



The Effect of Incentives on Real Effort: Evidence from the Slider Task

Felipe Augusto de Araujo / Erin Carbone /
Lynn Conell-Price / Marli W. Dunietz /
Ania Jaroszewicz / Rachel Landsman /
Diego Lamé / Lise Vesterlund /
Stephanie Wang / Alistair J. Wilson

CESIFO WORKING PAPER NO. 5372
CATEGORY 13: BEHAVIOURAL ECONOMICS
MAY 2015

An electronic version of the paper may be downloaded

- from the SSRN website: www.SSRN.com
- from the RePEc website: www.RePEc.org
- from the CESifo website: www.CESifo-group.org/wp

ISSN 2364-1428

The Effect of Incentives on Real Effort: Evidence from the Slider Task

Abstract

Real-effort experiments are frequently used when examining a response to incentives. For any particular real-effort task to be well-suited for such an exercise, subjects' cost for exerting effort must, for the range of incentives considered, result in an interior effort choice. The popular slider task in Gill and Prowse (2012) has been characterized as satisfying this requirement, and the task has been increasingly used to investigate the response to changes in both monetary and nonmonetary incentives. However, despite its increasing use, a simple between-subject examination of the slider task's response to incentives has not been conducted. We provide such an examination with three different piece-rate incentives: half a cent, two cents, and eight cents per slider completed. We find that participants in the three treatments completed on average 26.1, 26.6 and 27.3 sliders per round, respectively. The one-slider increase in observed performance is small, not only relative to the sixteen-fold increase in the incentives, but also relative to the observed heterogeneity across subjects, rates of learning, and even idiosyncratic variation. Our paper cautions that the slider task will be underpowered for uncovering a response to incentives in between-subject designs.

JEL-Code: C900, C910.

Keywords: slider task, real effort, experimental methodology.

*Felipe Augusto de Araujo,
Marli W. Dunietz,
Rachel Landsman,
Diego Lamé,
Lise Vesterlund,
Stephanie Wang &
Alistair J. Wilson (alistair@pitt.edu)
Department of Economics
University of Pittsburgh / USA*

*Erin Carbone
GSPIA / University of Pittsburgh / USA*

*Lynn Conell-Price &
Ania Jaroszewicz
Social and Decision Sciences
Carnegie-Mellon University / USA*

May 2015

This paper was written as a joint classroom project among faculty and students for a graduate course in experimental economics at the University of Pittsburgh.

1. INTRODUCTION

Early economic experiments examining labor effort in the lab relied on the stated-effort design (for example: Bull et al., 1987; Schotter and Weigelt, 1992; Nalbantian and Schotter, 1997; Fehr et al., 1993). Participants in the role of workers were given an endowment and asked to “purchase” a level of effort, which in turn benefited other participants in the role of principals. While stated-effort designs provided well-structured controls for participants’ costs of effort, the designs were seen as being abstract, overly distant from the types of labor effort the experiments were intended to capture.¹ Scholars subsequently began to use real-effort designs, where participants are instead paid for performing an actual task in the lab.

Real-effort designs achieve less abstraction by trading off experimental control, as participants’ effort costs are no longer directly tied to payoffs. This inability to manipulate the cost places restrictions on the types of tasks that can be used to study the response to incentives.

Take a simple decision-theoretic model of a real-effort task to make things concrete. In choosing her effort e , participant i solves the following problem:

$$e_i^*(w) = \arg \max_{e \in [0, E_i]} w \cdot e - c_i(e),$$

where w denotes a piece rate payment, $c_i(e)$ denotes the (differentiable) cost of effort she brings into the lab, and E_i represents the maximum effort level she can choose to exert (due to physical/time constraints). If we are to use a real-effort task to study the response to incentives in the laboratory, then it must be—at least for the set of incentives considered—that the cost function generates an interior solution to the above maximization problem. For example, a real-effort task that individuals see as enjoyable will not satisfy the restriction. If it is satisfying to perform the task, then the marginal cost is negative, $0 > c'_i(e)$, and the same optimal effort $e^*(w) = E_i$ will be chosen for all piece rates $w \geq 0$. Any subject who enjoys anagrams, math problems, sudoku puzzles, and/or throwing a ball into a basket will exert maximum effort on these tasks independent of the offered piece rate w . That is, their behavior will not respond to incentives.

The experimental community has been quick to develop creative real effort tasks. In considering easily implementable tasks that are short enough to be run repeatedly, the “slider task” has stood out as being sensitive to incentives. Gill and Prowse (2012, hereafter abbreviated to G&P) introduce the slider task in a study on disappointment aversion. Participants are shown a screen with 48 sliders, where each slider has a range of positions from 0 to 100. Each slider is solved by using the computer’s mouse to move the slider’s marker (initially placed at 0) to the midpoint of 50. Participants are given two minutes to solve as many sliders as possible, with the participant’s chosen effort given by the number of sliders correctly positioned at 50 by the end of the two

¹For example, Charness and Kuhn (2011, p.13) state “the advantage of using real effort is that the task is more in line with what most people consider labor, and so might be considered to be a better match to the field environment.”

minutes. The task is normally repeated T times—in many experiments $T = 10$ —and cumulative earnings across the entire experiment are given by $\sum_{t=1}^T w_t \cdot e_i^*(w_t)$.

