In What Mood Are You Today? An Analysis of Crowd Workers' Mood, Performance and Engagement

Mengdie Zhuang
University of Sheffield
Sheffield, United Kingdom
mzhuang1@sheffield.ac.uk

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

ABSTRACT

The mood of individuals in the workplace has been well-studied due to its influence on task performance, and work engagement. However, the effect of mood has not been studied in detail in the context of microtask crowdsourcing. In this paper, we investigate the influence of one's mood, a fundamental psychosomatic dimension of a worker's behaviour, on their interaction with tasks, task performance and perceived engagement. To this end, we conducted two comprehensive studies; (i) a survey exploring the perception of crowd workers regarding the role of mood in shaping their work, and (ii) an experimental study to measure and analyze the actual impact of workers' moods in information findings microtasks. We found evidence of the impact of mood on a worker's perceived engagement through the feeling of reward or accomplishment, and we argue as to why the same impact is not perceived in the evaluation of task performance. Our findings have broad implications on the design and workflow of crowdsourcing systems.

CCS CONCEPTS

• **Human-centered computing**; • **Information systems**;

ACM Reference Format:
Mengdie Zhuang and Ujwal Gadiraju. 2019. In What Mood Are You Today? An Analysis of Crowd Workers' Mood, Performance and Engagement. In *11th ACM Conference on Web Science (WebSci '19), June 30-July 3, 2019, Boston, MA, USA*. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3292522.3326010

1 INTRODUCTION

Crowdsourcing has been flourishing as a type of gig economy, and has found wide use in both industry and academia, fostering particular interest for conducting *Web Science* research [19]. The motivation for using crowdsourcing has mostly focused on the ease and cost-effective nature of acquiring human input; and a considerable number of prior works have addressed challenges to improve the effectiveness of this paradigm [8, 31].

WebSci '19, June 30-July 3, 2019, Boston, MA, USA

ACM ISBN 978-1-4503-6202-3/19/06…$15.00
https://doi.org/10.1145/3292522.3326010

Theoretical and empirical research suggests that people's moods have an effect on lexical decision making [5], learning and training in the workplace [59], the perception, encoding, storage, and retrieval of information [3, 56], and workplace outcomes, such as productivity [57], behavior and engagement [1]. The popular notion is that people with a positive mood perform better in their work assignments and are more likely to have increased intrinsic motivation. However, there have been diverging opinions among researchers.

Early studies suggest the existence of an indirect causality link between mood and work/task performance intermediated by a spectrum of psychosomatic states. Seibert and Ellis found that students in non-neutral moods displayed greater propensity towards ineffectual thinking, which in turn produced a negative impact on the outcome of tasks such as memory recall [54]. More recent work has been focused on sociability [58] as the mediator between mood and performance, in which they found that workers in positive moods may perform better through interpersonal processes such as helping and seeking help among co-workers. Apart from sociability, non-neutral mood is found to increase vulnerability to distractions [48], which might lead to unwanted behavior at work (e.g., long response time). On the other hand, unpleasant mood states also facilitate work withdrawal behavior due to the individual's need for mood repair [40].

An aspect that resonates throughout the aforementioned studies is the importance of establishing clear links with actionable implications between an individual's mood and its impact and efficacy in the workplace. However, these links and their implications have not previously been established within the context of microtask crowdsourcing, a branch of online labor which is gaining increased traction among the exponents of the current generation, and having an outreaching impact upon human data driven research.

The key difference in the context of microtask crowdsourcing stems from the nature of the work setting. It is both a form of gig work – hence temporary, and on demand – but also remote, which allows the workers to cater their workspace to their own needs [18]. Moreover, the overarching rules and laws that govern traditional workplace interactions fade in the online crowdsourcing context where there is a flexibility in time, space and with respect to organizational boundaries [15, 31]. To the best of our knowledge, prior work has not comprehensively studied the role of worker moods in shaping crowd work in microtask marketplaces.

In this paper, we aim to address this knowledge gap by presenting two studies that aim to understand the role of crowd worker moods in shaping their work outcomes. First, we explore how workers think their moods affect the task outcomes such as worker performance and engagement by using a survery addressing 100 workers. Our findings from the survey further inform and direct the second study, in which we collected data from 300 crowd workers completing information finding microtasks and analyzed how their moods affect their interaction, performance and behavior. We build upon the previously mentioned studies in order to inform and direct discussion together with work design theory on task design, workflows and interventions in crowdsourcing marketplaces, thereby improving the overall effectiveness of the paradigm and at the same time serving the interest of the growing crowd workers community.

Original Contributions. We make the following contributions:

- We enhance and enrich the existing understanding of workers' mental states within the context of microtask crowdsourcing, by quantifying task outcomes and identifying work characteristics in this novel setting in relation to worker moods.
- We find support for the hypothesis that there is an impact of mood on a worker's perceived reward, and shed light on contrary evidence of its impact on performance.
- This work challenges the status quo, by pitting quantitative evidence against the biased understanding of workers themselves in how mood is perceived and how it impacts their engagement and performance.

2 BACKGROUND AND RELATED LITERATURE

2.1 Mood and Emotion

Although both *mood* and *emotion* are valenced affective responses, prior work has elaborately discussed the difference between the two [9]. Firstly, moods last longer than emotions [2, 60]. 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 [33, 44]. Further, emotions are elicited by threats or opportunities [17], while moods are responses to one's overall position in general [49]. 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 [7].

The concept of workplace mood has been traditionally associated with workplace quality (e.g., productivity [57], engagement [1]). We build on the substantial prior works that have established an understanding of *moods* to unearth the background role that mood plays on workers within a microtask crowdsourcing environment. We investigate (a) the *moods* that crowd workers are typically in, while contributing to piecework, and (b) how particular mood dispositions affect task outcomes.

2.2 Exploiting '*Mood*' in Microtask Crowdsourcing

Recently, [55] proposed to leverage the relationship between people's mood and their productivity to operationalize a concept from workforce management research known as 'productive laziness'[47]. The authors argue that crowd workers need to efficiently schedule when to work and rest, to maximize their overall productivity and sustain the long-term function of the crowdsourcing system. The dynamic scheduling method introduced by the authors jointly minimizes the effort exerted by crowd workers while maximizing the overall throughput. In closely related work, [43] explored the relationship between the mood of crowd workers and their capacity for creative outcomes. The authors proposed two approaches for enhancing worker performance in creative tasks; affective priming, and affective pre-screening. Their findings suggest that workers in a positive mood exhibit enhanced creativity. Other empirical works have established that happiness makes individuals more productive [34, 41].

In contrast to these prior works, we use a robust tool called '*Pick-A-Mood*' (described later in Section 3), which only requires a single click from the participants, to explicitly gather self-reports of worker moods, and test their effect on task outcomes such as worker performance and engagement.

