Working Hard or Working Long? Responsiveness to Incentives in Real Effort Tasks

David Hagmann[1] and Alex Rees-Jones[2]

[1]Department of Social and Decision Sciences, Carnegie Mellon University
hagmann@cmu.edu
[2]Operations, Information, and Decisions Department
The Wharton School, University of Pennsylvania
alre@wharton.upenn.edu

September 21, 2018

Abstract

When the piece-rate wage for workers increases, we expect them to exert greater effort. A growing body of evidence from the field, however, suggests that this relationship is not so straight-forward: when wages increase, workers have repeatedly been shown to decrease their output – from cab drivers in New York City, to bike messengers in Switzerland and pear packers in California. In parallel, experimental economists have developed tasks in the laboratory that allow effort decisions to be studied in a controlled setting. Yet, results have not been much more promising for the standard model. Responses to incentives have been muted: a 10,000% increase in the wage rate in one study increased output by 5%. Despite the shockingly small effect sizes, this observation has received no systematic attention. Given the importance of wages in labor markets, understanding how people adjust their effort in response to low or high wages has wide-ranging implications. Workers frequently have two ways to respond to higher wages: they could work harder, increasing their productivity, or they could choose to work longer. We show, using three separate real effort tasks, that output is not responsive to wages when participants face a time constraint. Indeed, participants complete as many tasks when they earn a wage that is a multiple of their opportunity cost as when they earn nothing. With the time limit removed, however, participants in our between-subjects experiment are indeed sensitive to wage rates.

> I learned a long time ago that a person can stand just about anything for 10 seconds, then you just start on a new 10 seconds. All you've got to do is take it 10 seconds at a time.
>
> Unbreakable Kimmy Schmidt

In many professional industries, including academia, workdays are loosely defined. Workers choose when to begin their days and, importantly, when to end them. They may continue to work after returning home from the office and may engage in various tasks over the weekend and on holidays. In addition, their performance is not closely monitored. They may choose to exert a lot of effort while at work or they may spend their time at the proverbial water cooler. When high performance is needed, employees may be motivated via higher wages and performance bonuses. But how do they respond to these incentives? Do they become more productive and produce more output in the same amount of time – or do they instead work longer hours? Or do they, maybe quite sensibly, choose a mixture of both?

The question is challenging to answer with field data: the manual tasks in which output is easy to track often do not allow flexible working hours. A worker at a grocery store may stack produce more or less efficiently, but she generally has little influence over her working hours. And in jobs where flexible hours are common, output may be nearly impossible to alter. An Uber driver facing a higher wage, for example, can at most marginally increase the number of rides per hour (e.g. by speeding and running through red lights). She can best respond by working additional hours. These issues arise long before we consider self-selection into different industries and the difficulty of tracking work performed outside of the office (e.g. working on presentations at home or pursuing continuing education).

Economic theory, of course, is clear on what we should expect in the abstract: when wages increase, effort should correspondingly increase as well. Given the straight-forward prediction, it is all the more surprising that previous field and laboratory studies do not observe that. In the seminal paper on New York City taxi drivers, for example, Camerer et al. (1997) find that as the wage of cab drivers increases, they spend fewer hours working. It looks as if they set a daily income target and retire for the day after reaching it. Chang and Gross (2014) report that pear packers respond by working more slowly when their piece rate wage increases. Similarly, Fehr and Goette (2007) find that bike messengers in Switzerland work longer hours when their wage increases, but complete fewer rides per hour. Ariely et al.

(2009b) find that when they offer participants a massive incentive for performing well on tasks, they end up "choking" and performing substantially worse.

Experimental economists have developed a series of "real effort" tasks in which participants repeatedly push buttons for 10 minutes (DellaVigna and Pope, 2016; Ariely et al., 2009a), transcribe a fixed number of strings of blurry Greek characters (Augenblick et al., 2015), enter strings in reverse for up to 60 minutes (Huck et al., 2015), move sliders to various points (Gill and Prowse, 2012). Lezzi et al. (2015) find that effects can differ by task type, for example finding that men outperform women in the slider task, but not in two other tasks they had used. These papers generally impose a time limit and assume that participants' effort, that is their output per minute on the task, is a function of how they are compensated. Evidence for this is limited, however: introducing a piece-rate wage for the participant, in addition to donations to charity, appears to have an effect on output only when performance is private, but not when it is public (Ariely et al., 2009a).

However, with only two exceptions, previous work did not explicitly test the fundamental question of responsiveness to incentives. DellaVigna and Pope (2016) recruit about 550 participants for each of 18 experimental conditions via Amazon Mechanical Turk, an online labor platform. Of those, five conditions are of particular relevance: participants either earn no piece-rate wage, one cent for every 1,000 pairs of letters entered, one cent for every 100 pairs, 4 cents for every 100 pairs, or 10 cents for every 100 pairs. The number of tasks completed increases from 1,521 to 1,883 when increasing pay from zero to mean earnings of less than two cents total. They further increase to 2,029 when paying one cent per 100 pairs, to 2,132 at 4 cents per 100 pairs, and to 2,175 at 10 cents per 100 pairs. Thus, a 10,000% increase in the wage increases effort by approximately 5%. Araujo et al. (2016) explicitly pitch different incentive levels against each other in a task in which participants position sliders and find a similar 5% increase in effort following a 1,500% increase in the wage.

In a large study on incentives in psychological experiments, Buhrmester et al. (2011) found that recruitment speed increases when the wage goes up. Conditional on participation, however, there is no difference in observable demographics nor does data quality differ across wage rates. It appears that once people have decided to participate in a task, they try to perform it to the best of their abilities.

These findings leave us with two possibilities: first, and maybe a priori most plausibly, experimental tasks are largely unable to capture the essence of effort and participants, for whatever reason, always work hard when they get paid. This appears unlikely, however: in a meta-study of 39 psychology experiments, Jenkins Jr. et al. (1998) find that incentives do have an effect on the quantity of the work performed, though not on the quality. Second, and maybe more interestingly, experimental tasks capture exactly what goes on and people simply are not more productive when paid more in the commonly employed experimental paradigms.

Recall that participants could increase their earnings in one of two ways: they can work harder or they can work longer. Economists using experiments traditionally impose time limits, holding fixed the number of minutes participants in their experiments can work on a task. They generally do so for practical reasons, e.g. to ensure that participants move through the experiment at the same pace. The effect of incentives, then, is observed with regard to how productive they are in the given time. This matches many jobs in the real world, in which employees have fixed hours during which they perform their work. They may, however, both in experiments and real employment simply decide to stop performing additional tasks and letting the clock run out. Alternatively, some experiments allow participants to click a "stop" button at any time. In the experiments we have cited above, however, the authors do not discuss how long participants actually work on the task.[1] We suspect, and will investigate experimentally, that the timer is a fairly strong demand and participants will run out the clock rather than quit prior. Even salaried employees, who technically have the freedom to leave work early, may not do so and read up on news and social media posts before heading home at the end of the workday.

