The international platform of Ludwig-Maximilians University's Center for Economic Studies and the Ifo Institute

Working Papers

www.cesifo.org/wp

Combining "Real Effort" with Induced Effort Costs: The Ball-Catching Task

Simon Gächter
Lingbo Huang
Martin Sefton

CESIFO WORKING PAPER NO. 5345
CATEGORY 13: BEHAVIOURAL ECONOMICS
MAY 2015

ISSN 2364-1428

CESifo
Center for Economic Studies & Ifo Institute

CESifo *Working Paper No. 5345*

Combining “Real Effort” with Induced Effort Costs: The Ball-Catching Task

Abstract

We introduce the “ball-catching task”, a novel computerized real effort task, which combines “real” efforts with induced material cost of effort. The central feature of the ball-catching task is that it allows researchers to manipulate the cost of effort function as well as the production function, which permits quantitative predictions on effort provision. In an experiment with piece-rate incentives we find that the comparative static and the point predictions on effort provision are remarkably accurate. We also present experimental findings from three classic experiments, namely, team production, gift exchange and tournament, using the task. All of the results are closely in line with the stylized facts from experiments using purely induced values. We conclude that the ballcatching task combines the advantages of real effort with induced values, which is useful for theorytesting purposes as well as for applications.

JEL-Code: C910, C920, J410.

Keywords: experimental design, real effort task, induced values, incentives, piece-rate theory, team incentives, gift exchange, tournaments, online real effort experiments.

Simon Gächter
CeDEx and School of Economics
University of Nottingham
Sir Clive Granger Building
University Park
United Kingdom – Nottingham NG7 2RD
simon.gaechter@nottingham.ac.uk

Lingbo Huang
CeDEx and School of Economics
University of Nottingham
Sir Clive Granger Building
University Park
United Kingdom – Nottingham NG7 2RD
lingbo.huang@nottingham.ac.uk

Martin Sefton
CeDEx and School of Economics
University of Nottingham
Sir Clive Granger Building
University Park
United Kingdom – Nottingham NG7 2RD
martin.sefton@nottingham.ac.uk

30 April 2015

1 Introduction

Experiments using "real effort" tasks enjoy increasing popularity among experimental economists. Some frequently used tasks include, for instance, number-addition tasks (e.g., Niederle and Vesterlund (2007)), counting-zero tasks (e.g., Abeler, et al. (2011)) and slider-positioning tasks (Gill and Prowse (2012)).[1] In this paper, we present a novel computerized real effort task, called the "ball-catching task", which combines "real" efforts in the lab with induced material cost of effort. In the ball-catching task, a subject has a fixed amount of time to catch balls that fall randomly from the top of the screen by using mouse clicks to move a tray at the bottom of the screen. Control over the cost of effort is achieved by attaching material costs to mouse clicks that move the tray.

The ball-catching task shares an advantage of other real effort tasks in that subjects are required to do something tangible in order to achieve a level of performance, as opposed to simply choosing a number (as is done in experiments that implement cost of effort functions using a pure induced value method, where different number choices are directly linked with different financial costs). A drawback, however, of existing real effort tasks is that in using them the researcher sacrifices considerable control over the cost of effort function. As noted by Falk and Fehr (2003): "while 'real effort' surely adds realism to the experiment, one should also note that it is realized at the cost of losing control. Since the experimenter does not know the workers' effort cost, it is not possible to derive precise quantitative predictions" (p. 404). Incorporating material effort costs re-establishes a degree of control over effort costs and, as we shall demonstrate, allows researchers to manipulate observable effort costs and to make point predictions on effort provision.

Here, we report three studies aimed to evaluate the ball-catching task. In Study 1, we examine individual performance on the ball-catching task under piece-rate incentives. Subjects incur a cost for each mouse click and receive a prize for each ball caught. We first show that clicking behavior corresponds closely with comparative static predictions derived from piece-rate incentive theory. We then estimate the relationship between clicks and catches and use this to predict how the number of clicks will vary as the costs of clicking and the benefits of catching are manipulated. We find that average efforts, as measured by the number of mouse clicks, are close to the predicted number of clicks. These findings also add to the literature on empirical testing of incentive theories (Prendergast (1999)) by presenting real effort experimental evidence supporting basic piece-rate incentive theory. By comparison, the prominent field evidence reported by Lazear (2000) and lab evidence reported by

[1] To our knowledge, one of the first experimental studies to use a real effort task for testing incentive theory is Dickinson (1999) in which subjects were asked to type paragraphs in a four-day period. Other early studies implementing real effort tasks within typical laboratory experiments include van Dijk, et al. (2001); Gneezy and Rustichini (2000); Gneezy (2002) and Konow (2000).

Dickinson (1999) provide support for comparative static predictions of basic incentive theory, whereas we show that in the ball-catching task the theory also predicts effort levels accurately.

In Study 2, we demonstrate how the task can be implemented in some classic experiments. We administer the task in experiments used to study cooperation, fairness and competition, namely, team production (e.g., Nalbantian and Schotter (1997)), gift exchange (e.g., Fehr, et al. (1993)) and a tournament (e.g., Bull, et al. (1987)). In all three experiments, the results reproduce the stylized findings from previous experiments that used purely induced values. Moreover, behavior also follows equilibrium point predictions closely in those experiments where point predictions are available.

In Study 3, we introduce an online version of the ball-catching task and conduct the same experiment as in Study 1 using Amazon Mechanical Turk workers as participants. Comparative statics results are replicated, which we view as an important robustness check. Behavior is noisier than in the lab, however, which most likely is due to the more varied decision environment online compared to the lab. We use this observation to discuss some caveats of using the (online version of the) ball-catching task.

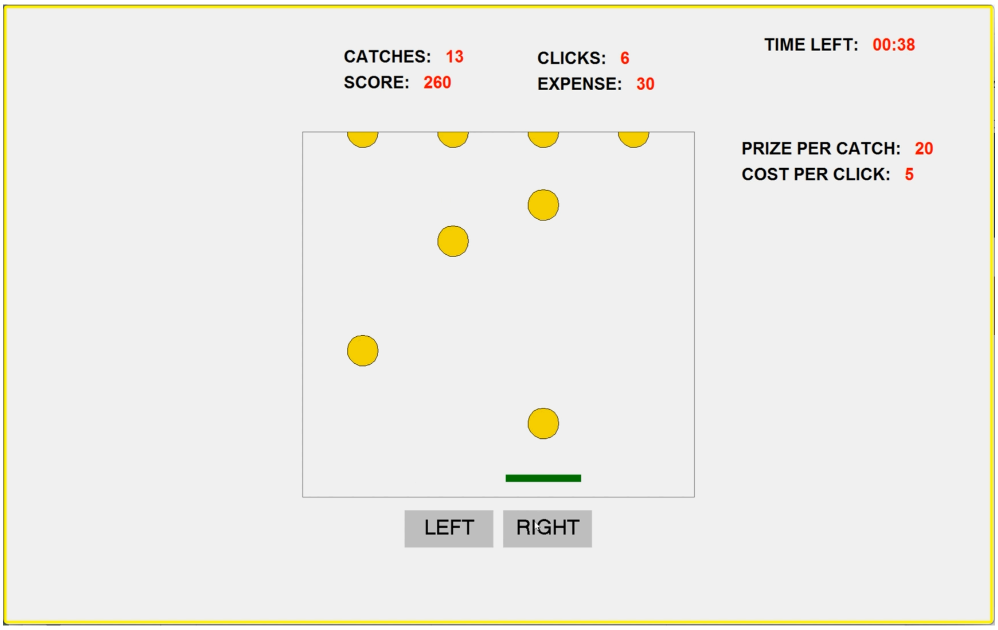
2 The ball-catching task

The lab version of the ball-catching task is a computerized task programmed in z-Tree (Fischbacher (2007)), and requires subjects to catch falling balls by moving a tray on their computer screens. Fig. 1 shows a screenshot of the task. In the middle of the screen there is a rectangular task box with four hanging balls at the top and one tray at the bottom. Once a subject presses the “Start the task” button at the lower right corner of the screen, the balls will fall from the top of the task box. In the version used in this paper, the timer starts and balls fall one after another in a fixed time interval. Balls fall at random in each column. The software allows adjusting the speed of falling balls and the time interval between falling balls. It is also possible to change the number of columns and fix a falling pattern rather than a random one. As will be discussed later, flexibility in all these parameters will allow tight control over the production function in this task, that is, the relationship between the number of balls caught and the number of clicks made.

To catch the falling balls, the subject can move the tray by mouse clicking the “LEFT” or “RIGHT” buttons below the task box. At the top of the screen, the number of balls caught (CATCHES) and the number of clicks made (CLICKS) are updated in real time. We will take the number of clicks as our observable measure of effort. As will become clear later, we acknowledge that other forms of effort (e.g., concentration, deliberation) may be exerted by the subject in this task.

What distinguishes our ball-catching task from other real effort tasks is that we attach pecuniary costs to mouse clicks. By specifying the relation between clicks and pecuniary costs we can implement any material cost of effort function. The most convenient specification might be to use a linear cost function by simply attaching a constant cost to every mouse click, but it is also possible to specify any nonlinear cost function (we will present an example in Section 4.2). In the example of Fig. 1 the subjects incurs a cost of 5 tokens for each mouse click. Accumulated costs (EXPENSE) are updated and displayed in real time. It is also possible to attach pecuniary benefits to catches. In Fig. 1 the subject receives 20 tokens for each ball caught and accumulated benefits (SCORE) are updated on screen in real time.

Fig. 1 A screenshot of the ball-catching task



In existing real effort tasks output and effort are typically indistinguishable. In the ball-catching task there is clear distinction between the catches and the clicks variables, with the natural interpretation being that the former represents output and latter input. Moreover, by choosing the time constraint, ball speed, etc., the researcher has flexibility in selecting the production technology.

Evidence collected in a post-experimental questionnaire suggests that subjects find the ball-catching task easy to understand and learn. In the next section we examine in more detail how subjects perform on the task under piece-rate incentives. In Section 5 we present a version of the ball-catching task that can be used for online experiments.

3 Study 1: Testing the ball-catching task under piece-rate incentives

3.1 Experimental design and comparative static predictions

Study 1 examined performance on the ball-catching task under piece-rate incentives. Each subject worked on the same ball-catching task for 36 periods. Each period lasted 60 seconds.[2] In each period one combination of prize-per-catch (either 10 or 20 tokens) and cost-per-click (0, 5 or 10 tokens) was used, giving six treatments that are varied within subjects (see Table 1). We chose a within-subject design to be able to observe reactions to changes in incentives at an individual level. The first 6 periods, one period of each treatment in random order, served as practice periods for participants to familiarize themselves with the task. Token earnings from these periods were not converted to cash. The following 30 periods, five periods of each treatment in completely randomized order (i.e., unblocked and randomized), were paid out for real. In all, 64 subjects participated in the experiment with average earnings of £13.80 for a session lasting about one hour.[3]

Table 1 Within-subject treatments in Study 1

Treatment No.	**1**	**2**	**3**	**4**	**5**	**6**
Prize per catch (P)	10	10	10	20	20	20
Cost per click (C)	0	5	10	0	5	10

Note: All subjects played five rounds of all six treatments in random order.

Given a particular piece-rate incentive, what level of effort would subjects exert? Basic piece-rate theory assumes that subjects trade off costs and benefits of effort in order to maximize expected utility. Assume that utility is increasing in the financial rewards, which are given by $PQ - Ce$, where Q is the number of catches and e is the number of clicks, and assume the relationship between Q and e is given by $Q = f(e, \epsilon)$, where f is a production function with $f' > 0$ and $f'' < 0$, and ϵ is a random shock uncorrelated with effort. Given these assumptions the expected utility maximizing number of clicks satisfies:

$$e^* = f'^{-1}\left(\frac{C}{P}\right). \tag{1}$$

This analysis posits a stochastic production function linking individual catches and clicks, and so an individual's optimal effort may vary from trial to trial as the marginal product of a click varies

[2] Unless otherwise stated, in the version of the ball-catching task we use in this paper a maximum of 52 balls can be caught within 60 seconds.

