Improving Learning through Achievement Priming in Crowdsourced Information Finding Microtasks

Ujwal Gadiraju
L3S Research Center
Leibniz Universität Hannover
Appelstr. 9a, Hannover, Germany
gadiraju@L3S.de

Stefan Dietze
L3S Research Center
Leibniz Universität Hannover
Appelstr. 9a, Hannover, Germany
dietze@L3S.de

ABSTRACT

Crowdsourcing has become an increasingly popular means to acquire human input on demand. Microtask crowdsourcing marketplaces facilitate the access to millions of people (called *workers*) who are willing to participate in tasks in return for monetary rewards or other forms of compensation. This paradigm presents a unique learning context where workers have to learn to complete tasks on-the-fly by applying their learning immediately through the course of tasks. However, most workers typically dropout early in large batches of tasks, depriving themselves of the opportunity to learn on-the-fly through the course of batch completion. By doing so workers squander a potential chance at improving their performance and completing tasks effectively. In this paper, we propose a novel method to engage and retain workers, to improve their learning in crowdsourced information finding tasks by using achievement priming. Through rigorous experimental findings, we show that it is possible to retain workers in long batches of tasks by triggering their inherent motivation to achieve and excel. As a consequence of increased worker retention, we find that workers learn to perform more effectively, depicting relatively more stable accuracy and lower task completion times in comparison to workers who drop out early.

CCS CONCEPTS

•Applied computing → Education; •Information systems → *World Wide Web;* **•Human-centered computing →** *Human computer interaction (HCI);*

KEYWORDS

Crowdsourcing, Microtasks, Learning, Retention, Crowd Workers, Information Finding, Achievement Priming

ACM Reference format:
Ujwal Gadiraju and Stefan Dietze. 2016. Improving Learning through Achievement Priming in Crowdsourced Information Finding Microtasks. In *Proceedings of ACM LAK Conference, Vancouver, BC, Canada, March 13-17, 2017 (LAK'17),* 10 pages.
DOI: 10.1145/nnnnnnn.nnnnnnn

DOI: 10.1145/nnnnnnn.nnnnnnn

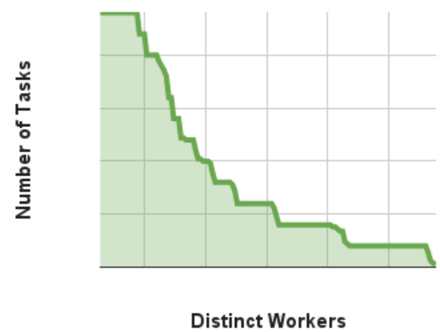


Figure 1: Typical task consumption in long batches of tasks, depicting low worker retention rate (amount of work on the y-axis, distinct workers on the x-axis). Very few workers complete all the available tasks in a batch, thereby maximizing the opportunity to *learn* through the course of the batch. Most workers drop out after completing only a few tasks in a long batch.

1 INTRODUCTION

In a technologically advanced world today, there are still several problems that cannot be solved by machines alone and require human intervention. In the last decade crowdsourcing has emerged as an effective paradigm that enables access to human intelligence at scale, thereby playing a pivotal role in hybrid man-machine systems. Paid crowdsourcing platforms provide a means to reach millions of people around the world (called *crowd workers*), creating endless opportunities to acquire human input at scale in return for monetary rewards. Amazon's Mechanical Turk (AMT[1]) and CrowdFlower[2] are popular examples of microtask crowdsourcing platforms, where *requesters*[3] can easily access labor on demand and leverage the prevalent wisdom and skills to satisfy varying requirements. The principle categories of crowdsourced work on such platforms are diverse; information finding, verification and validation, interpretation and analysis, content creation, content access and surveys [15].

Understanding various aspects of crowd workers such as their motivations, behavior, and capabilities has been pivotal in building crowd-powered systems that are efficient (in terms of costs incurred) and effective (in terms of quality of the results produced) [16, 24]. Crowd workers have been shown to participate in crowdsourcing

[1]https://www.mturk.com/
[2]https://www.crowdflower.com/
[3]Requesters are people who deploy tasks on crowdsourcing platforms to gather responses from crowd workers.

due to various intrinsic and extrinsic motivations; to earn their primary livelihood or as a secondary source of income [22], for community or enjoyment related motivations [2, 3], or due to the meaningfulness of their contributions [6]. Inherent human factors play an important role in the dynamics of crowd work on microtask platforms [25, 34].

Previous work in technology enhanced learning has brought to fore the unique *learning environment* that characterizes the crowdsourcing paradigm [14]. Authors reflected on the short-lived learning phase and the immediate application of learned concepts in crowdsourcing microtasks. While it is a general notion that crowdsourcing microtasks are short and can be completed quickly, a recent analysis of tasks on AMT over a five year period (from 2009-14) has shown that a majority of tasks are deployed in large batches [9]. The authors showed that such large batches of repetitive tasks are more attractive to crowd workers compared to shorter batches. However, it was found that long batches tend to starve towards the end; with a fewer number of available tasks to complete in such batches, fewer workers choose to contribute. Retaining workers in tasks for as long as possible is useful since workers who gain experience through the course of task completion become more effective. Empirical evidence supports the notion that workers who complete large batches of tasks exhibit a tendency to *learn* through the course of batch completion [8]. This suggests that by retaining workers longer in microtasks of a given type, it is possible to induce learning, resulting in improved worker performance over time. In this paper, we tackle the problem of retaining crowd workers in large batches of microtasks to facilitate learning among workers.

We aim to improve worker retention in the real-world crowdsourcing microtask category of *information finding*. Information finding tasks were found to be on the rise on AMT in a recent data-driven study [9]. Some of such typical tasks that were deployed on AMT in the past include finding contact details on a list of websites, email addresses of a series of people, TV-shows listed on different channels, and so forth. Often such tasks involve large batches where workers need to adopt and repeat the same workflow to accomplish the objectives. We explore the potential of triggering the motivation to achieve among crowd workers, with an aim to improve their retention in crowdsourcing tasks and thereby facilitate learning. To this end, we investigate the applicability of *achievement priming* to retain workers in information finding tasks. We measure the effectiveness of our approach in terms of (i) the *worker retention rate*, defined by the average number of tasks that workers complete in a given batch of available tasks, (ii) *worker performance*, i.e., the accuracy with which workers complete the tasks, and (iii) the worker *learning rate*, that describes the average change in worker performance through the course of batch completion.

