-point Likert scale – *1: Disagree* to *5: Agree*
(11)	**My general opinion on the task:** Open-ended
(12)	**What I would have liked to see improved next time I do this task:** Open-ended

worker moods before they began working on the tasks, and worker emotions in the last step of the workflow.

1) Measuring Worker Moods. We use the Pick-A-Mood scale (PAM) to measure worker moods. PAM is a character-based pictorial scale for mood expression and measurement [7]. Workers are first asked to self-report their mood (cf. **a1**), by selecting one out of 9 images that best corresponds to their current mood. The images represent 8 non-neutral moods and a neutral mood. We capture worker moods to evaluate whether the mood of the workers (i) influences their emotional response to rejection, or (ii) mediates the effectiveness of the self-reflection process. The PAM instrument can be used with different pictorial characters representing male, female, and gender-neutral robot images. As shown in Figure 2 (a), we employ the gender-neutral version of PAM to avoid any potential bias.

2) Microtask Execution. After self-reporting their moods, the workers are asked to complete a batch of 10 microtasks (cf. **a2**). The task itself is described in detail in subsection 3.4. Workers are informed of the potential to complete an additional 10 tasks and earn the associated rewards. Based on the performance of workers, measured using output accuracy, the workers are either approved or rejected from partaking in the additional set of 10 tasks. In line with the typical quality control thresholds, if workers are able to achieve an accuracy of at least 70%, they are accepted into the next stage with additional tasks, or rejected otherwise.

3) Task Acceptance/Rejection. On completing the 10 microtasks, workers are displayed a message that conveys whether they have been accepted or rejected. If the performance of workers is below the 70% threshold, the workers are shown a short message that their performance was not satisfactory, and that they would not receive access to the additional paid tasks. They are then asked to continue
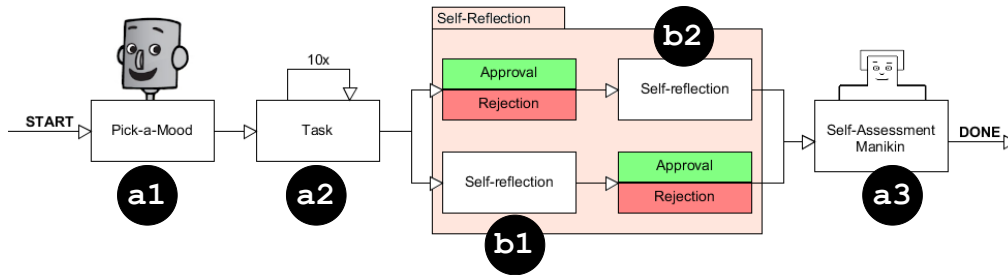
Figure 1: An overview of the workflow across the three conditions in our study: `Control`, `SR-Rej` and `Rej-SR`. Note that in the `Control` condition, there is no self-reflection step.

(a) PAM

(b) SAM
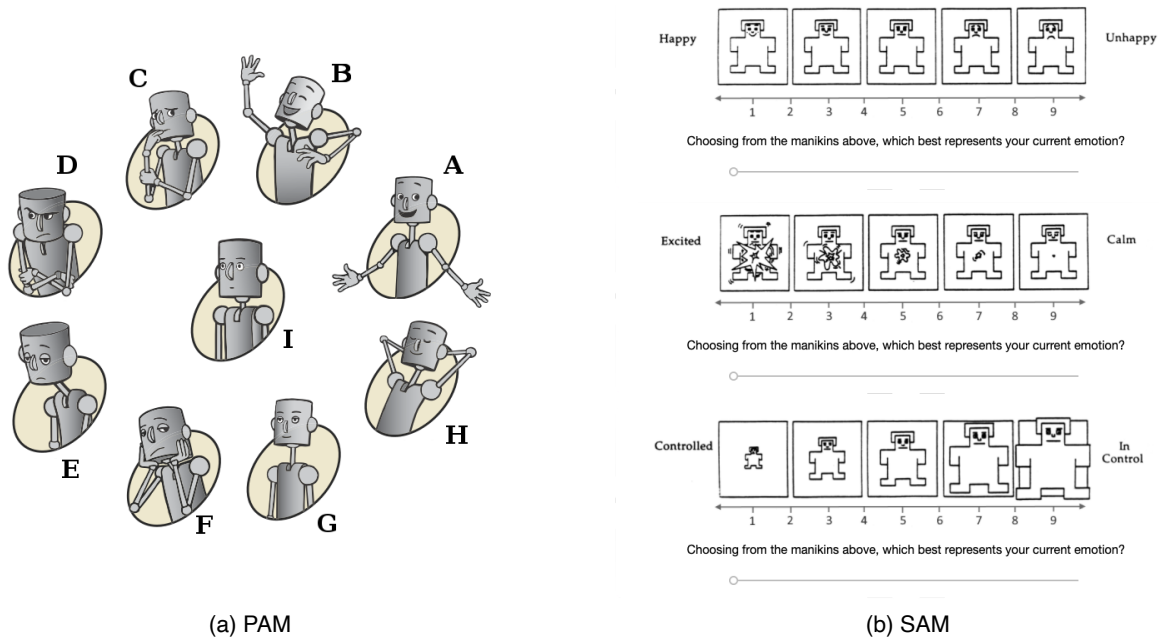
Figure 2: (a) Pick-A-Mood scale (PAM) for measuring worker moods; and (b) Self-Assessment Manikin (SAM) for measuring worker emotions.

to the next stage of the workflow. Otherwise, workers receive the message that their performance was satisfactory and that they could access the additional tasks. Through clear instructions, workers were made aware of the fact that their access to additional work in our task was dependent on their performance. This is consistent with how future access to work is dependent on the historical worker performance in popular crowdsourcing platforms such as MTurk or Appen.

