Understanding Worker Moods and Reactions to Rejection in Crowdsourcing

Ujwal Gadiraju
L3S Research Center
Leibniz Universität Hannover
Hannover, Germany
gadiraju@L3S.de

Gianluca Demartini
School of ITEE
University of Queensland
Brisbane, Australia
demartini@acm.org

ABSTRACT

Requesters on crowdsourcing platforms typically exercise the power to decide the fate of tasks completed by crowd workers. Rejecting work has a direct impact on workers; (i) they may not be rewarded for the work completed and for their effort that has been exerted, and (ii) rejection affects worker reputation and may limit their access to future work opportunities. This paper presents a comprehensive study that aims to understand worker moods and how workers react to rejections in microtask crowdsourcing. We experimentally investigate the effect of the mood of workers on their performance, and the interaction of their moods with their reactions to rejection. Finally, we explore techniques such as presenting social comparative explanations to foster positive reactions to rejection. We found that workers in pleasant moods significantly outperform those in unpleasant moods. Workers whose work is rejected due to narrowly failing pre-screening tests exhibited the most negative emotional responses.

CCS CONCEPTS

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

ACM Reference Format:
Ujwal Gadiraju and Gianluca Demartini. 2019. Understanding Worker Moods and Reactions to Rejection in Crowdsourcing. In *30th ACM Conference on Hypertext and Social Media (HT '19), September 17–20, 2019, Hof, Germany.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Microtask crowdsourcing has now become a popular means to acquire human input at a large-scale in return for monetary rewards. Despite the growth of platforms like Amazon's Mechanical Turk (AMT[1]) and FigureEight[2], microtask crowdsourcing workflows lack elegant ways of dealing with suboptimal work that is produced.

Task requesters in the microtask crowdsourcing paradigm have to contend with a large variance in the quality of work [28]. Platforms have therefore empowered requesters to determine the merit of responses from crowd workers, either qualifying them as acceptable and rewarding the workers or rejecting work due to poor quality among other reasons [35]. Depending on the type of the task however, it is not always straightforward for requesters to estimate the quality of the work in real-time. Often requesters adopt simple strategies such as accepting and rewarding all the responses first, and employing a post-hoc analysis for quality control thereafter. A substantial amount of prior work has dealt with quality control, and several methods ranging from pre-screening [15] to behavioral methods [18] have been proposed to ensure high-quality results [40]. In other cases, due to inaccurate assessment requesters reject work despite the workers satisfying the requirements. Evidence of this has been observed on crowd worker forums where workers discuss unfair rejections, and through tools such as TurkOpticon [29] where workers rate requesters based on their fairness.

The reality of crowd work in platforms such as AMT or FigureEight includes workers having to face unfair rejection with minimal explanation which negatively affects the user experience in crowdsourcing platforms. Recent work has highlighted the importance of a fair and transparent means of dealing with crowd work [22, 46]. In the absence of adequate policies, and task abandonment [25], unfair rejection of work can have detrimental effects on crowd workers. The effect of fair or unfair assessment of crowd work, leading to acceptance or rejection of work, and passing or failing pre-screening phases has not been studied in detail.

Prior research in work psychology has studied the nature of rejection letters and their impact on job applicants [30]. The authors found that the majority of rejection letters in their analysis expressed the rejection in an indirect style and provided some form of explanation for the rejection. On analyzing the impact of rejection letters on job applicants, the authors found that the structural and content characteristics of the letters and the directness of the statement of rejection, did not seem to have a large effect on applicants' self-perceptions. Others studied the impact of manuscript rejections in scientific communities. Cassey and Blackburn quantified the extent to which a sample of ecologists with the most successful publication careers experienced manuscript rejection [8]. Woolley and Barron reflected on the fact that rejections hurt but are not necessarily fatal, due to their finding that at least 50% of rejected manuscripts, if pursued are published within a time frame of 2 years [54]. Gilliland et al. studied explanations in employment rejection letters from the perspective of fairness theory [23]. The

[1] https://www.mturk.com/
[2] https://www.figure-eight.com/

HT '19, September 17–20, 2019, Hof, Germany
© 2019 Association for Computing Machinery.
ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

authors found that explanations detailing qualifications of the individual who received the job, significantly reduced perceptions of unfairness among the rejected applicants.

Inspired by these observations from previous works, and our own experience of finding solace in manuscript rejection letters that describe 'low-acceptance rates and high quality peer work' as an attributing factor for rejection at top-tier scientific venues, we postulate that relaying social comparative information regarding the performance of workers and their peers can improve their emotional handling of rejection from pre-screening phases on microtask crowdsourcing platforms. Thus, in this paper we focus on the worker experience on crowdsourcing platforms in fair and unfair rejection situations. In doing so, we aim to fill the existing knowledge gap pertaining to reactions to rejection in crowd work. We investigate how personalized feedback about their performances can be used to improve the rejection experience.

We present results from a controlled study on user reactions to rejections on a micro-work platform and analyze the impact of rejections on user performances. We adopt standard methodologies to measure user mood, rejection sensitivity, and reactions to rejections to understand how to communicate rejection decisions to crowd workers.

Research Questions. It is known that crowd work quality is affected by many factors such as task design, clarity, and complexity [31, 56]. In this work, we investigate how the emotional state in which crowd workers find themselves when they approach a task on the platform impacts the quality of the work they produce. Based on findings from social psychology work, we also look at the important aspect of work rejection that negatively impacts crowd work experience. To this end, we investigate the following research questions –

RQ#1 – What *moods* are crowd workers typically in while contributing to piecework? What *emotions* do crowd workers elicit during the course of task completion?

RQ#2 – How does fair and unfair rejection in the pre-screening phase affect workers on crowdsourcing platforms?

Hypothesis: Rejection with comparative explanations makes crowd workers relatively more receptive to the rejection and leaves a smaller affective perturbation on workers, in contrast to rejection without comparative explanations.

2 BACKGROUND AND RELATED WORK

In this section, we discuss four realms of background literature and related works. We first discuss how work quality has been studied to be affected by different factors in conventional work within the management research field. We then discuss the social dimension of crowd work and the role of feedback, training and peer assessment in the paid crowdsourcing paradigm. Finally, we discuss rejection sensitivity and the wealth of knowledge in understanding moods and emotions.

2.1 How is Work Quality Affected by Worker Moods?

According to the well-known happy/productive worker thesis from management research, workers who are happy on the job will exhibit a higher job performance in comparison to those who are less happy [55, 57]. Wright and Cropanzano advocated for the incorporation of the psychological well-being of workers into the happy/productive worker thesis [55]. Zelenski et al. argued that among the happiness indicators that they examined –job satisfaction, quality of work life, life satisfaction, positive affect, and negative affect– positive affect was most strongly tied to productivity at both the state and trait levels [57]. Although this decades old thesis has been studied in the context of conventional work, we draw inspiration from this field of work. Our recent work explored the impact of mood on worker performance as well as their engagement in relatively long batches of 20 HITs [58]. We found a statistically significant impact of worker moods on their engagement but not their performance. In contrast, within this paper we investigate whether workers in happier moods exhibit better quality related outcomes on microtask crowdsourcing platforms (i.e., higher work accuracies and lower task completion times).

