

Labor Allocation in Paid Crowdsourcing: Experimental Evidence on Positioning, Nudges and Prices*

Dana Chandler and John Horton

MIT
oDesk Corporation

Abstract

This paper reports the results of a natural field experiment where workers from a paid crowdsourcing environment self-select into tasks and are presumed to have limited attention. In our experiment, workers labeled any of six pictures from a 2 x 3 grid of thumbnail images. In the absence of any incentives, workers exhibit a strong default bias and tend to select images from the top-left (“focal”) position; the bottom-right (“non-focal”) position, was the least preferred. We attempted to overcome this bias and increase the rate at which workers selected the least preferred task, by using a combination of monetary and non-monetary incentives. We also varied the saliency of these incentives by placing them in either the focal or non-focal position. Although both incentive types caused workers to re-allocate their labor, monetary incentives were more effective. Most interestingly, both incentive types worked better when they were placed in the focal position and made more salient. In fact, salient non-monetary incentives worked about as well as non-salient monetary ones. Our evidence suggests that user interface and cognitive biases play an important role in online labor markets and that salience can be used by employers as a kind of “incentive multiplier”.

Introduction

In this paper, we conduct a *natural field experiment* (Harrison and List 2004) in a paid crowdsourcing environment to study how worker decision making is affected by the type and saliency of incentives. In particular, we compare the effects of direct monetary incentives (which workers presumably directly value) to non-monetary incentives that merely indicate employer preferences without directly affecting workers’ material payoffs. Further, we investigate how these two types of incentives interact both with each other and with worker default behaviors (related to the pattern by which they scan a document or interface) that can give certain kinds of incentives more or less salience.

In Section we discuss the conceptual framework needed to understand the experiments and the theory that motivates our designs. Section explains the experimental design, Section reports the results and Section concludes.

Conceptual Framework

The power of positional effects has been extensively studied by human factors researchers; one well-established regularity is that people who read “left-to-right” in their native language tend to scan information in a corresponding left-to-right and top-to-bottom manner. This tendency has been well-documented in the context of navigating web pages and selecting items from computer interfaces (Goldberg et al. 2002; Hornof 2004). Positional effects also play an important role in how workers search for tasks in Amazon’s Mechanical Turk (MTurk) marketplace: users overwhelmingly tend to select tasks that are positioned first (Chilton et al. 2010). These positional effects suggest that people evaluate their options with time-saving heuristics of “default” behaviors.

Heuristics, attention and salience

Because it is not free to gather or process information, rational agents frequently employ shortcuts and thus often *rationality* decide to overlook certain aspects that may be relevant to a decision.¹ In the real world where people have limited attention, it matters how choices are presented.

Several empirical papers have documented that people respond more to *salient* prices. In toll booths where tolls are collected electronically, the price of tolls tend to rise more rapidly (even after controlling for confounding factors) suggesting that toll operators raise prices more since drivers are less aware of price increases when they do not physically take money out of their pockets to pay tolls (Finkelstein 2009).² In the public policy domain, Thaler and Sunstein argue in their book *Nudge: Improving Decisions About Health, Wealth, and Happiness* that, “setting default options, and other similar seemingly trivial menu-changing strategies, can have huge effects on outcomes, from increasing savings to improving health care to providing organs for life-saving transplant operations.” (Thaler and Sunstein 2008).

¹Herbert Simon, extended the concept of “homo economicus” with a model of “bounded rationality” in his seminal 1955 article (Simon 1955).

²DellaVigna also provides an overview of other interesting economic field experiments that measure the impact of limited attention, salience, and other cognitive biases (DellaVigna 2009)



Figure 1: The 2 x 3 grid of images from our task interface

Notes: This figure shows the 2 x 3 layout used in our experiment. Depending on treatment, different levels of bonuses and progress bars were shown below each image. Above, we show Treatment 6 where the bottom-right position has a 5-cent bonus and a 10% progress bar as compared with a 1-cent bonus and 90% progress bar in every other position.