4) Measuring Worker Emotions. Finally, workers are asked to self-report their emotions using the Self-Assessment Manikin (cf. **a3**). The SAM is also a pictorial assessment to track the personal responses of individuals to an effective stimulus. This is done by presenting three dimensions (valence, arousal and dominance) with five images spanning the spectrum of the dimensions happy-unhappy, excited-calm, controlled-in control. Workers can then indicate their

current emotion by selecting a value on the scale of each dimension, using a slider (ranging from 1 to 9). As shown in Figure 2 (b), the lower the valence score is, the happier the worker is; the higher the arousal score is, the more calm the worker is; and a higher dominance score represents that the worker has a stronger sense of control.

3.3 Experimental Conditions

To investigate the effect of self-reflection on the emotional reactions of workers to rejection, and address **RQ1**, we consider three distinct experimental conditions: 1) **`Control`**, 2) Self-Reflection before Rejection (hereafter referred to as **`SR-Rej`**), and 3) Self-Reflection after Rejection (hereafter referred to as **`Rej-SR`**). This is illustrated in Figure 1, within the Self-reflection stage of the workflow (cf. **b1**,

b2). The distinction between the conditions `SR-Rej` and `Rej-SR`, is made to address **RQ2** and understand whether the point at which self-reflection is induced within the workflow has a significant impact on the emotional reactions of the workers.

Control. The `Control` condition is used to establish the baseline emotional reactions of workers, to acceptance and rejection decisions that are meted out. In the control condition, there is no self-reflection that is explicitly induced.

Self-Reflection before Rejection. The `SR-Rej` condition allows workers to reflect on their work before they know whether it will be accepted or rejected. This enables the workers to reflect on whether they think the work they delivered was satisfactory, and whether the task that they were given was 'doable' or 'fair'. If the answer to either of the two aforementioned questions was no, the reflection can potentially allow them to brace for the emotional impact of being rejected. The workflow for this condition is similar to the workflow associated with the `Control` condition. The only difference is the introduction of another step between stage **2) Microtask Execution** and **3) Task Acceptance/Rejection** of the workflow — the *Self-Reflection*. Once the workers have performed the assigned set of 10 microtasks, they are given the self-reflection questionnaire, to allow them to self-reflect on their delivered work (as discussed in Section 3.1). Note that the 6th question is hidden in this condition.

Self-Reflection after Rejection. Contrary to `SR-Rej` condition, the `Rej-SR` condition enables workers to reflect on their work after they have been either rejected or accepted. In the case that work has been rejected, workers will be able to reflect on the work they have delivered and explore why their work was unsatisfactory. Whether this transpires, and such an understanding does result in an improved emotional response to the rejection remains to be investigated. The workflow of this condition is identical to the `SR-Rej` condition, except that the *Self-Reflection* stage is introduced between stages **3) Task Acceptance** and **4) Measuring Worker Emotions**, rather than between stage 2 and stage 3. This means that the workers will know whether their performance was deemed sufficient before their self-reflection.

3.4 Task Design

In order to study the reactions to rejections in crowd work, we needed to consider a task with a reasonable level of difficulty, that had the scope to result in worker performance below the 70% accuracy threshold. We therefore employed the publicly available dataset of information finding tasks from recent work [13]. This manually curated dataset consists of information finding tasks that ask workers to find the middle names of famous people. The tasks are divided into three levels of difficulty, based on the number of the disambiguation constraints that workers have to consider. Examples of such aspects are the period in which a said celebrity was active, and their field of work. An example of a *Level − I* difficulty task would be finding the middle name of 'Daniel Craig', where as a *Level − III* task would be finding the middle name of the Ice Hockey player 'Brian Smith', who was active in 1972 [13]. For the 10 microtasks that workers were asked to complete (cf. **a2**), *Level−III* difficulty tasks were selected exclusively. Using a pilot study, this

difficulty level was shown to challenge the workers enough to result in an average worker accuracy of lower than 70%, while still allowing roughly 30% of workers to complete the tasks successfully (e.g. with an accuracy of 70% or above).

4 EXPERIMENTAL SETUP

4.1 Platform

We carried out our experimental study on Amazon Mechanical Turk, considering that it is one of the most popular crowdsourcing platforms. We continued recruiting workers in each experimental condition, until we had at least 50 workers who did not manage to perform with an accuracy of 70% in the set of 10 microtasks. This was to ensure that we had a fair sample of workers, lending statistical strength to our comparisons across experimental conditions. All workers that participate in our experiment are rewarded regardless of their accuracy. When a worker successfully completes the task including PAM, 10 microtasks, and SAM, the worker immediately receives a base payment of USD 0.75 (calculated according to our pilot experiment to meet the minimum hourly wage of the US). Workers who perform with an accuracy of above 70% and complete additional microtasks, are rewarded additionally with USD 0.50 using the *bonus* function provided by the MTurk platform.