[1] We want to double check the last two claims.

In this paper, we report results from a controlled experiment in which we will randomly assign participants to different wages and either impose a maximum time constraint or allow them to work freely on the task. We then have two measures of productivity: we can look at the number of tasks participants complete per minute and the number of minutes participants spent on the task. We check the robustness of our results (following Lezzi et al., 2015) by conducting the experiment using three different and commonly used experimental tasks.

Experiment

We designed an experiment that allows us to study performance in a real-effort task in the presence of different incentives and in the presence (or absence) of time constraints. Participants enrolling in our experiment receive a fixed payment of 20 cents and are told they may be able to earn an additional bonus based on their performance. We present them with an experimental task (described below) and observe a count of how many tasks they have successfully completed and how much time they spent on the screen with the task. They may, at any time, click a "Stop" button and advance to a final set of demographic questions. Participants then are randomly assigned to receiving either 2 cents per completed task ("Low incentive"), 25 cents per completed task ("High incentive"), or no piece rate payment ("No incentive"). Furthermore, participants are assigned to a "Timed" condition, in which they have at most 5 minutes during which they can work on the tasks, and an "Untimed" condition that imposes no time constraint. In all cases, however, the experiment will end once 25 tasks were completed. We did not disclose this termination rule to avoid creating a target, but instituted the limit to ensure there was an upper limit to their earnings. The screens displayed to participants are shown in Appendix C.

We recruit our participants via Amazon Mechanical Turk (MTurk), an online labor platform that has become popular among experimental researchers. On the platform, participants select assignments that may involve experimental research or routine tasks (e.g. image classification) in exchange for payment. Although MTurk limits the control we have over participants' environments, it offers three considerable advantages for this design: first, it allows us to recruit a sample size that would be unavailable in experimental laboratories. Across our conditions, we had more than 1,800 participants engage in our tasks. Second, and more importantly, people on the platform are particularly aware of the opportunity costs of their participation. Any minute spent completing tasks as part of this experiment means they are not earning money doing another task. Participants who came into the laboratory, by contrast, would already have scheduled time to be there and might exert effort merely because they enjoy the task, which would diminish any researcher's ability to study the role of incentives.

Experimental Tasks

For our experiment, we draw on three different real-effort tasks used in the literature. This allows us to test the robustness of our findings to different types of tasks that may be more or less painful for participants to engage in. We start by describing the two tasks that have previously been used in the economics literature.

In the "Reverse" task (Huck et al., 2015), participants are shown a randomly generated string containing 35 alphanumeric characters. The string can contain any digit and any letter except for capital "I" and small "l," which are too difficult to distinguish and may lead participants to infer that the task is broken. Participants are then asked to type the string into a text box in reverse. If the string is entered correctly, the count of completed tasks increases by one and a new string is generated. If the string is incorrect, a message stating that an error has been made appears and participants have to find and correct the error.

In the "Greek" task (adapted from Augenblick et al., 2015), participants are shown a randomly generated string of length 30, consisting of 10 different Greek letters: $\alpha, \beta, \chi, \delta, \epsilon, \Phi, \gamma, \eta, \iota, \zeta$.

As these letters cannot be typed on a standard keyboard, they further have on their screen 10 buttons, each corresponding to one of the letters. Clicking a button inserts the associated letter at the end of a text box. Once all letters have been entered, participants can "submit" the entry and, if they made no errors, a new string is randomly generated. Participants who make a mistake have to delete all letters back to where the mistake occurred, as clicking the button appends only to the end of the text entry.

Finally, in the "Matrices" task (adapted from Mazar et al., 2008) participants are shown a 4×4 table containing 16 numbers between zero and ten (not inclusive). Each number is printed to three decimal places. Their task is to find the (unique) pair of numbers that adds up to 10 and enter them (in any order) into two text boxes. We generated 50 matrices prior to the experiment and participants were randomly shown matrices from this set (with no repeats).

Across all three tasks, we disabled the option to copy or paste text and in the matrices task, the tables were displayed as images. Participants in all conditions had the option to stop completing tasks at any time by clicking a "STOP" button. The design, sample size, inclusion criteria, and analyses were preregistered on AsPredicted and we report all variables that were collected.

We recruited participants from Amazon Mechanical Turk and 1807 completed our experiment. Because of a coding error, 19 participants in two conditions (the timed and untimed versions of the high incentive Greek task) received incorrect instructions and were exclude from our analyses. As pre-registered, we include participants who did not complete the full experiment, but who proceeded past the consent form to read the instructions where they learned what their task and bonus payments would be ($n = 215$. The final sample consists of 2003 participants. We show an overview of all collected variables, separated by wage condition, in Table 1.

Table 1: Characteristics of Participants.

	0 cents N=675	2 cents N=671	25 cents N=657	p.overall
finished: no	74 (11.0%)	73 (10.9%)	68 (10.4%)	0.926
task:				0.918
greek	227 (33.6%)	228 (34.0%)	208 (31.7%)	
matrices	225 (33.3%)	224 (33.4%)	226 (34.4%)	
reverse	223 (33.0%)	219 (32.6%)	223 (33.9%)	
timed: Untimed	339 (50.2%)	335 (49.9%)	332 (50.5%)	0.976
gender: Female	342 (56.3%)	318 (52.8%)	322 (54.6%)	0.470
ethnicity:				0.334
White	496 (81.7%)	510 (84.7%)	463 (78.5%)	
Black	46 (7.58%)	36 (5.98%)	52 (8.81%)	
Indian and Alaskan	5 (0.82%)	8 (1.33%)	5 (0.85%)	
Asian	38 (6.26%)	34 (5.65%)	46 (7.80%)	
Pacific Islander	1 (0.16%)	0 (0.00%)	1 (0.17%)	
Other	21 (3.46%)	14 (2.33%)	23 (3.90%)	
education:				0.745
Less than high school	3 (0.50%)	2 (0.33%)	3 (0.51%)	
High school graduate	49 (8.13%)	61 (10.1%)	54 (9.15%)	
Some college	150 (24.9%)	147 (24.4%)	125 (21.2%)	
2 year degree	88 (14.6%)	67 (11.1%)	72 (12.2%)	
4 year degree	222 (36.8%)	237 (39.4%)	247 (41.9%)	
Professional or masters	81 (13.4%)	79 (13.1%)	80 (13.6%)	
Doctoral degree	10 (1.66%)	9 (1.50%)	9 (1.53%)	

