Incentive Complexity, Bounded Rationality, and Effort Provision

Johannes Abeler
David Huffman
Collin Raymond

Supplemental Appendix

Table of contents

A Theoretical appendix

We first provide an initial result that shows that our typical intuitions regarding labor supply apply in this setting: workers respond to increases in wages by working more under static contracts when wages increase.

Proposition 2. *If the Period-2 target rate does not depend on Period-1 effort, i.e.,* $\varsigma = 0$*, then an increase in marginal wage* w *increases effort in both periods.*

Proof of Proposition 2: We focus on a single individual and so suppress i subscripts and θ_i. If individuals face only static incentives ($\varsigma = 0$) we will show that the utility function features increasing differences. Because the utility function is additively separable in the efforts in each period, we can consider the maximization problem in each period separately. Focus on Period 1, and consider utility as a function of effort and wage: $U(e_1, w)$. Let $w' > w$ and $e_1' > e_1$. Then $[U(e_1', w') - U(e_1, w')] - [U(e_1', w) - U(e_1, w)] = [w' - w][\hat{g}(\frac{e_1'}{\eta_1}) - \hat{g}(\frac{e_1}{\eta_1})] \geq 0$. Standard monotone comparative statics imply the optimal choice of effort is increasing in w. The proof is analogous for Period 2. $\square$

A.1 Proof of Proposition 1

We now prove the result in the main text.

Proof of Proposition 1: We again focus on a single individual and so suppress i subscripts and θ_i. Consider utility as a function of effort and ς (remember that ς measures the effect of Period-1 effort on Period-2 target rate): $U(e_1, \varsigma)$. Let $\varsigma' > \varsigma = 0$ and $e_1' > e_1$. Moreover, denote the induced effort level in Period 2, given e_1 and ς as $e_2(e_1, \varsigma)$. We want to show that $[U(e_1', \varsigma') - U(e_1', \varsigma)] - [U(e_1, \varsigma') - U(e_1, \varsigma)]$ is negative. This expression is equal to

$$\delta[w\hat{g}(\frac{e_2(e_1', \varsigma')}{\varsigma' e_1' + (1-\varsigma')r_1}) - w\hat{g}(\frac{e_2(e'_1, \varsigma)}{\varsigma e_1' + (1-\varsigma)r_1}) + ae_2(e_1', \varsigma') - c(e_2(e_1', \varsigma')) - ae_2(e_1', \varsigma) + c(e_2(e_1', \varsigma))]$$

$$-\delta[w\hat{g}(\frac{e_2(e_1, \varsigma')}{\varsigma' e_1 + (1-\varsigma')r_1}) - w\hat{g}(\frac{e_2(e_1, \varsigma)}{\varsigma e_1 + (1-\varsigma)r_1}) + ae_2(e_1, \varsigma') - c(e_2(e_1, \varsigma')) - ae_2(e_1, \varsigma) + c(e_2(e_1, \varsigma))]$$

When $\varsigma = 0$, i.e., Period-1 effort does not affect Period-2 target rate, the agents with different Period-1 efforts face the same optimization problem in Period 2 (as it does not depend on e_1), so $e_2(e_1', 0) = e_2(e_1, 0)$, implying that $\hat{g}(\frac{e_2(e'_1,\varsigma)}{\varsigma e_1'+(1-\varsigma)r_1}) = \hat{g}(\frac{e_2(e_1,\varsigma)}{\varsigma e_1+(1-\varsigma)r_1})$. Thus, the expression reduces to

$$\delta[w\hat{g}(\frac{e_2(e_1',\varsigma')}{\varsigma' e_1'+(1-\varsigma')r_1})+ae_2(e_1',\varsigma')-c(e_2(e_1',\varsigma'))]-\delta[w\hat{g}(\frac{e_2(e_1,\varsigma')}{\varsigma' e_1+(1-\varsigma')r_1})+ae_2(e_1,\varsigma')-c(e_2(e_1,\varsigma'))].$$

