Task Sequence Design: Evidence on Price and Difficulty*

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

1 Introduction

Many requesters in online labor markets offer workers opportunities to complete multiple tasks of the same type in a *sequence*. It is thus interesting to ask how to properly design a task sequence. We experimentally study the effects of two design variables, *price* and *difficulty* of tasks, on the quality of work produced in a task sequence. Intuitively, manipulating task price may affect workers' incentives and varying task difficulty may change required skills and resources for completing the task, all of which can affect work quality.

For the effects of task price, Mason and Watts (2009) showed that the magnitude of *performance-independent payment* influenced the quantity but not the quality of work completed. They found that the appropriate payment level claimed by workers in a post-task survey correlated with the actual payment of the task, resonating with a prominent psychological bias, the *anchoring effect*. In a recent study (Yin, Chen, and Sun 2013), we examined the *performance-contingent payment* and found that while the magnitude of such payment alone didn't affect work quality, the change in the magnitude of the payment in a two-task sequence did: increasing (decreasing) the payment level for the second task led to higher (lower) work quality on the task. These results are consistent with the conjecture that workers anchor their perception of the appropriate payment on the actual payment of the first task and increase (decrease) their effort level when paid higher (lower) in the second task.

A practical implication of our previous results is that we can obtain higher overall work quality in a sequence of two tasks by increasing the price of the second task following a low price of the first task. As real-world task sequences are often longer, we are interested in understanding whether the effect of the initial price anchor can last longer and influence work quality of subsequent tasks in a longer sequence or the effect will die out as workers observe more prices. Our first research question is: How is the work quality of the current task affected by the price of *each* of the previous tasks?

For the effects of task difficulty, we ask a similar question: Will work quality of the current task be influenced by the difficulty level of the previous task in a sequence of tasks of the same type? There is a large literature in psychology studying people's performance in a task after they transiting from the previous task. While psychologists robustly observe that work quality is lower when the previous task is different from the current task (task switch) than when the previous task is the same (task repetition), which is referred to as *task switching effect*, some researchers also notice a phenomenon called *sequential difficulty effect*, which predicts the work quality to always be lower following a difficult task due to the depletion of physical or mental resources. Furthermore, it is conjectured that the sequential difficulty effect can lead to the asymmetric or even reversed switch costs (Schneider and Anderson 2010).

2 Our Approach

We design and conduct two sets of experiments in Amazon Mechanical Turk (MTurk) to answer our research questions.

The first set of experiments aims to understand the relationship between work quality and the history of task prices. In these experiments, we place three tasks of the same type (and same or similar difficulty level) in a HIT and focus on analyzing work quality in the last task. We consider two levels of prices (as performance-contingent payment) for individual tasks: low (4 cents) and high (32 cents). Based on these two price levels, we experiment on five different price sequences: 4 – 4 – 4, 32 – 4 – 4, 4 – 32 – 4, 4 – 4 – 32, and 4 – 32 – 32.

To understand whether work quality of the current task will be affected by the difficulty level of the previous task, we conduct the second set of experiments where task price is the same but task difficulty can vary in a sequence of two tasks. Specifically, we consider four levels of task difficulty: easy, medium, hard and very hard and create three sets of treatments: HITs where the difficulty level of the second task is (1) equal to, (2) higher than or (3) lower than that of the first task. With these, we answer the second research question by comparing treatments (2) and (3) with treatments (1).

Finally, to understand whether our results depend on specific nature of tasks, we consider three types of tasks. We implement: (a) *the button clicking task* (BC): workers are asked to click a target button that alternates its location as quickly as possible in 3 minutes and work quality is measured as the number of clicks on the target; (b) *the spotting difference task* (SD): workers are asked to find as many differences be-

*We thank the support of Xerox Foundation on this work.

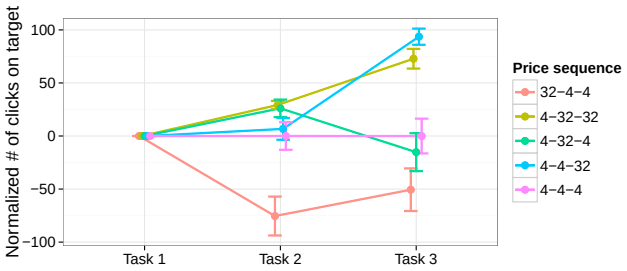


Figure 1: Work quality of BC tasks

tween two pictures as possible (five in total) and quality is measured as the number of differences correctly spotted; and (c) *the spatial N-back task* (NB): a colored square changes its location in a 3×3 grid every 3 seconds. Workers are asked to report the reappearance of the colored square at the same location as that of any of the most recent N trails and the report accuracy is used as the quality metric. Task (a) requires motor skills and is used in both sets of experiments (difficulty level is controlled by varying the distance between buttons). Both task (b) and (c) demand more cognitive skills like short memory. As it's easier to control the difficulty level for NB tasks by simply varying N, we use task (b) and (c) in the first and second set of experiments, respectively.