4.2 Quality control

Each worker is only allowed to participate in our study once throughout our entire experiment, to prevent learning bias. Extra Javascript code is used to acquire the unique Worker IDs, and record them on our own server. If a Worker ID already exists in the database, the task page turns to blank and kindly reminds the worker to "exit" the HIT since they had already completed the task. Furthermore, we exclude workers having 0% accuracy owing to obvious unreliable behavior. Correctly answering at least 1 question (out of 10) is the minimum criterion required to include the corresponding worker in the analysis. Furthermore, we carried out a manual inspection of the collected data to flag potential unreliable workers. It is important to note that we did not use any qualifications (such as worker approval rating, Master workers, etc.) to restrict participation. This design choice was motivated by our need to study the reactions to rejection of workers, and explore the effectiveness of self-reflection in general — not only with respect to a specific group of workers. Note that we were interested in ensuring reliability, but less concerned with accuracy of workers as long as they were completing the tasks with genuine intentions.

5 RESULTS AND ANALYSIS

All raw data, code, and analysis are made available online to promote open science. [1]

5.1 Worker Accuracy

We kept recruiting workers in our experiment and meanwhile manually excluded workers who exhibited unreliable behavior [11], until the number of rejected valid workers (who do not show unreliable behavior whilst the work accuracy is less than 70%). As described before, we use high-complexity tasks to increase the "rejection rate"

---

[1]https://osf.io/ub7d6/?view_only=a7a4bc39fd4844da9e24870327a622e7

of our experiment. Note that "rejection" in our experiments implies that workers were not given access to more microtasks in the HIT, and cannot acquire the corresponding additional bonus as a consequence of their performance. This design choice was motivated by the ethical constraints surrounding rejecting HITs submitted by workers, which can have negative repercussions. We therefore accepted their HITs on the MTurk platform, and fairly paid for their delivered work regardless of the quality of their work.

In case of the `Control` condition, we finally recruited 157 workers in total. Of the 157 workers, 50 workers were manually excluded due to malicious behaviors (0% accuracy or obvious treating activities), and 57 workers were accepted (accuracy $\geq$ 70%), which results in 50 rejected workers valid for analysis. The average accuracy of the 50 rejected workers was found to be 47.8% ($\pm$14.6%). In the `SR-Rej` condition, we recruited 148 workers in total. Of the 148 workers, 51 workers were manually excluded and 46 workers were accepted. Eventually, we have 51 valid rejected workers in this condition, and the average accuracy of the rejected workers was found to be 48.0% ($\pm$11.6%). Similarly, we recruited 155 workers in the `Rej-SR` condition. 61 out of 155 workers were excluded after our manual inspection, and 44 were accepted. Finally, we have 50 valid rejected workers whose average accuracy is 46.8% ($\pm$ 14.1%). Across these three conditions, we found that the percentage of valid rejected workers ranged between 32% to 34%. A one-way between subjects ANOVA revealed a lack of statistically significant difference in the average accuracy of rejected workers across the three conditions ($\alpha = 0.05$).

5.2 Worker Emotion Analysis

Table 2 shows emotion scores measured by SAM of rejected workers in three dimensions (valence, arousal, and dominance) and three experimental conditions. We apply ANOVA tests ($\alpha = 0.05$), with the null hypothesis that the mean value is the same over all three conditions (`Control`, `SR-Rej`, and `Rej-SR`), to explore the impact of self-reflection on worker emotions. We found a significant difference across the conditions in the dimension of valence ($p = 0.034$, $\eta^2 = 0.045$). However, we found no significant difference in dimensions of arousal ($p = 0.071$, $\eta^2 = 0.034$) and dominance ($p = 0.146$, $\eta^2 = 0.026$).

Considering reasonably small p-values across the three dimensions, we further conducted pairwise independent t-tests to find which experimental condition was the most effective in terms of improving workers' emotions. To account for Type-I errors due to multiple t-tests ($\alpha = 0.05$), we used Bonferroni correction. Our findings are described below with respect to each dimension.

Valence. In terms of the dimension of valence, which can be interpreted as the degree to which crowd workers feel happy/unhappy after being rejected, we found that using self-reflection before rejection (`SR-Rej`) could significantly decrease the valence score (meaning that workers felt happier), in comparison with the control condition ($p = 0.012$, Cohen's $d = 0.516$). Self-reflection after rejection (`Rej-SR`) could also reasonably enhance workers' happiness compared to the `Control` condition, however, we found no statistical significance in this regard ($p = 0.102$, Cohen's $d = 0.333$).

Arousal. The dimension of arousal explores whether workers feel calm or excited. We found no significant difference after Bonferroni

correction. Arousal scores of the `Control` condition and `Rej-SR` condition are similar. However, results show that, workers who self-reflected before the rejection were more excited than the workers in the `Control` condition ($p = 0.071$, Cohen's $d = 0.368$) and the workers who self-reflected after the rejection ($p = 0.039$, Cohen's $d = 0.420$).

Dominance. As for dominance, we found no significant difference according to independent t-tests corrected by Bonferroni method. However, once again, the dominance scores of the `Control` condition and `Rej-SR` condition are similar, and workers who self-reflected before the rejection reported that they felt less in control compared to the `Control` condition ($p = 0.049$, Cohen's $d = 0.400$) and `Rej-SR` condition ($p = 0.156$, Cohen's $d = 0.288$), without statistical significance though.

Emotion analysis of accepted workers. We also studied the effects of self-reflection on accepted workers. We reported the emotional data measured by SAM in Table 2. We use `SR-Acc` and `Acc-SR` to represent accepted workers who were assigned in the conditions of `SR-Rej` and `Rej-SR` respectively. We found no significant difference across all the dimensions and conditions according to significant tests ($\alpha = 0.05$), meaning self-reflection neither positively nor negatively affect accepted workers' emotions.