This expression is the (discounted) difference in Period-2 utility conditional on the optimal effort chosen in Period 2. Finding the sign of this is equivalent to asking whether Period-2 utility conditional on the optimal effort being chosen in Period 2, and $\varsigma' > 0$, is higher or lower when Period-1 effort was higher (all else being equal). Clearly, conditional on any choice of e_2, e_1 being larger reduces utility. Thus, this must be negative and the result follows from standard monotone comparative static results. □

A.2 Predictions for GROUP

We start from the model outlined for the INDIVIDUAL trial in Section 2. For the GROUP trial, we now suppose that there are a finite number of periods τ. We suppose there are n individuals, T of which are randomly allocated to the treatment, while $n - T$ are in the control (we will also use T to refer to the set of treatment workers). In order to simplify exposition we suppose that types are publicly known (so that there is no learning about others' types over time).

Most of the features of the utility function remain the same compared to the model in Section 2. However, workers can now also care about other workers via an altruism (or concern for others) coefficient $\omega \geq 0$. This could also capture social pressure motives, which might make collusion easier. The second difference, in line with the design of the GROUP trial, is that next period's rate $\eta_{i,t+1}$ is equal to the *average effort among treatment workers* in period t (and is thus the same for all workers): $\eta_{i,t+1} = \eta_{t+1} = \frac{\sum_{j \in T} e_{j,t}}{T}$ (in the first period the normalization rate is exogenous).[53]

[53] In order to construct the optimal policy when there is only a single individual, or when individuals coordinate on the same effort level, the normalization factor η must never be equal to 0. Thus, we can suppose that the equation holds so long as $\frac{\sum_{j \in T} e_{j,t}}{T} \neq 0$. If $\frac{\sum_{j \in T} e_{j,t}}{T} = 0$ we then suppose $\eta_{i,t+1} = \underline{\eta}$ for some small positive $\underline{\eta}$. This allows for the existence of an optimal policy.

Utility is then

$$\sum_{t=1}^{\tau} \delta^{t-1}[o + w\hat{g}(\frac{e_{i,t}}{\eta_t}) - c(e_{i,t}, \theta_i) + ae_{i,t} + \omega(\sum_{j\neq i} w\hat{g}(\frac{e_{j,t}}{\eta_t}))]$$

A key thing to note is that in any given period, the target rates for an individual, regardless of whether they are in Treatment or Control, are the same. The only difference is that in Treatment, effort in a given period helps determine the target rate (for everyone) next period. Our primary results is that we obtain the ratchet effect result in this setting:

Proposition 3. *In GROUP, fixing θ_i as well as the set of other workers $\theta_{j\neq i}$, Treatment puts in a lower effort in all periods than Control.*

The solution concept is a sub-game perfect Nash Equilibrium (we assume that all workers know the types of the other workers). The proposition does not state what is the optimal path of effort for either Treatment or Control, but only compares their effort levels to each other within a period. Computing the equilibrium path of effort for Treatment and Control is non-trivial, and depends on the size of the group, and the exact parameters. Numerical simulations show that, at least for a small number of periods, if the individuals in Treatment can coordinate, then the equilibrium path will feature cycling: effort by Treatment should drop to a very low level (potentially 0) in the first period of the cycle.[54] In the following period, Treatment will put in the minimal amount of effort to acquire the maximal bonus, repeating this until it is no longer optimal, at which point effort drops down to the starting point of the cycle again. However, the key insight for Proposition 3 is that, for every equilibrium path, a worker who is in Treatment or Control will always face the same contemporaneous incentives in a given period, but Treatment workers face additional dynamic incentives to hold back effort. Treatment workers will thus always work less in a given period.

Proof of Proposition 3: We solve via backwards induction. We want to show that in every period Treatment puts in less effort than Control. Consider two agents with the same θ_i (we will suppress this variable for the rest of the proof).