2.3 Worker Engagement and Performance in Crowdsourcing

Previous works have addressed the issue of boredom and fatigue in crowdsourcing marketplaces resulting due to the repetitive nature of long batches of tasks that workers often encounter. Thus, a variety of means to retain and engage workers have been proposed. [52] suggested introducing micro-breaks into workflows to refresh workers, and showed that under certain conditions micro-breaks aid in worker retention and improve their accuracy marginally. Similarly, [6] proposed to intersperse diversions (small periods of entertainment) to improve worker experience in lengthy, monotonous microtasks and found that such micro-diversions can significantly improve worker retention rate while maintaining worker performance. Other works proposed the use of gamification to increase worker retention and throughput [16, 50]. [36] studied worker engagement, characterized how workers perceive tasks and proposed to predict when workers would stop performing tasks. [12] introduced pricing schemes to improve worker retention, and showed that paying periodic bonuses according to pre-defined milestones has the biggest impact on retention rate of workers. More recently, [20] proposed the use of achievement priming to increase worker retention in crowdsourcing microtasks.

In contrast to these prior works in microtask crowdsourcing that measure worker engagement using the proxy of prolonged retention, we measure worker engagement using a standardized questionnaire [45], obtaining direct feedback from the workers. The concept of work engagement has been discussed in the field of work psychology [28], showing long term benefits on workers' development and well being [4]. However, crowd workers are gig economy workers who are fundamentally different from the workers discussed in previous work engagemnt studies; they do not have contracts or regular hours of work, and can pick and choose what jobs they complete and when. To the best of our knowledge, there is no engagement measurement developed specifically for piecework and gig economy workers. Since engaged employees experience a sense of reward and the presence of the necessary supporting resources for the job [28], we propose to measure work engagement of crowd workers using the concepts of *usability* and *reward* [45].

3 STUDY I: PRELIMINARY SURVEY

In this first study, we survey 100 distinct crowd workers on FigureEight[1] (a primary microtask crowdsourcing platform) to understand the following:

(1) How do crowd workers feel during task completion?
(2) What is the perception of crowd workers regarding the influence of their moods on their task performance and engagement?
(3) To what extent do crowd workers perceive their moods to effect their task performance and engagement?

We aim to reveal the distribution of worker moods, and juxtapose the *perceived impact* of moods on engagement and performance.

3.1 Study Design

3.1.1 Measuring the Mood of Workers. 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 [9]. Compared to other measures, this is ideal for the microtask crowdsourcing context where time is of the 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 PAM is designed on two principal dimensions: arousal and valence. In this study, we focus on the valence dimension. The eight non-neutral moods measured by PAM, can be grouped into two main mood groups [9]: **pleasant** (*excited, cheerful, relaxed, calm*), and **unpleasant** (*tense, irritated, bored, sad*) [62]. The *neutral* mood group stands on its own. The scale has been tested with a general population (with people from 31 different nationalities in the validation study)[9], revealing that the expressions presented by the visual characters are correctly interpreted (see Figure 1). This scale has also been used in other works [61]. In our study, workers were asked to select the pictorial representation of the mood that most closely resembled their current moods.

3.1.2 Survey Design. To achieve a comprehensive understanding of workers perspective on the effect of mood, the survey questions were developed to cover two broad areas (see examples of questions in Figure 2): 1) general perception and reasons for the perceived effect of mood on task performance, 2), perception and reasons for perceived effect of mood on detailed aspects of performance and engagement.

To begin with, the current mood of the workers was collected using the *Pick-A-Mood* scale. Next, the survey presented workers with questions regarding the demographics, educational and general background of the workers. Then, questions related to worker opinions on whether they believe their (current self-reported) mood would influence their work performance and engagement, are presented. The rationale behind using the mood that workers just self-reported is to enable the workers to associate the question to their own feelings. After collecting the general opinion, a set of statements related to their answer to the previous question (but diving into further details) was presented, and a binary response (agree/

[1]http://www.figure-eight.com/

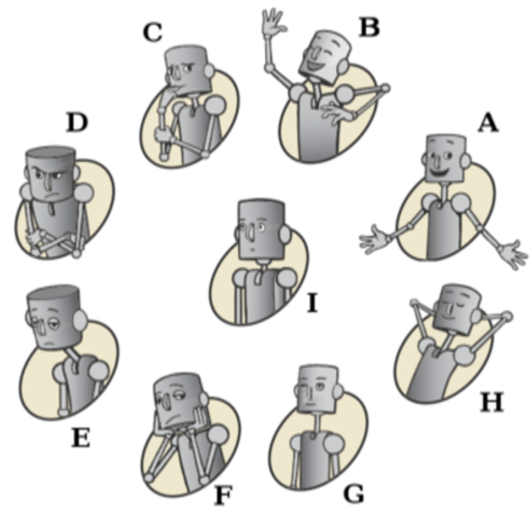


Figure 1: *Pick-A-Mood* scale to measure the self-reported mood of crowd workers in different conditions.

disagree) was required for each statement. For the statements pertaining to engagement, we adapted the three items belonging to *Perceived Usability* (PU) and *Reward* (RW) from the User Engagement Scale [45]. A detailed description of these two concepts is presented in the section 4.1.1. We finally added statements corresponding to each of task accuracy and completion speed which are the two main performance related attributes. In total the survey spanned a mixture of open-ended, multiple-choice, and Likert-type questions designed to filter out malicious workers and motivate workers to provide high quality responses; we collected the responses from 100 workers. In order to collect more specific answers, participants were asked to provide reasons for each of their response.

3.1.3 Study Procedure. We deployed the survey on FigureEight. To ensure high quality responses, we restricted the participation to *Level-3 workers*[2] on the platform. We presented workers with an overview at the beginning of the survey including the purpose of this study and the research impact, to motivate participants into providing genuine and detailed responses. We explicitly encouraged workers to report moods genuinely, assuring that it would have no bearings on their payment. We also used attention check questions to filter out untrustworthy workers [37]. In total, we collected responses from 100 workers and 95 workers passed the attention check. On average workers took ~10 mins to complete the survey and were compensated at a fixed hourly rate of 7.5 USD.

3.2 Results

Of the 95 workers after filtering, 34 were female and 61 were male. This is typical of crowdsourcing platforms, given the demographics of workers [11]. Most of the workers were found to be under 45 years old (N=83), of which 16 were between 18-25 years old. 61 of the workers reported to have at least a Bachelor's degree. Finally,

[2]*Level-3 contributors* on FigureEight comprise workers who completed over 100 test questions across hundreds of different types of tasks, and have a near perfect overall accuracy. They are workers of the highest quality.