2 RESEARCH QUESTIONS

Several studies have affirmed that the primary motivation of most workers to participate in crowd work is to earn monetary rewards that can contribute to their income and livelihood [15, 22]. We also know that crowd workers aim to maintain a high level of accuracy in their work and a good reputation to access more tasks, thereby maximizing their earnings [23]. We build on this understanding of the crowd workers' general attitude to accomplish good work, and investigate the potential of stimulating their motivation to elicit good quality and sustained work.

Achievement Motivation. Achievement motivation can be defined as the need for success or attainment of excellence [32]. A number of prior works in psychology have studied achievement motivation. McClelland first proposed that achievement motivation may be understood as a motivational process that involves the regulation of different social goals [28]. This was supported by more recent research corresponding to goal pursuits [35]. In this paper, we adopt the understanding of achievement motivation proposed by Hart and Albarracin [20]; the goal to achieve is an alternative to the goal to have fun or indulge in leisurely activity. The authors show that people choose to pursue excellence (at the expense of having fun) or pursue fun (at the expense of achieving) depending on their level of chronic achievement motivation (the amount of pleasure in achieving goals). People with high chronic achievement motivation exhibit goal-seeking behavior when they encounter motivational triggers, while those with low chronic achievement motivation exhibit a fun-seeking behavior in such cases.

We aim to address the following research questions with respect to information finding (*IF*) tasks.

RQ1 How can achievement priming be used to increase worker retention and facilitate learning in *IF* tasks?

RQ2 How is the learning process of workers effected when they are retained in long batches of *IF* tasks?

Based on prior works discussed earlier and crowd worker motivations, we construct the following hypotheses.

Hyp-I : Worker Retention On presenting crowd workers with motivational stimuli in the form of achievement primes, we can improve the worker retention rate and thereby facilitate learning.

Hyp-II : Worker Learning When workers are retained in long batches of information finding tasks, they learn more about the tasks and perform more effectively.

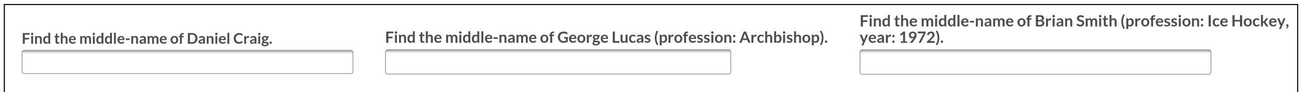
3 RELATED LITERATURE

The unconscious manipulation of participants' behavior (in the sense that subjects are not aware of being manipulated [21]) by means of semantic or pictorial cues is known as 'priming to behavior' [1]. A large number of studies have presented empirical evidence that it is possible to positively affect participants' achievements in different types of tasks by means of verbal ([12, 20]) or pictorial ([36]) cues.

We position and compare our work in this paper to the following two distinct realms of related work.

3.1 Priming in Crowdsourcing Environments

Researchers have also studied the impact of affective priming on creativity of people. Lewis et al. presented a method for manipulating affect by using pictures [26]. Through their experiments with crowd workers on Amazon's Mechanical Turk (AMT), the authors showed that positive affective priming helped to improve the quality of ideas generated in two tasks testing the creativity of workers. Similarly, Morris et al. explored the use of affective priming and affective pre-screening to improve the creativity of crowd workers [29, 30]. In these works, the authors used musical excerpts as postive and negative primes, and found that positively



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

Figure 2: Progressive difficulty-levels in the *information finding* task of finding the middle-names of famous persons.

valenced music can significantly enhance the creative performance of workers on AMT. Following research that suggested the influence of individual personality differences in performance with visualizations [18], Harrison et al. showed that affective priming can be used to influence user performance in classic graphical perception tasks [19].

Although these prior works use crowd workers and crowdsourcing platforms to show that affective priming can help in improving the performance of workers, the tasks that were used to gauge the impact of affective priming were largely cognitive tests requiring insight and creative problem-solving, that do not sufficiently correspond to the landscape of real-world microtasks. To our knowledge this is the first work that investigates the applicability of priming techniques in real-world microtask crowdsourcing.

3.2 Learning and Retention in Microtasks

In previous work, authors analyzed the task performance and learning outcomes in a real-world classroom setup, showing that appropriate learning conditions resulted in an improvement in task performance [10]. Other work proposed the use of crowdsourcing in classrooms to improve the learning process of students by receiving feedback on learning material [11]. In [14], authors introduced crowd workers as 'learners' in a unique learning environment and showed that implicit and explicit training can help to improve the performance of workers in crowdsourced microtasks.

It is known that a majority of crowdsourced microtasks are repetitive in nature and consist of batches of similar tasks [9]. Working for long periods on such tasks can lead to boredom and fatigue, resulting in a potential drop in worker performance and productivity. Previous works have addressed this issue and proposed different means to retain and engage workers. Rzeszotarski et al. suggested introducing micro-breaks into workflows to refresh workers, and showed that under certain conditions micro-breaks help to retain workers and improve their accuracy slightly [33]. Similarly, Dai et al. 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 [7]. Other works proposed the use of gamification to increase worker retention and throughput [13]. Mao et al. studied worker engagement, characterized how workers perceive tasks and proposed to predict when workers would stop performing tasks [27]. Difallah et al. 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 [8]. A side effect of workers dropping out early in long batches is the lack of opportunity to facilitate learning among participating workers.

In contrast to these prior works, we aim to improve worker retention and learning rate by relying on inherent characteristics of workers (i.e., their level of chronic achievement motivation), thereby triggering goal-seeking behavior. Moreover, channeling the workers' achievement motivation to improve worker engagement and retention is a relatively less intrusive approach due to no tangible change in the workflow, from a worker's standpoint.

4 METHODOLOGY AND SETUP

4.1 Task Design - Information Finding

Since there has been a steep rise in information finding tasks on the most popular microtask crowdsourcing platform, Amazon's Mechanical Turk (AMT), we consider this type of tasks [9]. We adopt the task of finding the middle-names of famous people, to emulate the workflow of real-world information finding microtasks where workers are typically asked to find contact details, addresses, or names of particular people, organizations or companies. Depending on the information that is to be searched for on the web, these tasks may comprise of varying difficulty. Recent work has shown how *task complexity* plays an important role in worker performance [37]. To account for varying levels of the inherent task difficulty in our information finding tasks and to study the impact of task difficulty on worker learning rate, 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 2. 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[4] or Wikipedia[5]. 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.