[54]More generally, it is not straightforward to have a calibrated rational benchmark model for GROUP as for INDIVIDUAL without coordination. In the GROUP trial, future rates depend on the interaction of many individuals' current efforts, and the induced game takes place over many time periods, opening the way for complicated equilibrium behavior.

First we show the result for the last period τ. In this period the two individuals face the same target rates and so will make the same effort decisions.

Second, in the penultimate period, $\tau - 1$, we can use the same technique as in the proof of Proposition 1 to show that workers in Treatment will exert less effort than those in Control.

Next we turn to the ante-penultimate period, $\tau - 2$. Fix Period $\tau - 1$ effort at $\hat{e}^T_{i,\tau-1}$ for any given individual i if they were in in Treatment, and $\hat{e}^C_{i,\tau-1}$ if they were in Control. For a given other player j, let $Z(j) \in \{T, C\}$ denote whether they are in Treatment or Control. Consider two situations.

1. First, suppose Treatment works 0 in $\tau - 2$. Then Control, by construction, must work weakly more.

2. Next suppose that Treatment works some positive amount. Notice that the objective function of both Treatment and Control is piece-wise differentiable (since c is differentiable and $\hat{g}$ is piecewise differentiable). Whenever defined, the derivative for Treatment for Period $\tau-2$ effort is $w\frac{1}{\eta_{i,\tau-2}}\hat{g}'(\frac{e_{i,\tau-2}}{\eta_{i,\tau-2}}) - c'(e_{i,\tau-2}) + a - \delta\frac{1}{T}\frac{\hat{e}^T_{i,\tau-1}}{(\frac{\sum_{k\in T} e_{k,\tau-2}}{T})^2} w\hat{g}'(\frac{\hat{e}^T_{i,\tau-1}}{\frac{\sum_{k\in T} e_{k,\tau-2}}{T}}) - \delta\frac{1}{T}\sum_{j\neq i}\frac{\hat{e}^{Z(j)}_{j,\tau-1}}{(\frac{\sum_{k\in T} e_{k,\tau-2}}{T})^2}\omega(\sum_{j\neq i} w\hat{g}'(\frac{\hat{e}^{Z(j)}_{j,\tau-1}}{\frac{\sum_{k\in T} e_{k,\tau-2}}{T}}))$ (fixing the effort levels in $\tau - 2$ for all $j \neq i$). The first and third terms capture the marginal benefits of extra effort, and the second, fourth and fifth terms capture the marginal costs of extra effort. Analogously for Control we get the first order condition $w\frac{1}{\eta_{i,\tau-2}}\hat{g}'(\frac{e_{i,\tau-2}}{\eta_{i,\tau-2}}) - c'(e_{i,\tau-2}) + a$. Denote the optimum effort level for Control as $e^C_{i,\tau-2}$, and for treatment as $e^T_{i,\tau-2}$. By way of contradiction, assume that $e^C_{i,\tau-2} < e^T_{i,\tau-2}$. Consider Control adjusting their effort from $e^C_{i,\tau-2}$ to $e^T_{i,\tau-2}$. Observe that integrating up along this path over the difference between marginal benefits and marginal costs generates a negative number for Control (by construction, since $e^C_{i,\tau-2}$ is an optimum for Control).

   Instead, consider moving from $e^C_{i,\tau-2}$ to $e^T_{i,\tau-2}$ as Treatment. Recall that Treatment and Control face the same target rates in Period $\tau - 2$ (as well as all future periods). Thus, Treatment has the same marginal benefit curve, but a higher marginal cost curve everywhere, compared to Control. This implies that along the path between $e^C_{i,\tau-2}$ to $e^T_{i,\tau-2}$ the integral of difference between marginal benefits and marginal costs must still be negative for Treatment. This means (by the Second Fundamental Theorem of

Calculus since the objective function is piecewise differentiable) that $e^T_{i,\tau-2}$ has a lower total payoff compared to $e^C_{i,\tau-2}$ for Treatment. This is a contradiction.