Experimental approach

We examine the consequences of default behavior and investigate whether workers persist in picking whichever image arbitrarily appears in the top left, even when the images are made to have different payoffs.

Our first series of experiments uses *progress bars* to indicate that images have different levels of completion and examines whether workers will choose the task that is most in need of completion. Although workers are provided with an indication of how close to completion the work is, they are *not* rewarded for selecting any particular image.

Our second series of experiments parallels the first, but instead of progress bars, the experiment uses *bonuses* between 1 and 5 cents in order to observe whether workers will choose the task that gives them the highest explicit monetary incentive.

Given the importance of default behavior and limited attention, we also test whether the *position*—i.e., salience—of each incentive type also has an effect. Several of our treatments are identical except that they are placed in either a “focal” position (i.e., in the top-left) or a “non-focal” position (i.e., the bottom-right); by holding the incentive types constant and varying position, we can identify the effect of salience. A full description of both sets of treatments can be found in section .

To summarize, our treatments were divided along the following dimensions:

- **type of incentive:** monetary, progress bars or both
- **size of monetary incentive:** large bonus (5-cents) or small bonus (1-cent)
- **salience/position of incentive:** focal or non-focal

Experimental Design

Subjects were recruited from Amazon Mechanical Turk (MTurk) and were offered a payment of 25 cents to label a single image. We recruited a total of 837 subjects. The subjects were stratified over the seven treatment groups.

Our job posting was designed to appear just like any other image-labeling task that workers ordinarily perform in this labor market. As a result, the behavior we observe is likely to be representative of ordinary behavior in this marketplace. This design is known as a “natural field experiment” (Harrison and List 2004) that uses the “experimenter-as-employer” paradigm (Gneezy and List 2006; Horton, Rand, and Zeckhauser forthcoming)³ (Mason and Watts 2009) provides an overview of financial incentives on MTurk and (Shaw, Horton, and Chen 2010) provides an overview of a variety of non-monetary incentives.

Description of task and task interface

In our experiment, workers/subjects were asked to provide labels for photographic images. This is a common task in the MTurk market, as the challenges of computer vision make image labeling a prototypical human computation task. (Huang et al. 2010; Von Ahn and Dabbish 2004)

Before beginning, workers were given treatment-specific instructions on how to complete their task. Afterwards, they were presented with a screen showing a 2 x 3 array of images where they chose an image to label. Figure 1 shows the array and one particular configuration of our experimental

³Other recent examples that use this design include (Chandler and Kapelner 2010; Horton and Chilton 2010; Shaw, Horton, and Chen 2010; Mason and Watts 2009; Pallais 2010).

parameters (i.e., progress bar, bonus and position). Images were randomly assigned to the six positions and randomized for each trial.⁴

Our primary outcome measure was the position a worker chose to label from a 2-row, 3-column array of thumbnail images.

Types of incentives

Progress bars Treatments 1, 2 and 3 examine whether an arrangement of progress bars affected task selection. Subjects were told that:

“[progress bars] indicate how near we are to our labeling goals for each image... 100% indicates that we have reached our goal and don’t need any more labels. 0% indicates that we have no labels for that image.”

Table 1: Treatments involving progress bars

Treatment Number	Name	Level of progress bars	Position of 10% bar
0 ($n = 118$)	Balanced	All at 10%	—
1 ($n = 117$)	Imbalanced/Non-salient	5 at 90%, 1 at 10%	bottom-right
2 ($n = 115$)	Imbalanced/Salient	5 at 90%, 1 at 10%	top-left

Table 1 lists all of our *progress bar* treatments.

Monetary incentive (bonus) treatments Our remaining treatments offered workers monetary bonuses in exchange for labeling images at certain positions. Each position was given a particular bonus level; a worker selecting an image received the bonus associated with that position (if any). Across treatments, we varied the size of the bonuses, the placement of the bonuses and whether bonuses were paired with progress bars.