continued on next page

Table 1 – *continued from previous page*

	0 cents N=675	2 cents N=671	25 cents N=657	p.overall
politics:				0.157
Extremely liberal	60 (9.95%)	87 (14.5%)	78 (13.2%)	
Liberal	159 (26.4%)	139 (23.1%)	164 (27.8%)	
Slightly liberal	82 (13.6%)	81 (13.5%)	83 (14.1%)	
Moderate	127 (21.1%)	137 (22.8%)	103 (17.5%)	
Slightly conservative	71 (11.8%)	74 (12.3%)	68 (11.5%)	
Conservative	86 (14.3%)	68 (11.3%)	70 (11.9%)	
Extremely conservative	18 (2.99%)	16 (2.66%)	24 (4.07%)	

Results

Task Difficulty

Because the nature of our tasks differs considerably, we begin by assessing their relative difficulty. One intuitive measure is to look at how many tasks participants could feasibly complete in the 5 minute timed conditions. Collapsing across all wages, we find that the most productive workers entered 7 Greek strings, found 9 pairs of numbers that added up to 10, and reversed 15 strings. Average performance was considerably lower: workers in the timed conditions completed on average 1.76 Greek tasks, found 2.04 pairs of numbers, and reversed 4.42 strings. This suggests that the Greek and Matrices tasks are approximately of similar difficulty, while the Reverse task appears to have been easier.

	Greek (timed)	Matrices (timed)	Reverse (timed)	Greek (untimed)	Matrices (untimed)	Reverse (untimed)
High Incentive	0.39	0.23	0.39	3.58^{***}	6.14^{***}	5.77^{***}
	(0.26)	(0.28)	(0.49)	(0.92)	(0.97)	(1.13)
No Incentive	-0.66^{**}	-0.06	-0.61	-0.64	-0.84	-3.16^{**}
	(0.26)	(0.28)	(0.49)	(0.90)	(0.97)	(1.13)
(Intercept)	1.87^{***}	1.98^{***}	4.49^{***}	3.42^{***}	4.17^{***}	7.21^{***}
	(0.18)	(0.20)	(0.35)	(0.64)	(0.69)	(0.80)
R^2	0.05	0.00	0.01	0.07	0.15	0.17
Adj. R^2	0.04	-0.00	0.01	0.06	0.15	0.16
Num. obs.	329	335	333	334	340	332
RMSE	1.93	2.11	3.65	6.82	7.31	8.37

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 2: Responsiveness to receiving a high incentive (25 cents) or no incentive, compared to a low incentive (2 cents) baseline. All regressions are OLS with the number of tasks completed as the dependent variable.

Average Effort

In Figure 1, we show the number of tasks completed in each of the 18 experimental conditions. For a table with means and standard errors, see Table 6. As is evident, we see much greater responsiveness to incentives when participants are not constrained by a time limit. In the Greek task, for example, effort in the timed condition increases from 1.2 tasks with no incentives to 1.87 tasks with the low incentive and 2.25 tasks with the high incentive.

In Table 2, we show the results of linear regression analyses. For each regression, our dependent variable is the number of tasks completed and our reference level is the 2 cent wage condition. This allows us to look at responsiveness from a low to a high wage (our pre-registered analysis) and also the change in effort from no to a low incentive. In the Greek task, for example, we see that the increase in effort is significant when going from no wage to a 2 cent per task wage, but not for an increase to a wage of 25 cents per task. In the reverse timed task, we see participants complete one additional task in the high wage condition compared to the no wage condition, but the remaining comparisons are not significant. Finally, in the timed matrix task, there is no difference across any of the three incentive levels.

The results look quite different in the timed conditions, however. Beginning again with the Greek task, participants complete 3.6 additional tasks in the high incentive condition