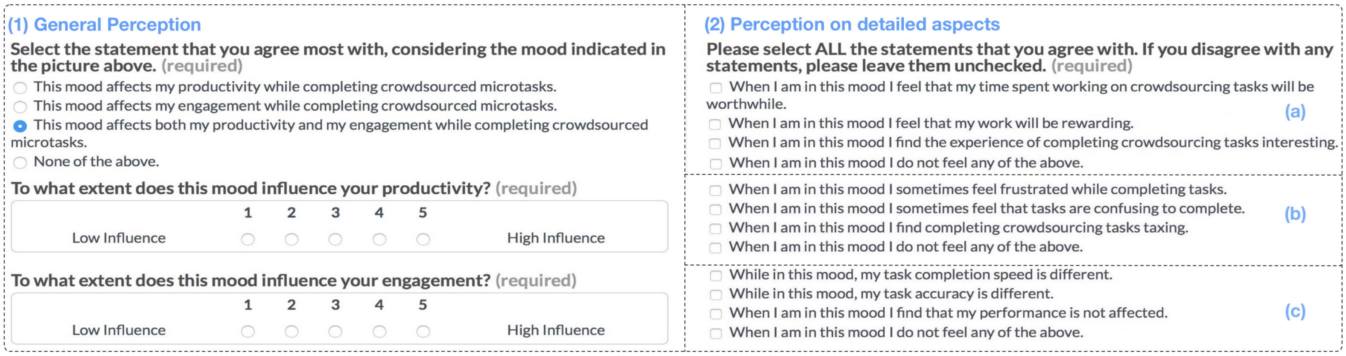


Figure 2: Main survey questions used. (a) Reward, (b) Preceived usability, and (c) Performance.

nearly half of the workers (N=45) reported crowdsourcing to be their primary source of income. The distribution of worker moods as collected using the PAM scale is presented in Figure 3. 63 workers reported pleasant moods, while 28 reported unpleasant moods.

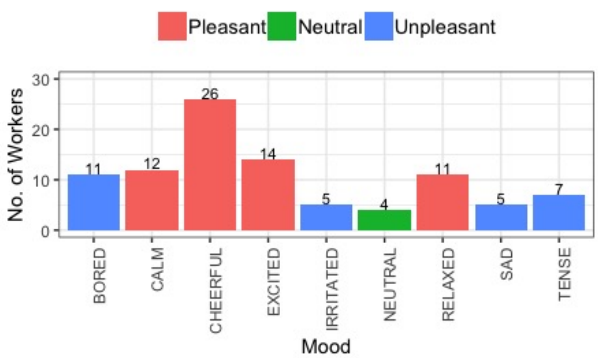


Figure 3: Mood distribution of workers who completed the survey in Study I.

Several important observations emerged from the responses collected using the survey. First, the distribution of moods reported by crowd workers is skewed towards pleasant, and very few people reported to have a neutral mood (N=4, 4.2%). This may be caused by the general conditions under which workers choose to complete crowdsourcing tasks, i.e., on their own terms. As three workers reported in their open-ended remarks, they typically do not feel like working on crowdsourcing tasks when they are in unpleasant moods. This stands in stark contrast to the typically rigid work hours that employees are expected to comply with in traditional workplaces.

Secondly, regarding the general perception of the effect of workers' moods, we found that workers' opinions diverge. As shown in Table 1, 33.68% workers (Pleasant: N= 22, Unpleasant: N= 7, Neutral: N= 3) believed that their mood would not affect either their task performance or work engagement, and 4 of them mentioned that regardless of their mood, they always wanted to perform well in their tasks. This zeal to constantly perform well is arguably due to the existence of the reputation system on FigureEight, where

Table 1: No. of worker agreed on the effect option of mood on their engagement and performance.

	None	Eng.	Perf.	Eng.&Perf.
Pleasant (N=63)	22	6	22	13
Neutral (N=4)	3	0	0	1
Unpleasant (N=28)	7	2	9	10
Total (N=95)	**32**	**8**	**31**	**24**

workers are awarded level badges based on their accuracy across several tasks, and their level goes on to dictate their general access to available tasks. Thus, workers often aim to maintain a high accuracy and good reputation [18, 39]. 63 workers expressed that their mood affects their engagement or performance. We found that 42 workers provided detailed reasons explaining their choice, or suggested the particular aspects of engagement and performance which are affected.

Table 2: The extent to which workers perceive their moods to effect their engagement and performance on a 5-point scale (Mean (SD), N= No. of responses).

	Pleasant	Unpleasant	Ple.& Un.
Eng.	4.21 (0.63), N=19	4.08 (0.99), N=12	4.18 (0.78)
Perf.	4.17 (0.78), N=35	3.84 (1.01), N=19	4.07 (0.88)

Those workers who believe that their moods do have an effect described attention as the mediator between mood and performance (N=8, keywords in the feedback: "attention", "concentrate", "focus" etc.). In line with our intuition, they mentioned that failing to concentrate on the task instructions in the past, resulted in a decrease in their performance. This is similar to the observation of ineffectual thinking in memory recall tasks [54] in which students with non-neutral mood exhibit more ineffectual thinking, which in turn decreases their performance. Lasecki et al. recognized this link as an upper bound on the memory capacity of an individual in crowdsourcing, and suggested to distribute the cognitive load required by the task among crowd workers [32]. Table 2 presents the 63 workers who rated the extent to which their mood affects their engagement and performance. Although fewer workers (no. of responses =32) indicate that mood has an effect on their engagement, the average value is larger than the average value of the effect of mood on performance (no. of responses =55) in all three groups (pleasant, unpleasant, and the combined non-neutral mood) with a smaller standard deviation.

Regarding the statements, within the 32 workers (Pleasant: N= 19, Unpleasant: N= 12, Neutral: N= 1) who agreed that mood has effect on engagement, 21 of them agreed on RW and 17 on PU. Only one worker didn't pick any of them and leave no more suggestions on the aspects of engagement. Within the 55 workers (Pleasant: N= 35, Unpleasant: N= 1, Neutral: N= 19) who agreed that mood has effect on performance, 36 agreed on the statements we provided and the rest of them did not provide details on the aspects of performance which are influenced.

Overall, we note that a predominant prior belief exists among crowdworkers as to how their mood influences their engagement and performance. However, crowdsourcing is fundamentally different from other types of work as the workers are self-motivated. These observations motivate our subsequent study, in which we collect and analyze data which tests this belief as a null hypothesis.

4 STUDY II: ANALYZING THE IMPACT OF MOOD

Given our findings in Study I, we hypothesize that the mood of crowd workers has an effect on their work engagement and performance. Thus, in Study II we aim to measure and analyze the *actual impact* of worker moods and whether it is consistent with worker beliefs. To this end, we designed a crowdsourcing task and collected data from 300 crowd workers on FigureEight.

Since several workers (N=8) cited attention as a mediator in the preliminary study, in Study II we gather and analyze workers' low-level behavioral traces (keypresses and mouse events) which have been shown to serve as indicators of attention while interacting with an interface [10, 53].