Table 2: Treatments involving bonuses

Treatment Number	Name	Level of bonuses	Position of largest bonus
3 ($n = 125$)	Large Bonus/Non-salient	5 at 1c, 1 at 5c	bottom-right
4 ($n = 125$)	Large Bonus/Salient	5 at 1c, 1 at 5c	top-left
5 ($n = 109$)	Small Bonus/Non-salient	5 at 0c, 1 at 1c	bottom-right
6 ($n = 128$)	Large Bonus/Salient	5 at 1c, 1 at 5c	bottom-right

⁴This allows us to disentangle whether a particular *position* was more popular and whether a particular *image* was more popular.

Design and predictions from bonus treatments Across our bonus treatments, we paid two levels of bonuses: a five cent (“large bonus”) and a 1 cent (“small bonus”). In Treatment 3, a 5-cent large bonus was placed on the lower right hand corner—the non-focal position. Every other position offered a 1-cent small bonus. In Treatment 4, a 5-cent bonus was placed on the focal position while every other position offered a 1-cent bonus. (Note the parallelism with Treatments 1 and 2.)

We next examined whether the size of the bonus matters. If workers do indeed have lexicographic preferences, they should select the image that offers the highest bonus, even if it is only 1-cent. Treatment 5 tested this by offering a 1-cent bonus in the non-focal position and no bonus (0 cents) at every other position.

In Treatments 3, 4 and 5, the levels of the progress bars were set to be neutral; specifically, all progress bars were set at 10% and the only dimensions of the experiment that changed were the size and position of the bonuses. In contrast, Treatment 6 used imbalanced progress bars, implicitly suggesting the non-focal image. Treatment 6 measured the double effect of using both progress bars and bonuses to draw people towards the bottom right. In essence, Treatment 6 combines Treatments 1 and 4 by employing both an imbalanced configuration of progress bars and a large (5-cent) bonus in the non-focal position.

Table 2 summarizes all of the *bonus* treatments.

Experimental results

In the subsections below, we present results relating to a number of hypotheses. The raw data for each of the seven experimental groups is shown in Figure 2—see the figure note for a detailed explanation.

In the top left grid of this figure, we can see that without any inducement, the fraction of subjects choosing the bottom-right position is only 8%, about half of what would be expected (1/6th or 16.7%—for a two-sided t-test: $p < .03$). Having imbalanced progress bars pushes workers to select the non-focal image. However, offering bonuses has a larger effect. The largest effect by a wide margin occurs when we combine both progress bars and monetary incentives, as shown in the bottom panel.

Workers exhibit a “default bias” and prefer tasks located in more focal positions

As hypothesized, our workers exhibit a strong preference for the image in top-left, focal position and select it 36.4% of the time in our baseline group, more than twice as much as would be expected by random ($p < .003$). The bottom-right (non-focal) position is selected only 8.5% of the time (or less than half as much as would be expected). Figure 2 shows the raw data for the baseline and all other treatments.

In addition to a preference for the focal position, workers also systematically preferred certain *images* over others. A formal test of whether all six images were equally popular is rejected using a chi-square goodness of fit test ($p < 0.014$).⁵

