Monetary Interventions in Crowdsourcing Task Switching

Ming Yin
Harvard University
mingyin@fas.harvard.edu

Yiling Chen
Harvard University
yiling@seas.harvard.edu

Yu-An Sun
Xerox Innovation Group
YuAn.Sun@xerox.com

Abstract

With a large amount of tasks of various types, requesters in crowdsourcing platforms often bundle tasks of different types into a single working session. This creates a *task switching* setting, where workers need to shift between different cognitive tasks. We design and conduct an experiment on Amazon Mechanical Turk to study how occasionally presented performance-contingent monetary rewards, referred as *monetary interventions*, affect worker performance in the task switching setting. We use two competing metrics to evaluate worker performance. When monetary interventions are placed on some tasks in a working session, our results show that worker performance on these tasks can be improved in both metrics. Moreover, worker performance on other tasks where monetary interventions are not placed is also affected: workers perform better according to one metric, but worse according to the other metric. This suggests that in addition to providing extrinsic monetary incentives for some tasks, monetary interventions implicitly set performance goals for all tasks. Furthermore, monetary interventions are most effective in improving worker performance when used at switch tasks, tasks that follow a task of a different type, in working sessions with a low task switching frequency.

Introduction

While workers in crowdsourcing platforms often choose to switch between different types of tasks to diversify their work or avoid fatigue or boredom, many task switches are initiated by requesters as a result of the design of the working sessions. For example, a requester may ask a worker to identify whether a pre-specified object (e.g. automobile or person) exists in each of a set of pictures by grouping the tasks according to the objects of interests; this results in task switches when the "target" object changes. Moreover, in many citizen science projects, workers are asked to go through a few tasks to complete some requested work. For instance, in Cell Slider[1], a worker is shown an image of blood cells and needs to identify the types of cells, count

the number of irregular cells and then estimate the brightness of the stained irregular cell cores; in Citizen Sort[2], a worker classifies the same group of moth pictures according to shape, color and forewing pattern respectively.

One of the challenges of the requester-initiated task switching is to ensure work quality of *all* types of tasks in a working session. It is well known that workers perform worse on *switch tasks*, tasks that follow a task of a different type, than on *repetition tasks*, tasks that follow a task of the same type (Rogers and Monsell 1995; Monsell 2003). Hence, a working session requiring task switches may result in lower worker performance on at least some tasks.

In this paper, we study whether and how *performance-contingent monetary rewards* can be used to influence worker performance in requester-initiated task switching settings. By "performance-contingent", we mean that a worker can only get the reward if her performance meets some pre-specified criteria. Instead of providing a performance-contingent reward for every task in a working session, we consider the impact of occasionally provided performance-contingent rewards, i.e. they are only placed on a few selected tasks, on worker performance in working sessions with requester-initiated task switches. We refer to these occasionally provided performance-contingent rewards as *monetary interventions*.

We conduct a between-subjects experiment with 1268 unique workers from Amazon Mechanical Turk (MTurk). Each worker is randomly assigned to an experiment condition where she is asked to complete a sequence of 96 tasks, with two types of tasks interleaving with each other in the sequence. Experiment conditions vary in either the task switching frequency or whether and where monetary interventions are used in the sequence. We adopt two competing metrics to evaluate worker performance on a task: reaction time and accuracy. Intuitively, shorter reaction time is often associated with lower accuracy, ceteris paribus.

When monetary interventions are used in a working session, the term *intervened tasks* refers to those tasks where the interventions are placed and the term *non-intervened tasks* refers to other tasks in the session. Our experimental results show that workers react significantly faster and

[1]http://www.cellslider.net/

[2]http://www.citizensort.org/

either are more accurate or have a larger accuracy improvement (from previous tasks) at intervened tasks in experiment conditions with monetary incentives than at corresponding tasks in experiment conditions without monetary incentives. This indicates that extrinsic financial incentives are effective at improving work quality, even to the extent to somewhat overcome the tradeoff between two competing performance metrics. In addition, when monetary interventions are used in a working session, they have a spillover effect on the non-intervened tasks in the session — for tasks where monetary interventions are not placed, workers still shorten their reaction time to a large degree with a small decrease in accuracy. Such externality of monetary interventions can be explained as workers interpreting the performance-contingency of extra rewards on intervened tasks as an implicit performance goal, and thus attempting to improve their performance on *all* tasks.

Comparing worker performance across experiment conditions, we find that monetary interventions incentivize overall performance improvement in a working session more effectively when tasks switch less frequently in the session. Furthermore, placing monetary interventions at switch tasks boosts the overall performance more significantly than introducing them at repetition tasks. These findings suggest that requesters who wish to elicit high-quality work in a task switching setting by using monetary rewards should pay close attention to both how tasks are interleaved and where to add extra bonuses.

Related Work

Multitasking, task interruption and resumption. Task-switching is closely related to a few other concepts, including multitasking, task interruption and resumption. Multitasking refers to either performing two or more types of tasks simultaneously or switching back and forth from one type to another (Salvucci and Taatgen 2010; Salvucci, Taatgen, and Borst 2009). Our setting in this paper is thus similar to the latter form of multitasking.

