Measuring costly effort using the slider task*

David Gill†, Victoria Prowse‡

November 13, 2018

Abstract

Using real effort to implement costly activities increases the likelihood that the motivations that drive effort provision in real life carry over to the laboratory. However, unobserved differences between subjects in the cost of real effort make quantitative prediction problematic. In this paper we present the *slider task*, which was designed by us to overcome the drawbacks of real-effort tasks. The slider task allows the researcher to collect precise and repeated observations of effort provision from the same subjects in a short time frame. The resulting high-quality panel data allow sophisticated statistical analysis. We illustrate these advantages in two ways. First, we show how to use panel data from the slider task to improve precision by controlling for persistent unobserved heterogeneity. Second, we show how to estimate effort costs at the subject level by exploiting within-subject variation in incentives across repetitions of the slider task. We also provide z-Tree code and practical guidance to help researchers implement the slider task.

Keywords: Experimental methodology; real effort; effort provision; cost of effort; slider task; design of laboratory experiments; unobserved heterogeneity.

JEL Classification: C91; C13.

*This paper replaces an earlier working paper, titled "A Novel Computerized Real Effort Task Based on Sliders," that did not include any estimation of the cost of effort function. Financial support from the George Webb Medley Fund at Oxford University and a Southampton School of Social Sciences Small Grant is gratefully acknowledged. We also thank the Nuffield Centre for Experimental Social Sciences for hosting our experiments.

†Department of Economics, Purdue University, gill53@purdue.edu

‡Department of Economics, Purdue University, vprowse@purdue.edu

1 Introduction

Laboratory experiments are a powerful tool for understanding drivers of agents' behavior and for testing the predictions of economic theories. In particular, laboratory experiments are often used to study how much effort subjects exert in a costly activity. For example, subjects may choose how much effort to exert when competing in a tournament (Bull et al., 1987), when producing output as part of a team (van Dijk et al., 2001), when responding to the wages set by an employer (Fehr et al., 1997), and when earning endowments that form the starting point for a bargaining game (Burrows and Loomes, 1994).

There are two ways of implementing costly activities in a laboratory experiment: via a monetary cost function that mimics effort by specifying output as a function of how much money the subject contributes (e.g., Bull et al., 1987); and using a real-effort task. The monetary cost function allows the experimenter full control over the cost of effort.[1] Increasingly, laboratory experiments have featured real-effort tasks, such as: (i) solving mazes (Gneezy et al., 2003), mathematical problems (Sutter and Weck-Hannemann, 2003; Niederle and Vesterlund, 2007) or word games (Burrows and Loomes, 1994); (ii) answering general knowledge questions (Hoffman et al., 1994); (iii) counting (Abeler et al., 2011), decoding (Chow, 1983), encrypting (Erkal et al., 2011) or entering (Dickinson, 1999) strings of characters; (iv) performing numerical optimization (van Dijk et al., 2001); and (v) filling envelopes (Konow, 2000), cracking walnuts (Fahr and Irlenbusch, 2000), or other physical tasks.

The main advantage of using a real-effort task over a monetary cost function is the greater external validity of the experiment: exerting actual effort makes the environment more realistic and less sterile, increasing the likelihood that the motivations that drive behavior outside the laboratory carry over to the laboratory. Real-effort tasks, however, suffer from two critical drawbacks. First, the cost of effort provision varies from subject to subject and is unknown to the experimenter. As Falk and Fehr (2003, p. 404) explain "Since the experimenter does not know the workers' effort cost, it is not possible to derive precise quantitative predictions." Second, since motivation and ability to complete the real-effort task vary considerably across subjects, data collected using real-effort tasks are noisy and, therefore, statistical analyses of real-effort data often lack precision. Charness et al. (2018) compare the stated- and real-effort experimental methodologies in more detail.

In this paper we present the *slider task*, which is a novel and simple computerized real-effort task that was designed by us specifically to overcome the drawbacks of real-effort tasks. In particular, the slider task provides a finely gradated measure of effort within a short time frame. The slider task thus allows the researcher to collect precise and repeated observations of effort provision by the same subjects. The resulting high-quality panel data on repeated effort choices allow sophisticated statistical analysis of effort provision in the laboratory using panel data methods. We illustrate these advantages in two ways. First, we show how to use panel data from the slider task to improve precision by controlling for persistent unobserved heterogeneity. Second, we show how to estimate effort costs at the subject level by exploiting within-subject

[1] In particular, the experimenter can control the extent of any convexity in the cost of the activity, and can also determine how the cost varies over individuals and over any repetitions of the task.

variation in incentives across repetitions of the slider task.

The slider task, first developed and used by us in Gill and Prowse (2012) to study disappointment aversion, consists of a single screen containing a number of "sliders" that subjects move to a specified position within an allotted time. One repetition of the slider task takes only 120 seconds and measures effort choices varying from 0 units to over 40 units. The slider task is now well established as a tool for experimental economists. Since its inception, the slider task has been used to study contract law (Depoorter and Tontrup, 2012), tax compliance (Fonseca and Myles, 2012; Doerrenberg et al., 2015), cheating in the workplace (Gill et al., 2013), gender differences in competition (Gill and Prowse, 2014), tax complexity (Abeler and Jäger, 2015), outside options (Goerg et al., forthcoming), downsizing (Drzensky and Heinz, 2016), social enterprises (Besley and Ghatak, 2017), volunteering (Brown et al., 2013), peer pressure (Georganas et al., 2015), social insurance (Ahlquist et al., 2014), delegation (Feess et al., 2014) and creativity (Bradler et al., 2015), among others.[2]

This paper describes the properties and advantages of the slider task, while also noting disadvantages of the task that might reduce its usefulness in particular research contexts. In so doing, we provide context for studies that have used the slider task, and we provide direction to experimental economists who are considering how to implement costly activities in their own experiments. In particular, the contribution of this paper is four-fold. First, we explain how the slider task overcomes many of the limitations of real-effort tasks. Second, we show that panel data collected from the slider task are of sufficient quality that they can be used to estimate precisely how subjects respond to financial incentives.[3] Third, we demonstrate how repeated observations of effort choices in the slider task can be used to recover information about each subject's cost of effort function.[4] Fourth, we provide some practical guidance for experimental economists who want to use the slider task in their own real-effort experiments.

The paper proceeds as follows: Section 2 outlines the design of our slider task; Section 3 details advantages and disadvantages of the slider task; Section 4 describes the data; Section 5 reports estimates of how effort in the slider task responds to financial incentives; Section 6 shows how data collected from the slider task can be used to estimate effort costs for each subject; Section 7 contains a practical guide for researchers wishing to implement the slider task; and Section 8 concludes. The z-Tree code that accompanies this paper is available online and can

[2]The slider task has also been used in other studies including Hetzel (2010), Eacret et al. (2011), Hammermann et al. (2011), Dolan et al. (2012), Kimbrough and Reiss (2012), Monahan (2012), Breuer (2013), Das et al. (2013), Djawadi and Fahr (2013), Fahr and Djawadi (2013), Ibanez and Schaffland (2013), Riener and Wiederhold (2013), Djawadi et al. (2014), Doerrenberg and Duncan (2014a), Doerrenberg and Duncan (2014b), Evdokimov and Rahman (2014), Fan and Zhao (2014), Friedl et al. (2014), Gärtner and Sandberg (2014), Gerhards and Siemer (2014), Khadjavi (2014), Kirchler et al. (2014), Lezzi (2014), Lindner (2014), Schmitz (2014), Straub et al. (2014), Teng et al. (2014), Avram (2015), Bonein and Denant-Boèmont (2015), Zeppenfeld (2015), Cettolin and Riedl (2016), Dorrough et al. (2016), Imas et al. (2016), Riener and Wiederhold (2016), Smith (2016), Brown et al. (2017), Gee et al. (2017), Meissner and Pfeiffer (2017), Kesternich et al. (2018) and Malik et al. (2018).