[3] The experiment was run in two sessions at the CeDEx lab at the University of Nottingham with subjects recruited using the online campus recruitment system ORSEE (Greiner (2004)). Instructions are reproduced in Appendix A1.

from trial to trial. This may reflect variability in the exact pattern of falling balls from trial to trial. We also recognize that the marginal product function might vary systematically across individuals. To make predictions at the aggregate level, we will estimate the production function (in Section 3.3) allowing for individual specific random effects and then use this estimate, evaluated at the mean of the random effects, along with our incentive parameters to predict average optimal effort. Before we proceed to this estimation, we discuss some features of the optimal effort solution and how they relate to our experimental design.

The first feature to note is that optimal effort is homogeneous of degree zero in C and P. That is, a proportionate change in both input and output prices leaves optimal effort unchanged. This feature reflects the assumption that there are no other unobserved inputs or outputs associated with working on the task that generate cognitive or psychological costs or benefits. In fact we can think of two plausible types of unobservable inputs/outputs. First, output may be a function of cognitive effort as well as the number of clicks. For example, output may depend not just on how many clicks a subject makes, but also on how intelligently a subject uses her clicks. If the production function is given by $f(e, t, \epsilon)$, where t represents cognitive effort, then e^* will reflect a trade-off between e and t. If *all* input and output prices were varied in proportion (including the "price" of cognitive effort), optimal effort would be unaffected. However, a proportionate change in just C and P would affect e^*. If e and t are substitute inputs then a proportionate increase in C and P will result in a decrease in e^* as the subject substitutes more expensive clicking with more careful thinking. Second, subjects may enjoy additional psychological benefits from catching balls. For example, suppose that in addition to the pecuniary costs and benefits there is a non-monetary benefit from a catch, and suppose this psychological benefit is worth B money-units per catch. Again, proportionate changes in P, C and B would leave optimal effort unchanged, but a change in just P and C would not. Maximization of $(P + B)Q - Ce$ implies that a proportionate increase in C and P (holding B constant) will decrease e^*.

Our experimental treatments allow us to test whether unobservable costs/benefits matter compared with induced effort costs in the ball-catching task. Our design includes two treatments that vary C and P while keeping the ratio $\frac{C}{P}$ constant (treatments 2 and 6 in Table 1). In the absence of unobserved costs/benefits, the distribution of clicks should be the same in these two treatments. The presence of unobserved costs/benefits could instead lead to systematic differences. Note that with this design the prediction that optimal effort is homogeneous of degree zero in C and P can be tested without the need to derive the underlying production function, f, since all that is needed is a comparison of the distributions of clicks between these two treatments.

A second feature of the optimal effort solution is that, for positive costs of clicking, optimal efforts decrease with the cost-prize ratio. Our design includes four further treatments that vary this ratio. Comparisons between treatments with different cost-prize ratios allow simple tests of the comparative static predictions of piece-rate theory, again without the need to estimate the production function. The variation in incentives across treatments serves an additional purpose: it allows us to recover a more accurate estimate of the underlying production function over a wide range of clicks.

A final feature of the solution worth noting is that when the cost-per-click is zero the optimal effort is independent of *P*. In this case, since clicking is costless the individual's payoff increases in the number of catches, and so regardless of the prize level the individual should simply catch as many balls as possible. Again, if there are psychological costs/benefits associated with the task this feature will not hold. Indeed, one could use the ball-catching task without material costs of effort, basing comparative static predictions (e.g. that the number of catches will increase as the prize per catch increases) on psychological costs of effort. However, as in many existing real effort tasks, the ball-catching task without material effort costs might exhibit a "ceiling effect", that is unresponsiveness of effort to varying prize incentives.[4] For this reason our design includes two treatments where the material cost of clicking is zero (treatments 1 and 4 in Table 1). These allow us to test whether there is a ceiling effect in the ball-catching task without induced effort costs.

3.2 Comparative statics results

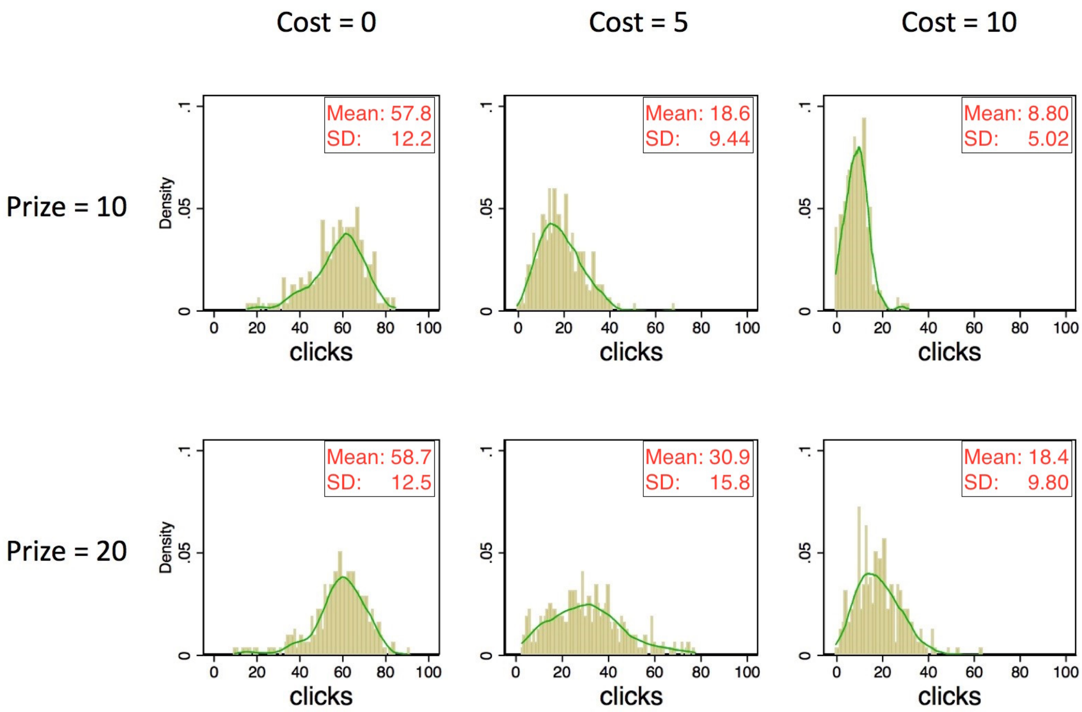
Fig. 2 shows the distributions of clicks for each treatment, pooling over all subjects and periods. Clear differences between panels show that efforts vary across incentive treatments. We begin by examining how these differences relate to the comparative static predictions based on the optimal effort solution (1). Consider first the comparison between treatments 2 ($P = 10$, $C = 5$) and 6 ($P = 20$, $C = 10$). These treatments vary the financial stakes without altering the cost/prize ratio. The basic piece-rate theory prediction is that this will not have a systematic effect on effort. As discussed in Section 3.1 however, unobserved psychological costs/benefits associated with the task will lead to systematic differences between the distributions of clicks in the two treatments. We find that the distributions of efforts are very similar, with average efforts of 18.6 under low-stakes and 18.4 under high stakes. Using a subject's average effort per treatment as the unit of observation, a Wilcoxon matched-pairs signed-ranks test ($p = 0.880$) finds no significant difference between treatments 2 and 6. Thus, we cannot reject the hypothesis that average efforts are invariant to scaling up the financial stakes.

Next we ask whether variation in the cost-prize ratio affects effort as predicted. Will increasing the cost-per-click, holding the prize-per-catch constant, reduce the number of clicks? And will the

[4] See an early review in Camerer and Hogarth (1999). Another possible reason for the "ceiling effect" is that subjects may also simply work on the paid task due to some experimenter demand effects (Zizzo (2010)), particularly in the absence of salient outside options (see Corgnet, et al. (2014) and Eckartz (2014) for discussions).

number of clicks depend on the prize level for a given clicking cost? First, we compare the top three panels of Fig. 2, where the prize is always 10. We observe a clear shift of the distribution of the number of clicks when moving across treatments with lowest to highest induced effort costs. Average efforts fall from 58.7 to 18.6 to 8.8 as the cost-per-click increases from 0 to 5 to 10. Friedman tests for detecting systematic differences in matched subjects' observations, using a subject's average effort per treatment as the unit of observations, show that the differences across treatments is highly significant ($p < 0.001$). A similar pattern is observed in the bottom three panels, where the prize is always 20, and again the differences are highly significant ($p < 0.001$).

Fig. 2 The distributions and the Kernel density distributions of the number of clicks in Study 1



Next, we perform two vertical comparisons between treatments 2 and 5 and between treatments 3 and 6. Holding the clicking costs constant, we find that a higher prize leads to higher number of clicks in both comparisons (Wilcoxon matched-pairs signed-ranks test: $p < 0.001$).

Finally, a comparison between treatments 1 and 4 offers an examination of whether a ceiling effect, observed in many other real effort tasks, is present in the ball-catching task. In these treatments the cost per click is zero, but the prize per catch is 10 in treatment 1 and 20 in treatment 4. If there is no "real" psychological cost/benefit associated with working on the task, subjects should simply catch as many balls as possible and we should observe the same distribution of the number of clicks in these two treatments, thus exemplifying the typical ceiling effect. Comparing the distributions of clicks of the zero-cost treatments illustrated in Fig. 2 suggest that distributions are very similar. Average efforts are 57.8 in the low prize treatment and 58.7 in the high prize treatment. The closeness of distributions between treatments 1 and 4 is statistically supported by a Wilcoxon matched-pairs signed-ranks test (p = 0.215), again using a subject's average effort per treatment as the unit of observation. The sharp contrast between the strong prize effect in treatments with induced effort costs and the absence of a prize effect in the zero-cost treatments illustrates that the ceiling effect can be avoided by incorporating financial costs in the ball-catching task.

For a more detailed overview of individual effort, we summarize in Appendix B the average number of clicks in each treatment for each individual. We also report the result of a test of whether each individual exhibited qualitatively consistent behavior, by which we mean that the subject's ranking of actual efforts is the same as the ranking of predicted efforts across all treatments. We predict the same effort level for treatments 2 and 6, and so we use an average of the actual efforts in these two treatments in our subject level test of consistency. Similarly, we predict the same effort level in treatments 1 and 4 and so the test uses the average effort over these two treatments. Only 3 of the 64 subjects (less than 5%) behaved inconsistently: in all three cases the effort in treatment 5 is lower than the average effort in treatments 2 and 6. [5]

In sum, we find that the comparative static predictions of basic piece-rate theory are borne out in the experimental data.[6] Our next goal is to derive point predictions about the number of clicks in the various treatments and to compare them to the data. To be able to do so, we next estimate the production function, which we will then use to derive the point predictions.

3.3 The production function

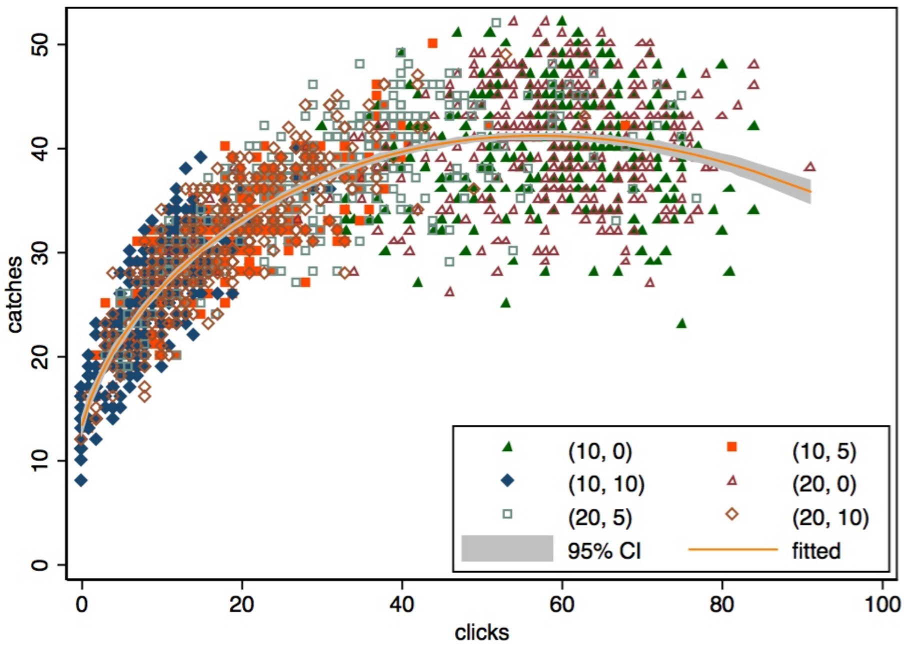
Our empirical strategy for estimating the production function is to first specify a functional form by fitting a flexible functional form to the catches-clicks data using the full sample. Next, we estimate the production function, allowing for persistent as well as transitory unobserved individual effects and fixed period effects. We then test whether the production function is stable across periods and invariant to varying prize levels. We will also examine the stability of the production function across experimental sessions.