Evidence on sensitivity to the incentives—in particular, that the slope of $e_i^*(w_t)$, assessed through a within-subject design, is positive—has led to the slider task being used more and more frequently in papers measuring the incentive effects associated with various mechanisms and work environments. However, in contrast to the sensitivity to monetary incentives uncovered in the initial study by G&P, more recent slider-task studies (in particular, those using between-subject designs) have found more modest or non-existent treatment effects. Our paper’s main result, obtained from a simple between-subject experimental design, suggests these recent papers might not be viewed as true null results, but stemming instead from performance in the slider task not being particularly sensitive to changes in the offered incentives.

Where other studies have varied more complex elements of the payoff environment—strategic elements within a game, the nature of feedback, the frame, etc.—ours will be very simple, focused only on assessing whether the slider task responds to monetary incentives.² Following G&P’s implementation of the slider task as much as possible, we conduct three treatments where we vary the piece-rate payment w that participants receive for each correctly positioned slider—a half cent at the low end, an intermediate two cent treatment, and eight cents at the high end. This sixteen-fold increase in the piece rate corresponds to dramatic differences in participants’ potential earnings, with maximum possible performance payments of \$2.40, \$9.60 and \$38.40, respectively. However, despite substantial differences in the incentives offered, we uncover fairly limited differences in average effort exerted: in order of increasing piece rates, we find that subjects complete 26.1, 26.6, and 27.3 sliders per two-minute round. This one-slider increase in performance is limited, not only relative to the order-of-magnitude increase in the incentives, but also relative to the levels of individual heterogeneity in ability in the task and rate of learning documented in our data.³

As a real-effort task, the slider task has many attractive characteristics. However, our paper indicates that the task is not well-suited for uncovering responses to incentives in between-subject designs.

2. LITERATURE REVIEW

Real effort has been used as a way of securing participants an earned endowment (Erkal et al., 2011), as a way of controlling for the individual’s ability (Niederle and Vesterlund, 2007), and as a way of examining the response to incentives (G&P). For the latter purpose, the concern in using real effort is that the cost of exerting effort is unknown, and so it may be more difficult to gauge

²Our emphasis is solely on extrinsic motivation. See Gneezy et al. (2011) for a careful account of the potential role played by intrinsic motivations across a varied set of environments.

³Individual-level heterogeneity can be thought of here as the expected variation with respect to the subject i for the maximal effort E_{it} , while learning effects can be thought of as within-subject shifts in this variable across time t .

the response to incentives. Indeed, such a response can only be uncovered if the marginal cost of effort is positive and sensitive to the level of effort exerted.

While numerous real-effort tasks have been conducted in the laboratory, the ‘slider task’ (G&P, Gill and Prowse 2015) has stood out as being a task with particularly desirable characteristics.⁴ Gill and Prowse (2015) note that the slider task is easy to implement, does not require pre-existing knowledge, is identical across repetitions, involves little randomness, leaves no scope for guessing, provides a finely gradated measure of effort within a short time scale, allows for repeated effort and thus controls for unobserved heterogeneity, and finally that the effort exerted responds to incentives. Highlighting that the repeated performance measures make it possible to estimate the individual implicit cost of effort, Gill and Prowse (2015, p.4) argue that the task overcomes “the principal drawback of using real effort up to now.”

The popularity of the slider task has indeed been impressive. Despite its recent development, the task has to date been used in more than 40 completed studies. The preeminent feature of the task is that performance within it is thought to be responsive to incentives, thus making it well-suited for studies that aim to examine how effort responds to changes in institutions and mechanisms, and to modifications that might influence the participant’s intrinsic or non-monetary incentives. Surprisingly, the response to treatment has been limited in a number of these studies.

One explanation for the absence of a treatment effect is of course that effort fails to respond to the particular treatments considered by these papers. Another more concerning explanation is that the slider task is not well suited for detecting a response. In particular it is of concern if the slider task is relatively insensitive to incentives.

In reviewing the many papers examining effort in the slider task, we noticed that significant responses to changes in the incentives offered arise more often in within-subject designs, where participants are aware of the differing incentives being offered. In particular, the original G&P study is within-subject, where subjects had full knowledge of the range of potential incentives, and they could judge each realization of the incentives relative to the reference range chosen by the experimenter. Similarly, in Besley and Ghatak (2013), subjects are told whether they would receive a high or low piece rate at the beginning of each round, with a clear reference to how the offered incentive related to the other possibility. Bradler et al. (2015) introduce additional reward mechanisms such as a tournament in the second round after a first round with just a piece-rate payment, and the subjects return to the piece rate scheme in the third round. That is, subjects in the second round are aware that their incentives are higher compared to the first round, and in the third round they are aware that the payoffs revert back to the low level. Thus, subjects may feel

⁴Examples of other real-effort tasks range from simple mental tasks such as solving anagrams (Charness and Villeval, 2009), arithmetic problems (Niederle and Vesterlund, 2007), mazes (Gneezy et al., 2003), or optimization problems (Dickinson and Villeval, 2008) to rote physical tasks such as pressing keys on a keyboard (Swenson, 1988), stuffing envelopes (Carpenter et al., 2010), or squeezing a hand dynamometer (Imas, 2014).

compelled to respond to the treatment and the control differentially, thinking that the experimenter expects different behavior for the different incentives.

Furthermore, much of the evidence that the slider task responds to incentives have come from studies with strategic externalities, where the subject's choice of effort affects the payoffs of others. That is monetary incentives along with other-regarding preferences, and strategic concern might jointly influence choices. In several studies, the incentive scheme varies between a piece rate and a tournament where factors like competitiveness must be considered (Besley and Ghatak, 2013; Bradler et al., 2015). In others, the treatment manipulates non-monetary incentives that could affect the behavior of subjects with other-regarding preferences (Hammermann et al., 2012; Feess et al., 2014), or concern for social image (Georganas et al., 2014).⁵ Studies on the effect of various tax institutions and tax rates also require that the subjects make a nuanced mapping between policy and direct monetary incentive differences (Abeler and Jäger, forthcoming; Doerrenberg and Duncan, 2014; Doerrenberg et al., 2015).