Summary: *We found a significant improvement in the valence of worker emotions due to self-reflection before rejection, indicating that it can be effectively used to improve workers' reactions to rejection. Particularly, using explicit self-reflection before the rejection can make workers feel significantly less unhappy about the rejection. Our results show that self-reflection does not negatively affect the emotions of workers whose work is accepted.*

5.3 Correlation Analysis

To study correlation between worker emotion and worker performance, we calculated the correlation coefficients R and corresponding p-values (cf. Table 3). We found that the work performance in terms of both accuracy and execution time does not significantly correlated with the worker emotion, across all the three conditions. With all the coefficients of determination R^2 less than 0.06, the presented effects are considered to be very small, meaning the worker emotion is dominated by more complex factors in crowd work, rather than simple factors such as worker accuracy and execution time. The largest correlation coefficient ($R = 0.234$) is given by the execution time and the valence score of `SR-Rej` condition ($p = 0.098$). We noticed that the R (and R^2) values of the worker performance (accuracy & execution time) and the dominance score of `Rej-SR` condition are extremely small, with quite large p-values (0.975 and 0.948 respectively). This can be interpreted that the worker performance cannot affect workers' emotional dominance (the degree to which workers feel in control) at all if they self-reflect after being rejected. Furthermore, we analyzed the correlation between SAM emotion scores and answers that workers reported in the self-reflection form. We selected the following four questions answered by 5-pt likert scales to represent four aspects: **Worker effort** is indicated by the 3rd question — *I worked at the best of my abilities on the task*; **Worker confidence** is indicated by the 4th question — *I have the abilities needed to succeed this task*; **Task clarity** is indicated by the 7th question — *The task was well explained*;

Table 2: Rejected and accepted worker emotion scores measured by SAM ($\mu \pm \sigma$) in three dimensions (valence, arousal, and dominance), of Control Group, Self-reflection before Rejection Group, Self-reflection after Rejection Group.

Measure	Rejected workers			Accepted workers		
	Control (N=50)	SR-Rej (N=51)	Rej-SR (N=50)	Control (N=57)	SR-Rej (N=46)	Rej-SR (N=44)
Valence	4.78 ± 2.71	3.47 ± 2.36	3.92 ± 2.45	2.30 ± 1.72	2.63 ± 2.25	2.34 ± 2.15
Arousal	4.72 ± 2.39	3.84 ± 2.38	4.86 ± 2.47	4.16 ± 2.46	3.89 ± 2.58	5.05 ± 2.50
Dominance	4.92 ± 2.16	4.04 ± 2.24	4.74 ± 2.62	5.47 ± 2.30	5.98 ± 2.55	5.32 ± 2.30

Criteria justifiability is indicated by the 10th question — *The task accept/reject criteria are reasonable.*

Calculated correlation coefficients and corresponding *p*-values are reported in Table 3. We found that for workers who self-reflected before rejection (SR-Rej), worker confidence and criteria justifiability seem to be correlated to emotional valance and arousal ($p < 0.08$). Particularly, for workers in the SR-Rej condition, the correlation between worker confidence and emotional arousal ($R = 0.294$, $R^2 = 0.086$, $p = 0.036$), and the correlation between criteria justifiability and emotional valence ($R = 0.278$, $R^2 = 0.077$, $p = 0.048$) are of statistical significance. However, once again, we found the corresponding coefficients of determination R^2 are very small (0.07-0.09). This is possibly due to the complex interaction during task execution, meaning the worker emotion could be determined by more complex factors. Moreover, workers' self-reported dominance was not affected by neither worker performance nor self-reflection.

Summary: The worker emotion is determined by complex factors during task execution. We found no evidence that the worker performance can significantly affect the worker emotion. However, we found that how workers self-reflected on their confidence and criteria justifiability could correlate to their emotions.

5.4 Mood Analysis

Of 151 rejected workers in total, 78% (N=118) reported pleasant moods, 13% (N=20) reported unpleasant moods, and only 13 reported a neutral mood. SAM scores of workers' emotions, categorized by worker moods, are reported in Table 4. Due to limited number of workers in unpleasant and neutral moods, we did not carry out statistical significance tests. In terms of valence, workers in the Control condition ended up with similar valence scores regardless of their initial moods. However, if self-reflection was introduced, either before or after rejection, workers who were initially in pleasant moods also tended to report that they were happier after the task execution. Furthermore, as for arousal, workers in pleasant moods seem to reported lower arousal scores (more excited), in the conditions of control and self-reflection before rejection. Finally, differences of dominance scores are small in the most conditions. We found that workers in pleasant moods reported lower dominance scores, when they self-reflected after the rejection.

Summary: Workers in a pleasant mood tend to react more positively to task rejections in general. However, this finding is inconclusive due to the limited number of workers who reported unpleasant moods.

6 DISCUSSION

To address the aforementioned research questions, we recruited 460 online workers on MTurk and conducted an experiment with three conditions. Our results revealed that using self-reflection in crowdsourcing can generally enable rejected workers to positively react to rejections. We found that, adding a self-reflection step for workers *before* meting out rejection decisions in particular, can significantly improve their emotions.

We performed a correlation analysis to explore how worker emotion relates to other common factors in microtask crowdsourcing. We found that the worker emotion is determined by rather complex factors during task execution. Results revealed that worker emotions did not follow a linear relationship with their performance (in terms of worker accuracy and execution time). However, the worker confidence and criteria justifiability reported in the self-reflection form were found to be significantly correlated to their emotions. Corroborating evidence from previous work, we found that workers in pleasant moods tended to have relatively positive reactions to rejections in general.