[5] We also administered a post-experimental questionnaire where we asked subjects to rate the difficulty, enjoyableness and boredom of the task. On average, subjects reported that the task was very easy to do and they had neutral attitudes towards the enjoyableness and boredom of the task. Along with the quantitative data on clicks and catches, these responses are consistent with our interpretation that in our implementation of the ball-catching task psychological costs/benefits are not so important relative to pecuniary costs/benefits.

[6] We also do not find any learning or fatigue effect: average catches, average clicks and average earnings all show no sign of increase/decrease over the 30 rounds.

Fig. 3 shows the observed catches-clicks data from all treatments along with a fitted production function based on a fractional polynomial regression.[7] The fitted production function has a clear concave shape, indicating a diminishing marginal rate of return to clicks. After a point, the production function is decreasing, indicating that there is a "technological ceiling" beyond which higher effort provision may actually lead to lower production levels. Observations in the decreasing range are predominantly from the treatments with a zero cost of clicking. As one of the main advantages of using the ball-catching task is precisely that clicking can be made costly, the decreasing part of the production function should be of little concern, since with positive clicking costs the number of clicks will be within the range where the empirical production function is concave and increasing.

Fig. 3 The relation between clicks and catches and the estimated production functional form from a fractional polynomial regression



Note: The first entry in (*,*) denotes the prize per catch and the second the cost per click. The fitted production functional form is given by $Q = 9.507 + 5.568e^{0.5} - 0.003e^{2}$, where Q denotes the number of catches and e the number of clicks. The estimates of coefficients are from a fractional polynomial regression.

[7] Fractional polynomials can afford more flexibility than conventional polynomials by allowing logarithms and non-integer powers in the models. The curve-fitting procedure used in the regression selects the best-fitting model with appropriate powers and/or logarithms. We also considered the possibility that the functional form might differ for $C = 0$ treatments, and so we fitted fractional polynomials excluding these data. We get the same specifications, and very similar coefficients.

Using the functional form suggested by the fractional polynomial regression, we move on to estimate the following random coefficients panel data model:

$$Catches_{i,r} = \beta_0 + \beta_1 Clicks_{i,r}{}^{0.5} + \beta_2 Clicks_{i,r}{}^{2} + (\delta_r + \omega_i + u_{i,r})Clicks_{i,r}{}^{0.5}, \tag{2}$$

where $Catches_{i,r}$ and $Clicks_{i,r}$ are respectively the number of catches and the number of clicks of subject i in period r. Period dummies δ_r (with the first period providing the omitted category), an individual random effect ω_i with mean zero and variance σ_ω^2, and a random error $u_{i,r}$ with mean zero and variance σ_u^2 are all assumed to be multiplicative with $Clicks_{i,r}{}^{0.5}$. Our specification of multiplicative heterogeneity and heteroskedasticity allows both persistent and transitory individual differences in the *marginal product function* which could also vary across periods. The model thus predicts heterogeneity in clicking both across and within subjects.[8] All equations are estimated using maximum likelihood and estimates are reported in Table 2.[9]

Columns (1), (2) and (3) in Table 2 reports the coefficient estimates for the full sample, the sub-sample with the prize of 10 and the sub-sample with the prize of 20 respectively. Note the similarity between the estimates of the parameters of the production function in all equations. The fitted production functions for the two sub-samples with different prizes are shown in Fig. B1 in Appendix B: the two production functions almost coincide. Furthermore, we find that both persistent and transitory unobserved individual effects are statistically significant, and that the transitory unobservables account for more of the variation in clicking than the persistent individual differences.

[8] The model specification of multiplicative terms with $Clicks_{i,r}{}^{0.5}$ implies that the conditional variation in catches is linear in clicks. We examined the relationship between clicks and squared residuals from a simple pooled regression of the model $Catches_{i,r} = \alpha_0 + \alpha_1 Clicks_{i,r}{}^{0.5} + \alpha_2 Clicks_{i,r}{}^{2} + \pi_{i,r}$. We then regressed squared residuals on $Clicks_{i,r}$ as well as a nonlinear term (either $Clicks_{i,r}{}^{0.5}$ or $Clicks_{i,r}{}^{2}$). The coefficients on the nonlinear terms are not statistically significant, supporting our modelling specification of a linear relationship between conditional variation in catches and clicks.

[9] To estimate the model (2), note that dividing both sides by $Clicks_{i,r}{}^{0.5}$ transforms the model to a standard random effects model: $Catches_{i,r}/Clicks_{i,r}{}^{0.5} = \beta_0/Clicks_{i,r}{}^{0.5} + \beta_1 + \beta_2 Clicks_{i,r}{}^{3/2} + \delta_r + \omega_i + u_{i,r}$. Usual econometric techniques then follow.

Table 2 Panel data regressions for model (2) in Study 1

Dep. Var.: **Catches**	Coefficient estimates (std. err.)		
	(1) full sample	(2) Prize=10	(3) Prize=20
$Intercept$	10.107***	10.477***	9.405***
	(0.230)	(0.308)	(0.423)
$Clicks^{0.5}$	5.495***	5.402***	5.660***
	(0.132)	(0.216)	(0.171)
$Clicks^{2}$	-0.003***	-0.003***	-0.003***
	(0.000)	(0.001)	(0.000)
σ_{ω}	0.366***	0.352***	0.384***
	(0.038)	(0.045)	(0.042)
σ_{u}	0.796***	0.870***	0.694***
	(0.013)	(0.021)	(0.016)
N	1905	946	959

Note: All period dummies are included and insignificant except for period 2 using the full sample. ***p < 0.01

To formally test whether the production function is invariant to different prize levels, we proceed to estimate an augmented model by adding interactions of the intercept, covariates $Clicks^{0.5}$ and $Clicks^{2}$ with a binary variable indicating whether the prize is 10 or 20. We then perform a likelihood ratio test of the null hypothesis that the coefficients on the interaction terms are all zero. We cannot reject the null hypothesis, indicating that the production function is stable across prize levels (χ^2(3)=4.70, p=0.195).

To test the stability of the production function across experimental sessions, we estimate an augmented model by adding interactions of the intercept, $Clicks^{0.5}$ and $Clicks^{2}$ with a session dummy. We cannot reject the null hypothesis that the production function is invariant across sessions (χ^2(3)=2.60, p=0.458). In fact the fitted production functions are barely distinguishable. [10]

3.4 Comparing predicted and actual effort levels

With the estimated production function from model (2) and treatment parameters, we are ready to see how quantitative predictions on effort perform. Table 3 compares the predicted effort that is derived from equation (1) given the estimated production function reported in column (1) of Table 2 and the

[10] See Appendix B for details of the results. Estimates of model (2) for each session are given in Table B2 and the fitted production functions are shown in Fig. B2.

cost-prize parameters, with the actual number of clicks for every treatment.[11] We find that average actual clicks are very similar to the predicted number of clicks in treatments 1, 2, 4 and 6 and near to, but statistically significantly different from, predicted values in treatments 3 and 5 (subjects seem to have over-exerted effort in treatments 3 and under-exerted effort in treatment 5).[12] Thus, overall, not only did they change their behavior in the predicted direction when incentives changed, but also for given incentives their choices were close, on average, to the profit maximizing choices. The results are surprising given that subjects cannot know the production function *a priori* and therefore they cannot calculate the optimal level of effort deliberatively. Nonetheless, on average, they behaved *as if* they knew the underlying structural parameters and responded to them optimally.

Table 3 Comparisons between the predicted number of clicks and the actual number of clicks

Treatment No.	**1**	**2**	**3**	**4**	**5**	**6**
Prize per catch (P)	10	10	10	20	20	20
Cost per click (C)	0	5	10	0	5	10
Predicted clicks	57.4	19.5	6.9	57.4	34.5	19.5
Av. actual clicks	57.8	18.6	8.8	58.7	30.9	18.4
(Std. Dev.)	(12.2)	(9.44)	(5.02)	(12.5)	(15.8)	(9.80)
p-value	0.723	0.367	0.000	0.276	0.040	0.294

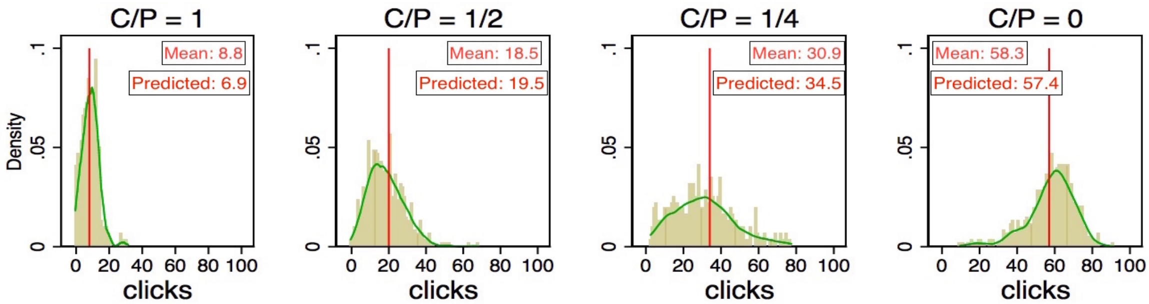
Note: P-values are based on two-tailed one-sample t-tests using a subject's average clicks per treatment as the unit of observation when testing against the predicted clicks.

Fig. 4 shows the predicted efforts and the distribution of actual efforts by combining categories whenever the treatments have the same predicted efforts. The distribution of efforts is approximately centered on the predicted effort in each case, but shows variability in choices for any given *C/P* ratio. As noted earlier, if the marginal product of effort is subject to individual-specific and idiosyncratic shocks variability in efforts is to be expected.

[11] Note that we have assumed a continuous production function. This assumption is made mainly for expositional and analytical convenience. In reality, the production function is a discrete relationship between catches and clicks.

[12] We also performed an out-of-sample test of predictions by comparing the actual number of clicks in an experimental session with the predictions derived from data from the other session. The results are essentially the same. See Appendix B for details.

Fig. 4 The distributions and the Kernel density distributions of the actual number of clicks and the predicted clicks



Note: The vertical line in each panel represents the predicted effort.

3.5. Discussion

Real effort tasks have the advantage that they offer subjects something tangible to do rather than just choosing among abstract options. The potential cost to the experimenter is loss of control because subjects might experience unobserved psychological benefits or costs. Thus, there is a tradeoff between "realism" and experimental control. The ball-catching task mitigates this tradeoff because it allows for real effort *and* control over important parameters, such as the production function and the cost function. This feature is particularly important if the experimenter wants to test theoretical predictions, in particular, point predictions. Existing real effort tasks typically allow at best for comparative static predictions, but not point predictions, because the latter requires full control over *all* costs and benefits, be they material or psychological.

Psychological costs and benefits always exist to some degree because *any* decision environment inevitably triggers emotions and requires some cognitive effort. Arguably, these psychological effects are stronger in real effort experiments than in abstract induced value settings. Smith (1982) (in particular pp. 930-934) was well aware of these non-monetary costs and benefits and argued that the "precepts" of induced value experiments will provide the necessary control of the experimental environment. The precepts are *non-satiation* in the reward medium (money), *salience* (rewards in the experiments should depend on decisions), and in particular *dominance* (the "reward structure dominates any subjective costs (or values) associated with participation in the activities of the experiment" (p. 934). It is the control over costs and benefits that renders experiments an informative tool to test economic theories -- be it an abstract induced value experiment or a real effort experiment. Satisfying *dominance* may be harder to achieve in real effort experiments than in induced value experiments.

Thus, the usefulness of the ball-catching task to test economic theories requires that dominance holds: psychological costs and benefits should be relatively small and dominated by pecuniary payoff considerations. In our piece-rate setting, "small" means that, in a statistical sense, efforts should be homogeneous of degree zero in those costs and prizes which the experimenter can manipulate. Our results unambiguously support this requirement. Thus, the ball-catching task has passed a first important test for its usefulness to test economic theories.

As a second test, we derived further comparative static predictions about how effort levels should vary with changing costs and prizes. The results strongly support the comparative static predictions. Theory also predicts that if effort costs are zero, people should catch as many balls as possible and prizes should therefore not matter, which is what we observe. Thus, the ball-catching task also passes this second test.