4.1 Study Design

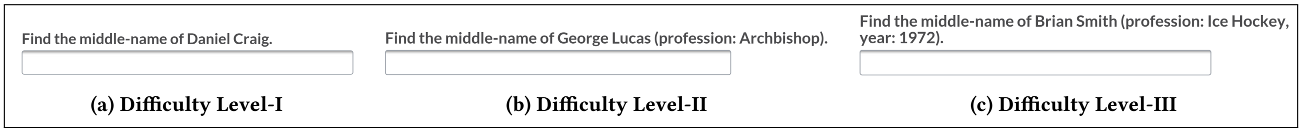
4.1.1 Measuring Engagement. Psychometric scales were used to capture the crowd workers engagement; the User Engagement Scale Short Form (UES-SF) [45], which contains four sub-scales with 12 items. Each item is presented as a statement using a 5 point scale from "*1: Strongly Disagree*" to "*5: Strongly Agree*". The reason we chose the UES-SF is that it has been validated in other HCI contexts, and to date, it is the most tested questionnaire that measures user engagement. Each of the four dimensions only has three items, which is practical in the microtask crowdsourcing context; in that it is easy to motivate workers to respond. Note that it is valid to sample sub-scales to fit the application [45].

To better measure engagement within the crowdsourcing context, we extracted two sub-scales which align with the concept of work engagement [28]: *Perceived Usability* (PU) and *Reward* (RW). Perceived usability measures the challenges workers face when performing the task, and whether the workers could conduct the task using the system the way they wanted to. Reward measures how well the experience with systems can satisfy worker needs, measures whether workers perceived the interaction as being 'successful', 'rewarding', and 'worthwhile'.

4.1.2 Task Design – Information Finding. We consider *information finding* tasks since prior work has shown a thriving market for information finding tasks on the popular microtask crowdsourcing platform, Amazon's Mechanical Turk (AMT) [13, 22]. We adopt the task of finding the middle-names of famous people, orchestrating the workflow of typical information finding microtasks where workers are asked to find contact details, addresses, or names of particular people, organizations or companies. Depending on the information that is to be searched for, these tasks possess various difficulties [63]. To account for varying levels of the inherent task difficulty in our tasks , we model task difficulty objectively into 3 levels, wherein workers need to consider an additional aspect in each progressively difficult level as shown in Figure 4. In *level-I*, workers are presented with unique names of famous persons, such that the middle-names can be found using a simple search on Google[3] or Wikipedia[4]. In *level-II* workers are additionally provided with the profession of the given person. We manually selected the names such that there are at least two different individuals with the given names in *level-II*, and the distinguishing factor that the workers need to rely upon is their profession. In *level-III* workers are presented names of persons, their profession, and a year during which the persons were active in the given profession. There are multiple distinct individuals with the given names, associated with the same profession in *level-III*. The workers are required to identify the accurate middle-name by relying on the year in which the person was active in the given profession.

4.1.3 Study Procedure. In this study, we deployed 3 crowdsourcing jobs on FigureEight corresponding to each of the 3 difficulty levels of the information finding tasks. In each case, we first administered the PAM scale, which was also used in our preliminary survey, to gather self-reports of worker moods. Following this, we asked participating workers to respond to a few general background questions regarding their age, gender, education, ethnicity, marital status and income. Next, workers received a batch of 20 information finding tasks at random corresponding to a difficulty level. The first 10 tasks in the batch were made mandatory, while the remaining 10 tasks were made optional. We offered workers a bonus of 5 USD cents for completing each of the optional tasks accurately. On completion of the information finding tasks, we administered the engagement questionnaire as described earlier. To ensure high quality responses, we also restricted the participation to *level 3 workers* on FigureEight. Enforcing reputation restrictions is a typically method adapted by requesters to ensure reliability [31]. We followed the guidelines laid down by prior work [24] and used attention check questions to label untrustworthy workers [23]. We examined the responses of workers to further flag those with an overall accuracy of 0% as being untrustworthy. We found 10, 18 and 27 untrustworthy workers in the difficulty levels I, II and III respectively. All workers were paid at an hourly rate of 7.5 USD.

[3] http://www.google.com/
[4] http://www.wikipedia.org/



(a) Difficulty Level-I **(b) Difficulty Level-II** **(c) Difficulty Level-III**

Figure 4: Progressive difficulty-levels in the *information finding* task of finding the middle-names of famous persons.

We implemented mousetracking using Javascript and the JQuery library, and logged user activity data ranging from mouse movements to keypresses. We took measures to distinguish between workers who use a mouse and those who use a touchpad. Apart from this data, we use a Javascript implementation[5] of *browser fingerprinting* [14] in order to prevent workers from participating in tasks multiple times (*'repeaters'*) by virtue of using different `worker-ids`. We take measures to avoid privacy intrusion of workers, by storing only the final hashed fingerprints. Workers were also given an opportunity to opt-out of the Javascript tracking. In this way, we gathered worker activity data from each of the jobs corresponding to the 3 difficulty levels deployed on FigureEight.

We recruited 300 workers in total; 100 distinct workers for each of the three difficulty levels. From the 300 workers, 55 workers (18.3%) did not pass the attention check questions or had 0% accuracy. In addition, some workers opted-out of our Javascript tracking and we could not track their mouse and keypress events thereby. After filtering out these workers, we were left with 216 reliable workers whose behavior we were able to track across the three levels of difficulty.

5 STUDY II RESULTS AND ANALYSIS

5.1 Demographics

Among the 216 workers after filtering, we observe a slightly unbalanced gender distribution (N=142, 65.7% male), which is typical of large crowdsourcing platforms depending on the country of origin of workers [11]. Most of the participants (N=191, 88.4%) are under 46 years old (18-24 years old: N=67 ;26-35 years old: N=72, ; 36-45 years old: N=52,). 119 of them have at least a Bachelor's degree. More than half (N=120, 55.5%) of the workers reported that crowdsourcing is their secondary source of income, and 40.3% of the workers (N=87) reported it is their primary source of income.

5.2 In What Mood Are You Today? Mood of Workers

Previous work has established that moods act as an affective background canvas of our actions and behaviour [7]. By using Pick-A-Mood (PAM) as described earlier, we obtained self-reported assessments of mood from workers before they began the actual tasks. We analyzed the responses of the 216 trustworthy workers and their mood distribution. Our findings are presented in the Figure 5. We found that on average across all conditions, most trustworthy workers claimed to be either *cheerful* (28.38%), *relaxed* (17.25%), *calm* (15.88%), or *bored* (12.5%). Relatively fewer workers were found to be *excited* (7.88%), *sad* (6.41%), *irritated* (3.56%), or *tense* (2.85%). Just over 5% of the trustworthy workers reported to be in *neutral* moods (see Figure 5).