[3]To be clear, although the magnitude of the effect of the prize on effort provision is of independent interest, the contribution is to illustrate how the slider task helps to provide precise estimates of the effects of incentives on effort provision, rather than to provide a specific estimate in the particular context of the dataset that we study. See Section 5 for more details.

[4]Gill and Prowse (2012) embed heterogeneous effort costs within a more complex structural model, but they do not recover individual-level estimates of the cost of effort.

be used to create real-effort laboratory experiments featuring the slider task.[5]

2 Design of the slider task

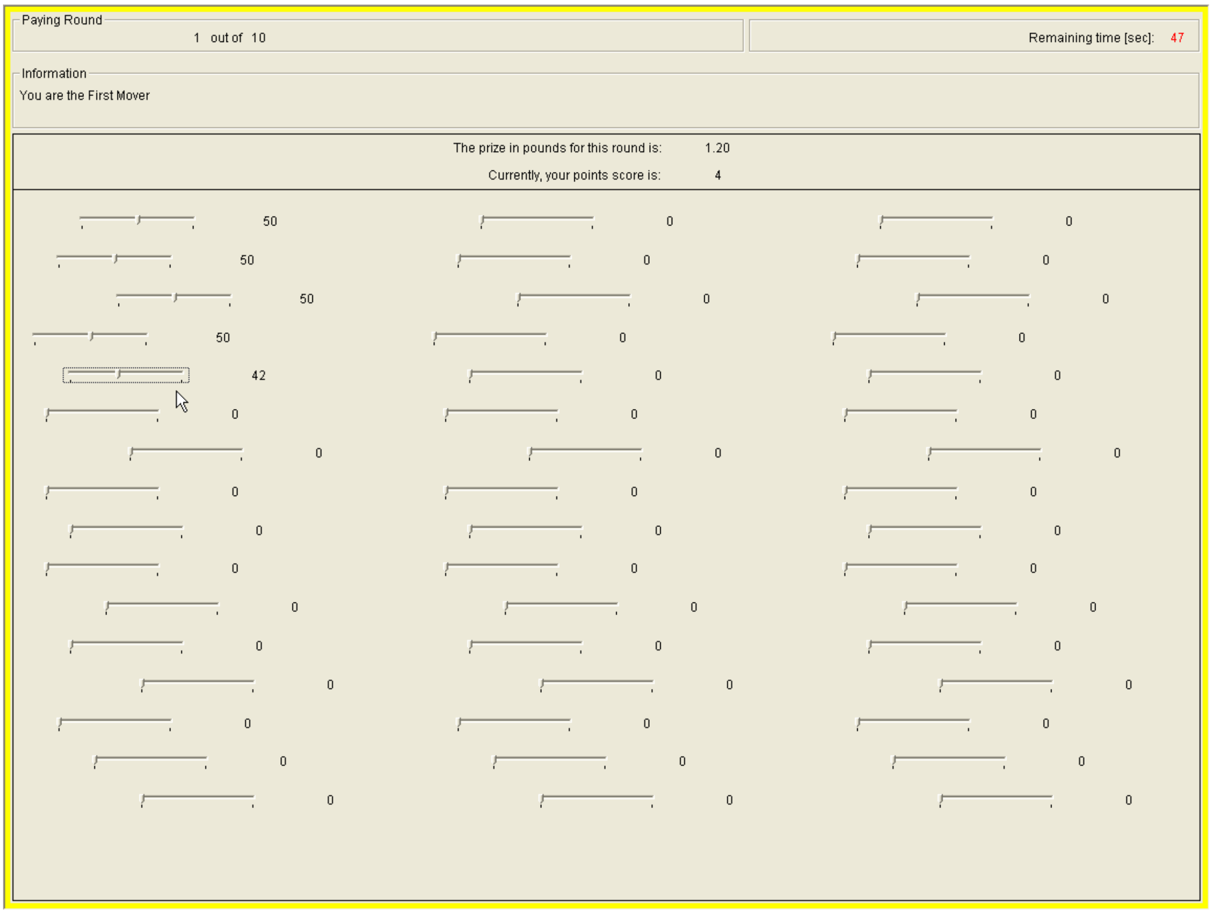
Our novel and simple real-effort task consists of a single screen displaying a number of "sliders" programmed in z-Tree (Fischbacher, 2007). This screen does not vary across experimental subjects or across repetitions of the task. A schematic representation of a single slider is shown in Figure 1. When the screen containing the effort task is first displayed to the subject all of the sliders are positioned at 0, as shown for a single slider in Figure 1(a). By using the mouse, the subject can position each slider at any integer location between 0 and 100 inclusive. Each slider can be adjusted and readjusted an unlimited number of times, and the current position of each slider is displayed to the right of the slider. The subject's "points score" in the task, interpreted as effort exerted, is the number of sliders positioned at 50 at the end of the allotted time. Figure 1(b) shows a correctly positioned slider. As the task proceeds, the screen displays the subject's current points score and the amount of time remaining.

(a) Initial position. (b) Positioned at 50.

Figure 1: Schematic representation of a slider.

Figure 2 shows a screen containing 48 sliders, as shown to the subject in the laboratory in Gill and Prowse (2012). In this example, the subject has positioned four of the sliders at 50 and a points score of 4 is shown at the top of the screen. A fifth slider is currently positioned at 42 and this slider does not contribute to the subject's points score as it is not correctly positioned. To ensure that all the sliders are equally difficult to position correctly, the 48 sliders are arranged on the screen such that no two sliders are aligned exactly one under the other. This prevents the subject being able to position the higher slider at 50 and then easily position the lower slider by copying the position of the higher slider. The number of sliders and task length can be chosen by the experimenter.

[5]The code is available from: `http://web.ics.purdue.edu/~vprowse/GillProwseSliderExample.ztt` and is also part of the Supplementary Materials that accompany this paper online. The code that we provide comes with no warranty or guarantee. In providing this code, we take no responsibility with regard to the use or modifications of the code.



Note: The screen presented here is slightly squarer than the one seen by our subjects.

Figure 2: Screen showing 48 sliders.

3 Advantages and disadvantages of the slider task

The slider task has a number of desirable attributes. First, the slider task is simple to communicate and to understand, and does not require or test pre-existing knowledge. Second, unlike solving mathematical problems, counting characters, solving anagrams, negotiating mazes or performing numerical optimization, the slider task is identical across repetitions. This feature reduces the noise in measured effort across repetitions of the slider task. Third, the task involves little randomness, and so the number of correctly positioned sliders corresponds closely to the effort exerted by the subject, again reducing noise. Fourth, there is no scope for guessing, which complicates the design and interpretation of some existing tasks such as those based on counting characters or numerical optimization.

These attributes are also shared by the envelope filling task, in which subjects stuff real envelopes with letters. Crucially, and in contrast to stuffing real envelopes, the slider task provides a finely gradated measure of effort within a short time frame. In Section 5 we see that with 48 sliders and an allotted time of 120 seconds, measured effort varies from 0 to over 40. By reducing measurement noise, the finely gradated measure of effort increases statistical precision. The ability to measure effort in a short time frame, meanwhile, makes it feasible for the subjects to repeat the identical task many times.

The resulting high-quality panel data on repeated effort choices allow sophisticated statistical analysis of effort provision in the laboratory using panel data methods. In Section 5, we show how to use panel data from the slider task to improve precision by controlling for persistent unobserved heterogeneity. In Section 6, we show how to estimate effort costs at the subject level by exploiting within-subject variation in incentives across repetitions of the slider task. The repeated observations of effort from the slider task can also be used to study the dynamics of effort over time: for example, Gill and Prowse (2014) study gender differences in how effort responds to the outcomes of earlier competitions. Furthermore, because the slider task is computerized, it allows flexible real-time subject interactions: Lee (2015) exploits this feature of the slider task to study the effect of real-time relative-performance feedback on effort provision.