The third and most demanding test is whether observed (average) behavior also follows point predictions. This is the case and thus the ball-catching task also passes this third test. We thus conclude that the ball-catching task is in principle suitable for theory testing purposes, if the researcher thinks that for his or her research question a real effort design is desirable.

A complementary way of looking at the experiments reported in this section is to see them as a test in its own right of piece-rate incentive theory. In its most simplified version, the first-order condition of optimal effort under piece-rate incentives is expressed in equation (1). Our experiment provides an environment to put the comparative static predictions from (1) as well as effort level predictions to a test. The experimental environment controls the production process (the ball dropping), the costs of clicking to catch balls, as well as the piece rates (the prizes) for each catch. Tests using field data, even those that have unusually detailed data such as Lazear (2000), typically do not have detailed information about effort costs that are necessary to predict effort levels. The ball-catching task can accommodate assumptions about effort costs (e.g. the cost consequences of ability differences) in the induced cost valuations given to subjects. The ability of the ball-catching task to control all aspects of the environment allows a complete behavioral characterization of all predictions of piece-rate theory, not just the comparative statics. Our results provide a comprehensive vindication of piece-rate theory.

In the next section, we provide further tests for the suitability of the ball-catching task by investigating its performance in well-known experimental settings that hitherto have typically used induced-value designs. This will be a further opportunity to see whether the ball-catching task produces behavior that is consistent with equilibrium point predictions, or with comparative static predictions.

4 Study 2: Applications - team production, gift exchange and tournament

The previous section has demonstrated the accuracy of predictions on effort using the ball-catching task in an individual decision making task. In this section, we use the ball-catching task in three classic interactive experiments that have been used to study cooperation, reciprocity, and competition. We chose these applications for several reason. First, they represent important classes of experimental games using induced value designs. Second, they allow for further tests of theoretical point predictions and/or of comparative static predictions in interactive settings. Third, they illustrate the versatility of the ball-catching task with regard to manipulations of the production function and the induced values for the cost function. We will utilize the estimated production function from Study 1 to derive predictions on effort whenever possible.

We ran five sessions, each with 32 subjects, for a total of 160 subjects. In each session two unrelated treatments were conducted, each involving ten repetitions of a task. Details of the treatments are specific to each session and will be explained separately below. Instructions for the second treatment were given after the first treatment was completed. At the end of each session, a post-experimental questionnaire was administered asking for subjects' perception of the ball-catching task, including its difficulty, enjoyableness and boredom. All the sessions were run at the CeDEx lab at the University of Nottingham. Sessions lasted no more than one hour and the average earnings were around £13.00.[13]

4.1 Team production

The understanding of free-riding incentives in team production is at the heart of contract theory and organizational economics (Holmstrom (1982)). A standard experimental framework for studying team production is the voluntary contribution mechanism in which the socially desirable outcome is in conflict with individual free-riding incentives (see a recent survey in Chaudhuri (2011) in the context of public goods).

Our team production experiment was run over three sessions. One session included a team production (TP) treatment, in which four team members worked on the ball-catching task independently over 10 periods. The same four subjects played as a team for the entire 10 periods. For each ball caught, the subject contributed 20 tokens to team production while he/she had to bear the cost of clicking, with a cost per click of 5 tokens. At the end of each period, total team production was equally shared among the four team members. Each member's earnings were determined by the share

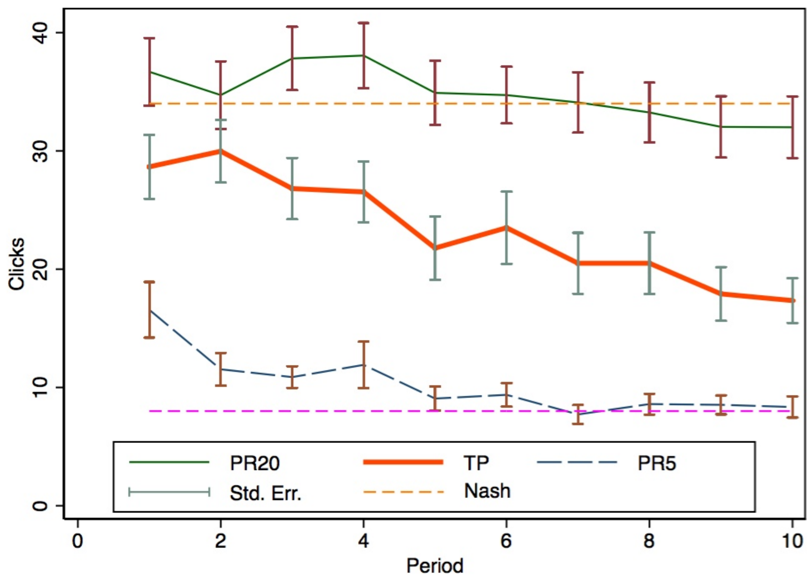
[13] Four of the treatments were unrelated to this paper and are not reported. Instructions of all reported experiments are reproduced in Appendix A2.

of the production net of the individual cost of clicking. Note that an individual's marginal benefit from another catch is 5 tokens, whereas the marginal benefit accruing to the entire group is 20 tokens. The other two sessions included control treatments where individuals play 10 periods according to a simple individual piece-rate. In the first treatment (PR20) an individual receives a prize per catch of 20 tokens and incurs a cost per click of 5 tokens. The second treatment (PR5) has a prize per catch of 5 tokens and a cost per click of 5 tokens.

Effort provision in PR5 gives a "selfish" benchmark for the TP treatment, while effort provision in PR20 gives an "efficiency" benchmark. If a subject in the TP treatment is only concerned about her own private costs and benefits from clicking and catching, and equates marginal costs to marginal private benefits, she should provide the same effort as in PR5. On the other hand, if she is concerned about total team production and equates marginal costs to marginal social benefits, then she should provide the same effort as in PR20. Our hypothesis is that free-riding incentives would drive effort towards the selfish benchmark, as is observed in many similar experiments using induced values (e.g., Nalbantian and Schotter (1997) and many public goods experiments using voluntary contribution mechanisms).

Fig. 5 displays the average numbers (± 1 SEM) of clicks in the three treatments. The two horizontal lines represent the Nash predictions on optimal effort levels in PR20 and PR5 respectively (using the estimated production function from Study 1 to compute the optimal effort levels).

Fig. 5 Average clicks over time in team production



The figure shows a clear declining average effort over time in TP. The average effort decreases from 30 clicks to just above 17 clicks in the last period. By comparison, the average effort in PR20 decreases from 38 to 32 and in PR5 from 16 to 8 and thus is consistent with our findings in Study 1.

Subjects in TP under-provide effort, relative to the efficiency benchmark, from the very first period and steadily decrease their efforts. Even in the final period, however, the average effort exceeds the extreme selfishly optimal level. This empirical result is qualitatively similar to previous findings from experiments using induced values, such as Nalbantian and Schotter's (1997) revenue sharing treatment and many public goods experiments, and also from some real effort experiments on team incentives (e.g., Corgnet, et al. (2015)).

4.2 Gift exchange

The gift exchange experiment (Fehr, et al. (1993)) examines reciprocal behavior between subjects in the role of firms and subjects in the role of workers. The gift exchange game using induced value techniques has been a workhorse model for many experimental investigations of issues in labor economics and beyond (see Gächter and Fehr (2002); Charness and Kuhn (2011) for surveys).

Our version of the bilateral gift exchange experiment follows Gächter and Falk (2002), except that they used induced values whereas we use the ball-catching task and slightly different parameters which we deem more suitable for the present purpose. In our experiment, in each period the firm offers a wage between 0 and 1000 tokens to the matched worker who then works on the ball-catching task. Each ball caught by the matched worker adds 50 tokens to the firm's payoff. The worker's payoff is the wage minus the cost of clicking. To compensate for possible losses, every firm and worker received 300 tokens at the beginning of each period. We implemented the gift exchange game in two sessions, one using a treatment with stranger matching over ten periods and the other using a treatment with partner matching over ten periods.

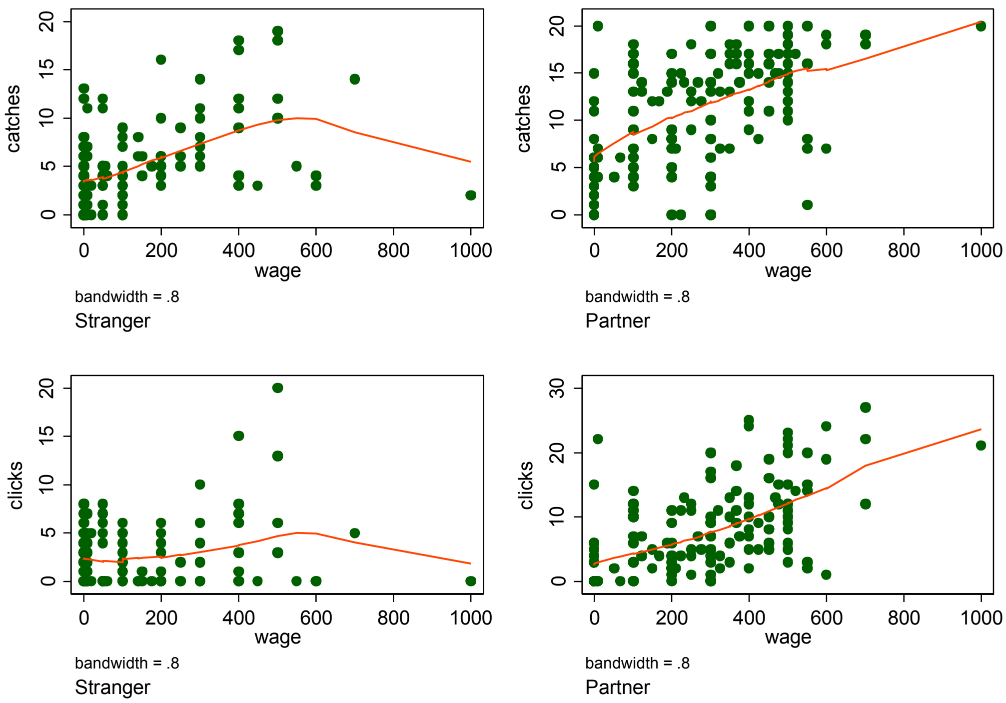
We made two key changes to the task compared with the version used in Study 1. First, we reduced the number of balls that could be caught within 60 seconds from 52 to 20 by increasing the time interval between falling balls. We made this change because we wanted to reduce the influence of random shocks as much as possible. The change makes it easy for a subject to catch every ball so that reciprocal behavior by workers could be reflected in their efforts as well as in their actual outputs. Second, the cost schedule was changed to a convex function in accordance with the parameters used in most gift exchange experiments. The cost for each click is reported in Table 5. For example, the 1st and 2nd clicks cost 5 tokens each, the 3rd click cost 6 tokens, etc., and finally the last column with No. 30+ means that the 30th and any further clicks cost 12 tokens each. Notice that if, for example, the worker makes a total of three clicks she will incur a total cost of 5 + 5 + 6 = 16 tokens.

Table 5 The cost schedule in gift exchange

No. of Click	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Cost	5	5	6	6	6	7	7	7	7	8	8	8	8	8	9
No. of Click	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30+**
Cost	9	9	9	9	10	10	10	10	10	11	11	11	11	11	12

Based on many gift exchange experiments and in particular the results by Gächter and Falk (2002) and Falk, et al. (1999) who also compared partners and strangers in gift exchange, we expect gift exchange and predict that the reciprocal pattern is stronger with partner matching where it is possible to build up a reputation between a firm and a worker. Fig. 6 confirms both predictions. It shows the relationship between outputs and wages on the upper panel and the relationship between efforts and wages on the lower panel. The data suggests a clear reciprocal pattern in both treatments and an even stronger pattern in the partner treatment whether we look at outputs or efforts.

Fig. 6 Reciprocal patterns in gift exchange



Note: the upper panel shows the relationship between outputs and wages in both treatments and the lower panel displays the relationship between efforts and wages. The relationship in the stranger matching treatment is shown in the left panels and in the partner matching treatment in the right panels. The fitted lines are estimated from non-parametric Lowess regressions.