⁵Incidentally, the most attractive image for workers was one

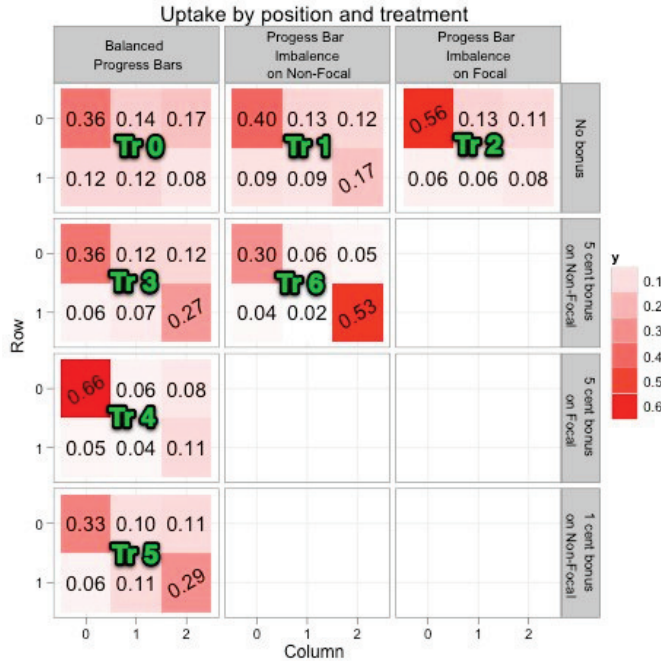


Figure 2: Fraction of subjects selecting different positions within each treatment

Notes: The above matrix includes seven different 2 x 3 grids, each of which represent the raw data showing the proportion of workers who selected each position for the seven treatments. The baseline group, Treatment 0, is shown at the top-left and illustrates the position bias that occurs when progress bars are balanced and without any bonuses. The other 2 x 3 grids are arranged according to the types of incentives employed. Within each 2 x 3 grid, if a number is oriented at 45-degrees, that particular position was incentivized as part of a treatment.

Saliency makes all incentives more effective

We tested the hypothesis of whether more salient incentives would have stronger effects. To evaluate monetary incentives, we either paid a 5-cent bonus in the non-focal position (Treatment 3) or the focal position (Treatment 4). Similarly, the progress bars indicated that the worker choose the non-focal position (Treatment 1) or the focal position (Treatment 2).

At baseline, 8.5% choose the non-focal position and 36.4% choose the focal position. For each position, we thus measure the effectiveness of a treatment by how much it increased the percentage relative to these bases. Our progress bar incentives increased the absolute percentages by 8.6% ($p < .03$) and 19.2% ($p < .002$) for the non-focal and focal positions, respectively; the monetary incentives increased the absolute percentages by 18.7% ($p < .0002$) and 30.0% ($p < .00001$) for the non-focal and focal positions.

which showed toy horses, the least popular image also had the lowest number of labels, suggesting that it may have been avoided due to its difficulty.

Figure 3 shows the size of the change relative to baseline for incentives depending on their type and saliency.

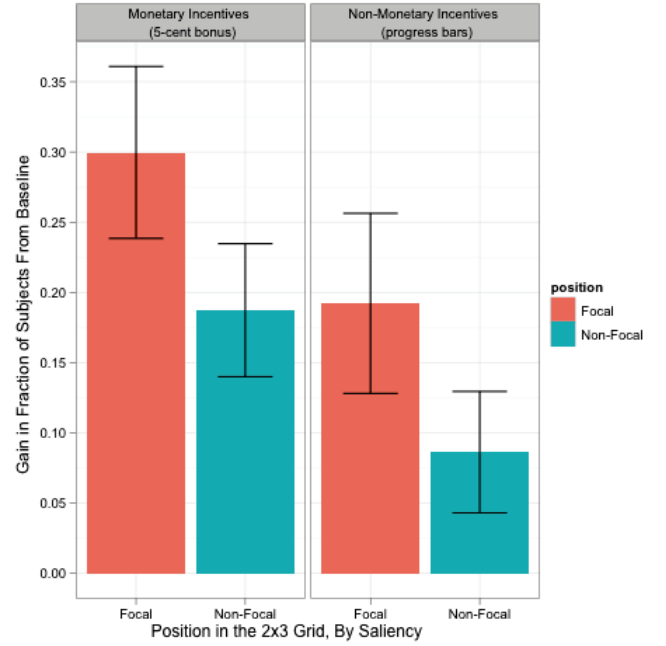


Figure 3: Comparison of treatment effects by saliency of position

Notes: This chart shows the increase in the fraction of people who selected each position relative to the baseline. The data are broken down by whether the incentive was placed in a focal or non-focal position. For example, the first bar indicates that an additional 30 percentage points of people chose the focal position when it had a 5-cent bonus. The above bars correspond to the treatment effects under Treatment 4, 3, 2 and 1 relative to Treatment 0, the baseline. Error bars report standard errors and show that more salient incentives had a significantly greater effect.