Finally, we note potential disadvantages of the slider task, which might reduce its usefulness in particular research contexts. First, unlike the envelope filing task but like most of the real-effort tasks used in the literature (see Section 1), the output of the slider task has no intrinsic value. Thus the slider task is not appropriate in settings where the inherent usefulness of the real-effort task is particularly important. Second, the slider task requires concentration and dexterity rather than higher-order cognitive skills. Thus, the task is not appropriate in settings where the researcher wants to study demanding cognitive behavior like creativity. Third, men tend to perform better on the slider task (see Gill and Prowse, 2014), and so the slider task is not appropriate in settings where the researcher requires gender neutrality.

4 Description of the data

The data analyzed in this paper are those collected by Gill and Prowse (2012) to study disappointment aversion.[6] We used the slider task in six experimental sessions conducted at the Nuffield Centre for Experimental Social Sciences in Oxford. The slider task included 48 sliders (as shown in Figure 2) and the task length was 120 seconds.[7] We use the terms "points score" and "effort" interchangeably to denote the number of sliders correctly positioned by a subject at the end of the 120 seconds.[8] The experimental instructions can be found in the Supplementary Web Appendix (or in Gill and Prowse, 2018, an earlier version of this paper).

Twenty subjects participated in each of the six sessions. At the beginning of every session half the subjects were told that they would be a "First Mover" and the other half told they would be a "Second Mover" for the duration of the session. At the beginning of each round, every First Mover was anonymously paired with a new Second Mover using the no-contagion algorithm of Cooper et al. (1996). A prize for each pair was randomly chosen from $\{£0.10, £0.20, \ldots, £3.90\}$ and revealed to the pair members. The First and Second Movers then completed the slider task

[6]Gill and Prowse (2012) explain how the specific experimental design choices outlined below allow clean identification of disappointment aversion from the behavior of the Second Movers.

[7]The sliders were displayed on 22 inch widescreen monitors with a 1680 by 1050 pixel resolution. To move the sliders, the subjects used 800 dpi USB mice with the scroll wheel disabled.

[8]As is common in the experimental literature on real-effort provision, we use the term "effort" to correspond to measurable performance in a work task rather than the cost associated with work effort (see, e.g., Abeler et al., 2011, Gill and Prowse, 2012, and Charness et al., 2014).

sequentially, with the Second Mover discovering the points score of the First Mover she was paired with before starting the task. The prize was then awarded to one pair member based on the relative points scores of the two pair members and some element of chance (see Table SWA.1 in the Supplementary Web Appendix). In particular, a subject's chance of winning the prize increased by one percentage point with every unit of her own effort and decreased by the same amount for every unit of her opponent's effort. Furthermore, if the First Mover and her opponent exerted the same effort then each pair member had an equal chance of winning the prize.[9]

In total we have data on 60 First Movers and 60 Second Movers, each observed during 10 rounds. For the purposes of illustrating the properties of the slider task, this paper looks only at the behavior of the First Movers (Gill and Prowse, 2012, focus on the behavior of the Second Movers, in particular analyzing how the effort of Second Movers responds to that of the First Movers, to identify disappointment aversion among Second Movers).

Table 1 summarizes the observed efforts of the First Movers in each of the 10 rounds. We see that the mean points score tended to increase over the 10 rounds, from an average of 22.2 sliders in the first round to 26.3 sliders in the final round. Given that the average prize was constant over rounds, this increase in effort is interpreted as a learning-by-doing effect. The maximum observed effort was 41, and therefore it appears that no subject was able to position correctly all 48 sliders in 120 seconds. We conclude that efforts were not constrained by the upper limit imposed by the design of the task. There are seven observations of 0s. Of these, five correspond to two subjects who appear to have had difficulty positioning sliders at exactly 50 until a few rounds into the session. The remaining two observations of 0 correspond to a further two subjects who chose to exert no effort toward the end of their session in response to low prizes of £0.10 and £0.30.

Figure 3 shows the distribution of points scores. This figure was drawn using all 600 subject-round observations of points score from the First Movers in our experiment. We see a substantial amount of variation in effort provision. Specifically, a small cluster of subjects exert zero or very low effort in a particular round, two-thirds of efforts lie between 20 and 30 inclusive, while around 20% of efforts exceed 30. Thus, despite subjects having only 120 seconds to complete the slider task, we see large differences in effort provision.

We further explore the heterogeneity in effort provision by decomposing the total variation in effort over the 600 subject-round observations into a component due to between-subject differences in average effort provision in the experiment (pooling together data from all 10 rounds) and a component due to within-subject fluctuations in effort provision over the 10 rounds of the experiment. We find that 65.4% of the total variation in effort in the experiment is due to between-subject differences in effort provision that persist over the experiment, while the remaining 34.6% of the total variation in effort in the experiment arises from within-subject fluctuations in effort over rounds. We further decompose the persistent between-subject differences in effort

[9] In addition to any prizes accumulated during the experiment, all subjects were paid a show-up fee of £4. The subjects also initially played two practice rounds against an automaton for which they were not paid. We do not include the practice rounds in the data analysis, with the exception of the regression reported in Column 4 of Table 2.

provision in the experiment into between-subject differences in initial effort provision, i.e., differences in effort provision in the first round of the experiment, and between-subject differences in the evolution of effort during the experiment, which we attribute to learning. We find that 84.6% of the between-subject variation in effort provision is due to differences in initial effort, while the remaining 15.4% of the between-subject variation in effort provision is due to learning.

Round	Subjects	Mean Effort	Median Effort	Minimum Effort	Maximum Effort
1	60	22.20 (6.07)	23	1	33
2	60	22.68 (6.66)	23.5	0	33
3	60	24.80 (6.03)	25.5	0	37
4	60	24.61 (5.90)	25	0	35
5	60	25.18 (6.94)	26	0	38
6	60	24.66 (7.45)	26	1	37
7	60	25.91 (5.81)	26	9	37
8	60	26.88 (5.82)	27	9	41
9	60	25.65 (8.48)	28	0	38
10	60	26.31 (6.72)	27	1	40

Note: Standard deviations in parentheses.

Table 1: Summary of First Movers' efforts by round.

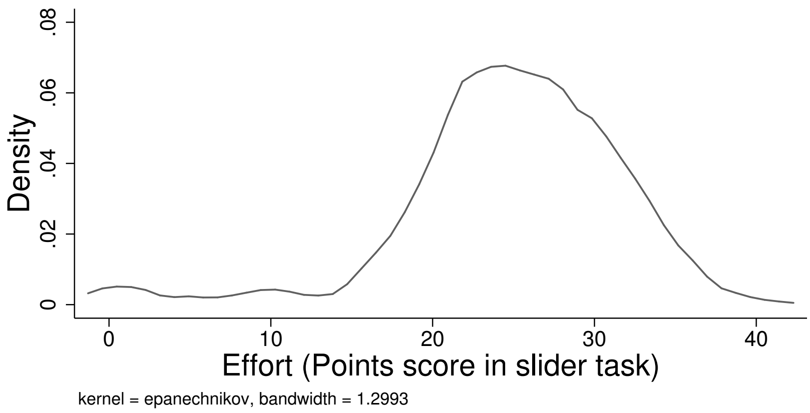


Figure 3: Distribution of First Movers' efforts.

5 Response to incentives in the slider task

In this section we use our panel data on repeated effort choices from the slider task to estimate how effort responds to financial incentives. We show that data generated by the slider task are of sufficient quality to estimate precisely the effect of monetary incentives on effort provision using panel data methods that control for unobserved heterogeneity.