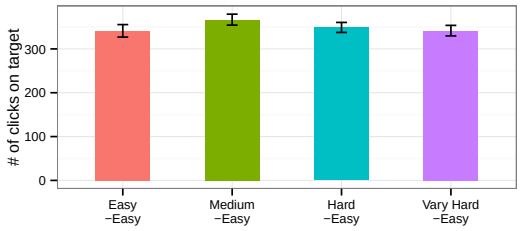
For each set of experiments, each worker is randomly assigned to one treatment upon arrival. We also restrict our experiments to US workers.

3 Preliminary Results

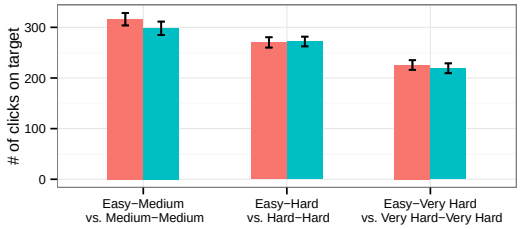
A few interesting results are observed in our initial experiments. We explain them below.

First, we find that task prices of all previous tasks have some effects on work quality of the current task. Figure 1 shows how quality changes when workers complete three BC tasks in a row after normalizing the data using performance in the 4 – 4 – 4 treatment as the reference. Consider work quality of the third task in the sequences. On the one hand, workers in the 32 – 4 – 4 treatment perform significantly worse than workers in the 4 – 4 – 4 treatment, indicating the perception of appropriate payment upon observing the price of the first task is carried over to the third task; on the other hand, work quality in the 4 – 32 – 4 (alternatively, the 4 – 32 – 32) treatment is lower than that in the 4 – 4 – 4 (alternatively, the 4 – 4 – 32) treatment, implying that the price of the second task also plays a role in influencing the work quality of the third task. Furthermore, since workers in the 4 – 32 – 4 treatment outperform their peers in the 32 – 4 – 4 treatment on the third task, we conjecture that the price of the first task exerts more influence than that of the second task on the quality of work produced in the third task. Results of sequences of three SD tasks are similar.

Surprisingly, for the second set of experiments, we observe no significant difference in work quality in the second task after the transition from tasks with different difficulty levels, which is inconsistent with the prediction of either task switching or sequential difficulty effect. Consider A and B as two difficulty levels with A being the easier one. For the transition from a difficult task to an easy task, i.e. treatment



(a) A – A vs. B – A



(b) A – B vs. B – B

Figure 2: Work quality of task 2 in the BC sequences after the difficulty level transition.

B – A, although the work quality of task 2 is predicted to be lower than that in treatment A – A by both the task switching and the sequential difficulty effects, we find workers actually have similar performance no matter how high the difficulty level B is (Figure 2(a)). For the transition from an easy task to a difficult task (treatment A – B), while the task switching effect and the sequential difficulty effect predict work quality in task 2 to be lower and higher, respectively, than that in treatment B – B, we didn't observe any significant difference (Figure 2(b)). Interestingly, according to our post-task survey, workers seem to adjust their perception of the appropriate payment corresponding to the change of difficulty levels over the subsequent two tasks: Workers' perception increases (or decreases) as the task difficulty increases (or decreases) in the sequence, which suggests a potential "anchoring" effect on the task difficulty. If such effect indeed exists, we expect workers to perform better (or worse) when the difficulty level decreases (or increases). Thus, the insensitivity of work quality to the transition of difficulty levels might be the result of a mixture of task switching, sequential difficulty and the "anchoring" effect. Similar results are observed from experiments with NB tasks.

References

Mason, W., and Watts, D. J. 2009. Financial incentives and the "performance of crowds". In *Proceedings of the ACM SIGKDD Workshop on Human Computation*, HCOMP '09.

Schneider, D. W., and Anderson, J. R. 2010. Asymmetric switch costs as sequential difficulty effects. *The Quarterly Journal of Experimental Psychology* 63(10):1873–1894.

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 '13.