6.1 Implications for Design

Our results show that the modular explicit self-reflection step we designed in this study is feasible and usable. We suggest that task requesters could consider using self-reflection to improve workers' reactions (to rejections), especially for tasks of high complexity and those with a low acceptance rate.

Managing Negative Reactions. For online crowd workers, negative emotions emanating due to prevalent dynamics of task rejection can be demotivating. Task rejection itself can affect worker profiles related to their overall (historical) performance, influence their reputation, and thereby curtail workers' job opportunities in the crowdsourcing marketplace (i.e., their future access to tasks). However, positive emotions can help workers build self-confidence and encourage continued high-quality work, and help workers in maintaining and improving their reputation. Our findings demonstrate that facilitating self-reflection has the potential to assist workers in improving their emotions.

In this work, we designed a self-reflection step to help workers reflect on their work. This can help workers quickly resume a positive working attitude, so that their future work quality may not be affected as a result of negative emotions stemming from task rejections. Crowdsourcing platforms could consider providing self-reflection features for workers before they receive rejections.

Tom Edixhoven et al.

Table 3: Linear relationship, measured by the correlation coefficient R, between emotion scores (valence, arousal, and dominance) and worker performance (accuracy and execution time), across three experimental conditions.

		Valence			Arousal			Dominance		
		`Control`	`SR-Rej`	`Rej-SR`	`Control`	`SR-Rej`	`Rej-SR`	`Control`	`SR-Rej`	`Rej-SR`
Worker Performance	Worker accuracy	$R = 0.104$ ($p = 0.472$)	$R = -0.060$ ($p = 0.678$)	$R = 0.120$ ($p = 0.405$)	$R = 0.074$ ($p = 0.610$)	$R = -0.211$ ($p = 0.137$)	$R = 0.149$ ($p = 0.303$)	$R = 0.185$ ($p = 0.200$)	$R = -0.118$ ($p = 0.408$)	$R = 0.005$ ($p = 0.975$)
	Execution time	$R = -0.168$ ($p = 0.244$)	$R = 0.234$ ($p = 0.098$)	$R = 0.090$ ($p = 0.536$)	$R = -0.194$ ($p = 0.177$)	$R = 0.067$ ($p = 0.638$)	$R = 0.049$ ($p = 0.737$)	$R = -0.021$ ($p = 0.883$)	$R = 0.224$ ($p = 0.115$)	$R = 0.009$ ($p = 0.948$)
Self-Reflection	Worker effort	-	$R = 0.148$ ($p = 0.300$)	$R = -0.191$ ($p = 0.183$)	-	$R = 0.197$ ($p = 0.165$)	$R = -0.062$ ($p = 0.666$)	-	$R = 0.139$ ($p = 0.332$)	$R = -0.190$ ($p = 0.185$)
	Worker confidence	-	$R = 0.255$ ($p = 0.071$)	$R = -0.180$ ($p = 0.211$)	-	$\mathbf{R = 0.294}$ ($\mathbf{p = 0.036}$)	$R = -0.048$ ($p = 0.739$)	-	$R = 0.140$ ($p = 0.328$)	$R = -0.216$ ($p = 0.132$)
	Task clarity	-	$R = 0.073$ ($p = 0.611$)	$R = 0.061$ ($p = 0.674$)	-	$R = 0.120$ ($p = 0.400$)	$R = 0.151$ ($p = 0.295$)	-	$R = 0.090$ ($p = 0.530$)	$R = 0.061$ ($p = 0.675$)
	Criteria justifiability	-	$\mathbf{R = 0.278}$ ($\mathbf{p = 0.048}$)	$R = -0.099$ ($p = 0.492$)	-	$R = 0.250$ ($p = 0.077$)	$R = 0.101$ ($p = 0.482$)	-	$R = 0.152$ ($p = 0.288$)	$R = 0.111$ ($p = 0.443$)

Table 4: Emotional reactions of rejected workers ($\mu \pm \sigma$) in 3 dimensions (valence, arousal, and dominance) across different moods (pleasant and unpleasant).

	Control		SR-Rej		Rej-SR	
Measure	**Pleasant (N=39)**	**Unpleasant (N=6)**	**Pleasant (N=38)**	**Unpleasant (N=8)**	**Pleasant (N=41)**	**Unpleasant (N=6)**
Valence	4.67 ± 2.78	4.50 ± 2.56	3.32 ± 2.26	3.88 ± 2.32	3.63 ± 2.33	5.67 ± 1.80
Arousal	4.54 ± 2.38	5.17 ± 2.48	3.63 ± 2.28	4.25 ± 2.28	5.00 ± 2.51	4.83 ± 2.19
Dominance	4.84 ± 2.29	5.00 ± 1.29	4.08 ± 2.07	4.00 ± 2.29	4.76 ± 2.67	5.67 ± 1.70

Our work has important implications on improving workers' negative reactions when the acceptance requirement is too strict or inappropriate. However, we continue to urge task requesters to take responsibility in avoiding unreasonable rejections.

Enhancing Requester-Worker Relationships. The sustainability of a crowdsourcing marketplace largely depends on the relationship between task requesters and workers. Task rejection, an issue that is discussed frequently in groups and forums where workers share their experiences stalls the formation of a healthy and strong requester-worker relationship. With the use of self-reflection, workers could potentially feel less unhappy when their tasks are rejected, and the requesters could potentially feel less stressed when they are dealing with poor quality, invalid, or suboptimal submissions. Giving workers an opportunity to self-reflect on their performances by actively including such as design intervention, requesters can hope to garner worker appreciation and in turn maintain their reputation. According to our findings, explicit self-reflection can significantly improve workers' emotions, which can pave way for long-term benefits.