A number of studies do vary the incentive mechanism or institution between subjects, and for the most part they find null results. Buser and Dreber (2014) find that subjects do not complete a significantly different number of sliders, regardless of whether they are paid by a piece rate, a single reward from a tournament, or paid via a lottery. Georganas et al. (2014) find that there is only a significant effect from being observed by others on the number of sliders solved in the first two rounds but not over all rounds. Ibanez and Schaffland (2013) find a treatment effect, and show that randomly selected leaders brought from outside the group complete fewer sliders than randomly selected in-group leaders (leaders only had social incentives here, as their effort only increases the payoff for the other group members). However, when leaders were chosen for being skilled in the slider task, they find no significant treatment difference between in- and out-group leaders. They remark (p.13) "This could indicate that skilled participants have an internal motivation to work on the task (i.e., they enjoyed it more than others, or wanted to increase performance) so their behavior was not affected by the treatment." Imas et al. (forthcoming) also find an effect, with more sliders completed when subjects are presented with a loss frame versus a gain frame, though no monetary incentives were involved in that experiment. Dolan et al. (2012) find that neither varying the size of the subjects' initial endowment nor varying whether subsequent payoffs were framed as gains/losses had any effect on the number of sliders completed.

In a few slider-task studies, the results do indicate an effect of changing incentives on effort, but only for a sub-population, or a subset of rounds. For sub-populations, Feess et al. (2014) find that only male subjects solve different numbers of sliders under different incentive schemes.

⁵A number of studies have used the slider task as a way for subjects to work for their endowment because of some of its aforementioned positive qualities (for example: Brown et al., 2013; Evdokimov and Rahman, 2014; Kesternich et al., 2014; Khadjavi, 2014). Other studies have used the slider task to generate a performance measure (for example, Lindner, 2014).

TABLE 1. Classifying the Slider Literature on Incentive Effects

Paper	Design (Incentives)	Monetary Incentive	Significant Effort?
Brown et al. (2013)	Between	No	No
Buser and Dreber (2014)		Yes	No
Doerrenberg et al. (2015)		Yes	No
Dolan et al. (2012)		No	No
Georganas et al. (2014)		No	No
Ibanez and Schaffland (2013)		No	Yes
Imas et al. (forthcoming)		No	Yes
Abeler and Jäger (forthcoming)	Within	Yes	Partial
Banuri and Keefer (2013)		No	Yes
Besley and Ghatak (2013)		Yes	Yes
Bradler et al. (2015)		Yes	Yes
Doerrenberg and Duncan (2014)		Yes	Partial
Feess et al. (2014)		Yes	Partial
Friedl et al. (2014)		Yes	Partial
Gerhards and Siemer (2014)		No	Yes
Gill and Prowse (2012)		Yes	Yes
Hammermann et al. (2012)		Yes	Yes
Straub et al. (2015)		Yes	No

Georganas et al. (2014) only find a significant effect from social observation in the first two rounds, while Friedl et al. (2014) only find a significant effect of a piece rate payment scheme versus a flat-rate one in the last five rounds, after subjects have been exposed to both payment schemes. For studying the response to incentives in the laboratory, experimenters would ideally want a task that produces an incentive effect for most subjects, and for all rounds of the experiment. Such a task would not only reduce the number of subjects the study needs to have sufficient power, but also reduce concerns that the population compositions are comparable across treatments.

Table 1 contains a summary of the previous studies that have examined incentives with the slider task, where we have categorized studies by: the type of design, within or between subject; whether the incentives being offered were monetary or not; and whether a significant response to effort was found over the treatment. For the seven between-subject studies, only two find a significant effort effect, where the remaining five can be broadly categorized as null results. In contrast, out of the eleven within-subject studies, ten find at least a partial response, while the majority (six) find a significant response for effort. Our paper seeks to determine whether the more frequent null results found with the slider task in between-subject designs results from the task itself, or the specific treatments examined with it. That is, the null results found have been for the

joint hypothesis that the slider task responds to incentives and that subjects respond to changes in incentives associated with a particular treatment. If the first part of the hypothesis is not true (or is underpowered statistically) inferences made on the second part will not be valid.

3. EXPERIMENTAL DESIGN

Our experiments were conducted at the Pittsburgh Experimental Economics Laboratory, using subjects recruited from the student population, randomly assigned to one of three possible treatments.⁶ Using a between-subject design the piece rate is held constant throughout an experimental session, so that each subject i receives a fixed payment per slider of $w_i \in \{0.5\text{¢}, 2.0\text{¢}, 8.0\text{¢}\}$.⁷ After instructions on the nature of the task, each session began with a two-minute practice round for subjects to become familiar with the slider task. This was followed by 10 paying rounds, each of which lasted two minutes. In each round, subjects saw a single screen displaying 48 sliders of equal length and offset from one another, as per G&P.⁸ At the end of each round there was a 10 second break during which subjects were reminded of how many rounds they had completed, the number of sliders completed, (Effort_{it}) and their corresponding earnings from that round ($w_i \cdot \text{Effort}_{it}$).