Many studies in the human-computer interaction community explored task switching from the perspective of task interruption and resumption. A subject was typically performing a primary task before being interrupted by a secondary task, and the effects of the interruption on the *primary* task were analyzed (Iqbal and Horvitz 2007; Mark, Gudith, and Klocke 2008; Bailey and Konstan 2006). Unlike such work, we care about work quality across *all* types of tasks rather than focusing on a single (primary) type of tasks, and our focus is on the effects of monetary interventions on worker performance in task switching settings.

Financial incentives in crowdsourcing. Prior work on the relationship between monetary rewards and work quality in the context of crowdsourcing was mostly conducted through experiments in which workers sequentially completed tasks of the *same* type. Harris (2011) showed that workers performed better when offered performance-contingent financial incentives. In addition, while the magnitude of financial incentives alone has little effect on worker performance (Mason and Watts 2010; Rogstadius et al.

2011; Yin, Chen, and Sun 2013), the changes in the magnitude of financial incentives over a sequence of tasks do (Yin, Chen, and Sun 2013).

The effects of monetary rewards on worker performance in a task switching setting was only studied in the labs. It was observed that if workers could earn additional rewards based on their *overall* performance in a working session, their performance on switch tasks was improved marginally (Nieuwenhuis and Monsell 2002). Our work focuses on interventions that only provide monetary bonuses on *selected* tasks.

Switch cost, learning and task specialization. A prominent psychological effect of task switching is observed in previous studies — workers usually have worse performance on switch tasks than on repetition tasks (Rogers and Monsell 1995; Monsell 2003). The performance difference between the switch and repetition tasks is called the *switch cost*, which is likely a result of the costly cognitive control processes triggered by the task switching (e.g. shift of attention and retrieval of task goals and rules into working memory) or task-set inertia, that is, the proactive interference between the competing old and new tasks (e.g. persistent activation of the old task and the involuntary inhabitation of the current task) (Mayr and Kliegl 2000; Allport, Styles, and Hsieh 1994; Kiesel et al. 2010). It is also known that more frequent task switching demands more cognitive resources, which may be mentally taxing or cause information overload for workers (Speier, Valacich, and Vessey 1999). In contrast, repetition tasks offer opportunities for workers to develop task-specific skills and strategies over time as a result of *learning* and *task specialization* and thus may lead to increased work quality. In this work, by placing monetary interventions at different positions in a task sequence (e.g. on switch tasks or on repetition tasks), we intend to understand whether performance-contingent financial rewards can incentivize performance improvement through mitigating switch cost or promoting faster learning and task specialization.

Goal setting. Informing a worker that performance-contingent bonuses will be offered for selected tasks could implicitly set a performance goal for the worker on all tasks. There is a large literature on *explicit goal setting* which demonstrates that setting specific and challenging goals often leads to better performance (Locke et al. 1981; Mento, Steel, and Karren 1987; Locke and Latham 2002) and when the explicit goals are combined with monetary incentives the performance may be further improved (Locke et al. 1981; Pritchard and Curts 1973). It is thus interesting to examine whether the implicit goals conveyed by monetary interventions have a similar effect as the explicitly stated goals. If they do, we expect that monetary interventions affect worker performance on not only intervened but also non-intervened tasks.

Experimental Design

Our experimental design is inspired by two classical task switching experimental paradigms: predictable task switching, where switches happen in a predictable way after a con-

Figure 1 (illustration of three treatments for the 24×4 sequence):

No Bonus:

Red	Green	...	Blue	Yellow	Red	...	Blue	Red	Magenta	...	Magenta	Green	Blue	...	Red
Task # 1	2	...	24	25	26	...	48	49	50	...	72	73	74	...	96
-	R-NI	...	R-NI	S-NI	R-NI	...	R-NI	S-NI	R-NI	...	R-NI	S-NI	R-NI	...	R-NI

Switch Bonus: (BONUS markers at tasks 25, 49, 73)

Red	Green	...	Blue	Yellow	Red	...	Blue	Red	Magenta	...	Magenta	Green	Blue	...	Red
Task # 1	2	...	24	25	26	...	48	49	50	...	72	73	74	...	96
-	R-NI	...	R-NI	S-I	R-NI	...	R-NI	S-I	R-NI	...	R-NI	S-I	R-NI	...	R-NI

Repetition Bonus: (BONUS markers at random repetition tasks)

Red	Green	...	Blue	Yellow	Red	...	Blue	Red	Magenta	...	Magenta	Green	Blue	...	Red
Task # 1	2	...	24	25	26	...	48	49	50	...	72	73	74	...	96
-	R-NI	...	R-NI	S-NI	R-I	...	R-NI	S-NI	R-NI	...	R-I	S-NI	R-NI	...	R-NI

← - - - Color - - - → ← - - - Word - - - → ← - - - - - Color - - - - - → ← - - Word - - →

Figure 1: An illustration of three treatments for the 24×4 sequence. S-NI denotes a switch task without monetary intervention, R-NI represents a repetition task without monetary intervention, S-I refers to a switch task with monetary intervention, and R-I is a repetition task with monetary intervention. The first task of a sequence is neither a switch nor a repetition task.

stant number of tasks in a sequence, and task cuing, where an explicit cue is presented before each task to specify the type of the current task (Kiesel et al. 2010).