Recall that the prize is announced at the start of each round, and a subject's chance of wining the prize increases with her effort in the round: a subject's financial incentive to exert effort therefore increases with the value of the prize for the round. Figure 4 provides initial evidence that effort provision in the slider task responds positively to financial incentives. The figure shows the results of a Lowess regression of the First Movers' efforts on the prize, estimated using all 600 subject-round observations of the First Movers in our experiment. As the prize increases from its lowest value of £0.10 to its highest value of £3.90, effort rises from about 22 sliders to about 26 sliders.

Figure 4: Response of First Mover's effort to the prize

Columns (1)–(3) of Table 2 present the results of a sequence of regressions of First Movers' efforts on the prize and round number. All regressions control for persistent unobserved heterogeneity by including subject fixed effects, which absorb subject-level differences in effort that persist over all 10 rounds of the experiment. Model (1) shows a positive and statistically significant round trend, which captures learning-by-doing. Model (2) includes a full set of round dummies.[10] Model (3) includes the prize as an additional explanatory variable. We see that the First Movers' efforts increase statistically significantly in the prize, with a £1 increase in

[10]The F statistic for the null hypothesis that the round trend is linear is 3.30, which corresponds to a p-value of 0.004. Thus round effects are non-linear. However, we are unable to reject linearity of the round effects starting from round 4: the F statistic for the null hypothesis that the round trend is linear from round 4 onward is 1.61 with a p-value of 0.160.

the prize causing an increase in effort of 0.7 of a slider.[11] In Column (4) of Table 2 we present the results of a regression of the proportional change in First Mover effort from the subject's average effort across the practice rounds on the prize and a full set of round dummies. We find that a £1 increase in the prize causes a 4.7% increase in effort and this effect is significant at the 1% level.

	(1)	(2)	(3)	(4)
	First Mover effort			Proportional change in First Mover effort
Prize	-	-	0.671*** (0.157)	0.047*** (0.011)
Round number	0.434*** (0.084)	-	-	-
Round 2	-	0.483 (0.557)	0.404 (0.540)	0.042 (0.038)
Round 3	-	2.600*** (0.545)	2.498*** (0.565)	0.159*** (0.046)
Round 4	-	2.417*** (0.562)	2.286*** (0.567)	0.166*** (0.056)
Round 5	-	2.983*** (0.694)	2.823*** (0.682)	0.215*** (0.059)
Round 6	-	2.467*** (0.797)	2.481*** (0.766)	0.154** (0.059)
Round 7	-	3.717*** (0.540)	3.694*** (0.516)	0.228*** (0.050)
Round 8	-	4.683*** (0.686)	4.676*** (0.682)	0.279*** (0.062)
Round 9	-	3.450*** (1.079)	3.482*** (1.044)	0.209*** (0.078)
Round 10	-	4.117*** (0.754)	4.355*** (0.767)	0.274*** (0.072)
Intercept	22.504*** (0.620)	22.200*** (0.466)	20.894*** (0.508)	0.276*** (0.058)
σ_α	5.494	5.494	5.491	0.447
σ_ϵ	3.971	3.938	3.873	0.256
Subject fixed effects	Yes	Yes	Yes	Yes
Subject-round observations	600	600	600	580
Subjects	60	60	60	58

Notes: The dependent variable in Columns (1)–(3) is the level of First Mover effort. The dependent variable in Column (4) is the proportional change in First Mover effort from the subject's average effort across the two practice rounds. The results in Column (4) exclude all subject-round observations from the two subjects with zero average effort across the practice rounds. σ_α denotes the standard deviation of the time-invariant subject-specific fixed effects and σ_ϵ is the standard deviation of the time-varying component of the subject-level error terms. Heteroskedasticity-robust standard errors with clustering at the subject level are reported in parentheses. *, ** and *** denote, respectively, significance at the 10%, 5% and 1% levels (two-sided tests).

Table 2: Fixed effects regressions for First Movers' efforts.

[11] We investigated the linearity of the prize effect by running a further regression (not reported) that also included the square of the prize. The results show that the square of the prize is not a statistically significant determinant of effort: the two-sided p-value for the coefficient on the square of the prize is 0.121. Effort therefore appears to increase linearly with the prize.

We explained in Section 3 that the main advantage of the slider task is that it generates high-quality repeated observations of effort in a short time frame. Table 3 explores the statistical precision afforded by the slider task when estimating the effect of monetary incentives on effort provision. Column (1) repeats the fixed effects regression from Model (3) of Table 2. Column (2) shows the results of an OLS regression on the full sample of 600 subject-round observations. Column (3) shows the results of an OLS regression using one randomly selected observation per subject.

	(1)	(2)	(3)
Prize	0.671^{***}	0.656^{***}	0.660
	(0.157)	(0.240)	(0.828)
Intercept	20.894^{***}	20.923^{***}	20.944^{***}
	(0.508)	(0.893)	(2.878)
Subject fixed effects	Yes	No	No
Round dummies	Yes	Yes	Yes
Subject-round observations	600	600	60
Subjects	60	60	60

Notes: Heteroskedasticity-robust standard errors with clustering at the subject level are reported in parentheses in Columns (1) and (2). Heteroskedasticity-robust standard errors are reported in parentheses in Column (3). *, ** and *** denote, respectively, significance at the 10%, 5% and 1% levels (two-sided tests).

Table 3: Further regressions of First Movers' efforts.

The three regressions reported in Table 3 deliver similar estimates of the effect of the prize on effort provision. However, comparing precision in Column (1) to that in Column (2) highlights the importance of being able to apply panel data methods to the high-quality repeated observations of effort generated by the slider task. The OLS regression on the full sample of 600 subject-round observations in Column (2) is inefficient compared to the fixed effects regression in Column (1): reflecting this, the standard error of the prize effect in Column (2) is substantially higher than that in Column (1) (0.240 versus 0.157). The reason for this loss of efficiency is that the OLS regression in Column (2) does not use the repeated observations of effort at the individual level to account for persistent unobserved heterogeneity when obtaining parameter estimates.

Furthermore, comparing precision in Column (1) of Table 3 to that in Column (3) illustrates the importance of collecting many observations per subject in a short time frame, which the slider task facilitates. When we use only one observation per subject in Column (3), precision falls dramatically: the standard error of the prize effect increases to 0.828, making the prize effect appear statistically insignificant.[12]

To summarize, effort provision responds positively to financial incentives, effort tends to increase over rounds, and the high-quality repeated observations of effort provision generated

[12]We eliminate the noise from the random selection of one observation per subject by repeating the random selection process 2,000 times and reporting the average parameter estimates and average standard errors over the 2,000 repetitions.

by the slider task allow us to use panel data methods to appreciably increase the precision of the statistical analysis.

Using between-subject designs with variation in strictly positive piece rates, Araujo et al. (2016) and Goerg et al. (forthcoming) also find that effort responds positively to financial incentives using the slider task. The magnitude of the response to financial incentives in Araujo et al. (2016)'s data is only about one-third of the one that we find, and the effect is only marginally statistically significant. However, using a similar design to Araujo et al. (2016), Goerg et al. (forthcoming) find that the magnitude of the between-subject response is substantial when subjects complete the slider task in the presence of an outside option (browsing the Internet), increasing three-fold relative to the standard case without the browsing option (even in the standard case, Goerg et al., forthcoming, find a bigger response to incentives than do Araujo et al., 2016).[13] Based on recent findings in other contexts (Corgnet et al., 2015; Eckartz, 2014), Araujo et al. (2016, p.11) correctly anticipated that incentive effects when using the slider task might be strengthened by the inclusion of an outside option.