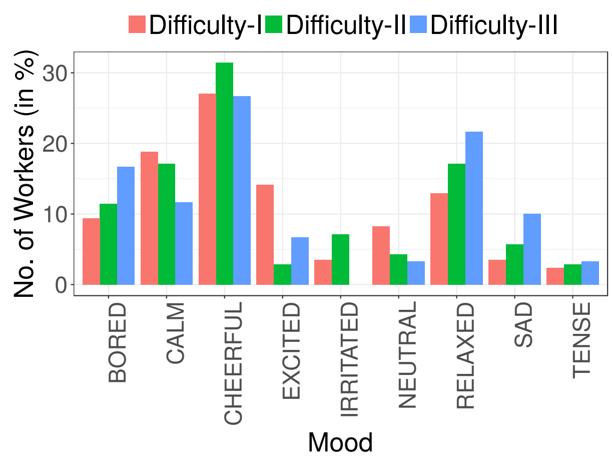


Figure 5: Mood distribution of workers who completed information findings tasks across the three different difficulty conditions (level-I, level-II and level-III).

Of the 216 trustworthy workers, 150 reported **pleasant** moods, 13 reported a **neutral** mood and 53 workers reported **unpleasant** moods.

5.3 Overview of Measures of Task Performance, Work Engagement and workers' behavior

In order to assess the task outcomes of workers in a comprehensive manner, we analyze our results through the distinct lens of *task performance*, *work engagement*, and*worker behavior*. We measured the task performance of workers using their accuracy in the first 10 mandatory information finding tasks (*Acc*@10), overall task accuracy in all the information finding tasks that they completed (*Acc*), and also time spent on completing the tasks (*TCT*). On average, we found that the workers achieved 75.6% *Acc*@10, 76.3% *Acc*, and spent 19.58 minutes to complete the entire crowdsourcing job.

Apart from the task performance, work engagement was measured through a subset of the UES-SF questionnaire [45] that covered two dimensions, Perceived Usability (*PU*) and Reward (*RW*), and number of tasks completed (*No. Tasks*). Workers completed 19.22 tasks by average, and 83.7% of them (N=181) completed all 20 tasks. *PU* measures whether workers think they complete the task the way they want, and *RW* measures whether the workers perceive the task as being successful and worthwhile. The Reliability Analysis (Cronbach's $\alpha = 0.83$ (*PU*), and 0.81 (*RW*)) indicates good internal consistency for both measures. To assign scores for

[5]http://github.com/Valve/fingerprintjs

the two sub-scales, we summed all the items within one sub-scale and divided the sum by the number of items for this sub-scale as recommended in [45]. The mean values of the two sub-scales were above 3 ($M = 3.8$ (*PU*), and 3.7 (*RW*)), signifying that on average crowd workers had a positive impression of the information finding microtasks. The correlation between these two dimensions (Spearman's $\rho = 0.324$, $p < .001$) is low, suggesting that these two dimensions do cover different aspects of engagement.

Previous works have shown that worker activity logs can be used to profile and model workers based on their behavior [51, 53], and their behavior is an indicator of attention [10]. We logged mouse events such as mouse movements[6], clicks, and the mouse entering or leaving the screen in which the job is open. We analyzed worker behavior, measured by *Mouse Events*, *Scroll Events*, *Window Events* and *Keypress Events*, based on their activity logs (mousetracking and keypresses). Mouse and Scroll events are the number of times that a mouse event or scroll event is detected. Mouse events include moving the mouse, and any clicks (right, left, middle). The scroll events we logged include scrolling up or down. We also logged window events such as the window blurring out of focus, or the window coming back into focus based on the worker activity. Every keypress executed by the workers was also logged.

To summarize, we have three measures for task performance (*Acc@10*, *Accuracy*, and *TCT* (in mins)); three measures for work engagement (*PU*, *RW*, and *No.Tasks*); and four measures for workers' behavior (*Mouse Events*, *Scroll Events*, *Window Events* and *Keypress Events*).

5.4 Effect of Mood on Task Performance and Work Engagement

We analyzed the mean values of crowd workers' task performance and user engagement with different moods, and compare the *Pleasant* mood group against the *Unpleasant* mood group using a corrected t-test. The rationale behind comparing only the two non-neutral groups is based on the small sample size ($N = 13$) of the *Neutral* mood group. Our analysis of statistical significance in this paper is comprised of 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 $\alpha < .05$.

Table 3 summarizes the mean value for the 6 measures, namely *Acc@10*, *Accuracy*, *TCT*, *PU*, *RW*, and *No.Tasks*, for three types of mood (Pleasant, Neutral, and Unpleasant). With the unbalanced sample size, crowd workers with neutral mood performed the best with a mean *Acc@10* of 80.7% ± 17.1% , and over all accuracy of 81.8% ± 18.6%. Crowd workers in pleasant moods performed better (*Acc@10*: 75.7% ± 20.3%, *Accuracy*: 76.5% ± 20.7%) than those in unpleasant ones (*Acc@10*: 74% ± 22.3%, *Accuracy*: 74.3% ± 23.6%), without significant differences on the two accuracy measures. Although workers in the neutral mood performed the best among the three groups, this result may be a reflection of the small sample size ($N = 13$). The mean value of *Accuracy* is higher than *Acc@10* across all three groups, suggesting a well-known learning effect in crowd work [12, 20]. Crowd workers in an unpleasant mood spent less time on the task, completed fewer tasks and reported a lower usability score than workers in a pleasant mood or those who reported neutral moods, but all those differences are not significant. Significant differences were observed between the *Pleasant* and *Unpleasant* groups on *RW* (t=3.061, $p < 0.01$), in which pleasant workers (*(M=3.86)*) felt that the tasks were more rewarding than unpleasant workers (*(M=3.45)*). This aligns with previous findings [28, 57] in traditional workplace settings, wherein workers in a positive mood perceived their work to be rewarding and tended to invest more in their role.

Table 3: Average values for the 10 task performance, work engagement, and task behavior measures for each of the three mood conditions (corrected t-test to compare the Pleasant group against the Unpleasant group, significance level (2-tailed): $^{**} = \alpha < 0.01$.