Building A Harmonious Online Work Environment. There has been a gradual increase in the number of people who moved their workplaces partly or totally online [28, 30]. Unlike traditional work, the mental/physical conditions, well-being and satisfaction of workers in online freelancing marketplaces (such as microtask crowdsourcing platforms) has not been well-studied. Recent research in the realm of crowdsourcing has focused on improving

quality-related outcomes to a greater extent than understanding and improving online work environments.

We empirically found that self-reflection can be an effective tool in bringing positive emotions to workers facing rejections, and does not negatively affect workers providing high-quality submissions. This can therefore serve as a useful tool to improve task dynamics and online work environments. In this study, we added the self-reflection step to microtask crowdsourcing using a modular design. This can make self-reflection a simple and generic approach to assist crowdsourcing stakeholders in building a harmonious online work environment with a little overhead.

6.2 Caveats, Limitations, and Future Work

Inspired by previous work, we designed a self-reflection form suitable for microtask crowdsourcing [6, 20, 34, 41, 44]. A correlation analysis revealed that workers' emotions can be significantly related to how workers self-reflect on their own performances and tasks. In this study, our designed self-reflection form is mainly used to guide workers to complete the self-reflection step. We are aware that a well-designed self-reflection form can be used to assess whether the workers blame the task or themselves for the rejection. Exploring and optimizing the self-reflection form itself was beyond the scope of this work. We did not investigate whether workers shouldered the blame for their rejection themselves or shifted their blame to the tasks (requesters). Gathering such data would enable us to understand whether those workers who shoulder the responsibility of their task rejections exhibit relatively more positive reactions to

rejection. Future work could focus on the design of self-reflection questionnaires to achieve an optimal effect on the worker emotion in microtask crowdsourcing.

Prior studies have shown the evidence that worker moods could significantly affect quality-related outcomes and worker experiences [13, 35, 50, 54]. In this study, results revealed that workers in a pleasant mood tend to positively react to rejections. However, due to the unbalanced distribution of moods (a phenomenon that was also reported in previous crowdsourcing studies), the limited number of workers in unpleasant and neutral moods limit statistical significance testing. Future experiments can reply on recruiting sufficient participants informed by an apriori power analysis to study the correlation between worker moods and worker reactions to rejection.

In this work, we did not find specific factors that could significantly affect workers' emotions. This is possibly due to workers' emotional management skills, meaning that some workers could manage and suppress their negative emotions even before the self-reflection session [19]. More work is required to study the influence of this potential confound.

A study exploring reaction to rejections of crowd workers needs to account for ethical implications. In our study, we only "reject" workers in terms of restricting their access to more microtasks, depriving them of the additional rewards. We however accepted their HITs on MTurk regardless of the output quality. Therefore, the overall approval rates and future job opportunities of these workers on MTurk was not influenced. In terms of the generalizability of the findings, we were constrained to studying the emotional reactions of workers when their work was rejected and the consequence was their lack of access to additional tasks within our setup, as opposed to the consequence that would directly hinder their reputation on the platform and their general access to tasks thereafter. We believe that studying the emotional reactions to rejection within our setup is a reasonable proxy — in comparison to platform-level rejection that would directly hinder workers' reputation and their general access to tasks thereafter — since the latter would reflect higher stakes and a potentially greater emotional impact. Since we found self-reflection to be effective in improving the emotions of workers in our setup with relatively lower stakes, we envision a similar impact in the latter context.

In this study, we did not consider monetary incentives, while prior work has shown that that workers generally value their work time and optimize for monetary rewards [47, 48]. However, inherent gratification obtained through successful task completion, irrespective of other material compensation interacts on a primal level with workers' motivation for tasking, and experiencing meaningfulness is a critical psychological state [54]. Due to the fact that self-reflection does not require a significant amount of time, we believe that there is a practical utility to this method. Future work could explore methods to effectively engage workers in self-reflection with improved trade-offs with regard to time, and explore the interesting suggestion of using a small monetary reward to improve reactions to rejection.

7 CONCLUSIONS

We recruited 460 unique crowd workers from MTurk to investigate the effect of self-reflection in improving worker reactions to rejections. We carried out a between-subjects crowdsourcing study with information finding HITs across three conditions (control condition without self-reflection, self-reflection before rejection, and self-reflection after rejection). We found that using self-reflection, especially before rejection, can effectively improve the emotion of rejected workers, while it does not negatively affect the emotion of accepted workers. We also investigated the influence of worker performances and worker moods. Our findings reveal that using self-reflection in crowdsourcing tasks has important implications on fostering an emotionally healthy crowdsourcing marketplace.

REFERENCES

[1] Christopher Beedie, Peter Terry, and Andrew Lane. 2005. Distinctions between emotion and mood. *Cognition & Emotion* 19, 6 (2005), 847–878.

[2] Alberto Betella and Paul FMJ Verschure. 2016. The affective slider: A digital self-assessment scale for the measurement of human emotions. *PloS one* 11, 2 (2016).

[3] Margaret M Bradley and Peter J Lang. 1994. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry* 25, 1 (1994), 49–59.

[4] Jennifer D Campbell, Paul D Trapnell, Steven J Heine, Ilana M Katz, Loraine F Lavallee, and Darrin R Lehman. 1996. Self-concept clarity: Measurement, personality correlates, and cultural boundaries. *Journal of personality and social psychology* 70, 1 (1996), 141.