4.2 Inspiring Quotes as Achievement Primes

In their studies of achievement behavior, Hart and Albarracin used words that emulated achievement-related meanings as achievement primes (such as `win, master, achieve, excel` and so forth) [20]. In this work, we hypothesize that inspirational quotes about achievement by famous and successful figures can instill a similar priming effect. We manually collected 100 quotes from http://www.brainyquote.com/ by searching for quotes related to 'achievement'. To pick the quotes which can be considered to emulate inspiration, we deployed a crowdsourcing task on CrowdFlower, gathering 10

[4]http://google.com/
[5]http://en.wikipedia.org/

judgments from distinct workers on a 5-point Likert scale (as shown in Figure 3) on each of the 100 quotes. We awarded workers with 4 USD cents for every 10 quotes that they rated, and controlled for quality by using attention check questions [16]. Based on the average aggregated rating corresponding to each quote, we consider the top 25 quotes as our achievement primes (all with an average rating >= 4.5).

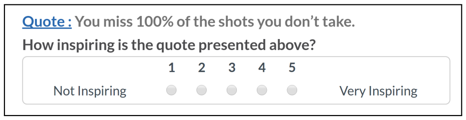


Figure 3: Task to standardize inspiring quotes.

4.3 Measuring Achievement Motivation

We measure the level of chronic achievement motivation of workers using the *excellence motivation* subscale introduced by Cassidy and Lynn to capture one's motivation to pursue standards of excellence [5]. The scale had good internal reliability by means of Cronbach's reliability coefficient, $\alpha = .71$. Workers rate seven questions[6] on a 5-point Likert scale ranging from '*1:Not at all like me*' to '*5:Extremely like me*'. We computed the level of chronic achievement motivation for each worker by adding their responses to each of the seven questions after appropriate reverse coding. Workers with an aggregated score greater than the scale's midpoint (21) were considered to be more achievement-oriented, and the remaining workers were considered to be more 'fun-oriented'. We refer to the two groups of workers as the *ACHIEVE* and *FUN* groups respectively.

4.4 Study Design

We consider the following variations in our studies.
(i). *Passive Achievement Priming* (**AP-Passive**) – In this setting, crowd workers are presented with the quotes amidst the actual information finding units of the task. Workers do not necessarily have to interact with quotes beyond reading them (see Figure 4). The quotes act as passive achievement primes, and are randomly interspersed among units of the actual task. The order in which the quotes appear is also randomized to prevent ordering bias effects.

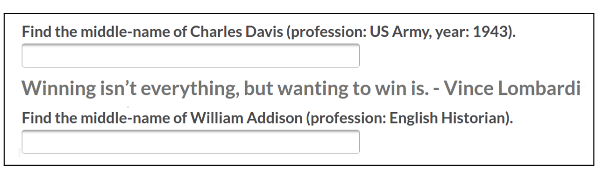


Figure 4: A passive achievement prime embedded between two units of the information finding task.

We deployed an initial task consisting of the *excellence motivation scale* to measure the achievement motivation of workers on CrowdFlower. The task consisted of a few background questions, an attention check question, and seven questions corresponding to measuring their level of achievement motivation. On completion of this task, workers were provided a link to participate in a follow-up information finding task if they wished (described earlier). In this setting, we gathered responses from a total of 240 distinct workers from the top-level on CrowdFlower. We gathered responses from these workers to additionally analyze and carry forward learnings to the other task setups about (i) the distribution of workers with high and low achievement motivation, (ii) the fraction of workers that would choose to participate in the follow-up task. Of these 240, 106 trustworthy workers[7] participated in the follow-up information finding task with embedded passive achievement primes.

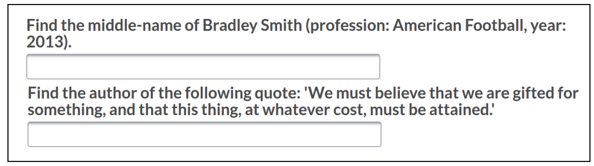


Figure 5: An active achievement prime embedded between two units of the information finding task.

(ii). *Active Achievement Priming* (**AP-Active**) – In this setting, crowd workers are presented with quotes and are asked to find the author of the quotes. By modeling the active achievement primes as information finding units, the workers treat these primes as a part of the actual task (the workflow in both cases is to search the web to find either a name or a middlename). These quotes act as active achievement primes since there is a direct interaction in the workflow with these primes. Thus, the active primes, masked as additional information finding units are randomly interspersed between the units of the actual task where workers are asked to find the middle-names of people (see Figure 5).

We followed the same setup as described in the previous case of **AP-Passive**. Here we collected responses from 100 distinct top-level workers in the intial task consisting of the *excellence motivation scale* deployed on CrowdFlower. These workers were then presented with an opportunity to proceed onto a follow-up information finding task (with embedded active achievement primes). 56 trustworthy workers went on to participate in the follow-up task.

(iii). *No Priming* (**NP-Baseline**) – To adequately gauge the impact of passive and active achievement priming using inspirational quotes in information finding tasks with varying levels of difficulty, we also consider the basic setting without any primes. All other task parameters remain the same as in case of passive and active achievement priming.

In this setting, since there were no primes embedded in the actual task, we directly deployed the information finding task on CrowdFlower and acquired responses from a total of 150 distinct top-level workers. Of these, 138 workers were trustworthy and we consider their responses alone in our analysis.

(iv) *Random Quotes for Achievement Priming* (**RQ-Passive** and **RQ-Active**) – To verify that any impact on worker retention is due to the inspiring nature of quotes which in turn triggers the chronic achievement motivation of workers, we run two additional experiments with exactly the same settings as in case of **AP-Passive** and

[6]https://sites.google.com/site/lak2017learning/

[7]Trustworthy workers are those who responded correctly to a very simple attention check question.

AP-Active primes but using randomly selected quotes unrelated to achievement or inspiration. To this end, we randomly select 100 quotes from http://www.brainyquote.com/. We deployed a similar task on CrowdFlower as shown in Figure 3 to collect 10 independent judgments for each quote on a 5-point Likert scale, based on how inspiring workers found these random quotes. Once again, workers were paid 4 USD cents for every 10 quotes that they rated. Here, we chose the top 25 quotes which corresponded to the least aggregated average rating as our primes. From earlier experiments in (i) and (ii), we note that most crowd workers have a high level of chronic achievement motivation. Thus, we do not differentiate between high and low level achievement motivation in this setup.