Once the ten paying rounds had concluded, subjects were asked to complete a survey.⁹ Only after completing the survey were respondents informed of their total earnings for the session. Subjects were then privately paid, receiving a \$10 participation payment on top of their earnings across the ten rounds $W_i = \sum_{t=1}^{10} (w_i \cdot \text{Effort}_{it})$.¹⁰

In order to measure the extent to which the slider task responds to incentives, our paper’s design adheres closely to that employed in G&P. There are four main differences: i) The G&P design is within subject, where ours is between subject. ii) G&P examine a game between two randomly matched subjects over a variable prize; ours examines a decision problem, removing any externalities over payment. iii) The marginal incentives in G&P work through a probability of winning a prize, where each additional slider completed leads to a one percent increase in the probability of winning a prize; in our experiment the marginal incentives work through a fixed piece rate per slider completed. iv) In G&P peer effects may be present, as subjects observe the other player’s effort at the end of each round; in our study there is no feedback on others’ effort levels.

⁶For consistency, one single member of the project read the instructions for all experimental sessions, and was assisted by another fixed experimenter. All data was collected over the course of two weeks in April of 2015, where all treatments were gender-balanced and interspersed across the data collection period. Initially, a total of three sessions were planned for each treatment; however, a computer error led to subjects’ terminals freezing in one round in one session. Another session was therefore added to have three complete sessions for each treatment.

⁷The effective marginal incentives for a risk-neutral subject in G&P varied within a session between 0.15¢ and 6.2¢ with an average of 3.1¢.

⁸The experiment was programmed in z-Tree (Fischbacher, 2007), and the program KeyTweak was used to disable all arrow keys on the keyboard, thereby ensuring that subjects only used the mouse to complete the slider tasks.

⁹Data from the survey available from the authors by request.

¹⁰Subjects in our 0.5¢ treatment had their final payoff W_i rounded up to the nearest whole cent.

4. RESULTS

Our experimental results are all provided in Table 2, where the first part reports summary statistics, and the second part reports regression results. Table 2(A) reports the average number of sliders completed (effort), the minimum and maximum effort levels for any subject-round, the total number of subjects N , and the effective average hourly wage rate (as the incentivized part lasts 20 minutes, this is simply $3 \cdot W_i$). On average, subjects across all of our experiments complete 26.7 sliders in each two-minute period. The lowest number of sliders solved by a subject in any round is ten, where the highest is 46 (two away from the 48 possible). Broken out across treatments, we see that effort increases with the piece rate: the average effort is 26.1 for the lowest incentive of 0.5¢, somewhat higher at 26.6 in the middle incentive, and at its highest of 27.4 in the 8¢ incentive.

Just from the averages in Table 2(A) it is apparent that the size of the incentive effect we observe in our data is small: going from a piece-rate of 0.5¢ to 2¢ leads to a half slider increase, and from 2¢ to 8¢ yields a 0.7 slider increase. Though the range of our incentives represents a substantial increase—from an effective hourly rate of about half the US Federal minimum to just over \$65 an hour¹¹—the average additional effort from this sixteen-fold increase in the incentive is just a single additional slider.

Across treatments and sessions, we observe substantial learning. Figure 1 presents the round averages for each of three treatments (where we have additionally provided bars indicating 95 percent confidence intervals, given subject variation). In round one, the average effort is 24.2 in both the 0.5¢ and 2¢ treatments, and 24.9 in the 8¢ treatment, though the variation across subjects is large. Across the session, the effort mostly increases, so that the final effort levels in round ten are 28.6 in the 0.5¢ treatments and 28.9 in both the 2¢ and 8¢ treatments. While the effort exerted in each treatment appears ordered according to incentives, it is noteworthy that the incentive order is only fully observed in six of the ten rounds.

To more exactly quantify the effects from incentives while controlling for learning and subject-level variation, we run the following regression

$$(1) \quad \text{Effort}_{it} = \beta \cdot \left(\frac{w_i - 0.5}{8 - 0.5} \right) + \sum_{s=2}^{10} \delta_s \cdot 1_{s=t} + \eta + u_i + \epsilon_{it},$$

where u_i is a subject-level random-effect, and ϵ_{it} an idiosyncratic error. The regressions include the treatment as a right-hand-side variable, rescaling the marginal incentive to run linearly from zero to one (0.5¢ at the low end, 8¢ at the high, with the 2¢ marginal taking the intermediate value 0.2), and additionally adds nine period dummies as regressors, $\{\delta_t\}_{t=2}^{10}$ and a constant η . The main content in Table 2(B) is the estimates for the incentive effect $\hat{\beta}$, the initial effort level $\hat{\eta}$ at the beginning of the session, and the average amount of learning across the sessions $\hat{\delta}_{10}$. In

¹¹By way of comparison, the average lawyer makes an hourly wage of \$64.17 according to the Bureau of Labor Statistics, while the average financial manager makes \$62.61.

TABLE 2. Result Tables

(A) Summary Statistics

Treatment	Effort			N	Hourly Rate
	Avg.	Min	Max		
0.5¢	26.1	12	41	42	\$3.92
2¢	26.6	14	46	43	\$15.95
8¢ [†]	27.3	10	46	63	\$65.46
	(27.4)	(14)	(46)	(45)	(\$65.68)
Total	26.7	10	46	148	

(B) Random-effect regression

Variable	Coef.	Std. Err	p-value
Incentive Effect, β	1.05 [†]	0.65	0.109
Initial Effort, $\hat{\eta}$	23.98	0.51	0.000
Learning Effect, $\hat{\delta}_{10}$	4.34	0.48	0.000
Between Std. Dev., $\hat{\sigma}_u$	3.47	0.29 [‡]	
Within Std. Dev., $\hat{\sigma}_\epsilon$	2.77	0.12 [‡]	

Note: †- A programming error led to one round in a single 8¢ marginal session hanging, so that data was not recorded for that round. Results are similar if we remove the entire session from our analysis (in parentheses below this line) and so we conduct our analysis on the entire data excluding this one round. The estimated incentive effect in Table 1(B) when dropping this session is 1.16 sliders, where the test for significance has a p -value of 0.107. ‡- Standard errors for between and within standard deviations are drawn from a bootstrap of size 1,000 that resamples across subjects, then subject-rounds.

addition, the table estimates the between-subject standard deviation, $\hat{\sigma}_u$, as 3.5 sliders; while the within-subject standard deviation, $\hat{\sigma}_\epsilon$, is estimated to be 2.8 sliders.