Because this is true for any $\hat{e}^T_{i,\tau-1}$ and $\hat{e}^C_{i,\tau-1}$ it is true for the actual chosen effort levels in Period $\tau-1$. Thus, we find that in Period $\tau-2$, a worker in Treatment will work less than a worker in Control. The proofs for periods prior to $\tau-2$ work analogously. □

A.3 Results for the SIMPLE contract

In our online experiments we denote the contract modeled in the body of the text (which is the one used in the field) as COMPLEX. In the lab we also have subjects respond to a different contract that we label SIMPLE. Recall that in SIMPLE, payments in the first period are subtracted from earnings in the second period. Moreover, in both periods the worker faces an exogenous target rate of η. Since effort in all periods is paid out at the same time, experimental periods are separated by a very short period of time, and participants are only paid if they complete all periods, we assume $\delta = 1$. Thus, the optimization problem becomes (suppressing subscript i and θ again)

$$\begin{aligned}\max_{e_1,e_2} \quad & ae_1 + w\hat{g}\left(\tfrac{e_1}{\eta}\right) - c(e_1) \\ + \quad & \delta[ae_2 + [w\hat{g}\left(\tfrac{e_2}{\eta}\right) - w\hat{g}\left(\tfrac{e_1}{\eta}\right)] - c(e_2)] \\ = \max_{e_1,e_2} \quad & ae_1 - c(e_1) + ae_2 + w\hat{g}\left(\tfrac{e_2}{\eta}\right) - c(e_2)\end{aligned}$$

The next result highlights that workers should also exhibit a ratchet effect when faced with a SIMPLE contract.

Proposition 4. *Individuals reduce effort in Period 1 in SIMPLE relative to a static contract.*

Proof of Proposition 4: Clearly SIMPLE is equivalent to workers getting a wage of 0 in Period 1 in a static contract. Since we assume a positive marginal wage $w > 0$, this leads to the result. □

A.4 Extensions to the model in Section 2

A.4.1 Fatigue spillovers

Now we suppose that individuals' effort costs across time may be non-separable. Thus, there is now a single general cost function that depends on type and on effort in both periods. Utility (suppressing subscript i and θ) is then:

$$o + w\hat{g}(\tfrac{e_1}{\eta_1}, w) + ae_1 + \delta(o + w\hat{g}(\tfrac{e_2}{\eta_2}) + ae_2 - c(e_1, e_2))$$

We suppose c is strictly increasing and jointly convex in e_1 and e_2, differentiable in all arguments, and the limits of the partial derivatives with respect to e_1 and e_2 are ∞. All other assumptions are the same as before.[55] We still find a ratchet effect, as the next results demonstrates.

Proposition 5. *Fixing θ, individuals for whom the Period-2 target rate depends on Period-1 effort, i.e., $\varsigma > 0$, put in less effort in Period 1 than those individuals for whom $\varsigma = 0$.*

Proof of Proposition 5: The proof of this mirrors the proof of the main proposition. Consider utility as a function of effort and ς (recall that ς measures the effect of Period-1 effort on Period-2 target rate): $U(e_1, \varsigma)$. Let $\varsigma' > \varsigma = 0$ and $e_1' > e_1$. Moreover, denote the induced effort level in Period 2, given period of effort of e_1 and ς as $e_2(e_1, \varsigma)$. We want to show that $[U(e_1', \varsigma') - U(e_1', \varsigma)] - [U(e_1, \varsigma') - U(e_1, \varsigma)]$ is negative. This expression is equal to

$$\delta[w\hat{g}(\tfrac{e_2(e_1', \varsigma')}{\varsigma' e_{i,1}' + (1-\varsigma')r_1}) - w\hat{g}(\tfrac{e_2(e'_1, \varsigma)}{\varsigma e_{i,1}' + (1-\varsigma)r_1}) + ae_2(e_1', \varsigma') - c(e_1', e_2(e_1', \varsigma')) - ae_2(e_1', \varsigma) + c(e_1', e_2(e_1', \varsigma))]$$

$$-\delta[w\hat{g}(\tfrac{e_2(e_1, \varsigma')}{\varsigma' e_{i,1} + (1-\varsigma')r_1}) - w\hat{g}(\tfrac{e_2(e_1, \varsigma)}{\varsigma e_{i,1} + (1-\varsigma)r_1}) + ae_2(e_1, \varsigma') - c(e_1, e_2(e_1, \varsigma')) - ae_2(e_1, \varsigma) + c(e_1, e_2(e_1, \varsigma))]$$