Tasks. Two types of tasks are used in our experiments: the *color naming* task and the *word reading* task. In a task of either type, a worker will see a stimuli word on the screen, which is the name of one of the five colors, blue, green, magenta, red and yellow. The word is displayed in a color that may or may not match the word, but the color is also limited to the five alternatives. For example, a stimuli word "red" can be written in blue. The two types of tasks are:

- *The color naming task* (Color): A worker is asked to indicate the color in which the word is written, regardless of whether or not that matches the word itself. In the above example, the answer is "blue".
- *The word reading task* (Word): A worker is asked to indicate what the word denotes, regardless of the color it is written. In the above example, the answer is "red".

In each task, the worker is instructed to report the answer by typing the initial of it in lower case. For example, the worker can report the answer "red" by typing 'r' on the keyboard.

Worker performance on each task is measured in two dimensions:

- *Reaction time* (RT): The elapsed time between the onset of the stimuli and the worker's response.
- *Accuracy* (or correctness): A binary value indicating whether the reported answer is correct or not.

These two metrics innately compete with each other as when workers shorten their reaction time, they are likely to be less accurate, ceteris paribus.

The two types of tasks were initially used in the Stroop test, which revealed the Stroop effect, i.e. subjects generally spend more time on naming the colors than reading the words (Stroop 1935). They are now widely used by psychologists in studying task switching (Wylie and Allport 2000; Gilbert and Shallice 2002; Allport and Wylie 1999).

Task Sequences. In our experiment, we put 96 tasks, which include 48 tasks of each type, in a human intelligence task (HIT). For different task sequences, the two types of tasks switch at different frequencies.

Specifically, we define a "segment" in a sequence as a consecutive chunk of tasks of the same type and the length of a segment is the number of tasks in it. Thus, for our experiment, if the length of each segment in a task sequence is N, there are $M = 96/N$ segments in that sequence, and the sequence is then referred to as an $N \times M$ sequence. Different types of tasks are assigned to neighboring segments in a sequence. By varying segment lengths, we can control the task switching frequency. We consider five task sequences in our study: $4 \times 24, 8 \times 12, 16 \times 6, 24 \times 4$ and 48×2.

Intervention Treatments. Each worker is asked to complete one of the five task sequences and receives a performance-independent payment of 3 cents for each task completed. Monetary interventions are performance-contingent monetary rewards: a worker can earn an extra bonus of 2 cents on a task with monetary intervention if her reported answer for that task is correct *and* her reaction time is less than 1 second. By varying whether and where the additional bonuses are placed in a sequence, we create three treatments for each of the five task sequences:

- *No Bonus (baseline)*: No bonus is placed on any task in a task sequence.
- *Switch Bonus*: Starting from the second segment in a task sequence, a performance-contingent bonus is offered at the first task in every segment, i.e. bonuses are placed at all switch tasks.
- *Repetition Bonus*: Starting from the second segment in a task sequence, a performance-contingent bonus is offered at a randomly selected non-switch task in every segment, i.e. a bonus is placed at one random repetition task in each segment (except the first segment).

Figure 1 gives a graphical example of the three treatments.

We call a combination of a task sequence and an intervention treatment an *experiment condition*. Thus, there are 15 experiment conditions in our experiment.

Procedure. We post our HITs on MTurk on weekdays around 12:00-14:00 and 16:00-18:00 (Eastern Standard Time) in a week. To avoid network latency, we restrict our HITs to U.S. workers. We suggest workers who have difficulties in seeing colors or perceiving color differences not take the HIT. Using a desktop or laptop computer with a keyboard to complete the HIT is recommended. Each worker is limited to take the HIT once (i.e. only work on one sequence of 96 tasks).

Upon arrival, a worker is randomly assigned to an experiment condition. The worker then goes through an instruction page, a task and interface tutorial and a qualification test. In the tutorial, the worker is instructed to report the answer to each task as quickly and accurately as possible. If she is assigned to a Switch Bonus or Repetition Bonus treatment, she is also informed of the opportunities to earn extra bonuses at some tasks in the sequence, contingent on her answer in those tasks being correct and given within 1 second. The worker can only proceed to the actual task sequence after passing the qualification test.

The actual task sequence starts with a task of a random type. For each task in the sequence, the worker will first see a cue word, either "Color" or "Word", shown in white on gray background and indicating whether the current task is the color naming or the word reading task. For the Switch Bonus and Repetition Bonus treatments, a bonus icon is displayed together with the cue word if monetary intervention is placed at the current task. Each cue is displayed for two seconds and then the worker is automatically redirected to the task page, where a stimuli word is displayed. Both the word and the printing color of the stimuli are randomly chosen from the five alternatives. The type of the current task is also shown on the top of the task page in case of unawareness. Once the worker reports her answer to the current task, she will be automatically redirected to the cue page for the next task. Finally, after completing all 96 tasks, the worker is asked to complete a post-task survey of demographic information.