2.2 The Social Dimension of Crowd Work, Feedback and Peer Assessment

Micro-work is an individual effort in the vast majority of cases [29]. Users of microtask crowdsourcing platforms like AMT most commonly exert individual effort through the course of task completion. While some social features are embedded into crowd work by workers themselves (e.g., using web forums or tools like TurkOpticon [29] to share experiences with other workers), task selection and completion on platforms like AMT is typically driven by workers themselves. Previous work has studied how crowd workers performing collaborative work [27] can be organized to improve their efficiency, reaching close to real-time answers.

In contrast to these previous works, we aim to explore the role of social comparison in managing reactions of crowd workers on their work being rejected either fairly or unfairly. Prior works have investigated the benefit of providing feedback to workers with an aim to improve worker performance. Oelson et al. proposed the programmatic creation of gold questions to provide targeted training and feedback to workers, with an aim to increase quality related outcomes [40]. Gadiraju et al. compared implicit training (where workers received feedback on providing erroneous responses) with explicit training (where workers were required to go through a training phase before they could attempt to complete a task) [19]. The authors found that their proposed methods resulted in improving worker accuracy and reducing the task completion time.

Other works have explored the idea of using peers for work quality assessment. Dow et al. showed that feedback in the form of self-assessment or external assessment can yield higher quality work. In the context of their results, the authors reflected that scheduling and variance in quality are the two main challenges pertaining to the use of peers for providing feedback and assessment [13]. Kulkarni et al. introduced PeerStudio, an assessment platform that leverages the large number of students' peers in online classes to enable rapid feedback on in-progress work [32]. Through controlled experiments, the authors observed an improvement in the grades of students who received timely peer feedback. Gadiraju et al. recently observed that some workers can learn from social comparison and improve their performance [20].

Based on these previous works, there is a consensus that providing feedback to workers can promote self-reflection, learning and improve worker performance. However, the effect of feedback in the form of peer comparison on worker emotions has been unexplored. We aim to fill this knowledge gap through a series of experiments in this paper.

2.3 Rejection Sensitivity

Rejection sensitivity (RS) is a cognitive-affective processing disposition to anxiously expect, perceive, or intensely react to rejection [14]. It is shaped by cognitive social learning history and is triggered in situations when either rejection or acceptance is possible. The RS-Adult questionnaire (A-RSQ) is an adaptation of the RSQ [14] that was developed for assessing RS in adults [3]. The questionnaire presents 9 different situations in which people sometimes ask things of others. Participants are asked to imagine they are in the situation, and answer two questions that follow in each case. The total rejection sensitivity score is measured as the mean of the rejection sensitivity scores for the 9 situations. Since crowd workers constantly face the dichotomy between rejection and acceptance in typical workflows, it is useful to understand their rejection sensitivity. The authors of A-RSQ showed that it captures meaningful differences in rejection sensitivity across diverse groups of adults, making it suitable to employ in the crowdsourcing context.

2.4 Mood and Emotion

Although both *mood* and *emotion* are valenced affective responses, prior work has elaborately discussed the difference between the two [11]. Firstly, moods last longer than emotions [1, 49]. Secondly, emotions are always targeted towards an event, person or object, while moods are globally diffused [17]. Emotions are triggered by explicit causes and monitor our environment, while moods have combined causes and monitor our internal state [34, 39]. Further, emotions are elicited by threats or opportunities [17], while moods are responses to one's overall position in general [42]. However, note that moods and emotions are not entirely independent; they interact with each other dynamically. Accumulated emotions can lead to specific moods, and moods can lower the degree of emotional arousal [9]. We build on these substantial prior works [1, 9, 17, 34, 39, 42, 49], that have established an understanding of *moods* and *emotions* to unearth the background roles they play on workers of microtask crowdsourcing platforms.

3 METHODOLOGY AND SETUP

To evaluate the effects of rejection in crowd work, rather than asking workers to complete microwork and abstain from rewarding them for completed work, we introduce two work phases: a *pre-screening* and a *follow-up* work opportunity. Pre-screeing phases are popularly employed as a method for quality control in crowdsourced work [31]. In our experimental setup, rejection decisions were communicated based on the pre-screening performance of workers and were said to determine whether or not workers could gain access to further tasks.

3.1 Measuring *Mood* and *Emotion*

To measure the *mood* of crowd workers in an intuitive and easy manner, we use *Pick-A-Mood* (PAM), a character-based pictorial scale for reporting moods [11]. Compared to other measures, this is ideal for the microtask crowdsourcing context where time is of essence, since it was specifically made to be suitable for design research applications in which people have little time or motivation to report their moods. The scale has been tested with a general population (people from 31 different nationalities in the validation study), revealing that the expressions presented by the visual characters are correctly interpreted (see Figure 1). PAM has been used in a variety of research settings including quality of experience research [50] and in education [48], illustrating the robustness of the tool.

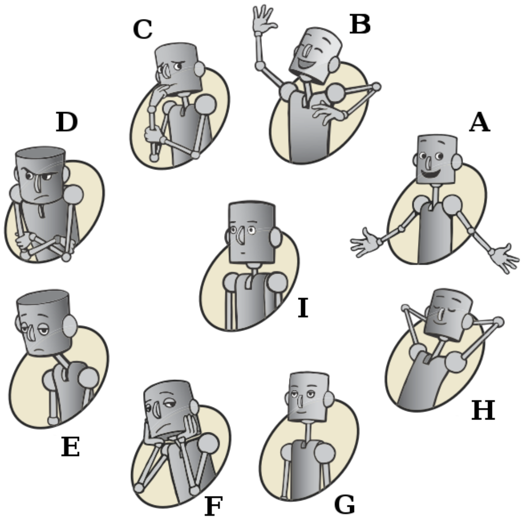
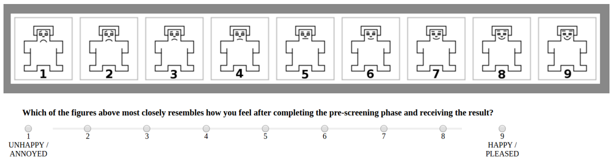


Figure 1: *Pick-A-Mood* scale to measure the self-reported mood of crowd workers in different conditions.



Which of the figures above most closely resembles how you feel after completing the pre-screening phase and receiving the result?

○ 1 UNHAPPY / ANNOYED ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 HAPPY / PLEASED

Figure 2: Example *SAM* scale to obtain the self-reported emotion of crowd workers on the dimension of '*valence*'.