Unsurprisingly, given the overall averages in Table 2(A), the estimated incentive effect is close to one slider. Once we have controlled for variation between and within subjects, as well as the across-session learning, the response to incentives is only marginally significant. Interestingly, even our participants appear to be aware that their performance is not motivated by the payment they received. On a survey at the end of the experiment, we find that three-quarters of the participants do not think that there is any lower piece-rate payment at which they would decrease their performance.

Despite a sixteen-fold increase in the piece-rate, the observed increase in effort is small relative to other variations for the task. In terms of heterogeneity in natural ability, a one slider increase represents under a third of a between-subject standard deviation. In terms of idiosyncratic variation, it represents slightly over a third of a standard deviation. Across the entire session subjects

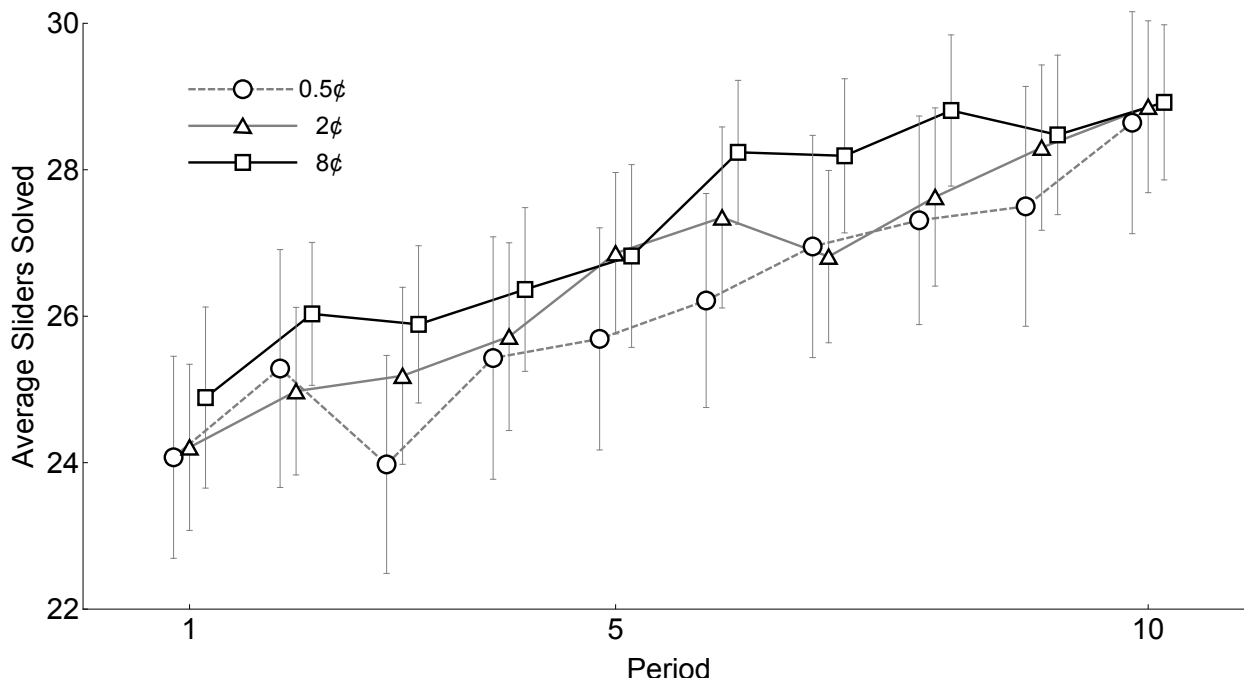


FIGURE 1. Effort across rounds

seem to learn to complete more than four additional sliders, relative to their effort in round one. The observed incentive effect therefore represents less than a quarter of the average learning effect.

To get a better sense for the subject-level heterogeneity in the data, Figure 2 provides a plot of the subject-level averages, expressed as a pdf. The figure is displayed using a kernel-regression (using a triangular kernel of bandwidth 2) to make comparison of the distributions easier to see. A comparison of the gray dotted line and the gray solid line represent a 300 percent increase in the incentives (0.5¢ to 2¢, respectively), while comparing the solid gray and solid black lines represents another 300 percent increase (from 2¢ to 8¢). The median subject in each distribution is indicated with a small circle on the relevant pdf, where we have additionally indicated the 95 percent confidence interval for the median (calculated through resampling). Continuing to look at the median, a one-sided Mann-Whitney test fails to reject equality of medians when comparing the 0.5¢ and 2¢ treatments ($p = 0.239$), and when comparing the 2¢ and 8¢ treatments ($p = 0.101$).¹² However, we do reject in favor of greater effort when comparing medians of the 0.5¢ and 8¢ treatments ($p = 0.024$); thus, using non-parametric tests the response to incentives is significant.¹³ Nonetheless, the magnitude of the effect over the median subject is economically small (1.6 sliders,

¹²Looking at the median subject's response is partly motivated by the fact that there is very little difference in the right-tails of the three distributions. The top quarter of subjects in each treatment are fairly consistently spread across averages between 30 and 40 sliders.