Each worker in our experiment gets a show-up fee of $0.20 and a performance-independent payment of $2.88 ($0.03×96) after submitting the HIT. Workers in Switch Bonus and Repetition Bonus treatments may earn extra bonuses depending on their performance in those tasks where monetary interventions are placed.

Data

We recruited 1305 workers in total from MTurk for our experiment. For each worker, we recorded: (1) the exact experiment condition the worker worked on, including both the task type and whether there was a monetary intervention for each task in the sequence; (2) the worker's reaction time for each task; and (3) the worker's accuracy for each task.

We noticed that some workers took an excessively long time to report their answers to some tasks, which might due to interruptions in their working environment. To eliminate the influences of these "outliers", we excluded the data from

a worker if her reaction time for any of the tasks in her sequence was longer than 20 seconds. Such elimination leaves us with 1268 valid workers. The data for these workers are then used in the subsequent analysis.

The average age of the valid workers is 30.8, 59.1% of them are male, and all of them use either a desktop or a laptop computer to complete the HITs. No significant demographic or equipment difference is observed for workers in different experiment conditions.

Results

To analyze the effects of monetary interventions on worker performance in task switching settings, we first examine their influences on *intervened tasks* and *non-intervened tasks* respectively. Then, we compare different experiment conditions to gain insights into how monetary interventions can be used most effectively in task switching settings.

In this section, we will use the same abbreviations (i.e. S-NI, S-I, R-NI, R-I) as in Figure 1 to describe properties of tasks. The Wilcoxon rank sum test is used to evaluate statistical significance unless otherwise stated.

Effects on Intervened Tasks

Our first effort is to understand whether introducing monetary interventions in a task switching setting can incentivize workers to improve their performance on tasks where the interventions are placed. We thus focus on comparing worker performance on intervened tasks in treatments with bonuses (i.e. Switch Bonus and Repetition Bonus treatments) with worker performance on corresponding tasks in the baseline treatment (i.e. No Bonus treatment).

We first analyze worker performance in terms of reaction time. To represent the expected reaction time of workers when there are no monetary interventions, five *baseline average reaction time sequences* are created from our data for the No Bonus treatment, one for each of the five task sequences. That is, for each of the five task sequences of the No Bonus treatment, we take all workers who worked on this task sequence and average their reaction time position-wise. For example, the value at position i in the baseline average reaction time sequence for the 4×24 sequence is obtained by averaging the reaction time for the i-th task across all workers of the 4×24 sequence in the No Bonus treatment.

Worker reaction time for intervened tasks in an experiment condition with monetary interventions is then compared with the values in the corresponding baseline average reaction time sequence. For example, consider the comparison between the No Bonus treatment and the Switch Bonus treatment. For each of the five task sequences, we use two buckets: the first bucket collects all intervened task reaction time (i.e. S-I task reaction time) for all workers in the Switch Bonus treatment for the task sequence, and for each reaction time of a S-I task at position x that we add to the first bucket, we put the reaction time value at position x in the corresponding baseline average reaction time sequence to the second bucket. We then calculate the average of each bucket. Figure 2(a) plots the differences of the average reaction time for the intervened tasks between the No Bonus treatment and
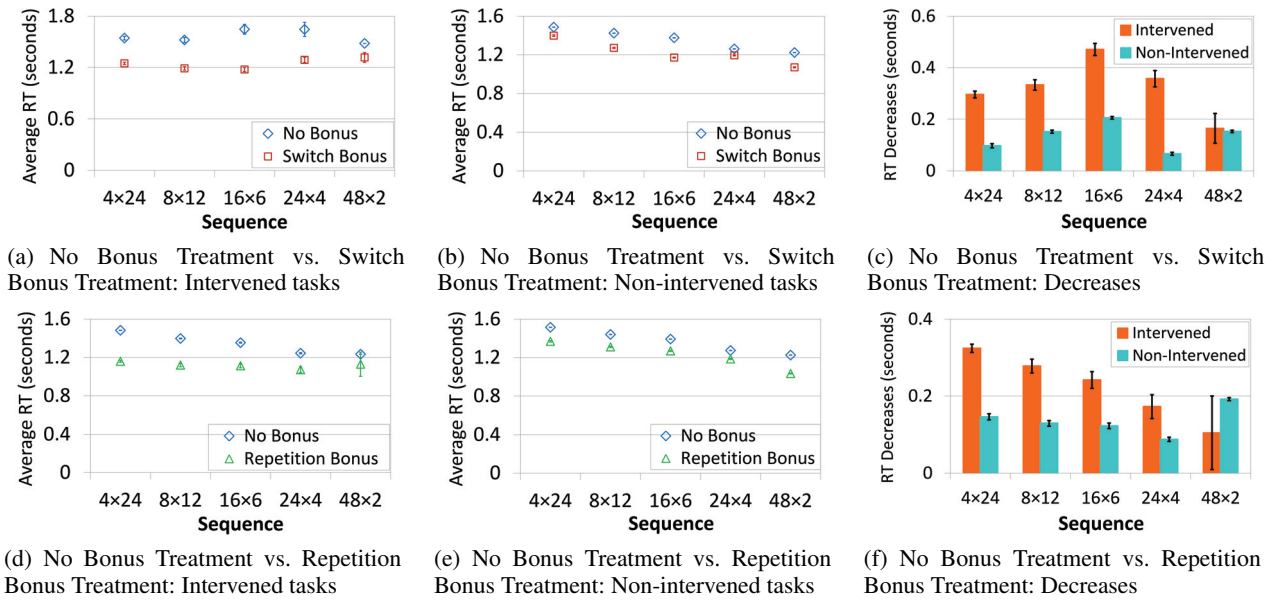
(a) No Bonus Treatment vs. Switch Bonus Treatment: Intervened tasks