To measure the *emotion* elicited by crowd workers, we use the popular *Self-Assessment-Manikin* (SAM) instrument [5]. The SAM is also a pictorial rating scale that is used to obtain self-assessments of the emotions experienced by people along the dimensions of *valence* (i.e., the degree of pleasure) *arousal* (i.e., the degree of excitement) and *dominance* (i.e., the degree of control). It has been shown to be reliable, valid and usable regardless of age, educational or cultural background of the responders [4, 33]. An example of the SAM scale used to obtain the emotion of workers on the dimension of '*valence*' is presented in Figure 2. Due to its brevity, it can be used to capture emotional responses [36] to a variety of emotion elicitation methods. For example, it has been used before and after biological challenges [16] to measure how emotional state changes in response to these procedures. The SAM has also been used to measure emotional state in numerous advertising studies [38]. We therefore decided to use the SAM to gather emotional responses from workers.

3.2 Task Design and Experimental Setup

We consider the microtask of *information finding* (IF), since a recent longitudinal study of task types on the AMT crowdsourcing platform revealed its increasing popularity [12]. To model task difficulty, we draw inspiration from Wood's seminal construct of task complexity [53]. Wood argued that tasks that have more steps are more complex. *Component complexity* refers to the number of distinct acts required to complete a task [24].

To account for the varying difficulty levels of information finding tasks, we manually created a dataset consisting of 3 difficulty-levels. In each case, workers were tasked with finding the middle-names of famous persons. However, with each progressive level of difficulty (from `Level-I` through `Level-III`), the number of aspects that workers had to account for and disambiguate increased by a factor of one, thereby making the task more difficult. For example, in case of tasks with `Level-I` difficulty, workers do not have to disambiguate; a Google search for 'Daniel Craig' reveals 1 famous person. In tasks with `Level-II` difficulty, workers are asked to disambiguate between 2 or more famous persons with the same name by using their profession as the criterion (for example, 'George Lucas- Archbishop'). Finally, in tasks with a `Level-III` difficulty, workers are required to disambiguate between multiple famous persons using not only their profession but also the year in which they were active (for example, 'Brian Smith- Ice Hockey, 1972').

To address the research questions and hypotheses stated earlier, we consider four different treatment conditions and one control condition. In all conditions, workers were first asked to indicate their mood using the *Pick-A-Mood* instrument as shown in Figure 1. We took care to follow all the recommendations suggested by the authors while using the PAM scale [10]. Next, workers were asked to respond to a few general background questions (age, gender, ethnicity, education, marital status). Then workers were administered the A-RSQ (adult rejection sensitivity questionnaire). Finally, workers were asked to complete 10 information finding tasks.

We deployed all conditions on the FigureEight platform, and restricted study participation to the highest quality workers using an inbuilt feature on the platform[3]. We use FigureEight since it is one of the primary commercial paid microtask crowdsourcing platforms. Our study of rejection in crowd work is focused on paid microtask crowdsourcing, since rejection in this context directly impacts worker earnings and future earning opportunities. To account for inter-task effects induced by task difficulty [7], we completely randomized the order of the tasks. We describe each of the conditions below. Each condition was deployed on separate days, and subjects were not allowed to participate in multiple conditions. We gathered responses from 50 distinct crowd workers in each of the 5 experimental conditions, resulting in a total of 250 workers. In each case, we rewarded workers for the tasks they completed at an hourly rate of 8 USD (by estimating the time needed to complete the tasks).

Control – After the 10 IF tasks, in this condition we gathered responses from workers regarding how they felt after completing the tasks, using the *Self-assessment Manikin* (SAM) as shown in Figure 2.

Accept-All – Here, workers were informed that the 10 information finding tasks would serve as pre-screening tasks, and that if they performed accurately they would gain access to 10 additional tasks of the same type along with the associated monetary rewards. After completing the 10 IF tasks, no matter how workers performed all of them were presented with the following message - "*Congratulations! You have successfully passed the pre-screening phase.*" Workers were then asked about how they felt after completing the pre-screening phase and receiving the result, using the SAM instrument. Finally, they were presented 10 additional follow-up IF tasks with varying difficulty in a randomized order.

Reject-All – In this condition, workers were also informed about the pre-screening phase and an opportunity to gain access to 10 additional IF tasks. However, after the pre-screening phase, immaterial of how the workers performed they were presented with the following message - "*Sorry! You did not succeed in passing the pre-screening phase.*" Once again, workers were asked about how they felt after completing the pre-screening phase and receiving the result, using the SAM instrument. They were then reassured that although they were not eligible to access additional follow-up IF tasks, they would receive the prescribed reward for the pre-screening phase.

Accurate Assessment with Individual Feedback ***(AAIF)*** – This condition was inspired by prior works which have shown positive effects of feedback on worker performance and learning [13]. Here, we explore whether transparent feedback on performance can improve worker reactions to rejections.

Workers in this condition were informed about the pre-screening phase and the opportunity to gain access to 10 additional IF tasks. Their performance in the pre-screening phase determined whether they gained access to the remaining tasks (as in the *Accept-All* condition), or whether they received the rejection message (as in the *Reject-All* condition). The accuracy threshold for passing the pre-screening phase was chosen to be 70% to reflect standard practice in real-world microtask crowdsourcing, such as on FigureEight[4]. On completing the pre-screening phase, workers received personalized messages that conveyed feedback regarding their performance. For example, workers who performed with an accuracy of X%, below the threshold of 70% received the following message - "*You have performed with an accuracy of X%. Sorry! You did not pass the pre-screening phase, the minimum accuracy required was 70%. You will not have access to the additional tasks.*" On the other hand, those who met the minimum accuracy criteria received the following message - "*You have performed with an accuracy of X%. Congratulations! You have successfully passed the pre-screening phase, the minimum accuracy required was 70%. You are now eligible to complete 10 additional information finding tasks.*"

Accurate Assessment with Social Comparison ***(AASC)*** – This condition is identical to the *AAIF* condition, except that on completing the tasks in the pre-screening phase, the feedback that

[3] *Level-3 contributors* on FigureEight comprise workers who completed > 100 test questions across hundreds of different types of tasks, and have a near perfect overall accuracy.

[4] By default FigureEight suggests a minimum accuracy of 70% in the pre-screening *Quiz Mode*

Table 1: Avg. accuracy of workers in the pre-screening IF tasks across different experimental conditions.

Condition	Avg. Accuracy (in%)	Std. Dev. (in%)
Control	65.20	17.76
Accept-All	71.20	11.36
Reject-All	66.81	17.95
AAIF	68.00	14.98
AASC	69.77	13.54
Overall	68.20	15.12

workers received was formulated by including social comparison cues. We used the performance distribution of workers in the *Control* condition to simulate real-time feedback including social comparison. For example, a worker who performed with an accuracy of 60% in the pre-screening phase received the following message - "*You have performed with an accuracy of 60%. Sorry! You did not pass the pre-screening phase and will not have access to the additional tasks. The minimum requirement for passing was 70% accuracy and you were very close! You performed on par with or better than 32% of the contributors who participated in this pre-screening phase.*"

In all the conditions, workers were provided with an optional field to leave any comments, feedback or suggestions. Since we aim to study the general distribution of worker moods within our subject population, we proceed without any additional filtering or control of workers beyond restricting participation to Level-3 workers in the conditions within this experimental setup. For example, balancing the number of workers who were fairly/unfairly accepted/rejected in the *Accept-All* and *Reject-All* conditions would not allow us to study the general distribution of worker moods. We leave a more detailed analysis of this comparison for future work, and choose this task design to help us understand the general distribution of worker moods in our population.