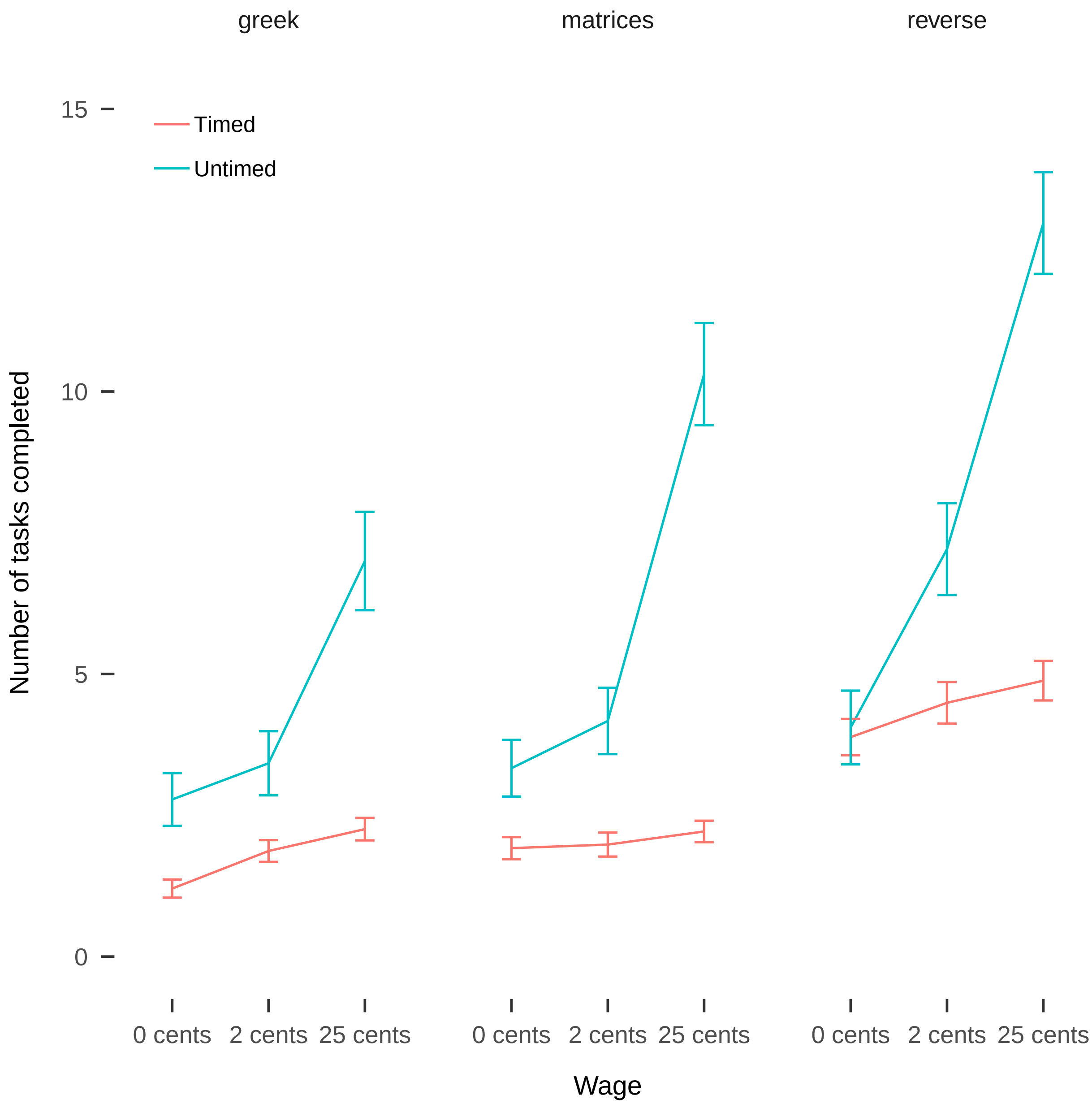

Figure 1: Number of tasks completed in each of the 18 experimental conditions.

compared to when they receive a low incentive. We observe no difference between the low or no incentive conditions. The same is true for the matrices task: participants complete the same number of tasks in the low and no incentive conditions, but complete an additional 6.1 tasks when incentives are high. Finally, in the untimed reverse task, we find that participants complete the least tasks (4.1) when they receive no incentive, more tasks with a low incentive (7.2) and the most tasks when the incentive is high (13.0).

While responsiveness to incentives with a time constraint is muted, at best, we see that all tasks see an increase in effort with higher pay when there is no such constraint imposed. We next look at one factor that might drive these results: the time participants spend completing tasks before choosing to stop (or running out of time).

Time on Task

A second measure of responsiveness to incentives is how much time participants spent completing tasks. We pre-registered the analyses with no predictions, but it allows us to explore whether participants work longer when the wage is high and serves as a robustness check for the analyses on the number of tasks completed. Participants might be intrinsically motivated to work on the task or feel that they are expected to exert some effort to earn their show-up fee.

Figure 2 shows the median number of minutes participants in each condition spent working before either running out of time (after 5 minutes in the timed conditions) or clicking the "Stop" button (in all conditions). The analysis is limited to the 1836 participants who completed this stage of the experiment. We do not have a measure for time on task for participants who terminated either prior to beginning the task or who quit the survey in the middle of the task.

When participants have a time limit imposed, we observe that the median completion time is close to the full 5 minutes available to them across wage rates. That is, even when participants are not paid for completing the tasks, they appear to work on it for the full

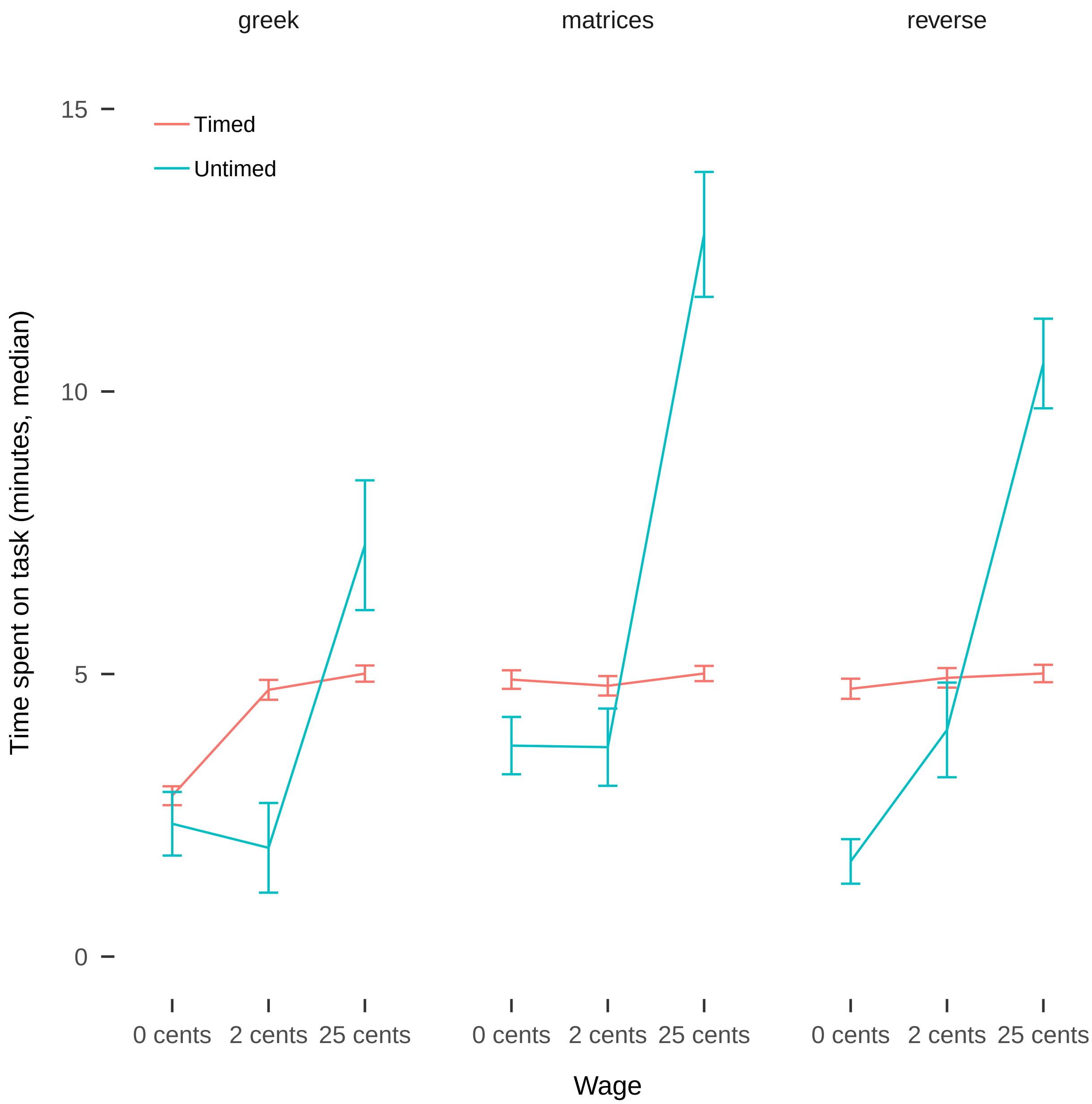

Figure 2: Minutes spent completing tasks across conditions and wage rates.