In the context of laboratory and online experiments that vary strictly positive monetary incentives between subjects using other real-effort tasks, the quantitatively modest response in Araujo et al. (2016)'s data is not surprising. A number of studies find a non-monotonic or monotonically decreasing relationship between incentives and effort (e.g.: Ariely et al., 2009; Pokorny, 2008; Takahashi et al., 2016).[14] Using a large sample of over 500 Mechanical Turk workers per treatment, DellaVigna and Pope (2017) find only a modest positive response to incentives.[15] Surveying the literature on piece rates and performance in laboratory experiments, Charness and Kuhn (2011, p.249) conclude that: "the effect of stronger incentives on performance, predicted to be monotonic by basic labor supply theory (at least when income effects are unimportant, which is expected for laboratory experiments), may in fact be highly non-monotonic." Loss aversion around a reference wage provides one explanation for these empirical findings (Pokorny, 2008).[16]

Furthermore, in the light of other slider task experiments, our stronger within-subject results are not surprising: the partial survey in Araujo et al. (2015) shows that within-subject experi-

[13]Goerg et al. (forthcoming) use a low and high piece rate that almost exactly replicate Araujo et al. (2016)'s low and medium piece rates. Moving from the low to the high piece rate, and using regressions with controls, Goerg et al. (forthcoming, pp.13-14 of the version dated 06/23/18) found that effort increased by 11% without the Internet browsing option ($p = 0.012$), but that it increased by 31% with the Internet browsing option ($p < 0.001$). By contrast, Araujo et al. (2016) found that effort increased by only 2% when moving from their low to medium incentives. In Goerg et al. (forthcoming) subjects were paid for sets of five completed sliders, while they were paid per slider in Araujo et al. (2016); translating Goerg et al. (forthcoming)'s incentives into a per-slider rate, the piece rate was 0.4c or 2c.

[14]Averaging across six tasks, Ariely et al. (2009) find a non-monotonic relationship between performance and incentives: relative to low incentives, moderate incentives modestly increased performance, but the effect was not statistically significant, while high incentives reduced performance substantially and statistically significantly. Using a counting task and a circle-clicking task respectively, Pokorny (2008) and Takahashi et al. (2016) find that higher incentives monotonically decrease performance.

[15]Using a button-pushing task, DellaVigna and Pope (2017) found that a 900% increase in the piece rate increased effort by about 7%. Using a ball-dragging task, Heyman and Ariely (2004) also find a positive response to incentives.

[16]Pokorny (2008) show theoretically that in a simple between-subject setting with only variation in piece rates, loss aversion around a reference wage moderates the magnitude of the response to incentives, and can cause a non-monotonic response.

ments that use the slider task in a variety of contexts generally produce statistically significant effects on performance. And, as explained above, Araujo et al. (2016)'s weaker between-subject results are consistent with the broader literature that uses between-subject designs together with a variety of real-effort tasks to measure the effect of monetary incentives on performance. Thus, the best way to understand the differences between our results and those of Araujo et al. (2016) is as an example of the more general pattern whereby within-subject experimental designs produce larger effects and greater precision. For example, Ariely et al. (2003, p.99) note that "the tendency for within-subject manipulations to produce larger effects than between subject manipulations is a common phenomenon," while Charness et al. (2012, p.2) note that "[within-subject designs] offer a substantial boost in statistical power." Charness et al. (2012) explore the issues surrounding the two types of design.[17]

6 Estimating the cost of effort at the subject level

In this section we show how to use the high-quality panel data on repeated effort choices from the slider task to estimate the cost of effort at the subject level. This exercise exploits within-subject variation in incentives across repetitions of the slider task.

We continue to focus on the sample of 60 First Movers, each of whom was observed for 10 rounds in our experiment. First Movers are indexed by $i = 1, ..., 60$, and rounds are indexed by $r = 1, ..., 10$. At the start of each round, each First Mover is informed of the prize for the round, $v_{i,r}$, which will be awarded to either the First Mover or her opponent. According to our experimental design, described above in Section 4, the First Mover's probability of winning the prize is given by:

$$\Pr(e_{i,r}, e'_{i,r}) \quad = \quad \left(\frac{e_{i,r} - e'_{i,r} + 50}{100} \right), \tag{1}$$

where $e_{i,r} \in \{0, ..., 48\}$ is the effort of First Mover i in round r and $e'_{i,r} \in \{0, ..., 48\}$ is the effort of First Mover i's opponent in round r. We assume that the First Mover is risk neutral and

[17] Araujo et al. (2016, p.11) note that within-subject designs better control for variation in individual-level ability, and Charness et al. (2012, pp.1-2) note that: (i) between-subject designs require that group assignment be random; while (ii) the internal validity of within-subject designs does not depend on random assignment. (On this point, we note that Araujo et al., 2016, provide no test of balance, and nor they do they include any demographic or ability controls (although their sessions were gender balanced by design).) Furthermore, as noted by Charness et al. (2012, p.2) "between designs typically have no natural anchor. Thus results inherently have substantial noise, and may miss important and real patterns." This is a particular problem in real-effort experiments, where subjects have little idea of what a fair or reasonable experimental wage might be, and empirically subjects work hard even in the absence of marginal pecuniary incentives (e.g., Gill et al., forthcoming). Thus, experimenter-demand effects might mute responses to monetary incentives in between-subject designs like Araujo et al. (2016), since subjects might interpret even modest wages as reasonable. On the other hand, experimenter-demand effects might amplify responses to monetary incentives in within-subject real-effort experiments, since subjects might think that the experimenter expects a response to wage variation. Having said this, experimenter-demand effects do not need to be invoked to explain why real effort varies more with the wage in within-subject designs: instead, subjects might rationally work hard in high-wage periods and rest in low-wage periods. Rather than being an artefact of the laboratory, this pattern is a prediction of standard intertemporal labor supply theory.

chooses her effort to maximize her expected utility

$$Eu_{i,r} = \left(\frac{e_{i,r} - e'_{i,r} + 50}{100} \right) v_{i,r} - C_i(e_{i,r}), \tag{2}$$

where $C_i(.)$ is First Mover i's cost of effort function. We include the subscript i on the cost of effort function because the cost of effort may be heterogeneous across subjects. Motivated by the finding reported in footnote 11 that effort increases linearly with the prize, we assume a quadratic cost of effort function $C_i(e_{i,r}) = c_i e_{i,r}^2/2$ with $c_i > 0$.[18] We also assume that the First Mover believes that she cannot affect her opponent's effort.[19] We then have the following expression for optimal effort provision:

$$e_{i,r}^* = \frac{v_{i,r}}{100c_i}. \tag{3}$$

We augment (3) to include an additive error term and then use non-linear least squares to estimate the subject-specific cost parameters $c_i > 0$ for $i = 1, ..., 60$.[20] In particular, for each First Mover, we estimate the cost parameter, c_i, by minimizing the sum of squares:

$$\sum_{r=1}^{10} \left(e_{i,r} - \frac{v_{i,r}}{100c_i} \right)^2. \tag{4}$$

Our estimation results reveal substantial heterogeneity in the cost of effort parameter: the Gini coefficient for c_i is 0.16 and the 90:10 ratio for c_i is 1.67. To facilitate interpretation of these findings, Figure 5 illustrates the estimated cost of effort function for a subject at the 90th percentile of the distribution of c_i, a subject with the median value of c_i, and a subject at the 10th percentile of the distribution of c_i.

Figure 6 shows how the estimated heterogeneity in costs translates into different levels of effort provision: at the average prize of £2, a subject at the 90th percentile of the distribution of c_i completes only 15 sliders, while a subject at the 10th percentile of the distribution of c_i completes around 26 sliders.

[18] In the data effort never reaches its maximum level of 48, and so we assume that $c_i > 0$.

[19] We make this assumption to abstract from other considerations that might have driven effort in the particular experimental setting in which these data were collected. From Gill and Prowse (2012) we know that in this setting Second Movers did in fact respond to First Mover effort. Thus we cannot rule out that sophisticated First Movers anticipated this response. However, at the average prize of £2, taking into account how First Mover effort influences Second Mover effort increases the marginal benefit of a unit of effort for First Movers by only a small amount (from £0.020 to c. £0.021).