¹³Of the three pairwise comparisons only the 0.5¢ and 8¢ distributions are significantly different from one another using a Kolmogorov-Smirnoff test for equality of distribution ($p = 0.004$).

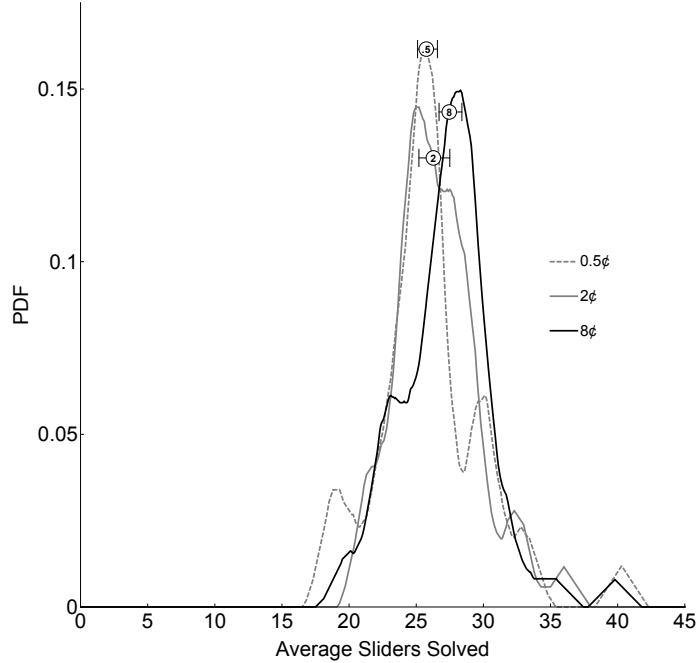


FIGURE 2. Average Subject Effort

Note: Each pdf recovered from a kernel regression with a triangular kernel (bandwidth of 2 sliders).

a six percent increase) while the change in incentives is large (1,500 percent, an order of magnitude larger).

5. DISCUSSION

Our between-subject design finds only a small effect of a single slider, for a large increase in the incentives. We now examine how this result compares to that of G&P.

In G&P, two players i (a first mover) and j (a second mover) are randomly matched and compete to win a common prize of size $100 \cdot w_{it}$ cents, drawn randomly from an interval. The probability of player i winning the prize is given by $\frac{1}{100} (50 + \text{Effort}_{it} - \text{Effort}_{jt})$, so for a risk-neutral participant the expected marginal incentive is w_{it} .¹⁴ The sequencing of the game was such that the first mover's effort (Effort_{it}) was observed by the second mover j , and the second mover's response was the main focus in G&P. In looking at the response to incentives, we follow Gill and Prowse (2015) and look only at the first movers.

As noted earlier, the first mover's task in G&P is different from that in our study: i) the session has within-subject variation over the incentive w_{it} ; ii) the tournament structure has own effort

¹⁴The raw prizes in G&P are drawn uniformly over $\{\pounds 0.10, \pounds 0.20, \dots, \pounds 3.90\}$. We transform these to expected marginal incentives for a risk-neutral agent, and then convert to US cents at a conversion rate of $\pounds 0.65 = 100\text{c}$.

TABLE 3. Random Effects Regression: G&P first mover

(A) Full Sample ($N = 60$)			
Variable	Coef.	Std. Err	p-value
Incentive Effect, $\tilde{\beta}^\dagger$	3.26	0.75	0.000
Initial Effort, $\tilde{\eta}$	21.11	0.89	0.000
Learning Effect, $\tilde{\delta}_{10}$	4.35	0.71	0.000
Between Std. Dev., $\tilde{\sigma}_u$	5.40	0.66 [‡]	
Within Std. Dev., $\tilde{\sigma}_\epsilon$	3.87	0.49 [‡]	

(B) Sub-sample ($N = 54$)			
Variable	Coef.	Std. Err	p-value
Incentive Effect, $\tilde{\beta}$	2.67	0.65	0.000
Initial Effort, $\tilde{\eta}$	22.70	0.72	0.000
Learning Effect, $\tilde{\delta}_{10}$	4.24	0.62	0.000
Between Std. Dev., $\tilde{\sigma}_u$	3.91	0.34 [‡]	
Within Std. Dev., $\tilde{\sigma}_\epsilon$	3.22	0.32 [‡]	

Note: †-Marginal incentive for G&P first movers is calculated as $(\frac{1\zeta}{40.65}\text{Prize}_{it} - 0.5\zeta)/(8\zeta - 0.5\zeta)$. Because of this $\tilde{\eta}$ has the interpretation of average effort in round one at a 0.5ζ incentive. ‡- Standard errors for between and within standard deviations are drawn from a bootstrap of size 1,000 that resamples across subjects, then subject-rounds.

inflicting a negative externality on the other player; iii) payment is incentivized only probabilistically; and iv) there is feedback on other participants' effort levels. Despite these differences, it is still of interest to compare the magnitudes of the incentive effect in the two studies.

Paralleling the regression results in Table 2(B), Table 3(A) provides a random-effects regression per the estimation equation (1) for the $N = 60$ G&P first-movers.¹⁵ The estimated incentive effect $\tilde{\beta}$ reflects the G&P prediction for the size of the effect in our experiment: a 3.26 sliders difference as the marginal incentive is raised from 0.5ζ to 8ζ . While our own incentive estimate $\hat{\beta}$ from Table 2(B) is only marginally significant when compared to zero, it is significantly different from the G&P estimate $\tilde{\beta}$ ($p = 0.000$).