	Greek (timed)	Matrices (timed)	Reverse (timed)	Greek (untimed)	Matrices (untimed)	Reverse (untimed)
High Incentive	0.56^{*}	0.70^{**}	0.45	6.24^{***}	9.20^{***}	3.82^{***}
	(0.24)	(0.24)	(0.24)	(1.29)	(1.21)	(1.02)
No Incentive	-0.37	0.19	-0.23	-1.03	-1.16	-4.18^{***}
	(0.24)	(0.24)	(0.24)	(1.24)	(1.21)	(1.04)
(Intercept)	3.42^{***}	3.50^{***}	3.61^{***}	5.68^{***}	6.52^{***}	7.76^{***}
	(0.17)	(0.17)	(0.17)	(0.88)	(0.86)	(0.73)
R^2	0.05	0.03	0.02	0.11	0.23	0.16
Adj. R^2	0.04	0.02	0.02	0.10	0.22	0.16
Num. obs.	304	297	324	299	300	312
RMSE	1.72	1.67	1.78	8.94	8.58	7.42

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 3: Responsiveness to receiving a high incentive (25 cents) or no incentive, compared to a low incentive (2 cents) baseline. All regressions are OLS with the time spent completing tasks (in minutes) as the dependent variable. Participants in the timed conditions were censored at 5 minutes.

amount of time allotted to them. A notable outlier is the unincentivized Greek task, where participants stop after less than 3 minutes on average.

Interestingly, we observe that for the no and low incentive conditions, the median time spent completing tasks is *lower* when there is no time limit, even though participants in all conditions could stop at any time. This is consistent with the notion that the timer imposes an expectation of spending the full allotted time on the task. In the high incentive conditions, on the other hand, participants spend substantially more time completing tasks. It appears that one channel through which greater effort materializes itself is the time spent working when that channel is available.

Table 3 shows the results of linear regressions, where the dependent variable is the time spent on the task (in minutes). We cannot rule out that participants merely kept open the tab in the background and switched to other websites. However, they had no incentive to do so and could simply have clicked the stop button at any time to advance to the demographic questions. Indeed, we made explicit in the instructions that there was no minimum amount of time that they were expected to spend on the task.

Our regression results find a small, but significant, effect of the high incentive condition on time spent in the timed Greek and Matrices tasks. On average across the three tasks, participants spent about 3.5 minutes (out of the available 5 minutes) working. Those who earned 25 cents per task, versus those who earned 2 cents, spent an additional 34 seconds working on entering Greek letters and an additional 42 seconds finding numbers adding up to 10 in a table. They did not work longer when the task was to reverse strings of letters. Providing no incentive whatsoever did not significantly reduce how long they worked.

We see much stronger effects across the untimed conditions. In the Greek task, participants with a low incentive worked for 5.7 minutes, while those who received the high incentive worked nearly 12 minutes. Similarly in the matrix task, where the high incentive increased the time spent on the task from 6.5 minutes to 15.7 minutes. For those tasks, too, we observe no effect of paying a small wage versus offering no piece-rate incentive. Finally, in the reversing string tasks, we see participants responding to both a low and a high incentive. When participants received no piece-rate bonus, they worked for only 3.6 minutes. With the low incentive, they worked 7.76 minutes and for as long as 11.58 minutes with the high wage.

Indeed, these estimates of the willingness to spend time on the task are likely to underestimate what people would be willing to do. Recall that we offered participants at most 25 tasks to complete. Across the untimed versions of the tasks, few participants with no piece-rate wage (3.8%) and the low wage (6.0%) reached this limit. In the high wage conditions, however, as many as 19.6% reached the ceiling and would likely have continued working.

Productivity

We next consider how productive participants were while they were working – that is how many tasks they completed per minute. We collapse the observations across the timed and untimed conditions and show productivity across the three tasks in Figure 3. We also fit a loess regression for each of the wage rates and each of the tasks. This approach allows the data to be fit by curves with multiple inflection points. As is visually apparent, however, the best-fitting estimate is linear for most of the range. Moreover, the lines overlap again for most of the range, suggesting that participants are equally productive per minute whether

	Greek	Matrices	Reverse
High Incentive	-0.05	-0.00	0.04
	(0.04)	(0.05)	(0.06)
No Incentive	-0.06	-0.02	0.02
	(0.04)	(0.05)	(0.06)
(Intercept)	0.77^{***}	0.76^{***}	1.24^{***}
	(0.03)	(0.03)	(0.04)
R^2	0.01	0.00	0.00
Adj. R^2	0.00	-0.00	-0.00
Num. obs.	380	448	527
RMSE	0.31	0.40	0.58

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 4: OLS regression on the number of tasks completed per minute working across incentive conditions.

they are paid a high wage or no wage at all. They diverge and become non-linear only for participants who spent more than 20 minutes on the task, where we have few observations.

To explore this further, we define a "productivity" measure equivalent to the number of tasks completed per minute. For the purpose of this analysis, we exclude those who completed no tasks and those who did not advance past the experimental task. We further collapse our observations across the timed and untimed conditions. This final productivity measure ranges from 0.07 tasks per minute to 3.2 tasks, with an average of 0.94. Table 4 shows that this measure does not differ across any of the experimental tasks.

Responsiveness to Incentives

Experimentalists may be particularly interested in exploring which of the tasks we used is most responsive to incentives. As pre-registered, we will compare the 2 cents to the 25 conditions only, across all 12 task, timing, and wage rate pairings. We interact dummy variables for high wages with dummies for each task and timing combination. The first column of Table 5 shows the pre-registered analysis in which the dependent variable is the number of tasks completed. The second column shows, as a robustness check, the same specification with the time spent working on the task as the dependent variable.