		Pleasant N=150	**Neutral** N=13	**Unpleasant** N=53
Perf.	*Acc.@10*	75.7%	80.7%	74%
	Acc.	76.5%	81.8%	74.3%
	TCT	19.68	19.61	19.28
Eng.	*PU*	3.81	3.82	3.67
	RW	3.86^{**}	3.59	3.45^{**}
	No. Tasks	19.26	19.46	19.03
Beh.	*Mouse Eve.*	666.53	616.85	637.87
	Scroll Eve.	928.18	811.77	765.1
	Window Eve.	87	75.92	89.23
	Keypress Eve.	50.23	25.08	57.56

To analyze the effect of mood on the behavior of workers, we conducted t-tests for each type of the events between two non-neutral mood groups. For all four types of events, no significant differences were observed between workers in pleasant moods when compared to those in unpleasant moods. The workers in a pleasant mood performed more mouse events *(M=666.53, SD= 352.70)* and scroll events *(M= 928.18, SD= 749.97)* than the workers in unpleasant moods (*mouse events : M=637.86, SD= 312.03; scroll events : M= 765.1, SD= 535.55*). This trend was reversed on the frequency with which workers left and re-entered the active window where the crowdsourcing job was open; here workers in pleasant moods *(M=87, SD=52.07)* corresponded to fewer events than those in unpleasant moods *(M=89.23, SD=59.01)*. Similarly, workers in pleasant moods *(M=50.23, SD=48.61)* performed fewer key presses than those in unpleasant moods *(M=57.56, SD= 60.26)*. The large standard deviation values of all behavior measures indicate that the data are spread out over a wide range of values. With a close look at the log files, we noticed that it is caused by workers' different interaction habits. For example, in order to check materials in the lower position of the web page, some workers scrolled down, which count as scroll events in the log, while others moved their mouse and clicked a lower position on the scroll bar, which count as mouse events. Similar in performing searching, instead of typing in the search box, some workers copy-pasted the names to the box.

6 DISCUSSION, CAVEATS AND LIMITATIONS

Our analysis provides several complex insights into the role that the mood of workers plays in relation to their performance, engagement

[6]Mouse movement events were logged at an interval of 500ms.

and outcomes in the microtask crowdsourcing context. We shape our main arguments around the following job characteristics model for crowdsourcing work. The Job Characteristics Model (JCM) [25], a dominant and widely tested model from work psychology (see detailed tests of JCM in [27]), identifies five core job characteristics: skill variety, task identity, task significance, autonomy, and feedback. The first three characteristics are related to the skills required for completing the task, task difficulty and task impact; thus, they are intrinsically task-centered. Autonomy is the degree to which the job provides substantial freedom, independence, and discretion to the individual in scheduling the work and in determining the procedures to be used in its performance, and is the only characteristic positively linked to objectively measured work performance. Feedback is the degree to which the individual obtains direct and clear information about the effectiveness of his or her performance. The five characteristics promote individual motivation, job satisfaction, and performance through critical psychological states such as experienced responsibility for the outcomes of the work.

6.1 Effects on Performance

No significant differences are observed between workers in a pleasant mood and those in an unpleasant mood with respect to their task performance, namely *Acc@10*, *Accuracy*, and *TCT*, although 55 out of 95 workers (57.9%) participating in the preliminary study reported they believe otherwise. This lack of impact can potentially be explained by the inherent characteristics of crowdsourcing. Contrary to traditional types of labour, gig work, including microtask crowdsourcing, represents an on-demand work activity that benefits from the worker's self-directed decision to work. This is in turn a multi-faceted act, comprising of several micro-choices on the aspect of time, environment and content. Such a latitude of options reflects the high autonomy of the subject in performing their tasks. Furthermore, the choice of whether to work or not represents a primer on the subject's behaviour, which could potentially dampen the effects of mood on the subject's interaction with the system and their task. Based on the tests performed on the JCM [27], we know that autonomy is principally related to the objective task outcome, and, in light of the previous discussion, it is likely that this component accounts for the relative insensitivity of performance metrics to the worker's mood. Moreover, the differences between these findings and prior work in affective computing, where authors found people in happier moods tend to be more creative [34] might be due to the fact that creativity is measured subjectively [34], and subjective performance is often influenced by a spectrum of psychological factors. Another explanation can draw from the drive for compensation among workers. It is well understood that crowd workers on paid microtask crowdsourcing platforms primarily seek monetary rewards, and platforms and requesters introduce various forms of quality control and may subsequently refuse payment if the work produced is suboptimal. Thus a worker's desire to be compensated will drive them to perform at least as well as necessary to achieve this goal. It is also worth noting that our selection of workers (all *level 3 workers*), motivated primarily by the need for high-quality data renders the underlying performance distribution rather sharp. Investigating unleveled crowdworkers from this perspective could be the subject of future work.

6.2 Effects on Engagement

As reported, we found a significant difference between workers in a pleasant mood in comparison to those in an unpleasant mood with respect to their perceived reward (*RW*). The perceived reward, i.e. the inherent gratification obtained through the task completion, irrespective of other material compensation interacts on a primal level with the subject's motivation for tasking (maximizing both the personal and monetary gain). Experiencing meaningfulness is a critical psychological state derived from the five job characteristics and it is thus unsurprising that workers in a pleasant mood are influenced more heavily (cf. [27]). This is furthermore inline with our intuition, due to the reported results of the preliminary survey.

Apart from perceived reward, no significant difference is found with respect to perceived usability (*PU*) and *No.Tasks*. *PU* is fundamentally a characteristic of the system and task design. Although surprising, it becomes apparent as an outcome of this study that mood has very little influence on this dimension, a fact which could be due to multiple factors. As the batch of tasks in Study II correspond to information finding, an ordinary crowdsourcing task [13], this may result in workers falling out of a *meta-critical* state of mind. This in turn can explain the very low variance in this dimension. The lack of variance in the *No. Tasks* can be accounted for by the fact that most workers completed the entire batch of tasks immaterial of the corresponding task difficulty (fewer than 17% failed to complete the entire batch). It is plausible that a larger batch would result in more variance in *No. Tasks*, but considering that workers spent almost 20 mins to complete 20 tasks on average, it is a large enough batch for microtasks. Prior works have revealed that paying workers well could help in worker retention [12], which is also the likely explanation for our observation.

6.3 Effects on Behavior

The lack of significant difference between the worker's low-level behaviour (comprising of *Mouse Events*, *Scroll Events*, *Window Events* and *Keypress Events*) carries perhaps a more trivial explanation, namely that multiple combinations of different actions can lead to the same outcome within the interaction (two *Keypress Events* can substitute for two *Mouse Events* and vice-versa). Analyzing behaviour at either higher or lower level (e.g., aggregating behaviour patterns based on intent or goal achieved) could prove far more fruitful. We defer this analysis to future studies.

6.4 Implications for Crowsdsourcing Systems

Our paper reflects on how different crowdsouring is from more traditional work settings. The effects of mood have been studied in order to capture this difference through its impact on performance and engagement. But how different is crowdsourcing really, and what are the commonalities it shares with all types of work? For one, the psychological factors present in workers seem to coincide at least at a macro-scale. Considering human factors in improving the effectiveness of the crowdsourcing results is the object of many studies such as [29–31]. But worker-centric studies of this nature could highlight how these goals could be better achieved while developing a system which is sensitive to the principles of work psychology we have been relying on for decades. Several studies have focused on improving the effectiveness of crowdsourcing results through matching eligible workers to the most appropriate task [38, 46]. The main orientation of these tools is derived as an extension of the task-setters' goals and perspective, and constitute a brute-force approach for harnessing more accurate results. Self-assessment has also been proposed as an effective ingredient in crowd work [21], however, none of these artifacts focus on the worker's well-being. Understanding the nature of the crowd worker's environment, interactions, incentives and gratification can serve to achieve the same goals but with a less heavy-handed, more *bottom-up* philosophy.