4 RESULTS AND ANALYSIS

4.1 Accuracy and Completion Time of Workers

Worker accuracy and task completion time are two of the most important quality related outcomes in paid microtask crowdsourcing [31]. Due to this reason, we conducted a one-way between workers ANOVA to explore the effect of the different conditions on the accuracy and task completion time of workers in the pre-screening IF tasks. We found no statistically significant differences across the different conditions (see Table 1). On average, the workers performed with an accuracy just short of the minimum requirement of 70% to pass the pre-screening phase (*M=68.2, SD=15.12*) and gain access to follow-up tasks.

Next, we investigated whether workers who passed the pre-screening IF tasks went on to maintain their accuracy in the additional follow-up IF tasks. We did not find statistically significant differences in the accuracy of workers who passed the pre-screeing tasks across different conditions; *Accept-All* (*M=70.5, SD=13.74*), *AAIF* (*M=67.27, SD=16.38*) and *AASC* (*M=68, SD=12.9*). This suggests that workers maintain their accuracy in the additional tasks. **Does guilt play a role in the *Accept-All* condition?** We checked whether workers who performed with a lower accuracy than the 70% threshold, and were still allowed to continue onto follow-up tasks in the *Accept-All* condition went on to improve their performance. We found that such workers did depict an improvement in their accuracy (in %) from the pre-screening IF tasks (*M=55, SD=7.9*) to the follow-up IF tasks (*M=61.67, SD=11.9*). However, a two-tailed T-test revealed a lack of statistical significance in the improvement.

4.2 How Do You Feel Today? Mood of Workers

We analyzed the mood of crowd workers who participated in all the experimental conditions described earlier. Previous work has established that moods are an affective background canvas of what we do [9]. By using *Pick-A-Mood* (PAM) as described earlier, we obtained self-reported assessments of mood from workers before they began the actual tasks in the pre-screening phase. Our findings are presented in the Figure 3. We found that on average across all conditions, most workers were either *cheerful* (29.4%), *calm* (15.6%), *relaxed* (11.4%), or *excited* (9%). Some workers reported to be *bored* (6.2%), while some others claimed to be *sad* (5.6%). 5.2% of workers on average, reported *neutral* moods. A small number of workers were either *tense* (4.2%) or *irritated* (3%).

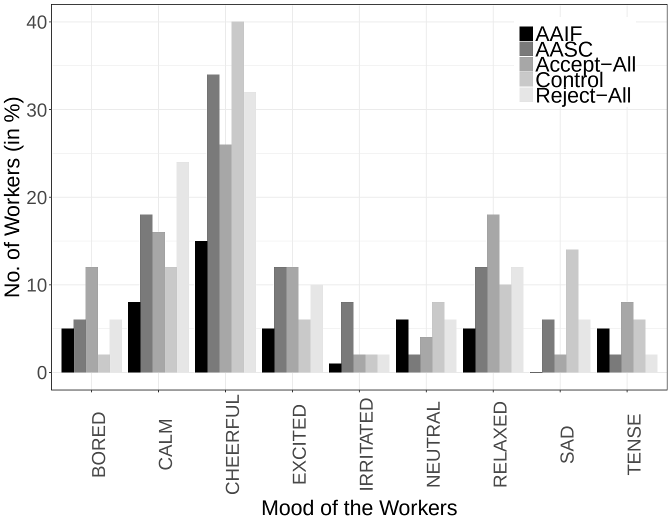


Figure 3: Mood distribution of crowd workers who participated in the different experimental conditions.

The eight non-neutral moods measured by PAM, can be grouped into four mood categories [52]; **activated-pleasant** (*excited, cheerful*), **deactivated-pleasant** (*relaxed, calm*), **activated-unpleasant** (*tense, irritated*), and **deactivated-unpleasant** (*bored, sad*). Thus, we found that on average 65.4% of crowd workers in our experiments were in pleasant moods, while 29.4% of the workers were in unpleasant moods. Next, we compared the performance in the pre-screening phase of workers in pleasant moods with respect to those in unpleasant moods. Using a two-tailed T-test, we found that workers in pleasant moods (*M=70.2, SD=12.8*) performed with significantly higher accuracy (in %) than those in unpleasant moods (*M=64.3,SD=17.2*); *t(224)=7.187, p<.01*. However, a two-tailed T-test revealed no significant difference in the task completion time of workers in pleasant moods (*M=18.69,SD=6.21*) when compared to those in unpleasant moods (*M=17.43,SD=6.50*); *t(224)=1.636, p =.20*.

4.3 How Sensitive Are Workers to Rejection?

All workers who participated in the different conditions completed the A-RSQ as described earlier. We analyzed the rejection sensitivity scores of the workers and our findings are presented in Figure 4. We did not find any statistically significant differences in the rejection sensitivity of workers who participated in the different conditions (see Table 2). Overall, we found that the average rejection sensitivity of workers was 9.75 ± 3.10. This is consistent with the range of rejection sensitivity observed among normal adults in prior works [2, 3]. This suggests that crowd workers are not any different from other comparable populations of adults studied in the past with respect to their rejection sensitivity.

To investigate whether there is a linear relationship between the rejection sensitivity scores of workers and their performance in the pre-screening phase, we computed the Pearson's **r** between all conditions. However, we did not find any significant correlation.

Table 2: Average *rejection sensitivity* scores of workers in different experimental conditions.

Condition	Avg. RS Score	Std. Dev.
Control	10.14	3.22
Accept-All	10.21	2.68
Reject-All	9.60	2.71
AAIF	9.14	3.95
AASC	9.66	2.94

4.4 Emotional Impact of Rejection on Workers

We obtained the emotional reactions of workers (by using the SAM pictorial scale as described earlier), after they completed the pre-screening phase and received the result regarding whether or not they passed. This workflow was motivated by prior work which suggests that emotions are momentary perturbations that can be triggered by events (such as passing/failing the pre-screening phase in this case) [9].

Figure 5 presents our findings across the different conditions, on the dimensions of arousal, dominance and valence. A combination of a medium to high valence, medium arousal and medium to high dominance can be interpreted as a joyful emotion [5]. On average we note that the *Accept-All* condition results in the most joyful emotions, followed by the *Control* condition. This can be intuitively understood since all workers pass the pre-screening phase in the *Accept-All* condition and gain access to 10 additional information finding (IF) tasks thereafter. In case of the *Control* condition, nothing hinges on the performance of the crowd workers in the IF tasks (since there are no additional tasks to gain access to, and all workers receive a reward independently of their performance). Similarly, it is intuitive that the *Reject-All* condition results in the least joyful emotions, since all workers are rejected from accessing the follow-up tasks. However, it is interesting to note the low degree of joyful emotions resulting from the individual feedback and social comparison interventions in *AAIF*, *AASC* respectively.