For formal statistical tests we estimate the following random effects panel data model for the number of clicks on the wage received:

$$Click_{i,r} = \beta_0 + \beta_1 wage_{i,r} + \omega_i + \delta_r + u_{i,r}$$

where ω_i is an individual-specific random effect identically and independently distributed over subjects with a variance σ_ω^2, δ_r denotes a period dummy for the r^{th} period (with the first period providing the omitted category), and $u_{i,r}$ is a disturbance term, assumed to be identically and independently distributed over subjects and periods with a variance σ_u^2.

Table 6 reports the estimates for both treatments and also for the pooled sample with an additional interaction term. Consistent with gift exchange reciprocity and the graphical evidence from Fig. 6, workers in both treatments respond to higher wages by exerting higher levels of effort which differ systematically from zero effort. Furthermore, the strength of reciprocity is stronger with partners than strangers as the interaction term between the wage received and the treatment dummy in the column (3) is highly significant. These results in our ball-catching gift exchange experiment are qualitatively similar to findings from induced value experiments in Falk, et al. (1999) and Gächter and Falk (2002). Our results from the stranger treatment are also consistent with another early real effort gift exchange experiment by Gneezy (2002) who used a maze solving task (without induced values) to measure worker's performance, although Gneezy's experiment was conducted in a one-shot setting.

Table 6 Random effects regressions for worker's clicks in gift exchange

Dep. Var.: **Clicks**	Coefficient estimates (std. err.)		
	(1)Stranger	(2) Partner	(3) Pooled
Wage	0.003** (0.001)	0.014*** (0.003)	0.004** (0.002)
Partner			1.154 (0.797)
Wage × Partner			0.014*** (0.003)
Intercept	3.279*** (0.681)	3.746*** (1.444)	2.200** (0.952)
σ_ω	1.753	3.346	2.293
σ_u	2.649	4.397	3.972
Hausman test for random versus fixed effects	df=10 p=1.000	df=10 p=0.956	df=11 p=0.984
N	160	160	320

Note: All period dummies are included and are all statistically insignificant. Partner is a binary indicator which equals 1 if the treatment is the partner matching and 0 if the stranger matching.
***p < 0.01, **p < 0.05.

4.3 Tournament

Tournament incentive schemes, such as sales competitions and job promotions, are an important example of relative performance incentives (Lazear and Rosen (1981)). One early laboratory experiment by Bull, et al. (1987) found that tournament incentives indeed induced average efforts in line with theoretical predictions. But the variance of behavior was much larger under tournament incentives than under piece-rate incentives. Many induced value tournament experiments have been conducted since (see Dechenaux, et al. (2014) for a survey).

In one session we included a simple simultaneous tournament treatment. The 32 subjects were randomly matched into pairs in a period and each pair competed in the ball-catching task for a prize worth 1200 tokens. The winner earned 1200 tokens net of any cost of clicking, whereas the loser received 200 tokens net of any cost of clicking. The cost per click was always 5 tokens. Each player's probability of winning followed a piecewise linear success function (Che and Gale (2000); Gill and Prowse (2012)): $prob\{win\}$ = (own output – opponent's output + 50)/100. This procedure was repeated over 10 periods.

We use this contest success function because it allows us to make a simple point prediction on the expected effort. This is because the specified piecewise linear success function implies that an additional catch increases the probability of winning by 1/100. Thus, the marginal benefit of clicking is equal to the prize spread between the winner prize and the loser prize, 1000, multiplied by 1/100, multiplied by the marginal product of a click. The marginal cost of clicks is 5 tokens. Once again, we simply utilize the estimated production function from Study 1 to compute the optimal effort level which turns out to be 20 clicks. Notice that while an additional catch increases earnings by 10 tokens in treatment 2 of Study 1, here an additional catch increases *expected* earnings by 10 tokens.

Fig. 7 Average clicks over time in tournament

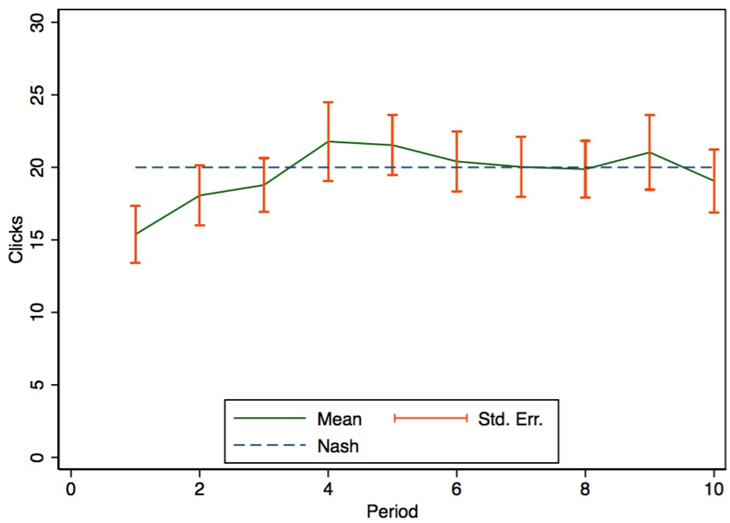


Fig. 7 displays the average efforts (± 1 SEM) across all subjects and periods. We observe quick convergence towards to the predicted effort level. The variance of effort in tournament also appears to be larger than that observed in treatment 2 of Study 1. The standard deviation of effort is around 12 in the former and 9.4 in the latter, perhaps reflecting the stochastic nature of the relationship between catches and earnings under tournament incentives.[14] Both results are qualitatively similar to previous findings from Bull, et al. (1987).

4.4 Discussion

The three experiments reported here showcase the implementation and versatility of the ball-catching task in three classic experimental paradigms that have been studied extensively in induced value experiments: team production, gift-exchange, and tournaments. In all three experiments the results are closely in line with findings from their induced value counterparts. Particularly noteworthy is that equilibrium predictions, derived from the production function of Study 1, are closely met in all cases where we could derive an equilibrium prediction (the piece-rate treatments of the team work experiment, and in the tournament). We also confirm the theoretical comparative static prediction that in the gift-exchange game a fixed matching should lead to stronger reciprocity than random matching. We see this as a strong encouragement for the suitability of the ball-catching task in potentially many more settings. The chosen experiments also demonstrate the versatility of the ball-catching task to manipulate the production technology and the cost function.

We conclude this section by discussing one central feature of the ball-catching task, namely its ability to control effort costs by inducing any effort cost function the experimenter deems appropriate. Recall that effort costs in economic models of labor supply denote any cost a worker might incur, physiological, psychological, or simply opportunity costs of foregone leisure. Existing real effort experiments have tried to model opportunity costs of effort by offering the subjects outside options, for example the opportunity to surf the Internet (Corgnet, et al. (2014)), to receive paid time-out for a few seconds (Mohnen, et al. (2008)), to work on other productive individual tasks (van Dijk, et al. (2001)), or to leave the task earlier than the deadline (e.g., Abeler, et al. (2011)). This method exploits the possibility of a trade-off between effort and off-the-job leisure and, indeed, there is experimental evidence that subjects make such a trade-off in response to different incentive schemes (see Corgnet, et al. (2014) and Eckartz (2014)). However, compared to the ball-catching task which in

[14] This difference in variability of effort between tournament and piece-rate incentives is smaller than that found by Bull, et al. (1987) in their induced value experiment, in which the standard deviation of effort under tournament incentives was more than double that under piece-rate incentives. Quantitative comparisons between their study and ours, however, should be treated cautiously as there are numerous differences between studies (e.g. we use a piece-wise linear contest success function, whereas they use a rank-order tournament).

its most minimal version may take only one minute to complete, the "outside options" method usually requires a rather long duration for it to work well (sometimes up to 60 minutes as in Abeler, et al. (2011)), thus preventing us from collecting repeated observations in the duration of a typical laboratory experiment. Moreover, while outside options imply some real effort costs, it is still unclear how subjects value them exactly. The ability of the ball-catching task to induce any cost function, be it linear, or non-linear as in the gift-exchange experiment discussed above (Table 5), circumvents the problem of unknown valuations and retains the possibility of making point predictions on effort choices.

5 Study 3: An online version of the ball-catching task

5.1 The ball-catching task on Amazon Mechanical Turk

As a third test of the versatility of the ball-catching task, we introduce an online version. This online version is programmed in PHP and has been designed to resemble the lab version as closely as possible.[15] The purpose of this section is to show the potential (and limitations) of using the ball-catching task in online experiments, which increasingly appear to be a valuable complement to the experiments in the physical laboratory.

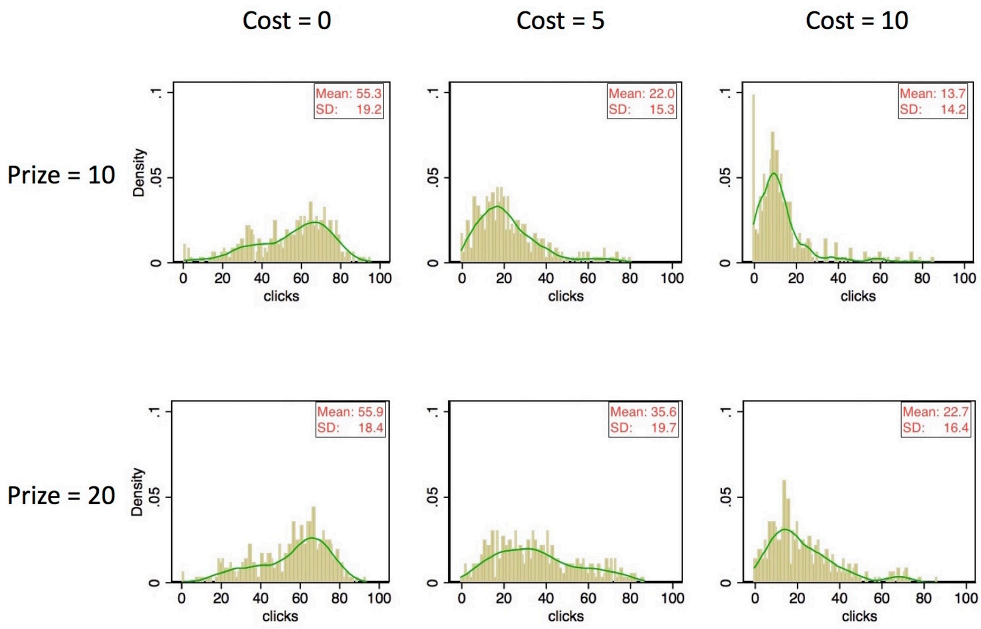
We ran the same experiment as in Study 1 on Amazon Mechanical Turk (MTurk).[16] In total, we recruited 95 subjects from MTurk and 74 of them finished the task. Recruitment took around 10 minutes. Given the unusually long duration of the task (50 minutes), the 78% completion rate suggests that our promised payment is sufficiently attractive to most of the workers on MTurk. The average payment, including a $3 participation fee, was around $5.90, which was well above what most MTurk tasks offered. The average age was 35 years, ranging from 20 to 66 years; and 52% were male.

Paralleling the presentation of Study 1 results, Fig. 8 summarizes the distribution and the Kernel densities of the number of clicks for each treatment. In general, we find that the comparative statics results are very similar to those in Study 1. A Wilcoxon signed-ranks test using a subject's average clicks per treatment as the unit of observations suggests that homogeneity of degree zero also holds here: the difference in clicks between the two treatments with the same C/P ratio is not systematic ($p=0.309$). The same is true for the difference in clicks between the two treatments with $C=0$ ($p=0.832$). Similarly, when comparing treatments with the same prize, Friedman tests indicate that comparative static predictions for different costs are supported ($p<0.001$ in both comparisons).

[15] See Appendix C for discussion of technical considerations associated with implementing the task.

[16] See Horton, et al. (2011) for a discussion of the usefulness of MTurk for experimental economists.

Fig. 8 The distributions and the Kernel density distributions of the number of clicks in Study 3



We observe some notable differences between the online and the lab version. The variance of clicking in each treatment for MTurkers appears to be higher than in the lab with student subjects. Moreover, we find that the production function is not invariant to prize levels, nor is it stable across sub-samples, thus preventing us from making meaningful point predictions.[17]

5.2 Discussion

The results from the online version strongly support the robustness of the ball-catching task with regard to all comparative statics predictions, including the crucial requirement of homogeneity of degree zero in C and P. This is encouraging and important support for the suitability of the ball-catching task.