Stimulating crowd workers to improve engagement, more than accuracy, which we have already scoped as relatively constant across worker experience levels, could be far more beneficial and could be as easily achievable as providing new task designs focused on enhancing perceived reward, or introducing a function which serves as a source of constant positive reinforcement such as a continuously updating performance dashboard.

To summarize, the results of such studies potentially carry keys to providing major improvements to the lives, well-being and mental health of crowdworkers, by recognizing their particular type of work economy as fundamentally innovative and disjoint from its more traditional counterparts. Worker-centric studies shift the focus from the task requester's goals by recognizing and identifying the virtues behind the particular types of interactions of the worker with their task. We show how maximizing task performance measures needn't be a concern when it comes to crowdsourcing in as much as it interacts with a worker's mood, and we remark how improving a worker's feeling of reward can be far more beneficial with an impact on the worker's long term career development. Tailoring tasks to this goal will also help create more quality crowdsouring jobs in the future.

7 CONCLUSIONS AND FUTURE WORK

We have presented a study in which the effects of crowdworkers' moods have been explored through the contrast between prior beliefs of workers, with actual quantitative evidence stemming from behavioral analysis of data gathered from exponents of the very same marketplace. In summary, we take away how mood provides a high impact on perceived reward or task gratification and far less so on performance or perceived usability.

We emphasize how our findings are central to better understanding this work sector, and encourage a shift in focus to worker-centric studies that would take these findings into account to improve the crowdsourcing system in a manner aligned with the worker's own interests, namely enhancing worker engagement without the need for negative reinforcement. We identify task design and system design (e.g., using positively valenced music can enhance the creative performance of workers on AMT [42]) as prime opportunities in achieving this goal.

In the imminent future, we aim to carry out additional studies across different types of crowdsourced tasks (for example, content creation tasks, verification and validation tasks, interpretation and analysis, etc.) to analyze how sensitive a type of task is to crowd worker moods based on the task type. Additionally, we will consider more worker types (e.g., unleveled workers, or workers with different culture background [35]) while collecting mood for the generalizability of the results. We will also explore methods that can robustly induce positive moods to improve task related outcomes in crowdsourced microtasks.

ACKNOWLEDGEMENTS

We would like to thank all the anonymous crowd workers who participated in our experiments. This research has been supported in part by the Erasmus+ project DISKOW (grant no. 60171990), and the EU Horizon 2020 transnational access program under SoBigData (grant agreement no. 654024).

REFERENCES

[1] Arnold B Bakker, Wilmar B Schaufeli, Michael P Leiter, and Toon W Taris. 2008. Work engagement: An emerging concept in occupational health psychology. *Work & Stress* 22, 3 (2008), 187–200.
[2] Christopher Beedie, Peter Terry, and Andrew Lane. 2005. Distinctions between emotion and mood. *Cognition & Emotion* 19, 6 (2005), 847–878.
[3] Borong Chen, Weiping Hu, and Jonathan A Plucker. 2016. The effect of mood on problem finding in scientific creativity. *The Journal of Creative Behavior* 50, 4 (2016), 308–320.
[4] Michael S. Christian, Adela S. Garza, and Jerel E. Slaughter. 2011. Work engagement: A quantitative review and test of its relations with task and contextual performance. *Personnel Psychology* 64, 1 (2011), 89–136.
[5] David M Clark, John D Teasdale, Donald E Broadbent, and Maryanne Martin. 1983. Effect of mood on lexical decisions. *Bulletin of the Psychonomic Society* 21, 3 (1983), 175–178.
[6] Peng Dai, Jeffrey M Rzeszotarski, Praveen Paritosh, and Ed H Chi. 2015. And now for something completely different: Improving crowdsourcing workflows with micro-diversions. In *Proceeding of The 18th ACM Conference on Computer-Supported Cooperative Work and Social Computing*. ACM, 628–638.
[7] Richard J Davidson. 1994. On emotion, mood, and related affective constructs. *The nature of emotion: Fundamental questions* (1994), 51–55.
[8] Gianluca Demartini, Djellel Eddine Difallah, Ujwal Gadiraju, Michele Catasta, et al. 2017. An introduction to hybrid human-machine information systems. *Foundations and Trends® in Web Science* 7, 1 (2017), 1–87.
[9] 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.
[10] Fernando Diaz, Ryen White, Georg Buscher, and Dan Liebling. 2013. Robust models of mouse movement on dynamic web search results pages. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*. ACM, 1451–1460.
[11] 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*. ACM, New York, NY, USA, 135–143.
[12] Djellel Eddine Difallah, Michele Catasta, Gianluca Demartini, and Philippe Cudré-Mauroux. 2014. Scaling-up the crowd: Micro-task pricing schemes for worker retention and latency improvement. In *Second AAAI Conference on Human Computation and Crowdsourcing*.
[13] 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.
[14] Peter Eckersley. 2010. How unique is your web browser?. In *Privacy Enhancing Technologies*. Springer, 1–18.
[15] Alek Felstiner. 2011. Working the crowd: employment and labor law in the crowdsourcing industry. *Berkeley J. Emp. & Lab. L.* 32 (2011), 143.
[16] Oluwaseyi Feyisetan, Elena Simperl, Max Van Kleek, and Nigel Shadbolt. 2015. Improving paid microtasks through gamification and adaptive furtherance incentives. In *Proc. WWW'15*. 333–343.
[17] Nico H Frijda et al. 1994. Varieties of affect: Emotions and episodes, moods, and sentiments. (1994).
[18] Ujwal Gadiraju, Alessandro Checco, Neha Gupta, and Gianluca Demartini. 2017. Modus operandi of crowd workers: The invisible role of microtask work environments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3, 49.
[19] Ujwal Gadiraju, Gianluca Demartini, Djellel Eddine Difallah, and Michele Catasta. 2016. It's getting crowded!: how to use crowdsourcing effectively for web science research. In *Proceedings of the 8th ACM Conference on Web Science*. ACM, 11–11.
[20] Ujwal Gadiraju and Stefan Dietze. 2017. Improving learning through achievement priming in crowdsourced information finding microtasks. In *proceedings of the*

Seventh International Learning Analytics & Knowledge Conference. ACM, 105–114.

[21] 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* 24, 4, Article 30 (Aug. 2017), 26 pages.

[22] Ujwal Gadiraju, Ricardo Kawase, and Stefan Dietze. 2014. A taxonomy of microtasks on the web. In *Proceedings of the 25th ACM conference on Hypertext and social media*. ACM, 218–223.

[23] 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*. ACM, 1631–1640.