We replicate the task designs of **AP-Passive** and **AP-Active** conditions, with the only exception of introducing random quotes as primes. In the **RQ-Passive** setting, we gathered responses from 100 distinct trustworthy workers. In the **RQ-Active** setting we gathered judgments from 60 distinct workers, who were also allowed to complete as much work as they wished. Of these workers, 58 were found to be trustworthy and are considered for our further analysis.

4.5 Important Design Considerations

We completely randomized the order of different units as well as the embedded primes (in case of **AP-Passive**, **AP-Active**) within the information finding tasks. The different variations of tasks were deployed on different days and we ensured that there was no overlap in the set of workers that participated across the three variations. To minimize timezone driven worker particpation effects, we deployed the tasks at the same time on the different days. Since the primary goal of our work is to improve worker retention and learning rate in information finding tasks, we chose to restrict participation to only the top-level workers on CrowdFlower to ensure reliability of our findings. By considering three different levels of objective difficulty of the information finding tasks, we can additionally measure the impact of task difficulty on worker dropout rates. It is important to note that in all cases (**AP-Passive**, **AP-Active**, **NP-Baseline**) a worker who entered the information finding task was allowed to complete as much work as she wished to (up to a maximum of 120 units), without any experimentally induced constraints. Due to this reason, we can reliably measure worker retention (or dropout) rates. We compensated workers with 5 USD cents for each set of 10 units that they completed in all the variations. To satisfy minimum wage requisites we paid additional bonuses after workers either completed the work, or dropped out.

4.6 Measuring Learning Rate

As crowd workers proceed to complete the tasks in a given batch, and then move on to complete more batches of the same type, it has been shown that workers learn from their experience and begin to perform better [8]. We aim to measure the overall worker learning rate across the different batches of tasks they complete. Let us consider, $B = \{b_1, b_2, ..., b_i, b_j, ..., b_n\}$ to be the set of available batches of tasks of a given type. In the given task setup, we can assume that the difference in worker performance from a given batch and that in the preceeding batch of tasks can be attributed to the *learning* (or a lack of learning) that has occurred. However, this alone does not define the rate at which the workers learn. For example, a worker can perform with a 100% accuracy across a sequence of batches while still learning from one batch to the next. This implies that the overall performance of the worker should also feature in measuring a worker's learning rate.

Hence, we define the worker *learning rate* as a linear combination of the average difference in worker performance from one batch to the next (called the 'learning constant', l), and the overall performance of the worker across the different batches. Thus, we measure the learning rate of workers (denoted by L_r) by using the following formula:

$$L_r = \frac{1}{n}\left\{\sum_{\substack{1<i<(n-1)\\ i+1<j<n}} (acc_j - acc_i) + \sum_{k=1}^{n} (acc_k)\right\} \quad (1)$$

where, n is the number of batches in B that the given worker has completed, $i = 1...(n-1)$, $j = (i+1)...n$, $k = 1...n$, and acc_i represents the accuracy of a worker in the i^{th} batch of tasks.

Note that the worker learning rate, L_r, can be either positive or negative. A positive learning rate indicates that a worker learns through the course of batch completion and depicts an improvement in performance, while a negative learning rate indicates that a worker does not learn through the course of batch completion and thereby depicts no improvement in performance.

5 RESULTS AND ANALYSIS

5.1 Achievement Motivation in the Crowd

Of the 240 workers that responded to the motivation excellence measurement task before entering the follow-up information finding tasks in the **AP-Passive** condition, 211 workers corresponded to being achievement-oriented (*ACHIEVE* group) as opposed to fun-oriented (*FUN* group). Similarly, of the 100 workers that responded to the motivation excellence measurement task before entering the follow-up information finding tasks in the **AP-Active** condition 90 workers were found to belong to the *ACHIEVE* group. We believe that these distributions, indicating a high achievement motivation among crowd workers, can be explained by the primary motivation of most workers to participate in crowdsourcing microtasks; to earn and maximize monetary rewards [16], despite typically facing obstacles such as unfair pay or rejection of work, power asymmetries [23], and sub-optimal worker environments [17].

5.2 Results in Different Priming Conditions

(i). *Passive Achievement Priming*– We note that of the 106 trustworthy workers that participated in the **AP-Passive** information finding tasks, 99 were found to be achievement-oriented and formed the *ACHIEVE* group, while 7 workers were found to be fun-oriented and constitute the *FUN* group.

Workers constituting the *ACHIEVE* group performed with a greater individual accuracy, depicted higher retention rates and produced a higher overall quality of work in comparison to the workers constituting the *FUN* group (see Table 1). However, the two groups are unbalanced and the differences are not statistically significant. The overall task accuracy is computed by considering

Table 1: Worker performance, retention, and task completion time (TCT) in AP-Passive information finding tasks.

Achievement Motivation	Avg. Acc. Per Worker (%)	Worker Retention Rate (%)	Overall Task Acc. (%)	Avg. TCT (in mins)
ACHIEVE	**79.60**	**36.78**	**80.96**	**10.74**
FUN	75.85	26.19	74.25	9.60

the accuracy and the amount of work completed by each worker (i.e., the weighted average accuracy of all workers in the task).

(ii). *Active Achievement Priming–* Of the 56 workers that participated in the **AP-Active** information finding tasks, 53 workers were found to be achievement-oriented and formed the *ACHIEVE* group, while 3 workers were found to be fun-oriented and constitute the *FUN* group. Similar to our findings in **AP-Passive** information finding tasks, workers in the *ACHIEVE* group performed with a greater individual accuracy, depicted higher retention rates and produced a higher overall quality of work (see Table 2). These differences however, were not found to be statistically significant.

Table 2: Worker performance, retention, and task completion time (TCT) in AP-Active information finding tasks.

Achievement Motivation	Avg. Acc. Per Worker (%)	Worker Retention Rate (%)	Overall Task Acc. (%)	Avg. TCT (in mins)
ACHIEVE	**76.98**	**45.60**	**80.45**	**11.41**
FUN	65.17	27.78	67	12.02

(iii). *No Priming–* Table 3 presents our findings with respect to worker performance, retention rate, and task completion time in the case where no primes are embedded in the information finding tasks, in comparison to the **AP-Passive** and **AP-Active** variations.