[20] Prior to estimation, we use the results from a regression of First Mover effort on a full set of round dummies to remove round effects from First Mover effort.

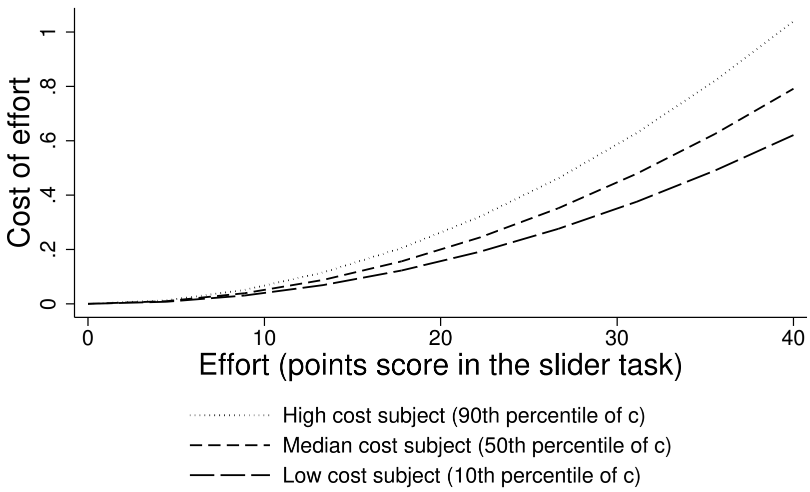


Figure 5: Heterogeneous effort costs.

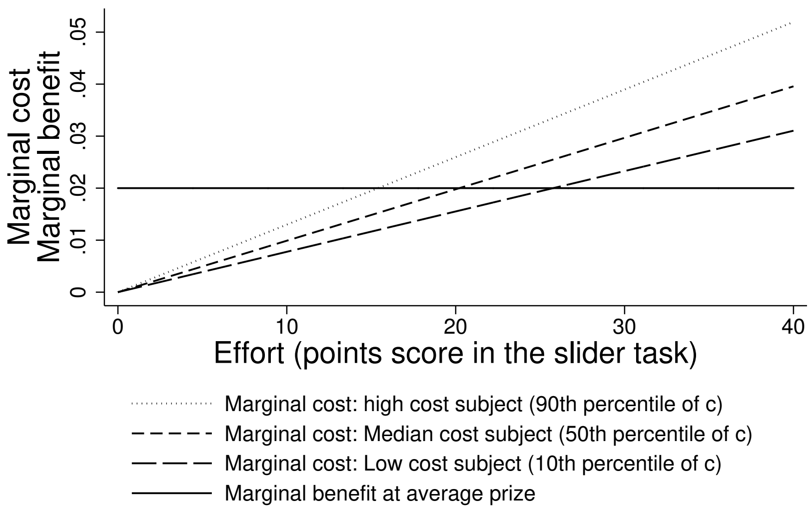


Figure 6: Heterogeneity in optimal effort provision.

7 A practical guide to using the slider task

This section provides a guide to researchers wishing to implement the slider task in the context of their own laboratory experiments. First, we describe the accompanying code, which allows researchers to implement easily the slider task in z-Tree (Fischbacher, 2007). We then list some practical considerations associated with the use of the slider task.

Code

The slider task was first developed and used by us in Gill and Prowse (2012) to study disappointment aversion.[21] To help readers create real-effort laboratory experiments that use the slider task, we have created a z-Tree code file that provides an implementation of the real-effort slider task. The code accompanies this paper online (see footnote 5). The code provided takes the form of a .ztt file and should be run in z-Tree. The code consists of a single file, named GillProwseSliderExample.ztt. This is a z-Tree treatment file. The program implements the slider task for a single subject, with the number of rounds set to one. This code can easily be embedded into an experimental design in which a real-effort task is required. Indeed, the code is based on the code used to program the real-effort task in the repeated sequential-move tournament of Gill and Prowse (2012). The treatment GillProwseSliderExample.ztt consists of three stages:

Stage 1 The subject is shown a screen informing her that the task is about to start. This screen is displayed for 5 seconds and then the program automatically moves to stage 2.

Stage 2 The subject is shown a screen displaying 48 sliders. The round number and the remaining time are shown at the very top of the screen, and between this information and the sliders there is a banner displaying the subject's current points score, i.e., the number of sliders currently positioned at exactly 50. This screen is displayed for 120 seconds and then the program automatically moves to stage 3.

Stage 3 The subject is shown a screen displaying her points score in the task. This screen is displayed for 20 seconds and then the program automatically ends.

We now give some more detail about this treatment. Prior to the treatment commencing a number of variables are created in the Background. First, the variable `Effort` is created. At any point during the treatment this variable equals the number of sliders currently positioned at exactly 50. Second, we create a set of 48 variables, denoted `qx` for `x` $= 1, ..., 48$. The variable `qx` is the current position of the x^{th} slider. Third, we create the variables `sx` for `x` $= 1, ..., 48$. The variable `sx` takes the value one if the current position of the x^{th} slider is equal to 50 and zero otherwise. All variables are initialized to zero.

Each time the position of a slider is adjusted, the values of `qx` and `sx` associated with the particular slider in question are updated. The value of `Effort` is then updated, and the banner

[21] We also use the slider task in Gill et al. (2013) and Gill and Prowse (2014). Section 1 lists a number of other studies that have used the slider task.

at the top of the screen is then refreshed to display the subject's new points score. The values of all the variables at the end of the 120 second task are stored in the Subjects table, and can be accessed at later stages.

Practical advice and guidance

Screen size

The average time taken to position a slider at exactly 50 depends on the size of the screen on which the task is displayed. We used relatively large screens, specifically 22 inch widescreen monitors with a 1680 by 1050 pixel resolution. 48 sliders and a 120 second task length was an appropriate configuration given the hardware employed, but may need adjusting if run on a different set-up. We believe that with our configuration it is impossible for any subject to position correctly all of the sliders (see Section 5). This ensures that the subject's effort choice is not constrained by the design of the task, so there is no incentive to work hard for the purpose of being able to rest at the end of the task.

Mice and keyboards

To treat all subjects equally, they should use the same specification of mouse. Our subjects used 800 dpi USB mice with the scroll wheel disabled (by removing them from the mice) to prevent subjects from using the scroll wheel to position the sliders. (Using the scroll wheel makes positioning the sliders much easier and requires less effort than a dragging and dropping technique using the left mouse button). Christopher Zeppenfeld (Cologne Graduate School) has kindly informed us that it is also possible to use an AutoHotKeys script in conjunction with z-Tree to disable the scroll wheel. Similarly, the keyboards were also disabled (by unplugging them) to prevent the subjects using the arrow keys to position the sliders. As well as dragging and dropping, it is possible to move the sliders in large fixed discrete steps by clicking the left mouse button with the cursor to the right or left of the current slider position. We did not point this out explicitly to our subjects, but told them that they could use the mouse in any way they liked to move the sliders.

Physical space and other environmental factors

Given subjects are being asked to complete a real-effort task it is important that they all have the same amount of physical space, i.e., all the booths are the same size, and that all subjects have the same equipment, e.g., mouse mats, chairs etc.

Practice rounds

Practice rounds, with the opportunity for questions at the end of each round, are recommended to allow subjects to become familiar with the task. We used two practice rounds.

8 Conclusion

This paper describes the properties and advantages of the real-effort slider task. The task was designed by us to overcome many of the drawbacks of real-effort tasks. In particular, our slider task provides a precise finely gradated measure of effort provision in a short time frame and, therefore, can be repeated many times in an experimental session. As we explained in detail above, the resulting high-quality panel data allow the experimenter to: (i) estimate the cost of effort at the subject level; and (ii) increase precision by controlling for unobserved heterogeneity. The paper provides direction to experimental economists who are considering how to implement costly activities in their own experiments. We also provide z-Tree code and practical guidance to help researchers implement the slider task in the laboratory.