Part of the very high incentive effect in G&P stems from a number of first-mover subjects who exert very low effort across the ten rounds of their experiment. There could be several reasons for exerting very low efforts, though one possibility that exists in G&P but not in our study is that subjects might be trying to pick the efficient outcome (both exerting zero effort and equally

¹⁵To distinguish between estimates on our data and G&P's we will use the notation $\hat{\beta}$, $\hat{\eta}$, etc., for estimates from our data, and $\tilde{\beta}$, $\tilde{\eta}$, etc., for estimates from the G&P data.

splitting the chance to win the prize).¹⁶ As a partial control for this, we re-ran the random-effects regression (1) where we exclude the G&P first-movers whose average effort across the ten rounds is *lower than the lowest subject average* in our between-subject data (18.5 sliders, from the 0.5¢ treatment). This excludes six subjects, representing ten percent of the G&P first mover subjects.¹⁷

The regression results for the subsample of $N = 54$ subjects are given in Table 3(B). Though the estimated incentive effect is lower—a 2.67 slider increase—our estimate is still significantly different ($p = 0.012$). Despite these large differences in the estimated incentive effects, the other regression coefficients are remarkably similar.

Looking at the Table 3(B)’s results (where we remove subjects in the left tail of the distribution), and comparing them to our results in Table 2(B), we find many commonalities. First, subjects on average get better across the sessions by approximately four sliders ($\hat{\delta}_{10}$ and $\tilde{\delta}_{10}$ are not significantly different).¹⁸ Second, though the initial effort level estimates of η are significantly higher in our sessions at 24 sliders in comparison to 22.7 in G&P, the size of the difference is quantitatively small.¹⁹ Third, between- and within-subject standard deviations for effort after controlling for the incentive effects (σ_u and σ_e , respectively) are very similar, though in both cases the estimated variation in our experiments is smaller than in G&P. The large estimated variation in Table 3(A) is mostly driven by the six subjects with unusually low average efforts.

Comparing our results to those of G&P, it is hard not to attribute the majority of the observed incentive effect to some combination of a within-subject effect (potentially a demand effect, potentially peer effects) and a strategic or social effect (with the negative externality potentially pushing subjects to exert low effort).

6. CONCLUSION

Using a between-subject design, we examine how performance on the slider task responds to changes in monetary incentives. Despite a sixteen fold increase in incentives, we find only a modest increase in effort—less than four percent for the average subject.

While our results point to the standard version of the slider task being poorly suited for the examination of response to incentives in between-subject designs, two recent studies point to techniques that can help regain control of the effort cost. Gächter et al. (2015) introduce a ball-catching task where the cost of effort is directly manipulated and they find in a within-subject design that there is a strong response to incentives when effort is costly. Similarly Eckartz (2014) finds that with a

¹⁶Gill & Prowse (2015) note that 2 subjects (whom we will also exclude) appear to have difficulty positioning sliders at exactly 50 until a few rounds into the session.

¹⁷Note that only subjects with low average performance are eliminated from the data. Data from subjects with particular rounds with less than 19 sliders completed are still included in the analysis, provided that the subject’s average across the session is above 18.5 sliders.

¹⁸All three of our treatments, as well as both movers in G&P show fairly consistent increases in average efforts across the session.

¹⁹A joint regression across both sets of data indicates no significant difference over the two constants ($p = 0.123$).

paid outside option, performance changes when presented with a flat-rate rather than a piece-rate incentive. While these types of modifications may be seen as more abstract, bringing the real-effort designs back toward those with stated-effort, for short and repeated tasks such modifications may be needed to secure a more powerful response to incentives. Certainly, as the slider task stands, the limited sensitivity to incentives renders it poorly suited for studying responses to changes in monetary or non-monetary incentives.

REFERENCES

- Abeler, Johannes and Simon Jäger**, “Complex Tax Incentives,” *American Economic Journal: Economic Policy*, forthcoming.
- Banuri, Sheheryar and Philip Keefer**, “Intrinsic motivation, effort and the call to public service,” 2013. World Bank Policy Research Working Paper 6729.
- Besley, Timothy and Maitreesh Ghatak**, “Profit with purpose? a theory of social enterprise with experimental evidence,” 2013. Economic Organisation and Public Policy Discussion Papers, EOPP 047.
- Bradler, Christiane, Susanne Neckermann, and Arne Jonas Warnke**, “Rewards and performance: A comparison across a creative and a routine task,” 2015.
- Brown, Alexander L, Jonathan Meer, and J Forrest Williams**, “Why do people volunteer? An experimental analysis of preferences for time donations,” 2013. NBER Working Paper w19066.
- Bull, Clive, Andrew Schotter, and Keith Weigelt**, “Tournaments and piece rates: An experimental study,” *Journal of Political Economy*, 1987, pp. 1–33.
- Buser, Thomas and Anna Dreber**, “The Flipside of Comparative Payment Schemes,” 2014. Tinbergen Institute Discussion Paper 13-190/I.
- Carpenter, Jeffrey, Peter Hans Matthews, and John Schirm**, “Tournaments and Office Politics: Evidence from a Real Effort Experiment,” *American Economic Review*, 2010, 100 (1), 504–17.
- Charness, Gary and Marie-Claire Villeval**, “Cooperation and Competition in Intergenerational Experiments in the Field and the Laboratory,” *American Economic Review*, 2009, 99 (3), 956–978.
- _____ **and Peter Kuhn**, “Lab labor: What can labor economists learn from the lab?,” *Handbook of Labor Economics*, 2011, 4, 229–330.
- Dickinson, David and Marie-Claire Villeval**, “Does monitoring decrease work effort?: The complementarity between agency and crowding-out theories,” *Games and Economic behavior*, 2008, 63 (1), 56–76.
- Doerrenberg, Philipp and Denvil Duncan**, “Experimental evidence on the relationship between tax evasion opportunities and labor supply,” *European Economic Review*, 2014, 68, 48–70.