However, the results from the online experiment also serve as an important caveat because they reveal that the environment where subjects take their decision might matter a great deal for the actual production function. In an online experiment, there are inevitably many differences compared to the physical laboratory: computer configurations (e.g., screen sizes and mice), speed of network connections, distractions in the working environment, etc. will vary strongly across online participants, but will typically be very similar for all subjects within a given physical laboratory. Physical labs might also differ, so the production function that can be used for deriving point predictions might also

[17] Analyses are available from authors upon request.

be lab specific. Hence, an important lesson is that for proper calibration of the production function pre-testing is necessary in whatever lab is used, physical or online.

6 Conclusion

In this paper we introduced a novel real effort task, the ball-catching task. Its greatest advantage over other real effort tasks lies in its versatility to manipulate the production technology and in particular in its ability to control effort costs. We presented three studies. Studies 1 and 3 showed that behavior in the ball-catching task in an individual decision making environment follows important comparative static predictions derived from incentive theory. Studies 1 and 2 suggest that the ball-catching task also has the potential to derive and test point predictions although Study 3 revealed that this most demanding feature of the ball-catching task requires careful calibration. Study 2 also showed that real effort behavior, elicited using the ball-catching task, strongly resembles behavior in experiments using induced value designs. Together, the three studies demonstrate that the ball-catching task is a potentially powerful tool for (theory testing) experiments in real effort environments.

Acknowledgements We are grateful for comments from Colin Camerer, Vincent Crawford, Robin Cubitt, Gianni De Fraja, Roberto Hernán-González and participants at several conferences and seminars. We also thank Lucas Molleman for help with the online experiment. Simon Gächter acknowledges support from the European Research Council Advanced Investigator Grant 295707 COOPERATION. Lingbo Huang acknowledges support from the ESRC Grant ES/J500100/1. Simon Gächter and Martin Sefton acknowledge support from the Network of Integrated Behavioral Science funded by the ESRC Grant ES/K002201/1.

References

Abeler, J., Falk, A., Goette, L., & Huffman, D. (2011). Reference points and effort provision. *American Economic Review*, *101*, 470-92.

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

Camerer, C. F., & Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty*, *19*, 7-42.

Charness, G., & Kuhn, P. (2011). Lab labor: What can labor economists learn from the lab? In *Handbook of labor economics*, ed. O. Ashenfelter, & D. Card. Amsterdam: Elsevier.

Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: A selective survey of the literature. *Experimental Economics*, *14*, 47-83.

Che, Y. K., & Gale, I. (2000). Difference-form contests and the robustness of all-pay auctions. *Games and Economic Behavior*, *30*, 22-43.

Corgnet, B., Hernán-González, R., & Rassenti, S. (2015). Peer pressure and moral hazard in teams: Experimental evidence. *Review of Behavioral Economics*.

Corgnet, B., Hernán-González, R., & Schniter, E. (2014). Why real leisure really matters: Incentive effects on real effort in the laboratory. *Experimental Economics*, 1-18.

Dechenaux, E., Kovenock, D., & Sheremeta, R. (2014). A survey of experimental research on contests, all-pay auctions and tournaments. *Experimental Economics*, 1-61.

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

Eckartz, K. (2014). *Task enjoyment and opportunity costs in the lab - the effect of financial incentives on performance in real effort tasks*. Jena Economic Research Papers No 2014-005

Falk, A., & Fehr, E. (2003). Why labour market experiments? *Labour Economics*, *10*, 399-406.

Falk, A., Gächter, S., & Kovacs, J. (1999). Intrinsic motivation and extrinsic incentives in a repeated game with incomplete contracts. *Journal of Economic Psychology*, *20*, 251-284.

Fehr, E., Kirchsteiger, G., & Riedl, A. (1993). Does fairness prevent market clearing? An experimental investigation. *Quarterly Journal of Economics*, *108*, 437-459.

Fischbacher, U. (2007). Z-tree: Zurich toolbox for readymade economic experiments. *Experimental Economics*, *10*, 171-178.

Gächter, S., & Falk, A. (2002). Reputation and reciprocity: Consequences for the labour relation. *Scandinavian Journal of Economics*, *104*, 1-26.

Gächter, S., & Fehr, E. (2002). Fairness in the labour market - a survey of experimental results. In *Surveys in experimental economics. Bargaining, cooperation and election stock markets*, ed. F. Bolle, & M. Lehmann-Waffenschmidt. Heidelberg: Physica Verlag.

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

Gneezy, U. (2002). *Does high wage lead to high profits? An experimental study of reciprocity using real effort*. The University of Chicago GSB, mimeo

Gneezy, U., & Rustichini, A. (2000). Pay enough or don't pay at all. *Quarterly Journal of Economics*, *115*, 791-810.

Greiner, B. (2004). An online recruitment system for economic experiments. In *Forschung und wissenschaftliches rechnen gwdg bericht 63*, ed. K. Kremer, & V. Macho. Göttingen: Gesellschaft für Wissenschaftliche Datenverarbeitung.

Holmstrom, B. (1982). Moral hazard in teams. *Bell Journal of Economics*, *13*, 324-340.

Horton, J. J., Rand, D. G., & Zeckhauser, R. J. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental Economics*, *14*, 399-425.

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

Lazear, E. P. (2000). Performance pay and productivity. *The American Economic Review*, *90*, 1346-1361.

Lazear, E. P., & Rosen, S. (1981). Rank-order tournaments as optimum labor contracts. *Journal of Political Economy*, *89*, 841-864.

Mohnen, A., Pokorny, K., & Sliwka, D. (2008). Transparency, inequity aversion, and the dynamics of peer pressure in teams: Theory and evidence. *Journal of Labor Economics*, *26*, 693-720.
Nalbantian, H. R., & Schotter, A. (1997). Productivity under group incentives: An experimental study. *American Economic Review*, *87*, 314-341.
Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *Quarterly Journal of Economics*, *122*, 1067-1101.
Prendergast, C. (1999). The provision of incentives in firms. *Journal of Economic Literature*, *37*, 7-63.
Smith, V. L. (1982). Microeconomic systems as an experimental science. *American Economic Review*, *72*, 923-955.
Van Dijk, F., Sonnemans, J., & Van Winden, F. (2001). Incentive systems in a real effort experiment. *European Economic Review*, *45*, 187-214.
Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, *13*, 75-98.

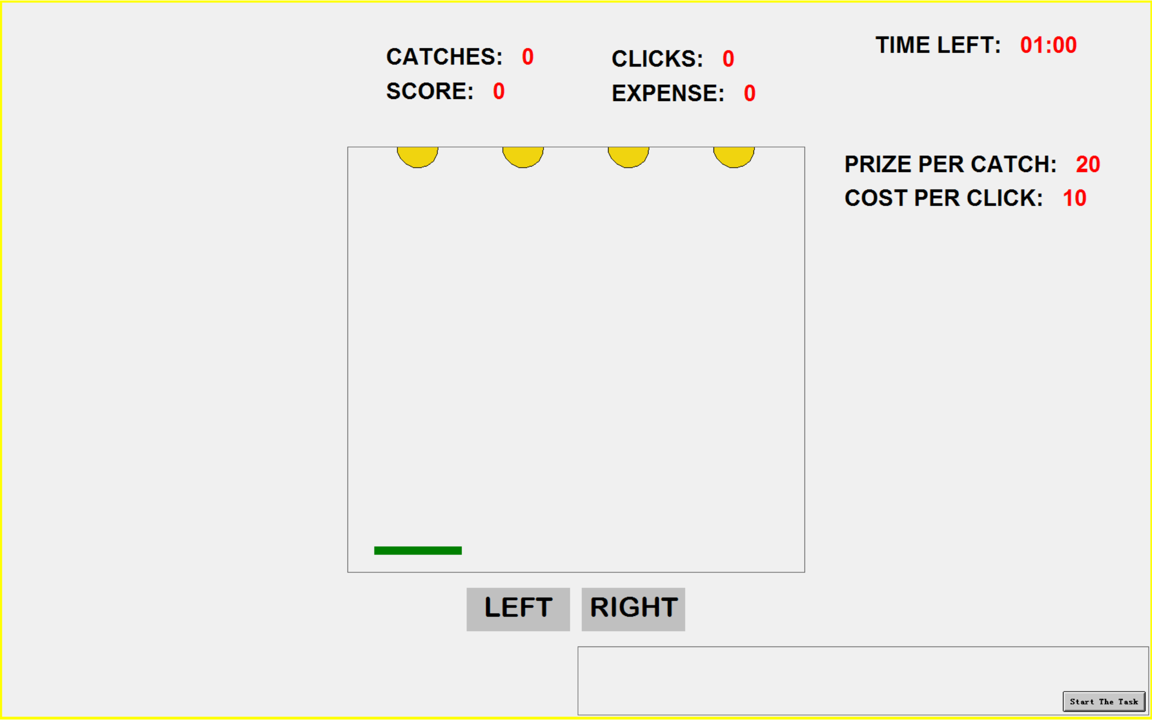
Appendix A. Experiment Instructions

Appendix A1. Instructions Used in Study 1

Welcome to the experiment. You are about to participate in an experiment on decision making. Throughout the experiment you must not communicate with other participants. If you follow these instructions carefully, you could earn a considerable amount of money.

For participating in this experiment you will receive a £3 show-up fee. In addition you can earn money by completing a task. Your earnings from the task will depend only on your own performance.

You will be asked to work on a computerized ball-catching task for 36 periods. Each period lasts one minute. In each period, there will be a task box in the middle of the task screen like the one shown below:



Once you click on the "Start the Task" button, the timer will start and balls will fall randomly from the top of the task box. You can move the tray at the bottom of the task box to catch the balls by using the mouse to click on the LEFT or RIGHT buttons. To catch a ball, your tray must be below the ball before it touches the bottom of the tray. When the ball touches the tray your catches increase by one.

You will receive a prize (in tokens) for each ball you catch and incur a cost (in tokens) for each mouse click you make. At the beginning of each period you will be informed of your prize for each ball caught, which will be either 10 or 20, and your cost for each click, which will be either 0, 5 or 10. In each period, the number of balls you caught so far (displayed as CATCHES), the number of clicks you made so far (CLICKS), your accumulated prizes so far (SCORE) and your accumulated costs so far (EXPENSE) are shown right above the task box. SCORE will be CATCH multiplied by your prize per catch for the period and EXPENSE will be CLICK multiplied by your cost per click for the period. At the end of the period your earnings in tokens for the period will be your SCORE **minus** your EXPENSE. Please note that catching more balls by moving the tray more often does not necessarily lead to higher earnings because both SCORE and EXPENSE matter for your earnings.

The first six periods will be practice periods which will not affect your earnings in the experiment in any way. At the end of the experiment, the tokens you earned from periods 7 to 36 will be converted to cash at the rate of 1200 tokens = 1 pound. You will be paid this amount in addition to your £3 show-up fee.

Appendix A2. Instructions Used in Study 2

General Information [All Treatments]

Welcome to the experiment. There will be two unrelated experiments for this session and there will be two separate instructions. The instructions for the second experiment will be distributed after the first experiment is ended. Throughout the session you must not communicate with other participants. If you follow these instructions carefully, you could earn a considerable amount of money. During the session your payment will be calculated in tokens.

At the end of each experiment tokens will be converted to cash at a rate of 1000 tokens = 1 pound. Your total payment for participating in this session will be the sum of your earnings in each experiment plus a £3 show-up fee. You will be paid this amount in cash at the end of the session.

If there is any question during the experiment, please raise your hand and someone will come to your desk to answer it.