Hence for both monetary and non-monetary incentives, we see that more salient positioning substantially increases its effectiveness. This is illustrated in Figure 3.

Progress bars weakly influence whether a worker selects an image

Using data from Treatment groups 0, 1 and 2, we tested whether the arrangement of progress bars causes workers to change which position they select. Surprisingly, the progress bar arrangements can be used as an effective way to influence how workers choose tasks (as shown in the bottom row of Figure 2). Using progress bars in the non-focal position increases the proportion of people who select it to 17.1%, about the level that would be expected if people did not have any bias or preference over positions.

Progress bars have a greater effect on which position a worker selects when placed in a salient, focal position. In the baseline, 36.0% of workers chose the focal image; when the focal image was incentivized using progress bars, this rose to 55.7%. For the non-focal position, the baseline was

8.5% and increased to 17.1% when incentivized. Figure 3 shows same information graphically.

Monetary bonuses strongly influence whether a worker selects an image

We hypothesized that by offering a 5-cent bonus to label the images at certain positions, we could induce workers to label those images. Further, we hypothesized that bonuses would be more effective when placed at the focal position as compared with the non-focal position simply because focal bonuses were more likely to be noticed.

Using data from Treatment groups 0, 3 and 4, we found that bonuses were highly effective, regardless of where they were offered. When the bonus was offered at the focal position, selection of the image at that position increased to 66.4% from the baseline level of 36.4%. When the bonus was offered at the non-focal position, uptake increased from 8.5% to 27.2%.

When monetary bonuses were combined with the progress bars in the non-focal position, the percentage of people choosing the bottom-right image increased from 8.5% to 53.1% for a total of almost 45%. Independently, progress bars raised the percentage almost 9% and bonuses by about 18%. Combining both increased the proportion that selected the non-focal position by an additional 18%.

Don't overpay—small bonuses worked as well as larger ones

We hypothesized that workers would choose the image that was in the position of the highest bonus. Our conjecture is supported by the data, but one caveat for would-be crowdsourcers is that our 5 cent bonus was unnecessary—1-cent was suffice to shift behavior, with no discernible difference in the 1-cent versus the 5-cent treatments.

Using data from treatment groups 0, 3, and 5, we varied the level of the bonus in the non-focal position and to determine whether paying 1 cent versus 5 cents affects the likelihood that workers will choose to work on the image in the non-focal position. As previously discussed, paying a 5-cent bonus in the non-focal position raised the proportion of people selecting that position to 27% relative to a baseline of 8%. When we pay a small bonus of 1 cent, 29% of workers are induced to select the non-focal position. Paying either a 1- or 5-cent bonus more than triples the proportion who select the non-focal position and there is essentially no difference between paying a 1-cent or 5-cent bonus.⁶

Conclusion

As crowdsourcing and human computation systems evolve and expand, the magnitude and efficacy of offered incentives will be of first-order importance: there can be no human

⁶The difference between 27% and 29% is not significant. However, there is a possibility that workers perceived the 1 cent bonus as larger given that it was positioned near a zero-cent bonuses. One way to test this hypothesis would be to introduce another treatment where Treatment 3 and 4 is modified to include all 1-cent bonuses and a single 2-cent bonus (rather than a 5-cent bonus).

computation without the humans, and humans will need to be given the proper incentives.

Our paper shows that incentives and signals—regardless of their type—need to be known and salient in order to be effective. This result is perhaps not too surprising: we would hardly expect workers to be motivated by bonuses they are unaware of or for community members to strive to earn badges that others do not recognize. However, our results have subtler and hence more easily overlooked implications—namely that combining incentives of different types can be very effective. In our experiment, the strongest effects were obtained by using hybrid incentives that combined non-monetary and monetary incentives or by placing monetary incentives in a salient position.