[5] Likang Chi and Joan L Duda. 1995. Multi-sample confirmatory factor analysis of the task and ego orientation in sport questionnaire. *Research Quarterly for Exercise and Sport* 66, 2 (1995), 91–98.

[6] Timothy J Cleary, Gregory L Callan, Jaime Malatesta, and Tanya Adams. 2015. Examining the level of convergence among self-regulated learning microanalytic processes, achievement, and a self-report questionnaire. *Journal of Psychoeducational Assessment* 33, 5 (2015), 439–450.

[7] Pieter MA Desmet, Martijn H Vastenburg, and Natalia Romero. 2016. Mood measurement with Pick-A-Mood: review of current methods and design of a pictorial self-report scale. *Journal of Design Research* 14, 3 (2016), 241–279.

[8] Djellel Difallah, Elena Filatova, and Panos Ipeirotis. 2018. Demographics and dynamics of mechanical Turk workers. In *Proceedings of the eleventh ACM international conference on web search and data mining*. 135–143.

[9] Djellel Eddine Difallah, Michele Catasta, Gianluca Demartini, Panagiotis G Ipeirotis, and Philippe Cudré-Mauroux. 2015. The dynamics of micro-task crowdsourcing: The case of amazon mturk. In *Proceedings of the 24th international conference on world wide web*. 238–247.

[10] Djellel Eddine Difallah, Gianluca Demartini, and Philippe Cudré-Mauroux. 2012. Mechanical cheat: Spamming schemes and adversarial techniques on crowdsourcing platforms.. In *CrowdSearch*. 26–30.

[11] Carsten Eickhoff and Arjen P de Vries. 2013. Increasing cheat robustness of crowdsourcing tasks. *Information retrieval* 16, 2 (2013), 121–137.

[12] Shaoyang Fan, Ujwal Gadiraju, Alessandro Checco, and Gianluca Demartini. 2020. CrowdCO-OP: Sharing Risks and Rewards in Crowdsourcing. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (2020), 1–24.

[13] Ujwal Gadiraju and Gianluca Demartini. 2019. Understanding Worker Moods and Reactions to Rejection in Crowdsourcing. In *Proceedings of the 30th ACM Conference on Hypertext and Social Media*. 211–220.

[14] Ujwal Gadiraju, Ricardo Kawase, Stefan Dietze, and Gianluca Demartini. 2015. Understanding malicious behavior in crowdsourcing platforms: The case of online surveys. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 1631–1640.

[15] Snehal Gaikwad, Durim Morina, Rohit Nistala, Megha Agarwal, Alison Cossette, Radhika Bhanu, Saiph Savage, Vishwajeet Narwal, Karan Rajpal, Jeff Regino, et al. 2015. Daemo: A self-governed crowdsourcing marketplace. In *Adjunct proceedings of the 28th annual ACM symposium on user interface software & technology*. 101–102.

[16] Stephen W Gilliland, Markus Groth, Robert C Baker IV, Angela F Dew, Lisa M Polly, and Jay C Langdon. 2001. Improving applicants' reactions to rejection letters: An application of fairness theory. *Personnel Psychology* 54, 3 (2001), 669–703.

[17] Anthony M Grant, John Franklin, and Peter Langford. 2002. The self-reflection and insight scale: A new measure of private self-consciousness. *Social Behavior and Personality: an international journal* 30, 8 (2002), 821–835.

[18] Mary L Gray and Siddharth Suri. 2019. *Ghost work: how to stop Silicon Valley from building a new global underclass*. Eamon Dolan Books.

[19] James J Gross and Oliver P John. 2003. Individual differences in two emotion regulation processes: implications for affect, relationships, and well-being. *Journal of personality and social psychology* 85, 2 (2003), 348.

[20] Douglas J Hacker, Linda Bol, and Kamilla Bahbahani. 2008. Explaining calibration accuracy in classroom contexts: The effects of incentives, reflection, and explanatory style. *Metacognition and Learning* 3, 2 (2008), 101–121.

[21] Stephanie J Hanrahan, Rachel Pedro, and Ester Cerin. 2009. Structured self-reflection as a tool to enhance perceived performance and maintain effort in adult recreational salsa dancers. *The Sport Psychologist* 23, 2 (2009), 151–169.

[22] Kotaro Hara, Abigail Adams, Kristy Milland, Saiph Savage, Chris Callison-Burch, and Jeffrey P Bigham. 2018. A data-driven analysis of workers' earnings on Amazon Mechanical Turk. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–14.

[23] J Gregory Hixon and William B Swann. 1993. When does introspection bear fruit? Self-reflection, self-insight, and interpersonal choices. *Journal of personality and social psychology* 64, 1 (1993), 35.

[24] Panagiotis G Ipeirotis, Foster Provost, and Jing Wang. 2010. Quality management on amazon mechanical turk. In *Proceedings of the ACM SIGKDD workshop on human computation*. 64–67.

[25] Lilly C Irani and M Six Silberman. 2013. Turkopticon: Interrupting worker invisibility in amazon mechanical turk. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 611–620.

[26] Gabriella Kazai, Jaap Kamps, and Natasa Milic-Frayling. 2011. Worker types and personality traits in crowdsourcing relevance labels. In *Proceedings of the 20th ACM international conference on Information and knowledge management*. 1941–1944.

[27] Aniket Kittur, Jeffrey V Nickerson, Michael Bernstein, Elizabeth Gerber, Aaron Shaw, John Zimmerman, Matt Lease, and John Horton. 2013. The future of crowd work. In *Proceedings of the 2013 conference on Computer supported cooperative work*. 1301–1318.