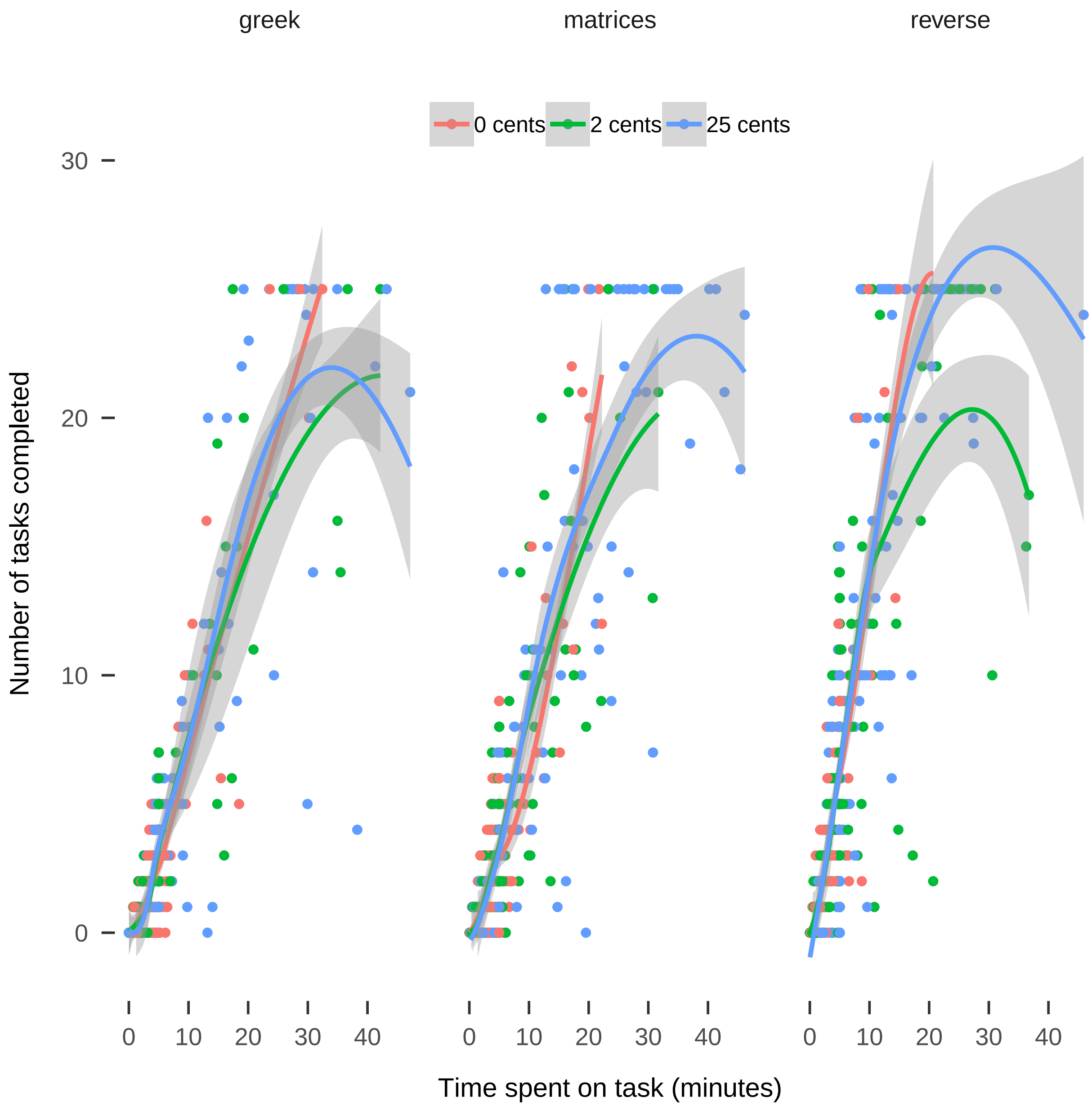

Figure 3: Tasks completed as a function of how many minutes participants in the untimed condition spent on the task. The overlapping loess regression curves suggest that (1) output is linear in the amount of time participants spend on the task and (2) higher incentives do not lead participants to complete the tasks any faster.

We find that in the timed conditions, participants are unresponsive to higher incentives, even as they increase more than 10-fold. In the untimed conditions, however, we see that participants complete between 3.19 (Greek) and 5.75 (Matrix) additional tasks. This observation similarly holds on time spent on the task: when the wage rate is high, participants do not spend more time on any of the tasks. However, the interaction between the wage rate and the untimed conditions is significant for all three tasks. Both in terms of number of tasks completed and time spent on the task, the matrices in which participants had to find the two numbers adding up to 10 showed the greatest responsiveness. This may be particularly noteworthy as the task has, to the best of our knowledge, not previously found applications in economics research.

Discussion

When faced with a high wage, participants may indeed be motivated to exert more effort. But are they able to do so? Settings that impose binding time constraints may be designed with the idea that participants will simply work harder. However, both prior research and our current findings suggest that this is difficult for participants to do. They appear to already be maximally productive even when no incentive is offered. When participants, however, are able to determine how long to spend on a task, they adjust their effort and work longer when incentives are high. This holds even as participants are not provided with a reference point in the study. That is, they do not know that they could have earned a high (or low) wage and are informed exclusively by drawing on their previous experience on the platform. Experimental tasks seeking to measure willingness to exert effort should consequently eliminate time limits and, if necessary, impose a cap on the number of tasks that can be completed instead.

Notably, we find that participants who have no time limit imposed and earn no or low wages terminate the experiment earlier than those who have a timer, even though the latter

	Tasks Completed	Time on Tasks
High wage	0.39	0.56
	(0.84)	(0.97)
Matrix (timed)	0.11	0.08
	(0.82)	(0.95)
Reverse (timed)	2.62^{**}	0.19
	(0.83)	(0.93)
Greek (untimed)	1.55	2.26^{*}
	(0.82)	(0.94)
Matrix (untimed)	2.30^{**}	3.10^{**}
	(0.82)	(0.95)
Reverse (untimed)	5.34^{***}	4.34^{***}
	(0.83)	(0.94)
High wage x Matrices (timed)	-0.15	0.15
	(1.18)	(1.37)
High wage x Reverse (timed)	0.01	-0.11
	(1.18)	(1.34)
High wage x Greek (untimed)	3.19^{**}	5.69^{***}
	(1.18)	(1.38)
High wage x Matrices (untimed)	5.75^{***}	8.64^{***}
	(1.18)	(1.36)
High wage x Reverse (untimed)	5.38^{***}	3.26^{*}
	(1.18)	(1.35)
(Intercept)	1.87^{**}	3.42^{***}
	(0.58)	(0.66)
R^2	0.23	0.25
Adj. R^2	0.23	0.24
Num. obs.	1328	1218
RMSE	6.18	6.81

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 5: Across all three tasks, participants in the untimed conditions were not responsive to an increase in the piece-rate wage from 2 cents to 25 cents, whereas participants without a time limit completed more tasks when payment was higher. The sample size for Time on Tasks is smaller because only those who completed this stage are included. We do not know when someone closed their browser window.

are explicitly told that they can quit at any time with no penalty to their earnings. Participants may nonetheless be motivated to exert effort either because they believe there to be a social norm, or because they treat the task as a game of skill and are motivated to perform well.