(b) No Bonus Treatment vs. Switch Bonus Treatment: Non-intervened tasks

(c) No Bonus Treatment vs. Switch Bonus Treatment: Decreases

(d) No Bonus Treatment vs. Repetition Bonus Treatment: Intervened tasks

(e) No Bonus Treatment vs. Repetition Bonus Treatment: Non-intervened tasks

(f) No Bonus Treatment vs. Repetition Bonus Treatment: Decreases

Figure 2: Effects of monetary interventions on reaction time for intervened tasks and non-intervened tasks. Error bars represent standard errors of the mean.

the Switch Bonus treatment for all five task sequences. The differences in reaction time between the No Bonus treatment and the Repetition Bonus treatment are calculated similarly and plotted in Figure 2(d). As the figures suggest, the presence of monetary interventions leads to shorter reaction time for the intervened tasks, no matter where the interventions are placed. Further statistical tests report $p<0.001$ for pairwise comparisons of all task sequences, indicating that the decreases are significant.

We then examine worker performance in terms of accuracy. Similar to the analysis on reaction time, for each of the five task sequences, we first create the *baseline average accuracy sequences* by taking all workers who worked on the sequence in the No Bonus treatment and averaging their accuracy position-wise. Then, for each worker who worked on this sequence in the Switch Bonus (or Repetition Bonus) treatment, we put her accuracy for each task into one of the three categories depending on whether that task appears before, at or after the placement of the monetary intervention in its task segment. Furthermore, for each accuracy value for a task at position x that we put into one of the three categories for the Switch Bonus (or Repetition Bonus) treatment, we also add the accuracy value at position x in the corresponding baseline average accuracy sequence to the same category for the No Bonus treatment. Finally, by taking the average of all data in each category, we can see in each treatment how accurate workers are before, at or after the monetary interventions within a segment and thus investigate whether worker's accuracy improves in the intervened tasks with the extra bonuses.

Figures 3(a) and 3(b) report how worker's accuracy changes within a segment for different treatments, with Figure 3(a) showing the comparison between the No Bonus treatment and the Switch Bonus treatment, and Figure 3(b) showing the comparison between the No Bonus treatment

and the Repetition Bonus treatment. Accuracy is plotted cumulatively: For example, in Figure 3(b), the average worker accuracy after monetary interventions is the sum of the average accuracy before interventions (green bar), the accuracy increment at intervened tasks (orange bar) and the accuracy increment after interventions (purple bar).

Figure 3(a) shows that the orange bar for the Switch Bonus treatment is longer than that for the corresponding No Bonus treatment, with the 4×24 sequence being the only exception. This indicates that the average accuracy at the switch tasks improves significantly ($p<0.001$) when monetary interventions are placed on these tasks in all but the 4×24 sequence. The exception of the 4×24 sequence may because workers are overwhelmed by the mentally-taxing frequent switches and thus find the additional bonuses disturbing rather than motivating.