Figure 6 draws a comparison between the worker retention curves in each of the different priming conditions, where workers can submit a minimum of 10 units and a maximum of 120 units, i.e, the entire batch. Through multiple t-tests and Bonferroni correction, we note that there is a significant improvement in the worker retention rate (over 8%) in case of the **AP-Active** (*M=44.64, SD=35.97*) setting when compared to the **NP-Baseline** (*M=36.23, SD=32.09*) setting; $t(192) = 1.735, p < .05$, and Hedges g=.3. In addition, **AP-Active** corresponds to a higher retention rate when compared to **AP-Passive** *(M=36.08, SD=31.53)*; $t(160) = 1.698, p < .05$. We did not observe a significant difference in the worker retention rate between **AP-Passive** and **NP-Baseline**. The average task completion time of workers is significantly lesser in case of the **NP-Baseline** *(M=9.11, SD=4.28)* setting in comparison to **AP-Passive** *(M=10.67, SD=5.19)*; ($t(242) = 2.551$, $p<.05$), and also in comparison to **AP-Active** *(M=11.44, SD=4.8)*; $t(192) = 3.309, p < .05$.

5.3 Do 'inspiring' quotes matter?

We analyzed the performance of workers in the presence of random quotes as primes, and our findings are presented in the Table 3. Through multiple t-tests and Bonferroni correction, we found that in the **RQ-Passive** setting, the worker retention rate *(M=28.58, SD=24.22)* is significantly lesser than in **AP-Active** *(M=44.64, SD=35.97)*; [$t(154)= 3.321$, $p < .01$], **AP-Passive** *(M=36.08, SD=31.53)*; [$t(204)= 1.907$, $p < .05$], and **NP-Baseline** *(M=36.23, SD=32.09)*; [$t(236)= 2.01$, $p < .05$]. In the **RQ-Active** setting,

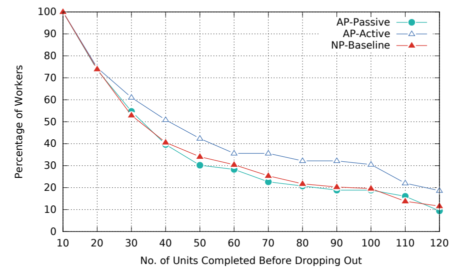


Figure 6: A comparison of the worker retention curves in each of the different priming conditions.

Table 3: Worker performance, retention, and average task completion time (TCT) in the information finding tasks with different priming conditions.

Priming Variation	Avg. Acc. Per Worker (%)	Worker Retention Rate (%)	Overall Task Acc. (%)	Avg. TCT (in mins)
NP-Baseline	76.71	36.23	78.88	**9.11***
AP-Passive	77.72	36.08	77.60	10.67
AP-Active	76.35	**44.64***	79.96	11.44
RQ-Passive	78.74	28.58	77.58	9.72
RQ-Active	74.81	32.61	75.41	9.87

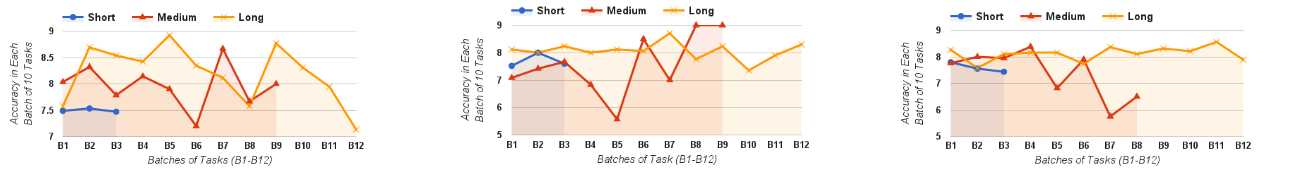
the worker retention rate is 32.61%. This was found to be significantly lesser than the retention rate in the **AP-Active** condition; *t(112)=1.892*, $p < .05$. We did not find significant differences between the worker retention rate in **RQ-Active** and either **AP-Passive** or **NP-Baseline**. We also did not find significant differences in the average worker accuracy and the overall task accuracy between each of the three priming conditions with respect to **RQ-Passive** and **RQ-Active**.

Based on our findings, we can confirm that the improvement in the worker retention rate observed in the **AP-Active** setting can be attributed to the inspiring nature of quotes which act as achievement primes and trigger the intrinsic achievement motivation of workers. Randomly selected quotes which are not inspiring in nature fail to have the same effect.

5.4 Effects of Task Difficulty

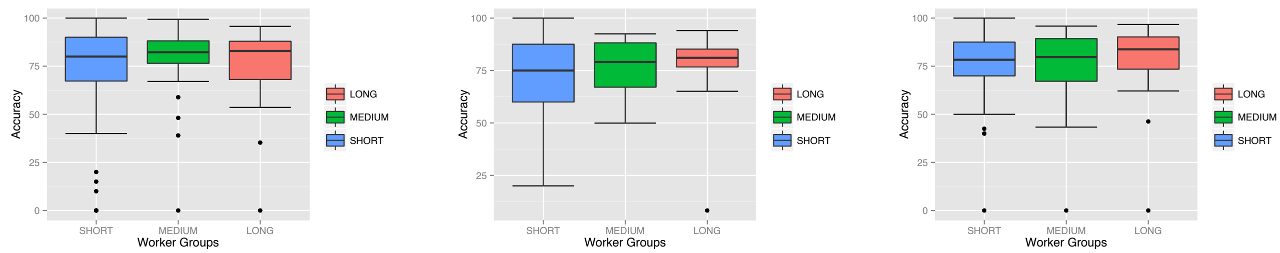
Recent work has shown the impact of intertask effects on the quality of responses from crowd workers [31]. The order of microtasks, especially with respect to their complexity, was shown to have a significant impact on the quality of work produced [4]. Thus, it was important to control for intertask effects by randomizing the order of units, and doing so further with respect to their level of difficulty. In this section, we describe our investigation of the impact of task difficulty on worker accuracy and on worker retention rate.

To understand the influence of task difficulty on worker dropout, we illustrate the average fraction of information finding tasks that workers completed with respect to each of the three difficulty levels (before dropping out) in Figure 7. Note that every worker that completed all the 120 units in each of the settings, completed an equal fraction of 40 units corresponding to each level of difficulty. We refer to such workers as '*finishers*', and the fraction of participating workers who complete all the 120 units as the '*finisher rate*'. We



(a) NP-Baseline **(b) AP-Active** **(c) AP-Passive**

Figure 8: Average accuracy of workers in each of the *short*, *medium*, and *long* groups through the course of batch completion in the different priming conditions. Note that each batch (from B1 through B12) consisted of 10 tasks.