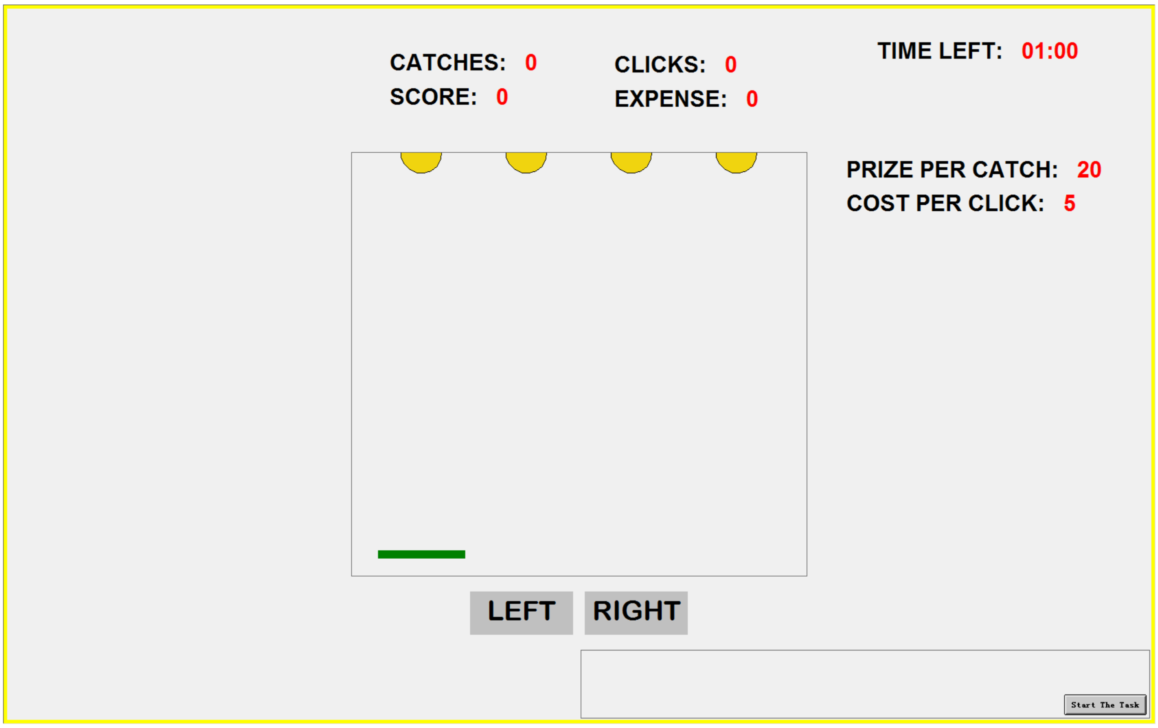
[Team Production]

The first experiment has 10 periods. Before the first period, you will be randomly assigned to a group of four participants. You will be in this group for the entire experiment.

In each period, you and each of the other three participants in your group will be asked to work on a computerised ball-catching task. Your earnings in each period will depend on the number of balls caught by you and the rest of your group as well as some personal expenses as detailed below.

Your Task in a Period

Each period lasts one minute. In each period, there will be a task box in the middle of the task screen like the one shown below:



Once you click on the "Start the Task" button, the timer will start and balls will fall randomly from the top of the task box. You can move the tray at the bottom of the task box to catch the balls by using the mouse to click on the LEFT or RIGHT buttons. To catch a ball, your tray must be below the ball before it touches the tray. When the ball touches the tray your catches increase by one.

For each mouse click you make you will incur a cost of 5 tokens.

For each ball you catch, you and the rest of your group will in total receive a prize of 20 tokens. Similarly, for each ball each of your group members catches, you and the rest of your group will in total receive a prize of 20 tokens.

In each period, the number of balls you have caught so far (displayed as CATCHES) and the number of clicks you have made so far (CLICKS) will be shown right above the task box. Also shown above the task box will be SCORE, which is CATCHES multiplied by the prize per catch, and EXPENSE, which is CLICKS multiplied by the cost per click.

How Your Earnings In Each Period Are Determined

When you and the other members of your group have finished the task, the computer will calculate the TOTAL SCORE of your group by adding up the four individual SCOREs in your group. Your earnings in tokens will be one-fourth of your group TOTAL SCORE minus your EXPENSE:

Your Earnings = (your group's TOTAL SCORE)/4 – EXPENSE.

Your SCORE and EXPENSE, your group's TOTAL SCORE, and your earnings for the period will be displayed on the screen at the end of each period.

Quiz

1. Suppose you have caught 20 balls by making 10 clicks. The rest of your group have caught 100 balls.
   What will be your EXPENSE?
   Answer: _____ tokens
   What will be your SCORE?
   Answer: _____ tokens
   What will be your group's TOTAL SCORE?
   Answer: _____ tokens
   What will be your earnings in this period?
   Answer: _____ tokens

2. Suppose you have caught 40 balls by making 40 clicks. The rest of your group have caught 20 balls.
   What will be your EXPENSE?
   Answer: _____ tokens
   What will be your SCORE?
   Answer: _____ tokens
   What will be your group's TOTAL SCORE?
   Answer: _____ tokens
   What will be your earnings in this period?
   Answer: _____ tokens

[Gift Exchange]

The first experiment has 10 paying periods. Before the first paying period, each participant will be randomly assigned to one of two groups: half will be "workers" and half "firms". You will remain either a worker or a firm throughout this experiment. In each paying period a firm will be randomly matched with a worker. Thus, [*Stranger treatment*: **you will be matched at random with another participant from period to period.** *Partner treatment*: **you will be matched with the same**

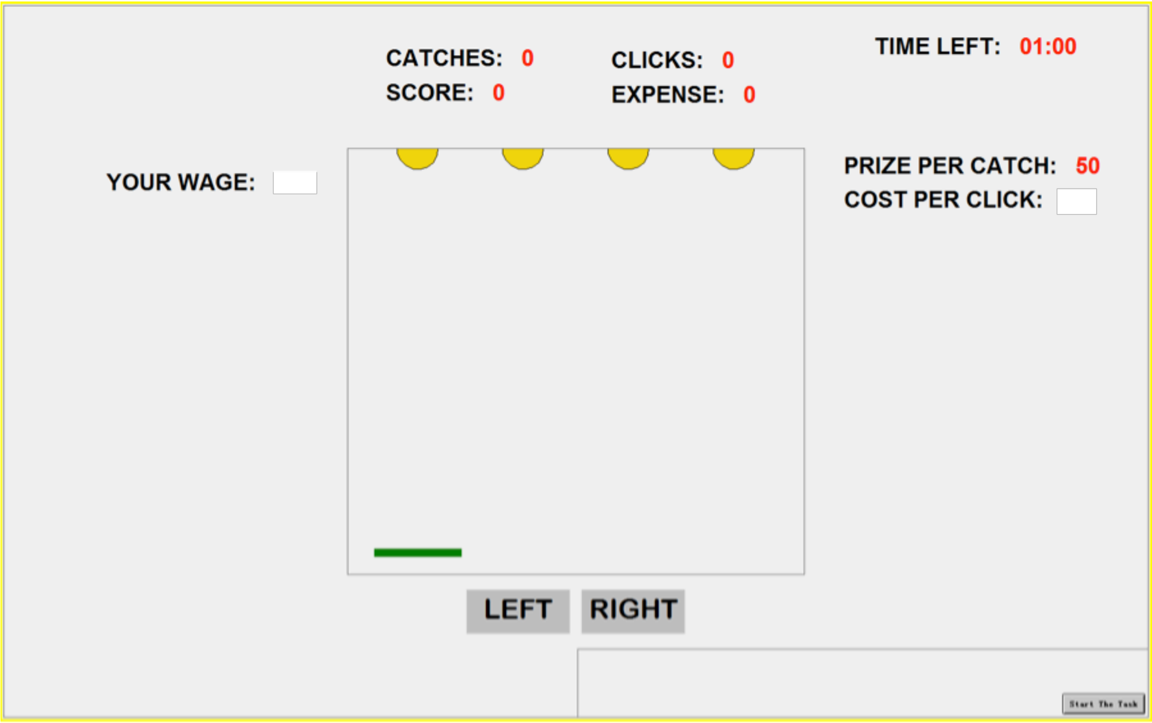
participant for the entire experiment.] All firms and workers and the information on pairings will remain anonymous throughout the experiment.

Each paying period consists of two stages:

Stage 1: A firm will make a wage offer to the worker.
Stage 2: The worker will work on a task. The exact procedure is described below.

How do you calculate the firm's and worker's earnings in each paying period?

1. In each paying period, every firm and worker will receive 300 tokens.
2. Each firm may choose any integer number between 0 and 1000 as the wage in tokens that the firm offers to her paired worker.
3. After the firm has made a wage offer to her matched worker, the worker will be asked to work on a computerized ball-catching task. In each period, there will be a task box in the middle of the worker's task screen like the one shown below:



4. Each task lasts one minute. Once the worker clicks on the "Start the Task" button, the timer will start and balls will fall randomly from the top of the task box. The worker can move the tray at the bottom of the task box to catch the balls by using the mouse to click on the LEFT or RIGHT buttons. To catch a ball, her tray must be below the ball before it touches the tray. When the ball touches the tray the worker's catches increase by one.
5. For each ball the worker catches, her matched firm will receive a prize of 50 tokens.
6. The worker will incur a specific cost for each mouse click she makes. The cost for each mouse click is shown below.

No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Cost	5	5	6	6	6	7	7	7	7	8	8	8	8	8	9

No.	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30+
Cost	9	9	9	9	10	10	10	10	10	11	11	11	11	11	12

For example, the column with No.6 means that the 6th click costs 6 tokens, the column with No.21 means that the 21st click costs 10 tokens, and finally the last column with No.30+ means that the 30th and any further click costs 12 tokens. Notice that if, for example, the worker makes a total of three clicks she will incur a total cost of 5 + 5 + 6 = 16 tokens.

7. The more clicks the worker makes, the more balls she may catch. Typically, if the worker decides to incur no cost by not moving the tray at all, she may still catch 4~6 balls. If the worker decides to catch every ball that she can, she may be able to catch 20 balls but she may need to click 20~30 times and incur a corresponding EXPENSE.
8. In each period, the number of balls the worker has caught so far (displayed as CATCHES), the number of clicks she has made so far (CLICKS) will be shown right above the worker's task box. Also shown above the task box will be SCORE, which is CATCHES multiplied by the prize per catch, and EXPENSE, which is the total cost of CLICKS.
9. After the worker has completed the task, her earnings in tokens for the period will be determined by the following formula:

   Worker's Earnings = 300 + Wage – EXPENSE.

10. The firm's earnings in tokens for the period will be determined by the following formula:

    Firm's Earnings = 300 – Wage + SCORE.

Practice periods

Before starting the paying periods, there will be three practice periods in which every participant will play the role of the worker. These practice periods are meant to familiarise yourselves with the task, and to see how CLICKS are translated into CATCHES. The earnings in these practice periods will not affect your total payment for the experiment.

In each practice period, you will receive a wage, which is either 100, 500, or 900 tokens. You will then be asked to work on the ball-catching task as described above. Your earnings in each practice period will be calculated following the same rule above.

You will discover whether you will be a firm or a worker in paying periods after these practice periods.

Quiz

1. Suppose that a firm offers a worker a wage of 800 tokens. The worker catches 20 balls by incurring the EXPENSE of 300 tokens. What will be the firm's and worker's earnings respectively?
   Worker's earning = ______tokens
   Firm's earning = ______tokens

2. Suppose that a firm offers a worker a wage of 300 tokens. The worker catches 15 balls by incurring the EXPENSE of 160 tokens. What will be the firm's and worker's earnings respectively?
   Worker's earning = ______tokens
   Firm's earning = ______tokens

[Tournament]

The second experiment has 10 periods. Before the first period, you will be randomly paired with another participant. You will be in this pair for the entire experiment.

Your Task in This Experiment

In each period, you and your paired participant will complete the same ball-catching task as in the first experiment.

You will receive a score of 10 tokens for each ball you catch and incur a cost of 5 tokens for each mouse click you make. That means SCORE = 10×CATCHES and EXPENSE = 5×CLICKS

How Your Earnings In Each Period Are Determined

In each period, the person with higher SCORE in each pair will have a higher probability of being the winner. If you are the winner, you will earn 1200 tokens minus your EXPENSE for the period. If you are the loser, you will earn 200 tokens minus your EXPENSE for the period.

Your probability of winning will depend on the difference between your SCORE and that of your paired participant and some element of chance.

Specifically, say that the SCOREs of you and your paired participant are S1 and S2 respectively. Then your probability of winning the award is calculated as $(S1 – S2 + 500)/1000$ and the probability of winning of your paired participant is correspondingly calculated as $(S2 – S1 + 500)/1000$. That means, if the SCOREs are the same, both of you will have a 50% chance of being the winner. If the SCOREs are not the same, the chance of winning for the pair member with the higher SCORE increases by 1 percentage point for every increase of 10 in the difference between the SCOREs, while the chance of winning for the pair member with the lower SCORE correspondingly decreases by 1 percentage point.

Your SCORE, the SCORE of your paired participant, your EXPENSE, the EXPENSE of your paired participant, your probability of winning, whether you were the winner or the loser of the period and your earnings will be displayed on the screen at the end of each period.