References

Felipe A. Araujo, Erin Carbone, Lynn Conell-Price, Marli W. Dunietz, Ania Jaroszewicz, Rachel Landsman, Diego Lamé, Lise Vesterlund, Stephanie W. Wang, and Alistair J. Wilson. The slider task: An example of restricted inference on incentive effects. *Journal of the Economic Science Association*, 2(1):1–12, May 2016. ISSN 2199-6784. doi: 10.1007/s40881-016-0025-7.

Dan Ariely, Anat Bracha, and Stephan Meier. Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially. *American Economic Review*, 99(1):544–555, February 2009a. ISSN 0002-8282. doi: 10.1257/aer.99.1.544.

Dan Ariely, Uri Gneezy, George Loewenstein, and Nina Mazar. Large Stakes and Big Mistakes. *Review of Economic Studies*, 76(2):451–469, April 2009b. ISSN 00346527.

Ned Augenblick, Muriel Niederle, and Charles Sprenger. Working Over Time: Dynamic Inconsistency in Real Effort Tasks. *The Quarterly Journal of Economics*, page qjv020, May 2015. ISSN 0033-5533, 1531-4650. doi: 10.1093/qje/qjv020.

Michael Buhrmester, Tracy Kwang, and Samuel D. Gosling. Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science*, 6(1):3–5, January 2011. ISSN 1745-6916. doi: 10.1177/1745691610393980.

Colin Camerer, Linda Babcock, George Loewenstein, and Richard Thaler. Labor Supply of New York City Cabdrivers: One Day at a Time. *The Quarterly Journal of Economics*, 112(2):407–441, May 1997. ISSN 0033-5533. doi: 10.1162/003355397555244.

Tom Chang and Tal Gross. How many pears would a pear packer pack if a pear packer could pack pears at quasi-exogenously varying piece rates? *Journal of Economic Behavior & Organization*, 99:1–17, March 2014. ISSN 0167-2681. doi: 10.1016/j.jebo.2013.11.001.

Stefano DellaVigna and Devin Pope. What Motivates Effort? Evidence and Expert Forecasts. Technical report, June 2016.

Ernst Fehr and Lorenz Goette. Do Workers Work More if Wages Are High? Evidence from a Randomized Field Experiment. *American Economic Review*, 97(1):298–317, March 2007. ISSN 0002-8282. doi: 10.1257/aer.97.1.298.

David Gill and Victoria Prowse. A Structural Analysis of Disappointment Aversion in a Real Effort Competition. *American Economic Review*, 102(1):469–503, February 2012. ISSN 0002-8282. doi: 10.1257/aer.102.1.469.

Steffen Huck, Nora Szech, and Lukas M. Wenner. More effort with less pay: On information avoidance, belief design and performance. Technical report, Working Paper Series in Economics, Karlsruher Institut für Technologie (KIT), 2015.

G Douglas Jenkins Jr., Atul Mitra, Nina Gupta, and Jason D Shaw. Are financial incentives related to performance? A meta-analytic review of empirical research. *Journal of applied psychology*, 83(5):777, 1998.

Emanuela Lezzi, Piers Fleming, and Daniel John Zizzo. Does it Matter Which Effort Task You Use? A Comparison of Four Effort Tasks When Agents Compete for a Prize. *SSRN Electronic Journal*, 2015. ISSN 1556-5068. doi: 10.2139/ssrn.2594659.

Nina Mazar, On Amir, and Dan Ariely. The Dishonesty of Honest People: A Theory of Self-Concept Maintenance. *Journal of Marketing Research*, 45(6):633–644, December 2008. ISSN 0022-2437. doi: 10.1509/jmkr.45.6.633.

Task	Time Constraint	0 cents	2 cents	25 cents
Greek	Timed	1.2 (0.16) (n = 113)	1.87 (0.19) (n = 114)	2.25 (0.2) (n = 102)
	Untimed	2.78 (0.47) (n = 114)	3.42 (0.57) (n = 114)	7 (0.87) (n = 106)
Matrices	Timed	1.92 (0.2) (n = 111)	1.98 (0.21) (n = 112)	2.21 (0.19) (n = 112)
	Untimed	3.33 (0.5) (n = 114)	4.17 (0.59) (n = 112)	10.31 (0.9) (n = 114)
Reverse	Timed	3.88 (0.32) (n = 112)	4.49 (0.37) (n = 110)	4.88 (0.35) (n = 111)
	Untimed	4.05 (0.65) (n = 111)	7.21 (0.81) (n = 109)	12.98 (0.9) (n = 112)

Table 6: Number of completed tasks across all 18 experimental conditions. Standard errors in parentheses.

A Data and R Code

The de-identified data from all our experiments and statistical code for all analyses and figures reported in the paper and the supplementary analyses will be available via Github and as an R library on CRAN following acceptance of the paper.

B Supplemental Analyses

C Experimental Materials

(To Be Included)