Although monetary incentives worked better than non-monetary incentives in general, a salient, well-positioned non-monetary incentive was as effective as a more costly monetary incentive that was poorly positioned. This suggests that task designers should recognize the role that saliency plays in enhancing or diminishing the power of incentives.

Acknowledgments

Thanks to the the Christakis Lab at Harvard Medical School and the NSF-IGERT Multidisciplinary Program in Inequality & Social Policy for generous financial support (Grant No. 0333403). Thanks to Lydia Chilton, Pat DeJarnette, Bruno Ferman, Adam Kapelner, Greg Little, Aaron Shaw, Daan Struyven, Robin Yerkes Horton and seminar participants at the Harvard Labor Economics seminar and for very helpful comments. Dana Chandler acknowledges support from the National Science Foundation in the form of a Graduate Research Fellowship. All plots were made using `ggplot2` (Wickham 2008). Thanks to John Comeau for excellent programming.

References

- Chandler, D., and Kapelner, A. 2010. Breaking monotony with meaning: Motivation in crowdsourcing markets. *University of Chicago mimeo*.
- Chilton, L.; Horton, J.; Miller, R.; and Azenkot, S. 2010. Task search in a human computation market. In *Proceedings of the ACM SIGKDD Workshop on Human Computation*, 1–9. ACM.
- DellaVigna, S. 2009. Psychology and economics: Evidence from the field. *Journal of Economic Literature* 47(2):315–372.
- Finkelstein, A. 2009. E-ZTAX: Tax Salience and Tax Rates*. *Quarterly Journal of Economics* 124(3):969–1010.
- Gneezy, U., and List, J. 2006. Putting behavioral economics to work: Testing for gift exchange in labor markets using field experiments. *Econometrica* 1365–1384.
- Goldberg, J.; Stimson, M.; Lewenstein, M.; Scott, N.; and Wichansky, A. 2002. Eye tracking in web search tasks: design implications. In *Proceedings of the 2002 symposium on Eye tracking research & applications*, number 650, 51–58. ACM.

- Harrison, G., and List, J. 2004. Field experiments. *Journal of Economic Literature* 42(4):1009–1055.
- Hornof, A. 2004. Cognitive Strategies for the Visual Search of Hierarchical Computer Displays. *Human-Computer Interaction* 19(3):183–223.
- Horton, J. J., and Chilton, L. B. 2010. The labor economics of paid crowdsourcing. *Proceedings of the 11th ACM Conference on Electronic Commerce 2010*.
- Horton, J. J.; Rand, D.; and Zeckhauser, R. J. forthcoming. The online laboratory: Conducting experiments in a real labor market. *Experimental Economics*.
- Huang, E.; Zhang, H.; Parkes, D.; Gajos, K.; and Chen, Y. 2010. Toward automatic task design: A progress report. In *Proceedings of the ACM SIGKDD Workshop on Human Computation (HCOMP)*.
- Mason, W., and Watts, D. J. 2009. Financial incentives and the ‘performance of crowds’. In *Proc. ACM SIGKDD Workshop on Human Computation (HCOMP)*.
- Pallais, A. 2010. Inefficient Hiring in Entry-Level Labor Markets. *MIT mimeo*.
- Shaw, A.; Horton, J.; and Chen, D. 2010. Designing Incentives for Inexpert Human Raters. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work*.
- Simon, H. 1955. A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*.
- Thaler, R., and Sunstein, C. 2008. *Nudge: Improving decisions about health, wealth, and happiness*. Yale Univ Press.
- Von Ahn, L., and Dabbish, L. 2004. Labeling images with a computer game. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 319–326. ACM.
- Wickham, H. 2008. ggplot2: An implementation of the grammar of graphics. *R package version 0.7*, URL: <http://CRAN.R-project.org/package=ggplot2>.