References

Abeler, J., **Falk, A.**, **Goette, L.**, and **Huffman, D.** (2011). Reference points and effort provision. *American Economic Review*, 101(2): 470–492

Abeler, J. and **Jäger, S.** (2015). Complex tax incentives. *American Economic Journal: Economic Policy*, 7(3): 1–28

Ahlquist, J.S., **Hamman, J.**, and **Jones, B.** (2014). Dependency status and social insurance: An experimental investigation. *Mimeo, University of Wisconsin*

Araujo, F.A., **Carbone, E.**, **Conell-Price, L.**, **Dunietz, M.W.**, **Jaroszewicz, A.**, **Landsman, R.**, **Lamé, D.**, **Vesterlund, L.**, **Wang, S.W.**, and **Wilson, A.J.** (2015). The effect of incentives on real effort: Evidence from the slider task. *CESifo Working Paper 5372*

Araujo, F.A., **Carbone, E.**, **Conell-Price, L.**, **Dunietz, M.W.**, **Jaroszewicz, A.**, **Landsman, R.**, **Lamé, D.**, **Vesterlund, L.**, **Wang, S.W.**, and **Wilson, A.J.** (2016). The slider task: An example of restricted inference on incentive effects. *Journal of the Economic Science Association*, 2(1): 1–12

Ariely, D., **Gneezy, U.**, **Loewenstein, G.**, and **Mazar, N.** (2009). Large stakes and big mistakes. *Review of Economic Studies*, 76(2): 451–469

Ariely, D., **Loewenstein, G.**, and **Prelec, D.** (2003). "Coherent arbitrariness": Stable demand curves without stable preferences. *Quarterly Journal of Economics*, 118(1): 73–106

Avram, S. (2015). Benefit losses loom larger than taxes: The effects of framing and loss aversion on behavioural responses to taxes and benefits. Technical report, ISER Working Paper Series

Besley, T. and **Ghatak, M.** (2017). Profit with purpose? A theory of social enterprise. *American Economic Journal: Economic Policy*, 9(3): 19–58

Bonein, A. and **Denant-Boèmont, L.** (2015). Self-control, commitment and peer pressure: A laboratory experiment. *Experimental Economics*, 18(4): 543–568

Bradler, C., **Neckermann, S.**, and **Warnke, A.J.** (2015). Incentives and creativity. *Mimeo, ZEW Centre for European Economic Research*

Breuer, L. (2013). Tax compliance and whistleblowing – The role of incentives. *Bonn Journal of Economics*, 2(2): 7–45

Brown, A.L., **Meer, J.**, and **Williams, J.F.** (2013). Why do people volunteer? An experimental analysis of preferences for time donations. *NBER Working Paper 19066*

Brown, A.L., **Meer, J.**, and **Williams, J.F.** (2017). Social distance and quality ratings in charity choice. *Journal of Behavioral and Experimental Economics*, 66: 9–15

Bull, C., **Schotter, A.**, and **Weigelt, K.** (1987). Tournaments and piece rates: An experimental study. *Journal of Political Economy*, 95(1): 1–33

Burrows, P. and **Loomes, G.** (1994). The impact of fairness on bargaining behaviour. *Empirical Economics*, 19: 201–221

Cettolin, E. and **Riedl, A.** (2016). Justice under uncertainty. *Management Science*, 63(11): 3739–3759

Charness, G. and **Kuhn, P.** (2011). Lab labor: What can labor economists learn from the lab? In O. Ashenfelter and D. Card, editors, *Handbook of Labor Economics, Vol. 4*, 229–330. Amsterdam: North Holland

Charness, G., **Gneezy, U.**, and **Henderson, A.** (2018). Experimental methods: Measuring effort in economics experiments. *Journal of Economic Behavior & Organization*, 149: 74–87

Charness, G., **Gneezy, U.**, and **Kuhn, M.A.** (2012). Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization*, 81(1): 1–8

Charness, G., **Masclet, D.**, and **Villeval, M.C.** (2014). The dark side of competition for status. *Management Science*, 60(1): 38–55

Chow, C.W. (1983). The effects of job standard tightness and compensation scheme on performance: An exploration of linkages. *Accounting Review*, 58: 667–685

Cooper, R., **DeJong, D.**, **Forsythe, R.**, and **Ross, T.** (1996). Cooperation without reputation: Experimental evidence from prisoner's dilemma games. *Games and Economic Behavior*, 12: 187–218

Corgnet, B., **Hernán-González, R.**, and **Schniter, E.** (2015). Why real leisure really matters: Incentive effects on real effort in the laboratory. *Experimental Economics*, 18(2): 284–301

Das, A., **Liang, J.**, **Liu, Y.**, and **Xu, H.** (2013). Other-regarding or self-fulfilling – An experiment on working after disappointment. *SSRN Working Paper 2180813*

DellaVigna, S. and **Pope, D.** (2017). What motivates effort? Evidence and expert forecasts. *Review of Economic Studies*, 85(2): 1029–1069

Depoorter, B. and **Tontrup, S.** (2012). How law frames moral intuitions: The expressive effect of specific performance. *Arizona Law Review*, 54: 673–717

Dickinson, D.L. (1999). An experimental examination of labor supply and work intensities. *Journal of Labor Economics*, 17(4): 638–670

Djawadi, B.M. and **Fahr, R.** (2013). The impact of tax knowledge and budget spending influence on tax compliance. *IZA Discussion Paper 7255*

Djawadi, B.M., **Fahr, R.**, and **Turk, F.** (2014). Conceptual model and economic experiments to explain nonpersistence and enable mechanism designs fostering behavioral change. *Value in Health*, 17(8): 814–822

Doerrenberg, P. and **Duncan, D.** (2014a). Distributional implications of tax evasion: Evidence from the lab. *Public Finance Review*, 42(6): 720–744

Doerrenberg, P. and **Duncan, D.** (2014b). Experimental evidence on the relationship between tax evasion opportunities and labor supply. *European Economic Review*, 68: 48–70

Doerrenberg, P., **Duncan, D.**, and **Zeppenfeld, C.** (2015). Circumstantial risk: Impact of future tax evasion and labor supply opportunities on risk exposure. *Journal of Economic Behavior and Organization*, 109: 85–100

Dolan, P., **Metcalfe, R.**, and **Navarro-Martinez, D.** (2012). Financial incentives and working in the education sector. *Department for Education Research Report DFE–RR251*

Dorrough, A.R., **Leszczyńska, M.**, **Barreto, M.**, and **Glöckner, A.** (2016). Revealing side effects of quota rules on group cooperation. *Journal of Economic Psychology*, 57: 136–152

Drzensky, F. and **Heinz, M.** (2016). The hidden costs of downsizing. *Economic Journal*, 126(598): 2324–2341

Eacret, N.D., **Lafferty, N.C.**, and **Jhunjhunwala, T.** (2011). When is inequality fair? *Mimeo, University of Texas at Dallas*

Eckartz, K.M. (2014). Task enjoyment and opportunity costs in the lab – the effect of financial incentives on performance in real effort tasks. *Jena Economic Research Papers 2014 – 005*

Erkal, N., **Gangadharan, L.**, and **Nikiforakis, N.** (2011). Relative earnings and giving in a real-effort experiment. *American Economic Review*, 101(7): 3330–3348

Evdokimov, P. and **Rahman, D.** (2014). Personality and bargaining power. *Mimeo, University of Minnesota*

Fahr, R. and **Djawadi, B.H.** (2013). The impact of risk perception and risk attitudes on corrupt behavior: Evidence from a petty corruption experiment. *IZA Discussion Paper, No. 7383*