To further understand the emotional impact of acceptance and rejection on the workers while considering the differences in feedback messages they received in the *AAIF* and *AASC* conditions, we divided the workers into three groups:

(1) *Unqualified*– This group of workers performed with an accuracy between 0-50% in the pre-screening IF tasks.
(2) *Near-misses*– This group of workers performed with an accuracy of 60% in the pre-screening IF tasks, missing out on the opportunity to gain access to follow-up tasks by a single correct answer.
(3) *Qualified*– This group of workers performed with an accuracy between 70-100% and qualified for completing the additional IF tasks.

Table 3: Distribution of workers (in %) in the 3 groups across the different conditions.

Condition	Unqualified	Near-Misses	Qualified
Control	14	22	64
Accept-All	8	16	76
Reject-All	20	10	70
AAIF	23	12	65
AASC	16	20	64
Overall	16.20	16	67.80

Table 3 presents the distribution of workers in the three groups across the different conditions. We found that the majority of workers passed the pre-screening phase (*qualified*), while the remainder formed a nearly even split between *near-misses* and *unqualified.* Next, we investigated how the different groups of workers elicited their emotions, as shown in Figure 6.

Consistent with our previous findings, all the 3 groups of workers in the *Control* and *Accept-All* conditions elicited highly joyful emotions (Fig. 6a, 6b). In the *Reject-All* condition (Fig. 6c), we understandably found that *qualified* workers exhibited the least joyful emotions, although the *near-misses* and *unqualified* workers also elicited less joy.

Interestingly in the *AAIF* condition (Fig. 6d), we found that providing *near-misses* with feedback that helped them to reflect on their individual performance and realize how close they were to passing the pre-screening phase resulted in them eliciting the least joy compared to the *qualified* and *unqualified* groups of workers. Additional feedback in terms of comparative explanations in the *AASC* condition also resulted in a similar outcome (Fig. 6e). This can be explained by the disappointment workers in the *near-misses* group may have felt on narrowly losing the opportunity to earn more money by completing the follow-up IF tasks. Prior work has established that such situations give rise to emotions and social ascriptions such as guilt, regret and blame [6].

We also analyzed the relationship between the rejection sensitivity scores of workers and their emotional reactions to acceptance and rejection across all conditions. However, we did not find any significant correlations using Pearson's **r** that could indicate an underlying linear relationship.

4.5 Worker's Lens: Fair and Unfair Assessment

Prior work has sufficiently established the prevalence of fair and unfair assessment in microtask crowdsourcing marketplaces [31, 35]. To further understand the emotional impact of passing and failing pre-screening phases on crowd workers, we turned to the optional comments that some workers left after receiving their

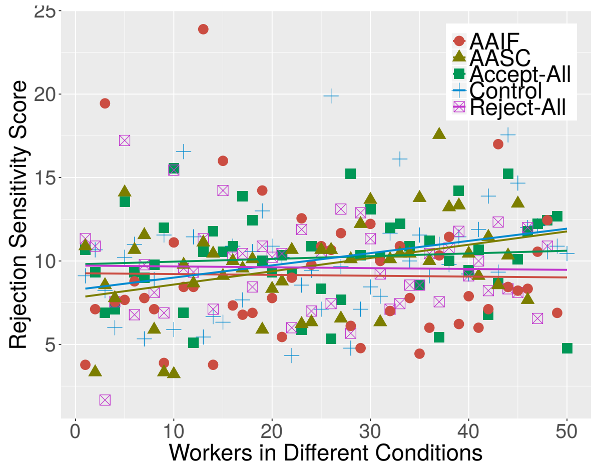


Figure 4: Rejection sensitivity score of workers who participated in different experimental conditions.

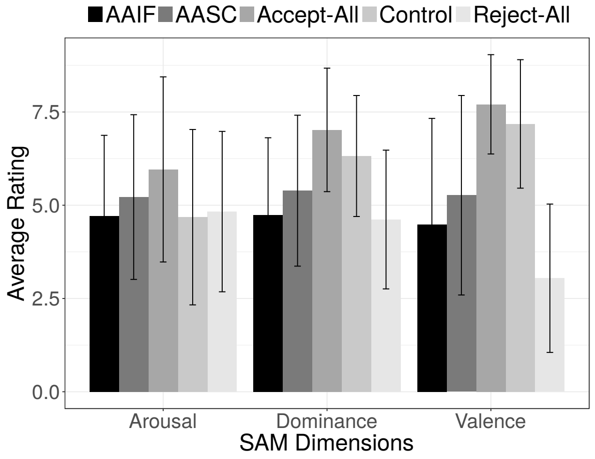


Figure 5: Emotional reactions of workers on passing/failing the pre-screening in different conditions.