(a) NP-Baseline **(b) AP-Active** **(c) AP-Passive**

Figure 9: Overall accuracy per worker and each of the *short*, *medium*, and *long* groups across the different priming conditions. The y-axis presents the accuracy per worker in percentage.

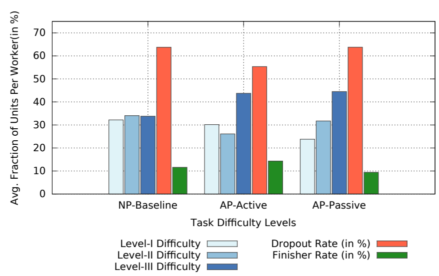


Figure 7: The average fraction of units with a given difficulty that workers encountered in different settings until they dropped out, corresponding dropout and finisher rates.

observe the highest finisher rate (14.29%) and the least dropout rate (55.36%) in the **AP-Active** setting.

Table 4: Average accuracy of workers on units with varying levels of difficulty in different priming conditions. '*' indicates statistical significance at $p < .05$, '' at $p < .01$.**

Priming Condition	Difficulty Level-I	Difficulty Level-II	Difficulty Level-III
NP-Baseline	**84.79***	79.28	66.46
AP-Passive	**88.86****	**88.77****	65.55
AP-Active	77.82	83.18	69.86
Overall	83.82	83.75	67.29

Table 4 presents the average accuracy of workers on units with a given difficulty level across the different priming conditions. To assess the relationship between the level of difficulty of the information finding tasks (modeled as an objective quantitative variable) and the average accuracy of workers across the different conditions, we computed Pearson' **r**. We found a moderate negative correlation between the difficulty level and the average accuracy of workers, $\mathbf{r}(894) = -.30$, $R^2 = .09$, $p < .001$.

To assess the interaction effects between the priming condition and the task difficulty on worker accuracy and retention rate, we computed a two-way MANOVA. This revealed a statistically significant interaction effect between the task difficulty and priming condition on the dependent variables (worker accuracy and retention rate); *F(4,887)=2.14, p<.05; Wilk's* $\lambda = 0.98$, $\eta_p^2 = .01$.

Given the significance of the overall test, the univariate main effects were examined for the worker accuracy and retention rate. We found a significant difference in the worker accuracies across the difficulty levels; *F(2,887)=58.51, p<.001*, but not across the different priming conditions; *F(2,887)=2.82, p=.06*. In contrast, we found a significant difference in worker retention rates across the different priming conditions; *F(2,887)=3.48, p<.05*, but not across the difficulty levels; *F(2,887)=1.85, p=.15*. This was confirmed by post-hoc comparisons using the Tukey HSD test. Our findings reveal the significant differences in impact of the different priming conditions on worker retention rates.

5.5 Learning Through the Batches

To analyze the learning that workers exhibit through the course of task completion, we investigated their performance. We divided workers into three groups based on the number of units completed (similar to [8]); the *short* group consists of workers who completed 25% or less units, the *medium* group consists of workers who completed more than 25% and less than or equal to 75% of the units, and the *long* group consists of workers who completed more than 75% of the units. Figure 10 presents the distribution of workers across the groups in each of the priming conditions. We note that the distributions are similar in case of **AP-Passive** and **NP-Baseline**. In case of **AP-Active**, we found that more workers belonged to the *long* group than *medium*, indicating the retention induced by the active achievement priming.

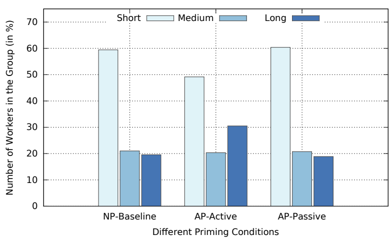


Figure 10: Distribution of workers into *short*, *medium*, and *long* groups as per the amount of work completed.

Since workers completed tasks in batches of 10 units, we analyzed their accuracy across the batches (from *B1* through a maximum of *B12*, as there were 120 available units in total). Figure 8 presents the average accuracy of workers belonging to the *short*, *medium* and *long* groups across the batches of tasks in each of the priming conditions. We observe that in case of the **NP-Baseline** condtion, the average accuracy of workers tends to drop through the course of the batches across all three groups. In contrast, we note that there is a steady increase in the average accuracy of workers in the *long* group in case of the **AP-Passive** condition. We analyzed the overall accuracy of each worker in the groups, and our findings are presented in Figure 9. We note that the average accuracy of workers does not vary across each of the three groups or the different priming conditions. However, in case of **AP-Active**, as workers complete more tasks the standard deviation in their accuracy becomes lower.

This is consistent with prior work, where authors found that workers who went on to complete more tasks in the best pricing scheme depicted an improvement in accuracy through the course of the tasks [8]. This suggests the impact of the active and passive priming conditions in retaining and engaging workers, enabling them to learn and apply learned concepts efficiently through the course of long batches.

5.6 Worker Learning Rate

We computed the learning rate of workers in the different priming conditions using the formula in Equation 1. Figure 11 presents our findings with respect to the learning rate of workers in each of the *short*, *medium* and *long* groups across the priming conditions.

We found that the learning rate of workers is maximum in case of workers constituting the *short* group, and this gradually decreases in case of the *medium* group, and further in case of the *long* group (see Table 5). This can be explained intuitively by the fact that when workers begin a first batch of new tasks of a given type, there is more to learn through the course of batch completion. Hence, those workers who complete only 25% of the tasks available, exhibit the most learning rate on average. As workers proceed through towards the completion of all batches available (i.e., workers who constitute the *medium* and *long* groups), due to their learning progress, the learning rate gradually decreases.

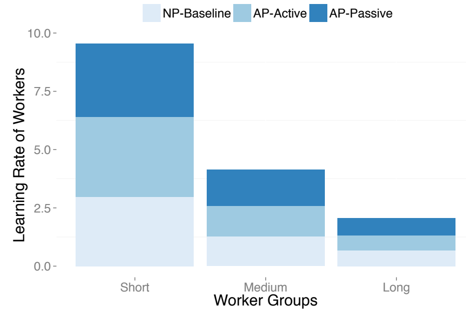


Figure 11: *Learning Rate* of workers in *short*, *medium* and *long* groups in the different priming conditions (stacked).

Table 5: *Learning Rate* of workers in the different groups across the priming conditions ('*' indicates significance).