Fahr, R. and **Irlenbusch, B.** (2000). Fairness as a constraint on trust in reciprocity: Earned property rights in a reciprocal exchange experiment. *Economics Letters*, 66(3): 275–282

Falk, A. and **Fehr, E.** (2003). Why labour market experiments? *Labour Economics*, 10(4): 399–406

Fan, L. and **Zhao, J.** (2014). Social identity and the theoretical justification of recusation system. *SSRN Working Paper 2519264*

Feess, E., **Lorz, O.**, and **Schramm, M.** (2014). Under-delegation in case of pure moral hazard: Evidence from the laboratory. *SSRN Working Paper 2464908*

Fehr, E., **Gächter, S.**, and **Kirchsteiger, G.** (1997). Reciprocity as a contract enforcement device: Experimental evidence. *Econometrica*, 65(4): 833–860

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2): 171–178

Fonseca, M.A. and **Myles, G.** (2012). Experimental evidence on taxpayer compliance: Evidence from students and taxpayers. *Her Majesty's Revenue and Customs Research Report 198*

Friedl, A., **Neyse, L.**, and **Schmidt, U.** (2014). Payment scheme changes and effort provision: The effect of digit ratio. *MPRA Working Paper 59549*

Gärtner, M. and **Sandberg, A.** (2014). Omission effects in fairness behavior. *SSE Working Paper in Economics 2014:1*

Gee, L.K., **Migueis, M.**, and **Parsa, S.** (2017). Redistributive choices and increasing income inequality: experimental evidence for income as a signal of deservingness. *Experimental Economics*, 20(4): 894–923

Georganas, S., **Tonin, M.**, and **Vlassopoulos, M.** (2015). Peer pressure and productivity: The role of observing and being observed. *Journal of Economic Behavior & Organization*, 117: 223–232

Gerhards, L. and **Siemer, N.** (2014). Private versus public feedback - The incentive effects of symbolic awards. *Aarhus Economics Working Paper 2014-01*

Gill, D. and **Prowse, V.** (2012). A structural analysis of disappointment aversion in a real effort competition. *American Economic Review*, 102(1): 469–503

Gill, D. and **Prowse, V.** (2014). Gender differences and dynamics in competition: The role of luck. *Quantitative Economics*, 5(2): 351–376

Gill, D. and **Prowse, V.** (2018). Measuring costly effort using the slider task. *IZA Discussion Paper 11411*

Gill, D., **Prowse, V.**, and **Vlassopoulos, M.** (2013). Cheating in the workplace: An experimental study of the impact of bonuses and productivity. *Journal of Economic Behavior and Organization*, 96: 120–134

Gill, D., **Kissová, Z.**, **Lee, J.**, and **Prowse, V.** (forthcoming). First-place loving and last-place loathing: How rank in the distribution of performance affects effort provision. *Management Science*

Gneezy, U., **Niederle, M.**, and **Rustichini, A.** (2003). Performance in competitive environments: Gender differences. *Quarterly Journal of Economics*, 118: 1049–1074

Goerg, S.J., **Kube, S.**, and **Radbruch, J.** (forthcoming). The effectiveness of incentive schemes in the presence of implicit effort costs. *Management Science*

Hammermann, A., **Mohnen, A.**, and **Nieken, P.** (2011). Whom to choose as a team mate? A lab experiment about in-group favouritism. *IZA Discussion Paper 6286*

Hetzel, S. (2010). The impact of financial incentives on individual performance: An experimental approach. *Diplomarbeit, University of Mannheim*

Heyman, J. and **Ariely, D.** (2004). Effort for payment: A tale of two markets. *Psychological Science*, 15(11): 787–793

Hoffman, E., **McCabe, K.**, **Shachat, K.**, and **Smith, V.** (1994). Preferences, property rights, and anonymity in bargaining games. *Games and Economic Behavior*, 7: 346–380

Ibanez, M. and **Schaffland, E.** (2013). The effect of outside leaders on the performance of the organization: An experiment. *Courant Research Centre Discussion Paper 149*

Imas, A., **Sadoff, S.**, and **Samek, A.** (2016). Do people anticipate loss aversion? *Management Science*, 63(5): 1271–1284

Kesternich, M., **Lange, A.**, and **Sturm, B.** (2018). On the performance of rule-based contribution schemes under endowment heterogeneity. *Experimental Economics*, 21(1): 180–204

Khadjavi, M. (2014). On the interaction of deterrence and emotions. *Journal of Law, Economics, & Organization*, 31(2): 287–319

Kimbrough, E.O. and **Reiss, J.P.** (2012). Measuring the distribution of spitefulness. *Plos ONE*, 7(8): 1–8

Kirchler, M., **Bonn, C.**, **Huber, J.**, and **Razen, M.** (2014). The inflow-effect: Trader inflow and bubble formation in asset markets. *Working Papers in Economics and Statistics 2014-22, University of Innsbruck*

Konow, J. (2000). Fair shares: Accountability and cognitive dissonance in allocation decisions. *American Economic Review*, 90(4): 1072–1091

Lee, J. (2015). How does competition affect people economically and emotionally? *Mimeo, Cornell University*

Lezzi, E. (2014). Agents' performance and emotions: An experimental investigation. *Doctoral Dissertation, University of East Anglia*

Lindner, F. (2014). Decision time and steps of reasoning in a competitive market entry game. *Economics Letters*, 122(1): 7–11

Malik, S., **Mihm, B.**, and **Timme, F.** (2018). An experimental analysis of tax avoidance policies. *International Tax and Public Finance*, 25(1): 200–239

Meissner, T. and **Pfeiffer, P.** (2017). Measuring preference over the temporal resolution of consumption uncertainty. *SSRN Working Paper 2654668*

Monahan, J. (2012). We're all Lockeans now: A test of the prevalence of just-desert considerations in a dictator game experiment. *M.Phil. Thesis, University of Oxford*

Niederle, M. and **Vesterlund, L.** (2007). Do women shy away from competition? Do men compete too much? *Quarterly Journal of Economics*, 122(3): 1067–1101

Pokorny, K. (2008). Pay - but do not pay too much: An experimental study on the impact of incentives. *Journal of Economic Behavior & Organization*, 66(2): 251–264

Riener, G. and **Wiederhold, S.** (2013). Heterogeneous treatment effects in groups. *Economics Letters*, 120(3): 408–412

Riener, G. and **Wiederhold, S.** (2016). Team building and hidden costs of control. *Journal of Economic Behavior & Organization*, 123: 1–18

Schmitz, J. (2014). One good deed a day? *SSRN Working paper 2538410*

Smith, J. (2016). The motivational effects of mission matching: A lab-experimental test of a moderated mediation model. *Public Administration Review*, 76(4): 626–637

Straub, T., **Gimpel, H.**, **Teschner, F.**, and **Weinhardt, C.** (2014). Feedback and performance in crowd work: A real effort experiment. *Mimeo, Karlsruhe Institute of Technology*

Sutter, M. and **Weck-Hannemann, H.** (2003). Taxation and the veil of ignorance - A real effort experiment on the Laffer curve. *Public Choice*, 115: 217–240

Takahashi, H., **Shen, J.**, and **Ogawa, K.** (2016). An experimental examination of compensation schemes and level of effort in differentiated tasks. *Journal of Behavioral and Experimental Economics*, 61: 12–19

Teng, J.C.Y., **Wang, J.T.Y.**, and **Yang, C.C.** (2014). Primary social goods and the Rawlsian difference principle. *Mimeo, National Taipei University*

van Dijk, F., **Sonnemans, J.**, and **van Winden, F.** (2001). Incentive systems in a real effort experiment. *European Economic Review*, 45(2): 187–214

Zeppenfeld, C. (2015). *Decisions under Uncertainty: Preferences, Institutions, and Social Interaction.* Ph.D. thesis, Universität zu Köln