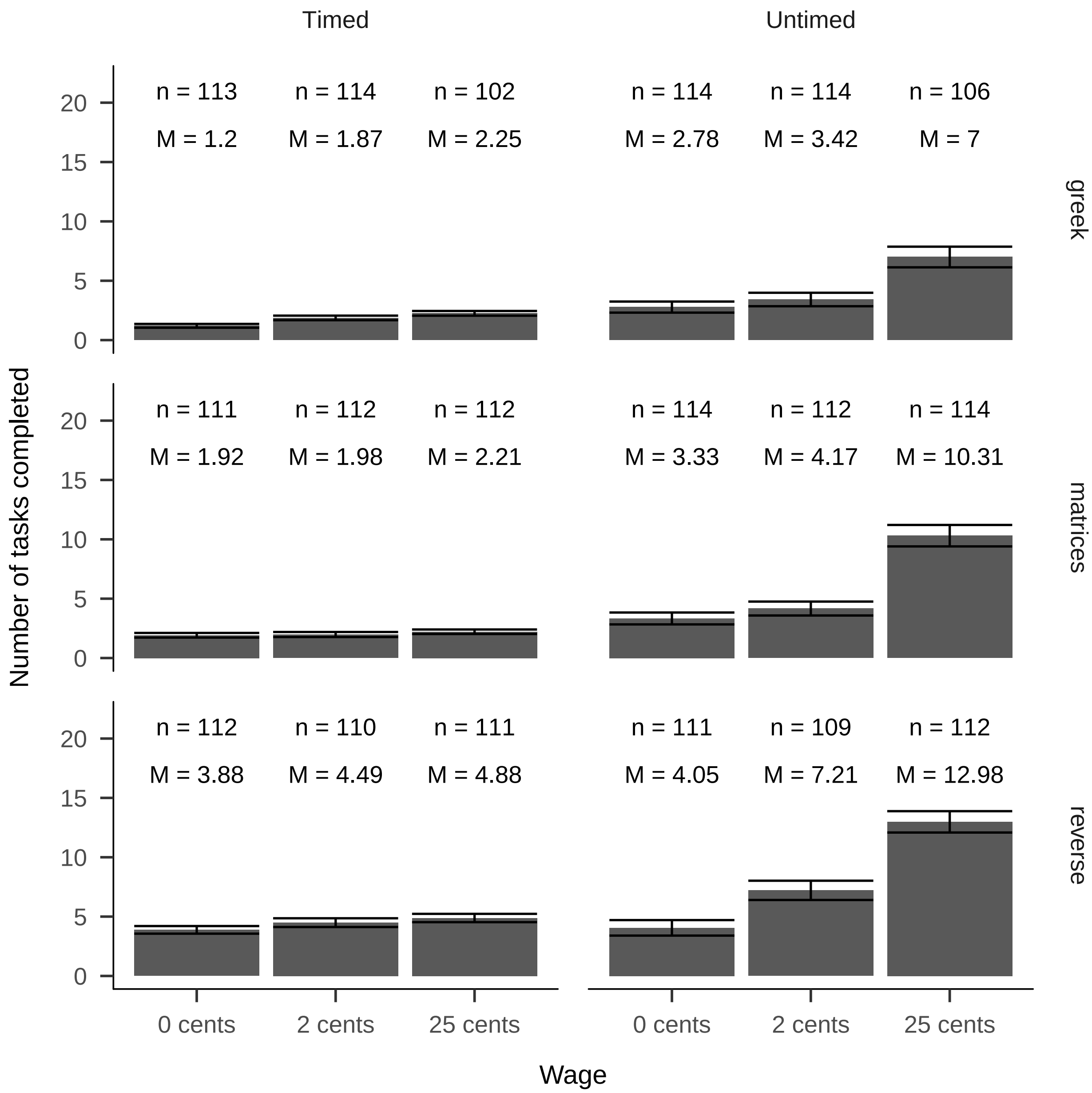

Figure 4: Number of tasks completed in each of the 18 experimental conditions.

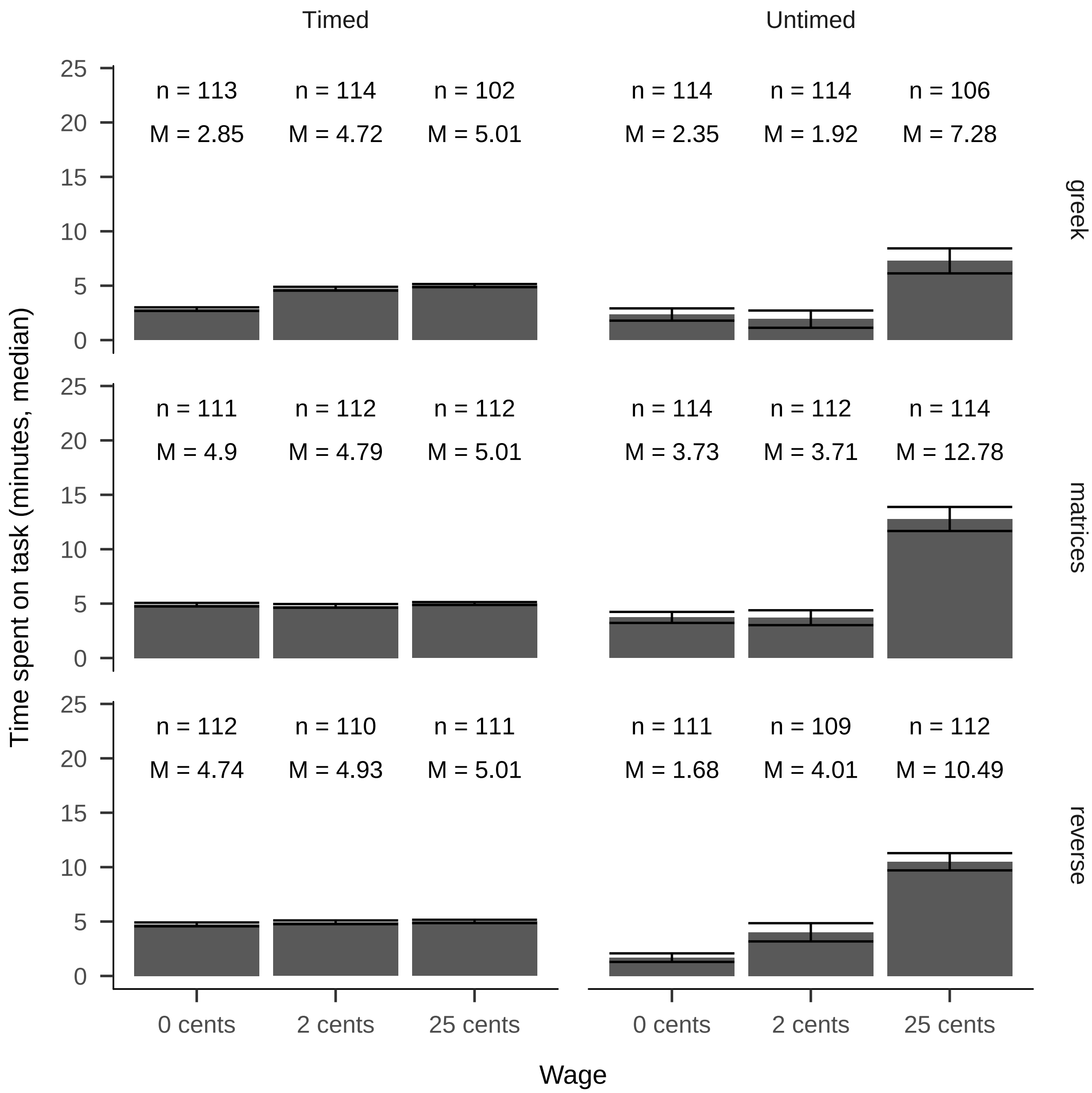

Figure 5: Minutes spent completing tasks across conditions and wage rates.