[28] Siou Chew Kuek, Cecilia Paradi-Guilford, Toks Fayomi, Saori Imaizumi, Panos Ipeirotis, Patricia Pina, and Manpreet Singh. 2015. The global opportunity in online outsourcing. (2015).

[29] Abigail Locke. 2010. *The mind-field of sport: Emotion, mind and accountability in athletes*. Ph.D. Dissertation. Loughborough University.

[30] Anoush Margaryan, Timothy Charlton, and Ujwal Gadiraju. 2020. Learning and skill development in online platform work. (2020).

[31] David Martin, Benjamin V Hanrahan, Jacki O'Neill, and Neha Gupta. 2014. Being a turker. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. 224–235.

[32] Brian McInnis, Dan Cosley, Chaebong Nam, and Gilly Leshed. 2016. Taking a HIT: Designing around rejection, mistrust, risk, and workers' experiences in Amazon Mechanical Turk. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 2271–2282.

[33] Reinhard Pekrun, Thomas Goetz, Anne C Frenzel, Petra Barchfeld, and Raymond P Perry. 2011. Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary educational psychology* 36, 1 (2011), 36–48.

[34] Paul R Pintrich et al. 1991. A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ). (1991).

[35] Sihang Qiu, Ujwal Gadiraju, and Alessandro Bozzon. 2020. Just the Right Mood for HIT! Analyzing the Role of Worker Moods in Conversational Microtask Crowdsourcing. In *International Conference on Web Engineering*. Springer, 381–396.

[36] Chris Roberts and Patsy Stark. 2008. Readiness for self-directed change in professional behaviours: factorial validation of the Self-reflection and Insight Scale. *Medical education* 42, 11 (2008), 1054–1063.

[37] Niloufar Salehi, Lilly C Irani, Michael S Bernstein, Ali Alkhatib, Eva Ogbe, and Kristy Milland. 2015. We are dynamo: Overcoming stalling and friction in collective action for crowd workers. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. 1621–1630.

[38] Shruti Sannon and Dan Cosley. 2019. Privacy, Power, and Invisible Labor on Amazon Mechanical Turk. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

[39] Sonja Schinkel, Dirk van Dierendonck, Annelies van Vianen, and Ann Marie Ryan. 2011. Applicant reactions to rejection. *Journal of Personnel Psychology* (2011).

[40] Keisuke Takano and Yoshihiko Tanno. 2009. Self-rumination, self-reflection, and depression: Self-rumination counteracts the adaptive effect of self-reflection. *Behaviour research and therapy* 47, 3 (2009), 260–264.

[41] Cathy R Thomas and Shannon A Gadbois. 2007. Academic self-handicapping: The role of self-concept clarity and students' learning strategies. *British Journal of Educational Psychology* 77, 1 (2007), 101–119.

[42] Paul D Trapnell and Jennifer D Campbell. 1999. Private self-consciousness and the five-factor model of personality: distinguishing rumination from reflection. *Journal of personality and social psychology* 76, 2 (1999), 284.

[43] William Tsai and Anna S Lau. 2013. Cultural differences in emotion regulation during self-reflection on negative personal experiences. *Cognition & emotion* 27, 3 (2013), 416–429.

[44] Joke H Van Velzen. 2004. Assessing students' self-reflective thinking in the classroom: The self-reflective thinking questionnaire. *Psychological reports* 95, 3_suppl (2004), 1175–1186.

[45] Philippe Verduyn, Iven Van Mechelen, and Francis Tuerlinckx. 2011. The relation between event processing and the duration of emotional experience. *Emotion* 11, 1 (2011), 20.

[46] Mark E Whiting, Grant Hugh, and Michael S Bernstein. 2019. Fair work: Crowd work minimum wage with one line of code. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 7. 197–206.

[47] Alex C Williams, Harmanpreet Kaur, Gloria Mark, Anne Loomis Thompson, Shamsi T Iqbal, and Jaime Teevan. 2018. Supporting workplace detachment and reattachment with conversational intelligence. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.

[48] Alex C Williams, Gloria Mark, Kristy Milland, Edward Lank, and Edith Law. 2019. The Perpetual Work Life of Crowdworkers: How Tooling Practices Increase Fragmentation in Crowdwork. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–28.

[49] Thomas A Wright and Russell Cropanzano. 2007. The happy/productive worker thesis revisited. (2007).

[50] Luyan Xu, Xuan Zhou, and Ujwal Gadiraju. 2019. Revealing the role of user moods in struggling search tasks. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1249–1252.

[51] Jie Yang, Carlo van der Valk, Tobias Hoßfeld, Judith Redi, and Alessandro Bozzon. 2018. How do Crowdworker Communities and Microtask Markets Influence Each Other? A Data-Driven Study on Amazon Mechanical Turk. (2018).

[52] Ming Yin, Mary L Gray, Siddharth Suri, and Jennifer Wortman Vaughan. 2016. The communication network within the crowd. In *Proceedings of the 25th International Conference on World Wide Web*. 1293–1303.

[53] John M Zelenski, Steven A Murphy, and David A Jenkins. 2008. The happy-productive worker thesis revisited. *Journal of Happiness Studies* 9, 4 (2008), 521–537.

[54] Mengdie Zhuang and Ujwal Gadiraju. 2019. In What Mood Are You Today? An Analysis of Crowd Workers' Mood, Performance and Engagement. In *Proceedings of the 10th ACM Conference on Web Science*. 373–382.