- _____, _____, and **Christopher Zeppenfeld**, “Circumstantial risk: Impact of future tax evasion and labor supply opportunities on risk exposure,” *Journal of Economic Behavior & Organization*, 2015, 109, 85–100.
- Dolan, Paul, Robert Metcalfe, and Daniel Navarro-Martinez**, “Financial incentives and working in the education sector,” 2012. Department for Education Research Report DFE–RR251.
- Eckartz, Katharina**, “Task enjoyment and opportunity costs in the lab: The effect of financial incentives on performance in real effort tasks,” 2014. Jena Economic Research Papers 2014-005.
- Erkal, Nisvan, Lata Gangadharan, and Nikos Nikiforakis**, “Relative earnings and giving in a real-effort experiment,” *American Economic Review*, 2011, pp. 3330–3348.
- Evdokimov, Piotr and David Rahman**, “Personality and Bargaining Power,” 2014.
- Feess, Eberhard, Oliver Lorz, and Markus Schramm**, “Under-Delegation in Case of Pure Moral Hazard: Evidence from the Laboratory,” 2014.
- Fehr, Ernst, Georg Kirchsteiger, and Arno Riedl**, “Does fairness prevent market clearing? An experimental investigation,” *Quarterly Journal of Economics*, 1993, pp. 437–459.
- Fischbacher, Urs**, “z-Tree: Zurich toolbox for ready-made economic experiments,” *Experimental economics*, 2007, 10 (2), 171–178.
- Friedl, Andreas, Levent Neyse, and Ulrich Schmidt**, “Payment Scheme Changes and Effort Provision: The Effect of Digit Ratio,” 2014. MPRA Paper 59549.
- Gächter, Simon, Lingbo Huang, and Martin Sefton**, “Combining “Real Effort” with Induced Effort Costs: The Ball-Catching Task,” May 2015. IZA Discussion Paper 9041.
- Georganas, Sotiris, Mirco Tonin, and Michael Vlassopoulos**, “Peer Pressure and Productivity: The Role of Observing and Being Observed,” 2014. CESifo Working Paper Series 4572.
- Gerhards, Leonie and Neele Siemer**, “Private versus Public Feedback-The Incentive Effects of Symbolic Awards,” 2014.
- Gill, David and Victoria L Prowse**, “A novel computerized real effort task based on sliders,” 2015.
- _____ and **Victoria Prowse**, “A structural analysis of disappointment aversion in a real effort competition,” *American Economic Review*, 2012, 102 (1), 469–503.
- Gneezy, Uri, Muriel Niederle, Aldo Rustichini et al.**, “Performance in competitive environments: Gender differences,” *Quarterly Journal of Economics*, 2003, 118 (3), 1049–1074.
- _____, **Stephan Meier, and Pedro Rey-Biel**, “When and Why Incentives (Don’t) Work to Modify Behavior,” *Journal of Economic Perspectives*, 2011, 25 (4), pp. 191–209.
- Hammermann, Andrea, Alwine Mohnen, and Petra Nieken**, “Whom to choose as a team mate? A lab experiment about in-group favouritism,” 2012. IZA Discussion Paper 6286.
- Ibanez, Marcela and Elke Schaffland**, “The Effect of Outside Leaders on the Performance of the Organization: An Experiment,” 2013. Courant Research Centre: Poverty, Equity and Growth -

Discussion Papers 149.

- Imas, Alex**, “Working for the “warm glow”: On the benefits and limits of prosocial incentives,” *Journal of Public Economics*, 2014, 114, 14–18.
- _____, **Sally Sadoff**, and **Anya Savikhin Samek**, “Do People Anticipate Loss Aversion?,” *Management Science*, forthcoming.
- Kesternich, Martin, Andreas Lange, and Bodo Sturm**, “The impact of burden sharing rules on the voluntary provision of public goods,” *Journal of Economic Behavior & Organization*, 2014, 105, 107–123.
- Khadjavi, Menusch**, “On the interaction of deterrence and emotions,” *Journal of Law, Economics, and Organization*, 2014, p. ewu012.
- Lindner, Florian**, “Decision time and steps of reasoning in a competitive market entry game,” *Economics Letters*, 2014, 122 (1), 7–11.
- Nalbantian, Haig R and Andrew Schotter**, “Productivity under group incentives: An experimental study,” *American Economic Review*, 1997, pp. 314–341.
- Niederle, Muriel and Lise Vesterlund**, “Do Women Shy Away from Competition? Do Men Compete Too Much?,” *Quarterly Journal of Economics*, 2007, pp. 1067–1101.
- Schotter, Andrew and Keith Weigelt**, “Asymmetric tournaments, equal opportunity laws, and affirmative action: Some experimental results,” *Quarterly Journal of Economics*, 1992, pp. 511–539.
- Straub, Tim, Henner Gimpel, Florian Teschner, and Christof Weinhardt**, “How (not) to Incent Crowd Workers,” *Business & Information Systems Engineering*, 2015, pp. 1–13.
- Swenson, Charles W**, “Taxpayer behavior in response to taxation: An experimental analysis,” *Journal of Accounting and Public Policy*, 1988, 7 (1), 1–28.