Appendix B. Additional Tables and Figures in Study 1

Table B1: Average number of clicks in each treatment by subject

Subject	*1:(10,0)*	*2:(10,5)*	*3:(10,10)*	*4:(20,0)*	*5:(20,5)*	*6:(20,10)*	*Consistent behavior?*
101	59.8	27.4	12	64.6	45	27.6	√
102	62.6	18.4	12.2	63.6	30.2	20	√
103	51	14.8	9.4	65	22.8	14.2	√
104	21.6	6	3	20.2	5.4	4.4	√
105	62	15.8	3.6	60.6	18	16	√
106	56.8	18.8	9	58.6	34.8	19.2	√
107	57.2	20.8	11.2	57.8	29.6	26	√
108	47.4	13.4	11.6	54.6	25.8	10.8	√
109	62.6	18.8	9.6	65	26	14.2	√
110	60.2	20.8	7.2	61.4	39	23.8	√
111	61.4	5.8	5	63.8	9	7	√
112	48.6	6.4	2.8	49.2	8.4	6.8	√
113	64.6	25	16	69.8	48.2	29.6	√
114	49.2	17.4	14.6	59	24.2	16.8	√
115	64	18.8	9.2	64.4	30.2	21.4	√
116	63.8	29.8	11	66	52.6	29.8	√
201	56.4	36.2	13.2	53	39.2	34.6	√
202	38.2	9.4	4.6	38	11.6	7	√
203	61.8	19.2	11.4	58.4	27.6	17	√
204	64	24.4	6.8	55	37.8	23.2	√
205	65.4	21	10.6	58.4	37.2	16.6	√
206	64.6	24.6	11.4	61.6	45.6	25.8	√
207	67.2	31.4	14.8	60.8	53.2	32.2	√
208	53	19.6	6.8	53	27.8	17.4	√
209	51.2	28.6	5.4	49.2	43.6	22.2	√
210	72.8	10.6	8	65.2	16.8	10.6	√
211	62.4	20.8	8.8	63	40	22.4	√
212	63.2	23.4	10.2	68.4	53	28.6	√
213	63.4	20	12.2	58	31.4	19.2	√
214	65.8	30.4	18.8	64.2	24.2	25.4	×
215	59.2	24.8	13.6	53.4	36.6	25.2	√
216	58.4	7.6	3.8	59.6	31.8	12.2	√
301	61.2	3.4	3.2	60.8	5.6	13.4	×
302	63.2	16.4	9	70	21.8	13	√
303	46.8	17.2	11	50.8	32.2	11.6	√
304	38.2	5	1.4	30.6	5.6	1.4	√
305	52.4	14.2	7.2	60	23.6	13.2	√
306	67	14.6	7.4	63.2	36.2	13.6	√
307	56.6	16.8	5.6	64.4	29.6	17.2	√
308	44.6	29.4	8.6	48.6	27.8	28.6	×
309	52.8	7.4	4	56.6	37	9.4	√
310	44	17.2	9.8	50.2	29.8	21.8	√
311	52.6	15.8	7.4	66.4	29.6	11.8	√
312	69.4	13.8	6.2	62.4	26.4	9.8	√
313	60.8	17	7.4	71.6	40.4	23.6	√
314	56.6	23.6	10	64.6	56.8	24	√
315	58.4	21.8	8.2	63	23	16.2	√
316	63.2	15.8	15.2	59.8	23.4	18.6	√
401	59.6	13.2	13.2	64.4	15.6	16.8	√
402	65	20.8	8.6	66.2	42.2	23.2	√
403	41.2	16.2	10.4	42.8	15.4	13	√

404	57.2	24.4	5.2	60.2	38.4	27.6	√
405	51.6	16.8	8	54	25.8	15	√
406	62.4	25.8	11.6	69	58.4	29	√
407	67	8.2	2.8	69	10	6.8	√
408	67.6	17.2	7.2	62.6	28.6	13.8	√
409	62.4	16	4.8	58.6	42	14.6	√
410	47.4	11.4	8.6	35.6	13.6	11.6	√
411	54.4	13	6.4	49.4	20.4	13.6	√
412	66.4	34.8	0	62.6	52.2	41.4	√
413	71.8	13.8	13.4	67.6	23.2	16.8	√
414	62	27.6	9.8	64.2	45	30	√
415	48.4	16	11.8	55.6	22.8	15.2	√
416	67.8	37.2	11.8	70.4	66	17.8	√

Note: we consider the average clicks of treatment 2 and 6 and average clicks of treatment 1 and 4 when evaluating consistency at the individual level.

Fig. B1 Fitted production functions for sub-samples with different prizes using model (2)

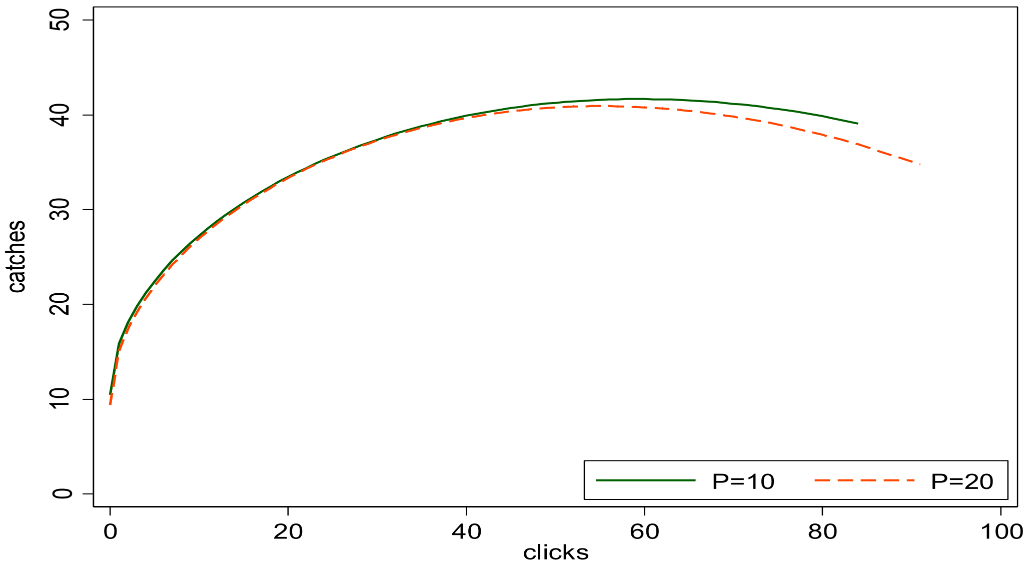


Fig. B2 Fitted production functions for sub-sample from different sessions using model (2)

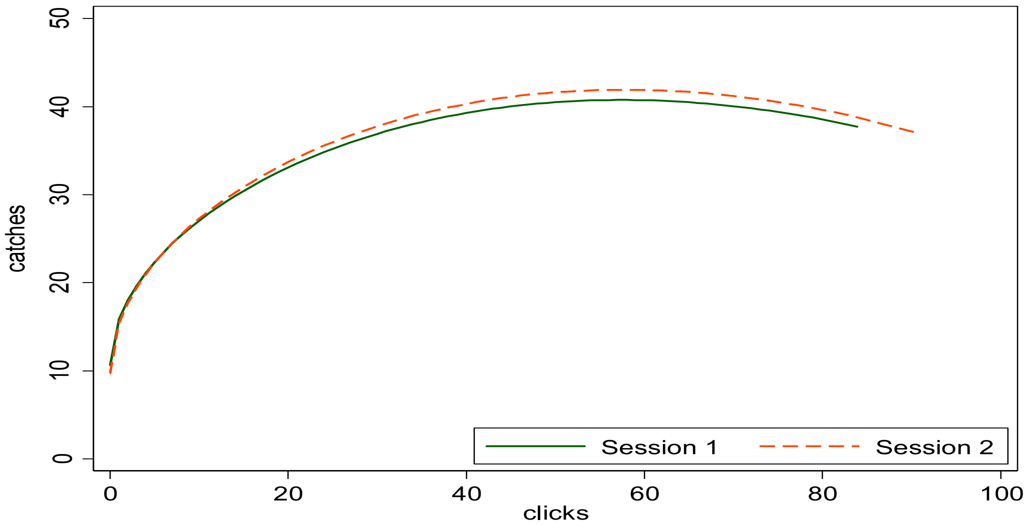


Table B2 Panel data regression estimates of model (2) for separate sessions

Dep. Var.: **Catches**	Coefficient estimates (std. err.)	
	(1) Session 1	(2) Session 2
$Intercept$	10.544*** (0.324)	9.730*** (0.326)
$Clicks^{0.5}$	5.315*** (0.189)	5.650*** (0.183)
$Clicks^{2}$	-0.003*** (0.000)	-0.003*** (0.000)
σ_{ω}	0.427*** (0.060)	0.287*** (0.046)
σ_{u}	0.785*** (0.018)	0.795*** (0.019)
N	959	946

Note: All period dummies are included and insignificant in both sessions. ***$p < 0.01$

Table B3 Comparisons between the predicted and actual number of clicks by session

Treatment No.	**1**	**2**	**3**	**4**	**5**	**6**
Prize per catch (P)	10	10	10	20	20	20
Cost per click (C)	0	5	10	0	5	10
Predicted clicks (S2)	57.6	20.1	7.3	57.6	35.1	20.1
Av. actual clicks (S1)	58.1	19.7	9.6***	58.2	31.5	19.6
(Std. Dev.)	(12.0)	(9.11)	(5.06)	(12.4)	(14.8)	(9.90)
Predicted clicks (S1)	57.4	18.7	6.5	57.4	33.9	18.7
Av. actual clicks (S2)	57.5	17.6	8.0***	59.2	30.3	17.3
(Std. Dev.)	(12.4)	(9.66)	(4.86)	(12.6)	(16.8)	(9.60)

Note: The full sample is equally separated into two halves by session: S1 and S2. We test the average actual clicks from S2 against the predicted clicks based on S1 and the average actual clicks from S1 against the predicted clicks based on S2. P-values are based on two-tailed one-sample t-tests against the predicted clicks using a subject's average clicks per treatment as the unit of observation. ***$p < 0.01$

Appendix C. Technical Appendix

Here, we describe the working and functionality of the ball-catching task, both the z-Tree version and the online version, in more detail and also give suggestions about how to implement the task in experiments. Both the z-Tree version and the online version are available from the authors upon request.

The *z-Tree version* of the ball-catching task allows experimentalists to manipulate the speed of falling balls and the time interval between falling balls directly in the global table in z-Tree. Changes to the layout of the task, such as the number of columns, height and width of the task box and the falling pattern, however, require more involved re-programming of the task. In the version used in this paper, the falling pattern is random. There are in fact four independent balls falling within a fixed time interval. Once a ball is caught or touches the bottom of the task box, it will reappear in a randomly selected column and fall again.

The z-Tree version has been tested using z-Tree 3.3.8 and later versions. The ball-falling and the tray-moving may become more sluggish with an increase in the number of z-Leafs simultaneously running the ball-catching task. In our experiments, we connected at most 16 z-Leafs to one z-Tree. A session with 32 subjects as in our Study 1 was accomplished by simultaneously opening two z-Trees in two separate master computers, each of which is connected with 16 z-Leafs. By affecting the level of sluggishness subjects may experience the number of connected z-Leafs may affect the production function. Other factors that may affect subjects' performance include the size of the task displayed on the specific computer screen, pixel resolutions of computer monitors, mouse configurations, etc. It is, therefore, advisable to test the software thoroughly in the lab where the actual experiment will be run. This will help for calibration of the production function to allow for accurate point predictions.

The *online version* of the ball-catching task can be administered using a PHP/MySQL compatible server controlled by the experimentalist and a participant can enter the experiment using a JAVASCRIPT-enabled browser (modern browsers such as Firefox, Chrome, Safari and IE). As in the z-Tree version, the speed of falling balls and the time interval between falling balls can be easily changed in the program. The online version works differently from the z-Tree version in that there is a ball-generating mechanism that produces each ball with a fixed time interval from a randomly selected column. Therefore, unlike the z-Tree version, the distance between two balls falling near to each other is always the same. Because of the different engine behind the online version, participants typically do not experience any sluggishness in the ball falling and tray moving, although it may happen due to network connection issues or not fully JAVASCRIPT-compatible browsers.

The actual implementation of online experiments using this online version requires additional considerations compared to laboratory experiments. As discussed in the main text, performance of online participants, such as MTurkers, may be affected by technological and environmental considerations that are not observed by the experimenter. These include details of computer configurations (e.g., screen sizes and mice), conditions of network connectivity, as well as environmental distractions, etc..