When monetary interventions are placed on repetition tasks, the average accuracy at these tasks is often not higher than that in the corresponding No Bonus treatment. To see this, we compare the combined length of the green and orange bars in Figure 3(b) for the Repetition Bonus and No Bonus treatments. The combined length for the Repetition Bonus treatment is shorter than that for the corresponding No Bonus treatment ($p<0.001$), except for the 48×2 sequence. However, the lower average accuracy at intervened tasks for the Repetition Bonus treatment can be largely attributed to the low average accuracy at the non-intervened tasks before the intervention, i.e. the green bar is shorter in the Repetition Bonus treatment than in the corresponding No Bonus treatment for most sequences. This is due to faster reaction on non-intervened tasks when extra bonuses are used and the competition between reaction time and accuracy, which we will detail in the next section. When focusing on the accuracy improvement at intervened tasks and thus comparing the length of orange bars between the two
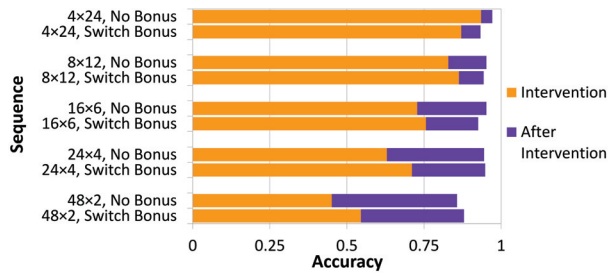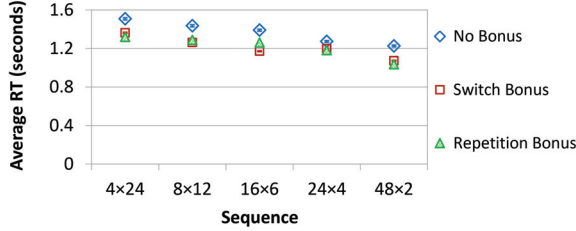
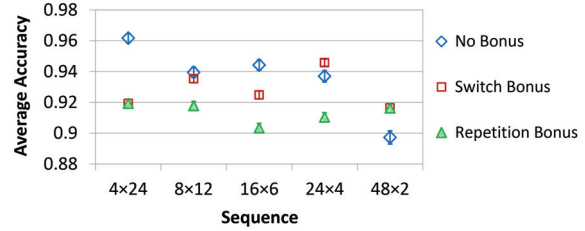(a) No Bonus Treatment vs. Switch Bonus Treatment
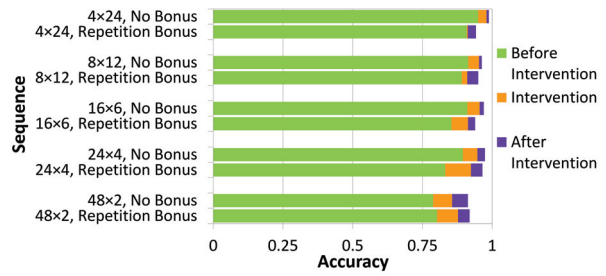
(b) No Bonus Treatment vs. Repetition Bonus Treatment

Figure 3: Comparison of the average worker accuracy within a segment

(a) Reaction time

(b) Accuracy

Figure 4: Comparison of average worker performance for all 96 tasks in a sequence across different treatments. Error bars represent standard errors of the mean.

treatments in Figure 3(b), we find that with monetary interventions, the accuracy improvement at the intervened repetition tasks is significantly larger for sequences with moderate to low task switching frequencies ($p < 0.05$).

To summarize, introducing monetary interventions incentivizes better performance on intervened tasks — workers complete the intervened tasks not only faster but also with either higher accuracy or a larger accuracy improvement. Recall that to earn the bonuses workers need to both react quickly and be accurate. While it may be easy for a worker to submit a response faster, the improved performance in accuracy suggests that workers are indeed motivated by the extrinsic financial incentives to improve her performance along *both* dimensions. The incentives to some degree help to overcome the innate tradeoff between the two performance metrics. Since the extra bonuses are placed on either switch tasks or repetition tasks, our observation provides supporting evidence for the effectiveness of performance-contingent financial rewards in mitigating switch cost or promoting faster learning and task specialization.

Effects on Non-intervened Tasks

Our next effort is to understand the effects of monetary interventions on non-intervened tasks.

The comparisons of the average reaction time for non-intervened tasks in the Switch Bonus treatment and the Repetition Bonus treatment against that in the baseline No Bonus treatment are displayed in Figure 2(b) and Figure 2(e) respectively. Interestingly, we find that although workers cannot earn extra rewards by completing the non-intervened tasks quickly, they still show a clear tendency in shortening their reaction time significantly ($p < 0.001$) for these tasks. On the other hand, while workers are still very accurate, their accuracy decreases at the non-intervened tasks:

for the Switch Bonus treatment, the average worker accuracy for non-intervened tasks is 93.75% across all task sequences, which is slightly lower (by 0.74%) than that for the No Bonus treatment; and for the Repetition Bonus treatment, the average worker accuracy for non-intervened tasks across all task sequences is 91.34%, which is 2.03% lower than that for the No Bonus treatment. The accuracy decreases for non-intervened tasks are statistically significant ($p < 0.001$). These results indicate that with the additional bonuses, workers try to improve their performance in reaction time while maintaining their performance in accuracy even when monetary rewards are not directly applied to the tasks. Yet, the competitive nature of the two performance metrics seems to still dominate at non-intervened tasks, which means that faster reaction comes with a cost in accuracy for these tasks.

As monetary interventions lead to decreases in reaction time for both intervened and non-intervened tasks, we further compare the magnitude of the decrease between these two categories of tasks. Results are reported in Figures 2(c) and 2(f). It is clear that no matter where the monetary rewards are placed, the decrease in reaction time for intervened tasks is significantly larger ($p < 0.05$) than that for non-intervened tasks, with the 48×2 sequence in the Repetition Bonus treatment being the only exception (the decrease in reaction time for non-intervened tasks there is marginally larger, with $p = 0.077$).