	NP-Baseline	AP-Active	AP-Passive
Short	2.96	**3.43***	3.15
Medium	1.28	1.30	**1.55***
Long	0.69	0.64	0.73
Overall Avg.	1.64	1.79	1.81

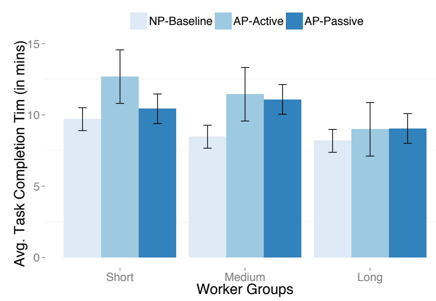


Figure 12: *Task Completion Time* (TCT) of workers in *short*, *medium* and *long* groups in different priming conditions.

The gradual decrease in learning rate through the course of batches coincides the effectiveness with which workers complete tasks within the batches. Figure 12 presents the average task completion time (TCT) of the three groups of workers in the different priming conditions. This supports our hypothesis (**Hyp-II**) that with an increase in worker retention, workers learn to perform more effectively in information finding microtasks, i.e., worker accuracy stabilizes as seen in Figure 9 and the TCT of workers decreases.

6 DISCUSSION

In our experiments, by using the excellence motivation scale we observed that a vast majority of crowd workers belong to the *ACHIEVE* group, i.e., they exhibit a prioritization of achievement over having fun, in contrast to the *FUN* group. This observation can be explained from two complementary standpoints. It is a well understood notion that crowd work can be tedious, monotonous and is often rewarded with meager pay. It is in these circumstances that crowd workers seek to earn monetary rewards, indicating an inherent motivation and a will to excel. We enforced the CrowdFlower filter of recruiting the highest quality crowd workers. For these workers to consistently complete several tasks while maintaining a high accuracy (leading to their highest quality level qualification on CrowdFlower[8]), we argue that it takes more than skill and competence, indicating further motivation. Although we found that workers belonging to the *ACHIEVE* group depicted better performance (accuracy and retention rate) in comparison to the *FUN* group, due to the highly skewed sample sizes the differences were not statistically significant. This however, means that a majority of crowd workers can be positively motivated and retained in long batches of information finding tasks, enabling them to learn and improve their performance.

Based on our experimental findings, we observe a significant impact of active achievement priming on worker retention in information finding tasks. The **AP-Active** priming condition leads to an improvement of over 8% in worker retention when compared to the **NP-Baseline** and **AP-Passive** conditions. This supports our hypothesis that achievement priming can improve worker retention in information finding tasks (**Hyp-I**). However, we note that the latter conditions correspond to a faster task completion time. This can be explained by the fact that (i) in the **AP-Passive** condition, workers don't necessarily have a direct interaction with the primes and can proceed in the tasks by ignoring them, and (ii) in the **NP-Baseline** condition workers do not encounter primes; they neither have to read quotes nor find author names that otherwise constitute additional units of work. We also found that the improvement in worker retention in the **AP-Active** priming condition was due to the inspiring quotes stimulating the inherent achievement motivation among workers. Worker retention rate deteriorates when using random primes that are not inspiring, thereby serving only as a distraction to the workers.

The **AP-Active** priming condition led to an improvement in the average learning rate of workers by nearly 8.5% compared to the **NP-Baseline**. Similarly, the **AP-Passive** priming condition improved the worker learning rate by nearly 10.5%.

By acknowledging the fact that information finding tasks that are typically crowdsourced have varying levels of difficulty, we investigated and found significant effects of task difficulty on worker accuracy. However, we found task difficulty did not effect the worker retention rate. Worker retention rate was effected by the priming conditions significantly, notably by the **AP-Active** condition. The improvement in worker retention rate due to active achievement priming using inspirational quotes and the impact on work quality (wherein workers who complete more work improve their accuracy) support **Hyp-II**.

[8]http://crowdflowercommunity.tumblr.com/post/80598014542/introducing-contributor-performance-levels

6.1 Caveats and Limitations

We investigated the average number of primes that workers encounter in each batch of 10 units. We found that in both conditions there were 2 primes on average across each batch of 10 units, with *SD=.55* in case of **AP-Passive** and *SD=.44* in case of **AP-Active**, indicating no bias due to the frequency of primes.

In our experiments, we considered a pool of 25 (active or passive) achievement primes that were randomly distributed over a batch of 120 units. We found that 2 primes in each set of 10 units on average triggered workers sufficiently (in **AP-Active** condition) to complete more work in the batch. However, further experiments are required to draw conclusions regarding the optimal frequency of achievement primes across the batch of units to maximize worker retention. We took care to ensure that workers were unaware of the achievement primes that are embedded into the actual task. We attempted to achieve this by integrating the primes into the design and not disrupting the workflow in case of **AP-Passive**, and by creating the same interaction with the active prime as with other units of work in the **AP-Active** condition.

Our measure for worker learning rate is simplistic, in that we do not consider external factors such as worker fatigue and boredom, previous experiences with similar tasks and so forth, that may effect the learning rate of workers.

6.2 Ethical Considerations

We must consider the ethical implications of using achievement priming in crowdsourcing microtasks. Achievement primes should not just be used to increase worker retention rates, but also with an aim to help improve the workers performance and widen the corresponding monetary opportunities and returns. If achievement primes are used to retain workers for relatively longer periods of time, workers should be adequately compensated. Paying heed to the ethical aspects of design and compensation with respect to achievement priming, we believe that in the digitally immersive current age, we are inadvertently but constantly primed by several aspects around us. It is the duty of task requesters to ensure that workers are not adversely affected due to irresponsible task design. However, due to the short-lived effects of priming we believe that if achievement priming is used responsibly it can be a useful means to improve crowd work and positively effect task consumption in crowdsourcing marketplaces.

7 CONCLUSIONS AND FUTURE WORK

In this paper, we investigated the use of achievement priming techniques to improve worker retention and learning rate in crowdsourced information finding tasks. We found that a vast majority of workers who participated in our tasks from the highest quality level on CrowdFlower are driven by achievement as opposed to fun. Thus, we proposed the use of inspirational quotes from famous people as achievement primes, and showed that active interaction with these primes in an inadvertent manner within information finding tasks led to a significant improvement in the worker retention rate (over 8% on comparison to the baseline method devoid of primes).

Further investigation of work quality between groups of workers who completed varying amounts of work in different priming conditions, revealed that workers who encountered active achievement primes show an improvement in their accuracies during the course of the tasks and exhibit relatively more stable performance (low standard deviation). We proposed worker *learning rate* as a metric to measure the learning that occurs when workers complete tasks or batches of tasks of a given type. We found that in addition to improving worker retention, achievement priming improves the learning rate of workers (thereby answering **RQ1, RQ2**).