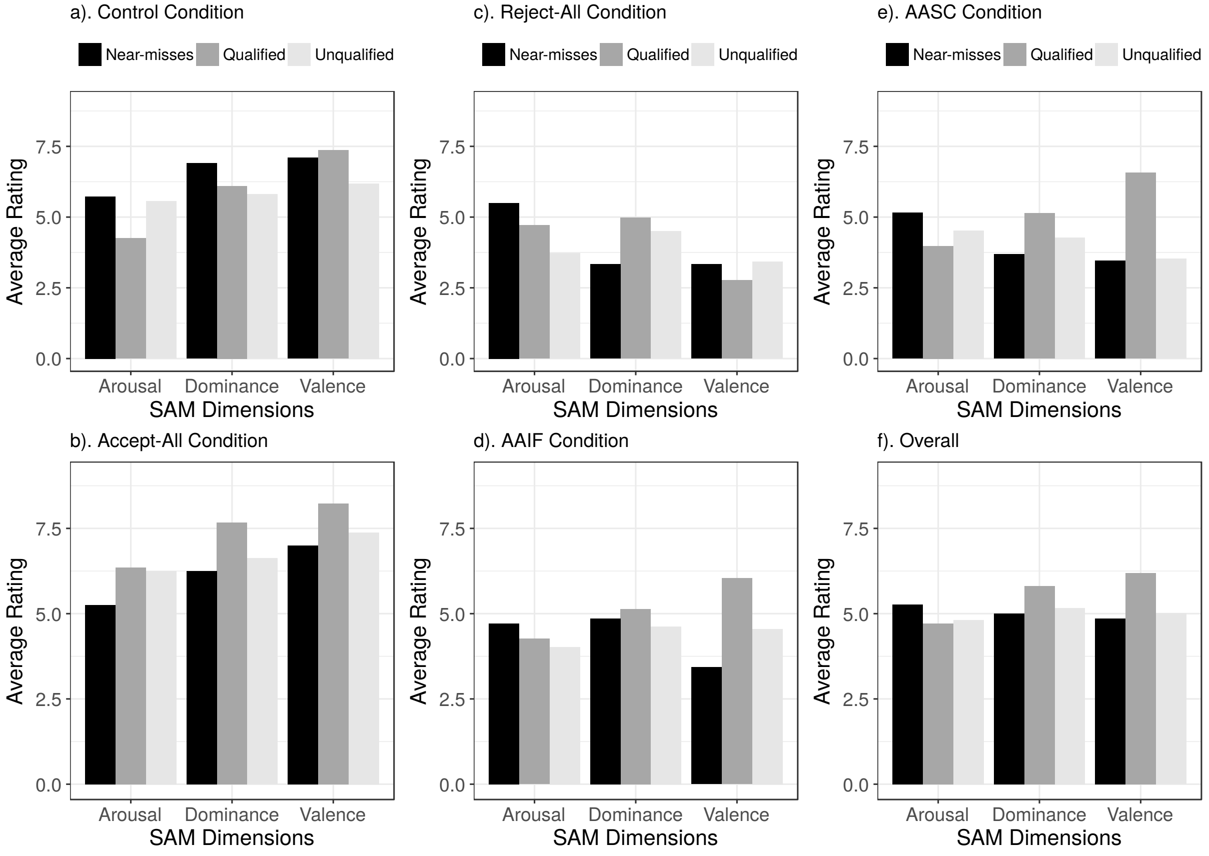


Figure 6: Emotional reactions of *near-misses*, *qualified*, and *unqualified* groups of workers on passing or failing the pre-screening phase in the different conditions.

results on the pre-screening phase. Table 4 presents descriptive statistics of workers who left comments at the end of the tasks in each condition. We notably found that workers in the *Reject-All* condition showed the least inclination to comment with only about 1 in 4 workers choosing to leave some comments.

Table 4: Workers who commented after receiving their results from the pre-screening IF tasks.

Condition	% of Workers	% Avg. Accuracy	% Std. Dev.
Control	52	68.84	9.52
Accept-All	36	72.22	13.96
Reject-All	26	63.52	70.00
AAIF	42	62.85	10.18
AASC	46	68.00	9.41

With an aim to accurately analyze the emotional states that workers were in at the end of the tasks, we grouped workers according to the fairness with which they were pre-screened. The *Fair Accept* group of workers consists of workers who passed the accuracy threshold in the *Accept-All*, *AAIF*, and *AASC* conditions. The *Fair Reject* group consists of workers who were fairly rejected after the pre-screening phase due to accuracies that did not meet the threshold (in the *Reject-All*, *AAIF*, and *AASC* conditions). Workers in the *Unfair Reject* group are those who were rejected in the *Reject-All* condition despite meeting the accuracy threshold in the pre-screening phase. Workers in the *Unwarranted Accept* group are those who were accepted despite not meeting the accuracy threshold in the pre-screening phase of the *Accept-All* condition.

Based on the comments received from workers in free-form natural language, we used the LIWC lexicon [47] to measure affective processes (i.e. positive emotions and negative emotions) across the different groups. LIWC has been reliably used in various related contexts in social science studies [41], for analyzing the language of fake news [43], and to classify the trustworthiness of tweets [51].

Table 5: Fraction of positive and negative emotions exhibited by workers on average at the end of different treatment conditions, as measured from their comments using LIWC.

Grouping	% Positive Emotions	% Negative Emotions
Fair Accept	21.15	2.63
Fair Reject	16.25	1.84
Unfair Reject	2.44	8.16
Unwarranted Accept	18.37	0.00

Our results are presented in Table 5. We note that workers generally depict positive emotions when they are dealt with fairly (*Fair Accept*, *Fair Reject* groups), or if they are accepted despite not meeting the accuracy threshold and therefore not deserving of it (*Unwarranted Accept* group). This is understandable since workers are given an opportunity to proceed past the pre-screening phase and earn more money. Our results also show that workers who are unfairly rejected (*Unfair Reject* group) generally depict negative emotions. We reason that this is due to the fact that workers are then not allowed to proceed to the actual tasks where they could have earned more money, despite providing high quality work and deserving to do so. In contrast, those workers who were accepted despite not meeting the required threshold of accuracy (*Unwarranted Accept* group) did not exhibit any negative emotions whatsoever in their comments.

5 DISCUSSION

Pre-screening and qualification tasks have become common practice in crowdsourcing platforms. In this study we have shown how rejection affects worker emotions in microtask crowdsourcing platforms. We observed an overall wish of users to perform well and a common desire to receive feedback when rejected, to understand their errors and to improve their performances during the next work opportunity. We also observed that subjects whose work is accepted tend to demonstrate better performances in follow-up tasks even if they performed poorly in the initial ones.

5.1 Workers in Pleasant Moods Perform Better

We found that workers who began the information finding tasks in pleasant moods performed with a significantly higher accuracy. This can have important implications on the design of crowdsourcing workflows and crowdsourced experiments. For example, a simple mood measure such as *Pick-A-Mood* can be used to select desired workers in pleasant moods with the expectation of improved quality of data collected via the crowdsourcing platform. We found no significant difference in the task completion time of workers in pleasant moods when compared to those in unpleasant moods. Moreover, in our tasks we found that most workers (over 65%) are in pleasant moods. This potentially suggests that throughput of large batches of tasks can remain reasonable when mood-based pre-selection is employed.

5.2 Understanding the Case of Near-misses

Considering our hypothesis that social comparative explanations would result in less negative reactions to rejection, we rather observed that social comparison can have the most negative effects on workers' emotions, especially in the case of the *near-misses* group (i.e., workers who get rejected by not qualifying past the acceptance threshold by a narrow margin). Our findings with respect to the case of near-misses can be explained by the psychological theory of counterfactual thinking [44]. Roese posed that counterfactuals are mental models of alternatives to the past and produce consequences that are both beneficial and aversive to an individual. Roese and Olson's recent social psychology work on counterfactual thinking has established that the emotional responses of people to events are influenced by their thoughts about "what might have been" [45]. Due to this, it has been found that situations in which people who may objectively be better off nonetheless feel worse. Medvec et al. analyzed the emotional reactions of bronze and silver medalists at the 1992 Summer Olympics at the end of their respective events as well as at the medal ceremony, and found that bronze medalists tended to be happier than silver medalists. The authors attributed these results to the fact that the most compelling counterfactual alternative for the silver medalist was winning the gold, whereas for the bronze medalist it was finishing without a medal [37]. In our context, the near-misses are comparable to the silver medalists, in that the most compelling counterfactual alternative for these workers is passing the threshold for acceptance by getting merely an additional correct response.

5.3 Other Key Findings

Contrary to our hypothesis, we observed that providing workers with individual feedback and comparative information regarding their performance with respect to that of a group of peers, can lead to a decrease in joyful emotions. While we found no strong support for our hypothesis, worker comments on receiving rejection decisions indicate that their disappointment arose from the lack of specific explanations on the errors they made and on the possibility of self-improvement. An implication of this is finding is that requesters can improve the reactions of workers to rejections by providing detailed explanations of the errors made rather than simply reject work, as is commonly done [35]. Further experiments are required to better understand the attributes of explanations that can improve worker reactions to rejections.