A unified explanation for our observations on both intervened and non-intervened tasks is that workers first interpret the performance-contingency of extra rewards on some selected tasks as setting an implicit performance goal, which has a similar effect as an explicit goal. Thus, workers attempt to improve their performance for *all* tasks in the sequence (subject to the innate tradeoff between the two performance

Table 1: Average worker reaction time in different experiment conditions and differences of reaction time between conditions. The statistical significance of the Wilcoxon rank sum test is marked as a superscript, with *, **, and *** representing significance levels of 0.05, 0.01, and 0.001 respectively. (Unit: seconds)

Task Sequences	Reaction Time Mean Values			Reaction Time Differences		
	No Bonus (NB)	Switch Bonus (SB)	Repetition Bonus (RB)	SB − NB	RB − NB	RB − SB
4×24	1.5078	1.3628	1.3193	-0.145***	-0.188***	-0.044***
8×12	1.4355	1.2625	1.2893	-0.173***	-0.146***	0.027
16×6	1.3904	1.1717	1.2615	-0.218***	-0.129***	0.090***
24×4	1.2731	1.1976	1.1824	-0.075***	-0.091***	-0.015
48×2	1.2261	1.0725	1.0352	-0.154	-0.191***	-0.037***

Table 2: Average worker accuracy in different experiment conditions and differences of accuracy between conditions. The statistical significance of the Wilcoxon rank sum test is marked as a superscript, with *, **, and *** representing significance levels of 0.05, 0.01, and 0.001 respectively.

Task Sequences	Accuracy Mean Values			Accuracy Differences		
	No Bonus (NB)	Switch Bonus (SB)	Repetition Bonus (RB)	SB − NB	RB − NB	RB − SB
4×24	0.9617	0.9193	0.9193	-0.042***	-0.042***	0.000
8×12	0.9395	0.9353	0.9178	-0.004***	-0.022***	-0.018***
16×6	0.9442	0.9248	0.9034	-0.019***	-0.041***	-0.021***
24×4	0.9370	0.9458	0.9104	0.009*	-0.027***	-0.036***
48×2	0.8973	0.9164	0.9163	0.019***	0.019***	-0.000

metrics), regardless of whether monetary interventions are placed at the tasks. For the intervened tasks in the sequence, workers are further incentivized by the extrinsic financial incentives and therefore improve their performance at these tasks to a larger degree by responding faster and more accurately.

More Effective Interventions

Finally, we seek to gain some insights into how to more effectively use monetary interventions in a task switching setting when we care worker performance on all tasks. Figure 4(a) and Table 1 report the comparisons of the three treatments in terms of the average worker reaction time over all 96 tasks for each of the five task sequences. Figure 4(b) and Table 2 present similar comparisons for worker accuracy.

First, we look into the baseline No Bonus treatment. As shown in Figures 4(a) and 4(b), when monetary interventions are not available, as task switching becomes less frequent, worker reaction time gets shorter and worker accuracy also exhibits a downward trend. One-way analysis of variance (ANOVA) further confirms that the differences in reaction time and accuracy across task sequences are statistically significant (p<0.001). In other words, without monetary interventions, by controlling how frequently tasks switch in a sequence, a requester may trade off better performance in average reaction time for better performance in overall accuracy.

When comparing worker performance in treatments with additional bonuses (i.e. the Switch Bonus or Repetition Bonus treatment) and that in the baseline treatment, we have an interesting observation: while workers can be incentivized to improve their performance in reaction time significantly regardless of the task switching frequency in the

sequences (i.e. negative differences in columns "SB − NB" and "RB − NB" of Table 1), similar improvement in accuracy can only be achieved when the task switching frequency is low (i.e. positive differences in columns "SB − NB" and "RB − NB" of Table 2 only for sequences with a low task switching frequency). This observation implies that adding monetary interventions to sequences with a low task switching frequency could be more effective: instead of trading off speed for accuracy or vice versa, workers perform better according to both metrics; in particular the incentives boost worker's overall accuracy in the sequence significantly.

Next, we consider where to place monetary interventions in a task sequence and examine worker performance in Switch Bonus treatment and Repetition Bonus treatment. We find that while both treatments have similar efficiency in improving performance in reaction time, placing the performance-contingent rewards on switch tasks generally leads to better performance in accuracy compared to providing extra bonuses at repetition tasks (i.e. negative differences in the "RB − SB" column of Table 2). This indicates that it is more efficienct to use monetary interventions right at the the switching points. With a closer look, this phenomenon can be attributed to two reasons: (1) accuracy improvement at the intervened tasks is significantly larger when bonuses are placed at switch tasks than when they are placed at repetition tasks (+3.49% vs. +0.34%, p<0.001); (2) combining extra bonuses with task switches makes workers focus more on the new type of tasks earlier — compared to the baseline treatment, the average number of tasks it takes for a worker to first submit a correct answer in a segment is decreased by 0.18 (not significant) for the Switch Bonus treatment while increased by 0.15 (p<0.05) for the Repetition Bonus treatment, leading workers in the Switch Bonus treatment to outperform workers in the Repetition Bonus treatment in the

early stage of each task segment (e.g. 88.52% vs. 85.30% for the average accuracy comparison of the first half of tasks in each segment, p<0.001).

In sum, monetary interventions can be most effective in motivating better worker performance when they are placed at switch tasks in a sequence with a low task switching frequency. We conjecture that this is due to that monetary interventions are less interruptive in sequences with a low task switching frequency and the demand for extra attention is highest at switch tasks.