Our findings and the novel method for improving learning in information finding tasks enrich the current understanding of crowd work and structuring workflow. In the imminent future, we will investigate the use of achievement priming in other domains of crowd work. We also aim to investigate how task framing can influence goal prioritization of crowd workers.

Acknowledgments

This research has been supported in part by the European Commission within the H2020-ICT-2015 Programme (AFEL project, Grant Agreement No. 687916), and the Erasmus+ Agreement 2015-1-LI01-KA203-000041 (*Learning Analytics and Learning Process Management for Small Size Organisations in Higher Education*). We also thank Peter Holtz for his valuable advise and feedback from the Psychology perspective.

REFERENCES

[1] John A Bargh, Peter M Gollwitzer, Annette Lee-Chai, Kimberly Barndollar, and Roman Trötschel. 2001. The automated will: nonconscious activation and pursuit of behavioral goals. *Journal of personality and social psychology* 81, 6 (2001), 1014.

[2] Daren C Brabham. 2008. Moving the crowd at iStockphoto: The composition of the crowd and motivations for participation in a crowdsourcing application. *First monday* 13, 6 (2008).

[3] Daren C Brabham. 2010. Moving the crowd at Threadless: Motivations for participation in a crowdsourcing application. *Information, Communication & Society* 13, 8 (2010), 1122–1145.

[4] Carrie J Cai, Shamsi T Iqbal, and Jaime Teevan. 2016. Chain reactions: The impact of order on microtask chains. In *Proceedings of the 34th Annual ACM Conference on Human Factors in Computing Systems (CHI'16). ACM*, Vol. 6.

[5] Tony Cassidy and Richard Lynn. 1989. A multifactorial approach to achievement motivation: The development of a comprehensive measure. *Journal of Occupational Psychology* (1989).

[6] Dana Chandler and Adam Kapelner. 2013. Breaking monotony with meaning: Motivation in crowdsourcing markets. *Journal of Economic Behavior & Organization* 90 (2013), 123–133.

[7] 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 *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 628–638.

[8] 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*.

[9] Djellel Eddine Difallah, Michele Catasta, Gianluca Demartini, Panagiotis G Ipeirotis, and Philippe Cudré-Mauroux. 2015. The dynamics of micro-task crowdsourcing: The case of amazon mturk. In *Proceedings of the 24th International Conference on World Wide Web*. ACM, 238–247.

[10] Son Do-Lenh, Patrick Jermann, Sébastien Cuendet, Guillaume Zufferey, and Pierre Dillenbourg. 2010. Task performance vs. learning outcomes: a study of a tangible user interface in the classroom. In *European Conference on Technology Enhanced Learning*. Springer, 78–92.

[11] Steven Dow, Elizabeth Gerber, and Audris Wong. 2013. A pilot study of using crowds in the classroom. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 227–236.

[12] Stefan Engeser. 2009. Nonconscious activation of achievement goals: moderated by word class and the explicit achievement motive? *Swiss Journal of Psychology* 68, 4 (2009), 193–200.

[13] Oluwaseyi Feyisetan, Elena Simperl, Max Van Kleek, and Nigel Shadbolt. 2015. Improving paid microtasks through gamification and adaptive furtherance incentives. In *Proceedings of the 24th International Conference on World Wide Web*. ACM, 333–343.

[14] 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.

[15] 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.

[16] 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, CHI 2015, Seoul, Republic of Korea, April 18-23, 2015*. 1631–1640.

[17] Neha Gupta, David Martin, Benjamin V Hanrahan, and Jacki O'Neill. 2014. Turk-life in India. In *Proceedings of the 18th International Conference on Supporting Group Work*. ACM.

[18] Lane Harrison, Ronald Chang, and Aidong Lu. 2012. Exploring the impact of emotion on visual judgement. In *Visual Analytics Science and Technology (VAST), 2012 IEEE Conference on*. IEEE, 227–228.

[19] Lane Harrison, Drew Skau, Steven Franconeri, Aidong Lu, and Remco Chang. 2013. Influencing visual judgment through affective priming. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2949–2958.

[20] William Hart and Dolores Albarracín. 2009. The effects of chronic achievement motivation and achievement primes on the activation of achievement and fun goals. *Journal of Personality and Social Psychology* 97, 6 (2009), 1129.

[21] E Tory Higgins and John A Bargh. 1987. Social cognition and social perception. *Annual review of psychology* 38, 1 (1987).

[22] 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.

[23] 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.

[24] 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.

[25] Walter S Lasecki, Samuel C White, Kyle I Murray, and Jeffrey P Bigham. 2012. Crowd memory: Learning in the collective. *arXiv preprint arXiv:1204.3678* (2012).

[26] 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.

[27] 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*.

[28] David C McClelland. 1965. Toward a theory of motive acquisition. *American psychologist* 20, 5 (1965), 321.

[29] Robert R Morris, Mira Dontcheva, Adam Finkelstein, and Elizabeth Gerber. 2013. Affect and creative performance on crowdsourcing platforms. In *Affective computing and intelligent interaction (ACII), 2013 humaine association conference on*. IEEE, 67–72.

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

[31] Edward Newell and Derek Ruths. 2016. How One Microtask Affects Another. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM.

[32] Scott T Rabideau. 2005. Effects of achievement motivation on behavior. *Retrieved from Personality Research: http://www.personalityresearch.org/papers/rabideau.htm* (2005).

[33] 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*.

[34] Harini Alagarai Sampath, Rajeev Rajeshuni, Bipin Indurkhya, Saraschandra Karanam, and Koustuv Dasgupta. 2013. Effect of Task Presentation on the Performance of Crowd Workers–A Cognitive Study. In *First AAAI Conference on Human Computation and Crowdsourcing*.

[35] James Y Shah and Arie W Kruglanski. 2008. The challenge of change in goal systems. *The Handbook of Motivation Science* (2008), 217–229.

[36] Amanda Shantz and Gary Latham. 2011. The effect of primed goals on employee performance: Implications for human resource management. *Human Resource Management* 50, 2 (2011), 289–299.

[37] Jie Yang, Judith Redi, Gianluca Demartini, and Alessandro Bozzon. 2016. Modeling Task Complexity in Crowdsourcing. In *Proceedings of The Fourth AAAI Conference on Human Computation and Crowdsourcing*. AAAI.