5.4 Caveats and Limitations

In this section we discuss some caveats and the limitations of our work. First, in this study we did not consider more complicated experimental models that take into account multiple factors (e.g., worker mood at the beginning of the task) and interaction between factors. Not controlling for these factors enables us to also gain an understanding of the distribution of such factors in the subject population we recruited through the crowdsourcing platform. Another limitation is focusing only on crowd workers. While work rejection is a broader issue, in this work our hypotheses, research questions, and experimental design is applied to crowd workers in paid micro-task crowdsourcing platforms and may not generalize beyond that. The closest context in which work rejection occurs, is that of voluntary crowdsourcing platforms such as Wikipedia, where work rejection is manifested in the form of reverts performed by other editors on the platform.

Note that our analysis of statistical significance in some parts of this paper included multiple t-tests. To control for Type-I error inflation in our multiple comparisons, we used the Holm-Bonferroni correction for family-wise error rate (FWER) [26], at the significance level of $\alpha < .05$.

The general threats and limitations of research conducted on crowdsourcing platforms have been identified and discussed in the past [21]. In this work, we are specifically interested in understanding the reactions to fair and unfair rejections in paid microtask crowdsourcing. Due to this, the threats of validity of our findings are limited to our methodological choices for which we have provided the rationale in Section 3.2. Moreover, the tools we have used in this work for measuring moods and emotions are robust and reliable. Hence, we believe that our findings are reliable and replicable under these settings and our sampling frame. More experiments are required to analyze the impact of worker moods on their performance across different task types.

6 CONCLUSIONS AND FUTURE WORK

The impact of crowd worker moods on the quality of work produced has not been studied in the past. In this paper, we studied the moods of workers who participated in our information findings tasks and found that over 65% of workers began tasks in pleasant moods. Workers in pleasant moods performed with a significantly better accuracy, depicting an increase of ~10% over workers in unpleasant moods (**RQ#1**).

Crowd workers are negatively affected by unilateral rejection decisions both in terms of lost reward as well as in terms of negative reputation which may limit their access to work in the future. For these reasons, rejections should not be an easy decision to take. However, it is exclusively in the power of requesters to decide when to accept or reject certain work and whether to provide explanations or not for such decisions.

We delved into the emotional responses of workers on FigureEight, to fair and unfair rejection decisions meted out across different conditions. Our main findings indicate that while workers whose work is accepted respond with more positive emotions than those whose work has been rejected, the most emotionally affected workers are the *near-misses*, that is, those workers who narrowly failed to pass the acceptance threshold (**RQ#1**).

Analyzing self-reported emotions of workers using the SAM scale revealed that *near-misses* depicted the least joyful emotions when compared to workers who passed the pre-screening phase and those who failed the pre-screenig phase by a larger margin (**RQ#2**). An analysis of worker comments on receiving rejection decisions using LIWC revealed that a far greater fraction of workers who were fairly rejected depicted positive emotions than negative emotions. In contrast, a larger fraction of those who were unfairly rejected depicted negative rather than positive emotions. Comments from all workers who were accepted despite having failed to meet the threshold accuracy, depicted positive emotions (**RQ#2**). Albeit without statistical significance, we found that workers who were accepted despite failing the pre-screening phase depicted an improvement in their accuracy.

We can conclude that although it is useful to provide rejection explanations to workers whose work is rejected, in certain cases (*near-misses*) explanations deteriorate the emotional responses to rejections, possibly due to the triggering of counterfactual thinking.

Our results highlight an important opportunity to foster better interactions between requesters and workers on paid microtask crowdsourcing marketplaces. In the imminent future, we will investigate the effect of worker moods on their performance in different types of tasks. We will also investigate different 'explanations types' and their role in further easing reactions to rejections.

Acknowledgements. We thank all the workers who participated in our experiments. This research has been supported in part by the Erasmus+ project DISKOW (grant no. 60171990), and the EU's H2020 research and innovation programme (grant no. 732328).

REFERENCES

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

[2] Kathy R Berenson, Geraldine Downey, Eshkol Rafaeli, Karin G Coifman, and Nina Leventhal Paquin. 2011. The rejection–rage contingency in borderline personality disorder. *Journal of abnormal psychology* 120, 3 (2011), 681.

[3] Kathy R Berenson, Anett Gyurak, Özlem Ayduk, Geraldine Downey, Matthew J Garner, Karin Mogg, Brendan P Bradley, and Daniel S Pine. 2009. Rejection sensitivity and disruption of attention by social threat cues. *Journal of research in personality* 43, 6 (2009), 1064–1072.

[4] Margaret M Bradley, Mark K Greenwald, Margaret C Petry, and Peter J Lang. 1992. Remembering pictures: pleasure and arousal in memory. *Journal of experimental psychology: Learning, Memory, and Cognition* 18, 2 (1992), 379.

[5] 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.

[6] Ruth MJ Byrne. 2002. Mental models and counterfactual thoughts about what might have been. *Trends in cognitive sciences* 6, 10 (2002), 426–431.

[7] Carrie J Cai, Shamsi T Iqbal, and Jaime Teevan. 2016. Chain reactions: The impact of order on microtask chains. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 3143–3154.

[8] Phillip Cassey and Tim M Blackburn. 2003. Publication rejection among ecologists. *Trends in Ecology & Evolution* 18, 8 (2003), 375–376.

[9] Richard J Davidson. 1994. On emotion, mood, and related affective constructs. *The nature of emotion: Fundamental questions* (1994), 51–55.

[10] Pieter MA Desmet, Vastenburg, and Natalia Romero. 2016. Pick-A-Mood manual: Pictorial self-report scale for measuring mood states. (2016).

[11] 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.

[12] 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.

[13] Steven Dow, Anand Kulkarni, Scott Klemmer, and Björn Hartmann. 2012. Shepherding the crowd yields better work. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*. ACM, 1013–1022.

[14] Geraldine Downey and Scott I Feldman. 1996. Implications of rejection sensitivity for intimate relationships. *Journal of personality and social psychology* 70, 6 (1996).

[15] Julie S Downs, Mandy B Holbrook, Steve Sheng, and Lorrie Faith Cranor. 2010. Are your participants gaming the system?: screening mechanical turk workers. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 2399–2402.

[16] Matthew T Feldner, Michael J Zvolensky, Georg H Eifert, and Adam P Spira. 2003. Emotional avoidance: An experimental test of individual differences and response suppression using biological challenge. *Behaviour research and therapy* 41, 4 (2003), 403–411.

[17] Nico H Frijda et al. 1994. Varieties of affect: Emotions and episodes, moods, and sentiments. (1994).