Conclusions

We experimentally study the effects of monetary interventions on worker performance in crowdsourcing task switching settings. We show that the occasional provided performance-contingent monetary rewards in a task sequence not only lead to an improved performance in the intervened tasks, but also cast a spillover effect on the non-intervened tasks. Such observations can be explained as that workers are motivated by both the implicit performance goal conveyed by the performance contingency of the bonuses and the extrinsic financial incentives. Finally, we find that monetary interventions are more effective in eliciting better worker performance when used at switch tasks in a sequence with a low task switching frequency.

The practical implication of our results is that in crowdsourcing task switching settings, monetary rewards can be an effective method for motivating high-quality work. Yet, the effectiveness of monetary interventions may depend on both the exact sequences that the interventions are applied to and the places that the interventions are provided at. Requesters need to consider these subtleties when trying to improve the quality of crowd work.

Acknowledgments

References

Allport, A., and Wylie, G. 1999. Task-switching: Positive and negative priming of task-set. In Humphreys, G.; Duncan, J.; and Treisman, A., eds., *Attention, space, and action: Studies in cognitive neuroscience*. Oxford: Oxford University Press. 273–296.

Allport, D. A.; Styles, E. A.; and Hsieh, S. 1994. Shifting intentional set: Exploring the dynamic control of tasks. In Umilta, C., and Moscovitch, M., eds., *Attention and Performance XV*. Cambridge, MA: MIT Press. 421–452.

Bailey, B. P., and Konstan, J. A. 2006. On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate, and affective state. *Computers in Human Behavior* 22(4):685–708.

Gilbert, S. J., and Shallice, T. 2002. Task switching: A pdp model. *Cognitive psychology* 44(3):297–337.

Harris, C. 2011. Youre hired! an examination of crowdsourcing incentive models in human resource tasks. In *Proceedings of the Workshop on Crowdsourcing for Search and Data Mining (CSDM)*

at the Fourth ACM International Conference on Web Search and Data Mining (WSDM), 15–18.

Iqbal, S. T., and Horvitz, E. 2007. Disruption and recovery of computing tasks: field study, analysis, and directions. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 677–686. ACM.

Kiesel, A.; Steinhauser, M.; Wendt, M.; Falkenstein, M.; Jost, K.; Philipp, A. M.; and Koch, I. 2010. Control and interference in task switching — a review. *Psychological bulletin* 136(5):849.

Locke, E. A., and Latham, G. P. 2002. Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American psychologist* 57(9):705.

Locke, E. A.; Shaw, K. N.; Saari, L. M.; and Latham, G. P. 1981. Goal setting and task performance: 1969–1980. *Psychological bulletin* 90(1):125.

Mark, G.; Gudith, D.; and Klocke, U. 2008. The cost of interrupted work: more speed and stress. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, 107–110. ACM.

Mason, W., and Watts, D. J. 2010. Financial incentives and the performance of crowds. *ACM SigKDD Explorations Newsletter* 11(2):100–108.

Mayr, U., and Kliegl, R. 2000. Task-set switching and long-term memory retrieval. *Journal of experimental psychology. Learning, memory, and cognition* 26(5):1124–1140.

Mento, A. J.; Steel, R. P.; and Karren, R. J. 1987. A meta-analytic study of the effects of goal setting on task performance: 1966–1984. *Organizational Behavior and Human Decision Processes* 39(1):52–83.

Monsell, S. 2003. Task switching. *Trends in cognitive sciences* 7(3):134–140.

Nieuwenhuis, S., and Monsell, S. 2002. Residual costs in task switching: Testing the failure-to-engage hypothesis. *Psychonomic Bulletin & Review* 9(1):86–92.

Pritchard, R. D., and Curts, M. I. 1973. The influence of goal setting and financial incentives on task performance. *Organizational Behavior and Human Performance* 10(2):175–183.

Rogers, R. D., and Monsell, S. 1995. Costs of a predictable switch between simple cognitive tasks. *Journal of experimental psychology: General* 124(2):207.

Rogstadius, J.; Kostakos, V.; Kittur, A.; Smus, B.; Laredo, J.; and Vukovic, M. 2011. An assessment of intrinsic and extrinsic motivation on task performance in crowdsourcing markets. In *ICWSM*.

Salvucci, D. D., and Taatgen, N. A. 2010. *The multitasking mind.* Oxford University Press.

Salvucci, D. D.; Taatgen, N. A.; and Borst, J. P. 2009. Toward a unified theory of the multitasking continuum: From concurrent performance to task switching, interruption, and resumption. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1819–1828. ACM.

Speier, C.; Valacich, J. S.; and Vessey, I. 1999. The influence of task interruption on individual decision making: An information overload perspective. *Decision Sciences* 30(2):337–360.

Stroop, J. R. 1935. Studies of interference in serial verbal reactions. *Journal of experimental psychology* 18(6):643.

Wylie, G., and Allport, A. 2000. Task switching and the measurement of switch costs. *Psychological research* 63(3-4):212–233.

Yin, M.; Chen, Y.; and Sun, Y.-A. 2013. The effects of performance-contingent financial incentives in online labor markets. In *Proceedings of the 27th AAAI Conference on Artificial Intelligence, AAAI*, volume 13.