[24] Ujwal Gadiraju, Jie Yang, and Alessandro Bozzon. 2017. Clarity is a worthwhile quality: On the role of task clarity in microtask crowdsourcing. In *Proceedings of the 28th ACM Conference on Hypertext and Social Media*. ACM, 5–14.

[25] J Richard Hackman and Greg R Oldham. 1976. Motivation through the design of work: Test of a theory. *Organizational behavior and human performance* 16, 2 (1976), 250–279.

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

[27] Stephen E Humphrey, Jennifer D Nahrgang, and Frederick P Morgeson. 2007. Integrating motivational, social, and contextual work design features: a meta-analytic summary and theoretical extension of the work design literature. *Journal of applied psychology* 92, 5 (2007), 1332.

[28] William A Kahn. 1990. Psychological conditions of personal engagement and disengagement at work. *Academy of management journal* 33, 4 (1990), 692–724.

[29] Nicolas Kaufmann, Thimo Schulze, and Daniel Veit. 2011. More than fun and money. Worker Motivation in Crowdsourcing-A Study on Mechanical Turk.. In *AMCIS*, Vol. 11. 1–11.

[30] Gabriella Kazai, Jaap Kamps, and Natasa Milic-Frayling. 2013. An analysis of human factors and label accuracy in crowdsourcing relevance judgments. *Information retrieval* 16, 2 (2013), 138–178.

[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 19th ACM Conference on Computer Supported Cooperative Work*. ACM, 1301–1318.

[32] Walter S Lasecki, Samuel C White, Kyle I Murray, and Jeffrey P Bigham. 2012. Crowd memory: Learning in the collective. In *Proceedings of the Collective Intelligence*.

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

[34] Sheena Lewis, Mira Dontcheva, and Elizabeth Gerber. 2011. Affective computational priming and creativity. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 735–744.

[35] Nangyeon Lim. 2016. Cultural differences in emotion: differences in emotional arousal level between the East and the West. *Integrative Medicine Research* 5, 2 (2016), 105–109.

[36] Andrew Mao, Ece Kamar, and Eric Horvitz. 2013. Why stop now? predicting worker engagement in online crowdsourcing. In *First AAAI Conference on Human Computation and Crowdsourcing*.

[37] Catherine C Marshall and Frank M Shipman. 2013. Experiences surveying the crowd: Reflections on methods, participation, and reliability. In *Proceedings of the 5th Annual ACM Web Science Conference*. ACM, 234–243.

[38] Panagiotis Mavridis, David Gross-Amblard, and Zoltán Miklós. 2016. Using hierarchical skills for optimized task assignment in knowledge-intensive crowdsourcing. In *Proceedings of the 25th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 843–853.

[39] 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.

[40] Andrew G. Miner and Theresa M. Glomb. 2010. State mood, task performance, and behavior at work: A within-persons approach. *Organizational Behavior and Human Decision Processes* 112, 1 (2010), 43 – 57.

[41] Robert R Morris, Mira Dontcheva, Adam Finkelstein, and Elizabeth Gerber. 2013. Affect and creative performance on crowdsourcing platforms. In *Proceedings of the 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*. 67–72.

[42] Robert R Morris, Mira Dontcheva, Adam Finkelstein, and Elizabeth Gerber. 2013. Affect and creative performance on crowdsourcing platforms. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*. IEEE, 67–72.

[43] Robert R Morris, Mira Dontcheva, and Elizabeth M Gerber. 2012. Priming for better performance in microtask crowdsourcing environments. *IEEE Internet Computing* 16, 5 (2012), 13–19.

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

[45] Heather L. O'Brien, Paul Cairns, and Mark Hall. 2018. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. *Intl. J. Human-Computer Studies* 112 (2018), 28 – 39.

[46] Peter Organisciak, Jaime Teevan, Susan Dumais, Robert C Miller, and Adam Tauman Kalai. 2014. A crowd of your own: Crowdsourcing for on-demand personalization. In *Second AAAI Conference on Human Computation and Crowdsourcing*.

[47] Andrew J Oswald, Eugenio Proto, and Daniel Sgroi. 2015. Happiness and productivity. *Journal of Labor Economics* 33, 4 (2015), 789–822.

[48] Antonia Pilar Pacheco-Unguetti and Fabrice B. R. Parmentier. 2016. Happiness increases distraction by auditory deviant stimuli. *British Journal of Psychology* 107, 3 (2016), 419–433.

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

[50] Markus Rokicki, Sergiu Chelaru, Sergej Zerr, and Stefan Siersdorfer. 2014. Competitive game designs for improving the cost effectiveness of crowdsourcing. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*. ACM, 1469–1478.

[51] Jeffrey Rzeszotarski and Aniket Kittur. 2012. CrowdScape: interactively visualizing user behavior and output. In *Proceedings of the 25th annual ACM symposium on User interface software and technology*. ACM, 55–62.

[52] Jeffrey M Rzeszotarski, Ed Chi, Praveen Paritosh, and Peng Dai. 2013. Inserting micro-breaks into crowdsourcing workflows. In *First AAAI Conference on Human Computation and Crowdsourcing*.

[53] Jeffrey M Rzeszotarski and Aniket Kittur. 2011. Instrumenting the crowd: using implicit behavioral measures to predict task performance. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software & Technology*. ACM, 13–22.

[54] Pennie S Seibert and Henry C Ellis. 1991. Irrelevant thoughts, emotional mood states, and cognitive task performance. *Memory & Cognition* 19, 5 (1991), 507–513.

[55] Han Yu Zhiqi Shen, Simon Fauvel, and Lizhen Cui. 2017. Efficient scheduling in crowdsourcing based on workers' mood. In *Agents (ICA), 2017 IEEE International Conference on*. IEEE, 121–126.

[56] Melanie J Taylor and Peter J Cooper. 1992. An experimental study of the effect of mood on body size perception. *Behaviour research and therapy* 30, 1 (1992), 53–58.

[57] Elizabeth R Tenney, Jared M Poole, and Ed Diener. 2016. Does positivity enhance work performance?: Why, when, and what we donâĂŹt know. *Research in Organizational Behavior* 36 (2016), 27–46.

[58] Wei-Chi Tsai, Chien-Cheng Chen, and Hui-Lu Liu. 2007. Test of a model linking employee positive moods and task performance. *Journal of Applied Psychology* 92, 6 (2007), 1570.

[59] Viswanath Venkatesh and Cheri Speier. 1999. Computer technology training in the workplace: A longitudinal investigation of the effect of mood. *Organizational behavior and human decision processes* 79, 1 (1999), 1–28.

[60] 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.

[61] Bjørn Villa, Katrien De Moor, Poul Heegaard, and Anders Instefjord. 2013. Investigating quality of experience in the context of adaptive video streaming: findings from an experimental user study. In *Norsk informatikkonferanse NIK 2013, Universitetet i Stavanger, 18.-20. november 2013*. Akademika forlag, 122–133.

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

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