[18] Ujwal Gadiraju, Gianluca Demartini, Ricardo Kawase, and Stefan Dietze. 2018. Crowd Anatomy Beyond the Good and Bad: Behavioral Traces for Crowd Worker Modeling and Pre-selection. *Computer Supported Cooperative Work (CSCW)* (2018), 1–27.

[19] Ujwal Gadiraju, Besnik Fetahu, and Ricardo Kawase. 2015. Training workers for improving performance in crowdsourcing microtasks. In *Design for Teaching and Learning in a Networked World*. Springer, 100–114.

[20] Ujwal Gadiraju, Besnik Fetahu, Ricardo Kawase, Patrick Siehndel, and Stefan Dietze. 2017. Using worker self-assessments for competence-based pre-selection in crowdsourcing microtasks. *ACM Transactions on Computer-Human Interaction (TOCHI)* 24, 4 (2017), 30.

[21] Ujwal Gadiraju, Sebastian Möller, Martin Nöllenburg, Dietmar Saupe, Sebastian Egger-Lampl, Daniel Archambault, and Brian Fisher. 2017. Crowdsourcing versus the laboratory: towards human-centered experiments using the crowd. In *Evaluation in the Crowd. Crowdsourcing and Human-Centered Experiments*. Springer, 6–26.

[22] Snehalkumar Neil S Gaikwad et al. 2016. Boomerang: Rebounding the consequences of reputation feedback on crowdsourcing platforms. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. ACM, 625–637.

[23] Stephen W Gilliland, Markus Groth, Robert C Baker IV, Angela E 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.

[24] Thorvald Hærem, Brian T Pentland, and Kent D Miller. 2015. Task complexity: Extending a core concept. *Academy of Management Review* 40, 3 (2015), 446–460.

[25] Lei Han, Kevin Roitero, Ujwal Gadiraju, Cristina Sarasua, Alessandro Checco, Eddy Maddalena, and Gianluca Demartini. 2019. All those wasted hours: On task abandonment in crowdsourcing. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*. ACM, 321–329.

[26] Sture Holm. 1979. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics* (1979), 65–70.

[27] Ting-Hao Kenneth Huang, Walter S Lasecki, and Jeffrey P Bigham. 2015. Guardian: A crowd-powered spoken dialog system for web apis. In *Third AAAI conference on human computation and crowdsourcing*.

[28] 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*. ACM, 64–67.

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

[30] Fredric M Jablin and Kathleen Krone. 1984. Characteristics of rejection letters and their effects on job applicants. *Written communication* 1, 4 (1984), 387–406.

[31] 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*. ACM, 1301–1318.

[32] Chinmay E Kulkarni, Michael S Bernstein, and Scott R Klemmer. 2015. PeerStudio: rapid peer feedback emphasizes revision and improves performance. In *Proceedings of the Second (2015) ACM Conference on Learning@ Scale*. ACM, 75–84.

[33] Peter J Lang. 1985. The cognitive psychophysiology of emotion: Anxiety and the anxiety disorders. *Lawrence Eribaum, Hillsdale* (1985).

[34] Richard Lazarus. 1994. The stable and the unstable in emotion. *The nature of emotion: Fundamental questions* (1994), 79–85.

[35] 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*. ACM, 2271–2282.

[36] Mary W Meagher, Randolph C Arnau, and Jamie L Rhudy. 2001. Pain and emotion: effects of affective picture modulation. *Psychosomatic medicine* 63, 1 (2001), 79–90.

[37] Victoria Husted Medvec, Scott F Madey, and Thomas Gilovich. 1995. When less is more: counterfactual thinking and satisfaction among Olympic medalists. *Journal of personality and social psychology* 69, 4 (1995), 603.

[38] Jon D Morris. 1995. Observations: SAM: the Self-Assessment Manikin; an efficient cross-cultural measurement of emotional response. *Journal of advertising research* 35, 6 (1995), 63–68.

[39] William N Morris. 2012. *Mood: The frame of mind*. Springer Sci. & Business Media.

[40] David Oleson, Alexander Sorokin, Greg P Laughlin, Vaughn Hester, John Le, and Lukas Biewald. 2011. Programmatic Gold: Targeted and Scalable Quality Assurance in Crowdsourcing. *Human computation* 11, 11 (2011).

[41] James W Pennebaker, Ryan L Boyd, Kayla Jordan, and Kate Blackburn. 2015. *The development and psychometric properties of LIWC2015*. Technical Report.

[42] Jesse J Prinz. 2004. *Gut reactions: A perceptual theory of emotion*. Oxford UP.

[43] Hannah Rashkin, Eunsol Choi, Jin Yea Jang, Svitlana Volkova, and Yejin Choi. 2017. Truth of varying shades: Analyzing language in fake news and political fact-checking. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. 2931–2937.

[44] Neal J Roese. 1997. Counterfactual thinking. *Psychological bulletin* 121, 1 (1997), 133.

[45] Neal J Roese and James M Olson. 2014. *What might have been: The social psychology of counterfactual thinking*. Psychology Press.

[46] Niloufar Salehi, Lilly C Irani, Michael S Bernstein, Ali Alkhatib, Eva Ogbe, Kristy Milland, et al. 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*. ACM, 1621–1630.

[47] Yla R Tausczik and James W Pennebaker. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology* 29, 1 (2010), 24–54.

[48] Cesar Vandevelde, Francis Wyffels, Maria-Cristina Ciocci, Bram Vanderborght, and Jelle Saldien. 2016. Design and evaluation of a DIY construction system for educational robot kits. *International Journal of Technology and Design Education* 26, 4 (2016), 521–540.

[49] 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.

[50] Bjørn J Villa, Katrien De Moor, Poul E Heegaard, and Anders Instefjord. 2013. Investigating Quality of Experience in the context of adaptive video streaming: findings from an experimental user study. *Akademika forlag Stavanger, Norway* (2013).

[51] Svitlana Volkova, Kyle Shaffer, Jin Yea Jang, and Nathan Hodas. 2017. Separating facts from fiction: Linguistic models to classify suspicious and trusted news posts on twitter. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Vol. 2. 647–653.

[52] David Watson and Auke Tellegen. 1985. Toward a consensual structure of mood. *Psychological bulletin* 98, 2 (1985), 219.

[53] Robert E Wood. 1986. Task complexity: Definition of the construct. *Organizational behavior and human decision processes* 37, 1 (1986), 60–82.

[54] Karen L Woolley and J Patrick Barron. 2009. Handling manuscript rejection: insights from evidence and experience. *Chest* 135, 2 (2009), 573–577.

[55] Thomas A Wright and Russell Cropanzano. 2007. The happy/productive worker thesis revisited. In *Research in personnel and human resources management*. Emerald Group Publishing Limited, 269–307.

[56] Jie Yang, Judith Redi, Gianluca Demartini, and Alessandro Bozzon. 2016. Modeling task complexity in crowdsourcing. In *Fourth AAAI Conference on Human Computation and Crowdsourcing*.

[57] 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.

[58] 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*. ACM, 373–382.