When $\varsigma = 0$, i.e., Period-1 effort does not affect Period-2 target rate, the agents with different Period-1 efforts face the same optimization problem in Period 2 (as it does not depend on e_1), so $e_2(e_1', 0) = e_2(e_1, 0) = \hat{e}_2$, which means $\hat{g}(\tfrac{e_2(e'_1, \varsigma)}{\varsigma e_{i,1}' + (1-\varsigma)r_1}) = \hat{g}(\tfrac{e_2(e_1, \varsigma)}{\varsigma e_{i,1} + (1-\varsigma)r_1})$ and so the expression simplifies to

[55] While we suppress θ for notational simplicity, we assume the cross partial of the first two arguments of $c(e_1, e_2, \theta)$ with θ is positive (so that higher types have higher marginal costs) and that the partial derivative at $(0, 0, \theta)$ with respect to the first two argument is 0.

$$\delta[w\hat{g}(\tfrac{e_2(e_1',\varsigma')}{\varsigma' e_{i,1}'+(1-\varsigma')r_1}) + ae_2(e_1',\varsigma') - c(e_1', e_2(e_1',\varsigma'))] - \delta[w\hat{g}(\tfrac{e_2(e_1,\varsigma')}{\varsigma' e_{i,1}+(1-\varsigma')r_1}) + ae_2(e_1,\varsigma') - c(e_1, e_2(e_1,\varsigma'))] + \delta[c(e_1', \hat{e}_2) - c(e_1, \hat{e}_2)]$$

As before, the sign of the first two terms in brackets is the (discounted) difference in Period-2 utility conditional on the optimal effort chosen in Period 2. Finding the sign of this is equivalent to asking whether Period-2 utility conditional on the optimal effort being chosen in Period 2, and $\varsigma' > 0$, is higher or lower when Period-1 effort was higher (all else being equal). Clearly, conditional on any choice of e_2, e_1 being larger reduces utility. Moreover the third term must be negative, and so the overall function is negative. Thus, the result follows from standard monotone comparative static results.□

A.4.2 Multiple tasks

In the actual warehouse, each worker has multiple tasks they work on. In our model in the body of the paper, we suppose that workers only have a single task. Now we extend our model to allow for multiple tasks and show that we still observe a ratchet effect in terms of total effort. Each worker i works across tasks $\gimel = 1, ..., J$ over time periods $t = 1, 2$. The worker spends hours $H_{i\gimel t}$ on task $\gimel$ out of a total of H hours. $H_{i\gimel t}$ is exogenously given, i.e., the worker does not decide how much time to spend on each task. This is in line with the practice in the firm, where the firm decides on which task a worker works. Given a task and time, workers choose per-hour effort level of $\tilde{e}_{i\gimel t}$ to produce output $\tilde{e}_{i\gimel t} H_{i\gimel t}$. $\eta_{i\gimel t}$ is the target rate used for normalizing effort in Period $t = 1$ for output, and then in Period 2 the target rate is a function of Period-1 output (just as before). We assume that the weight placed on past effort in the Period-2 normalization is the same across tasks (i.e., ς does not depend on the task), and that the normalization is concerned about per-hour effort in Period 1, not total output, for a given task (since total output at a given task is only partially under the control of the worker, as the time $H_{i\gimel t}$ spent on the task is exogenous). This is in accordance with the actual policy followed by the firm. In particular, recall that the firm normalizes effort for each task individually, and then sums up the normalized efforts (as opposed to summing up efforts and then normalizing).

We need to be more careful in notation and assumptions here compared to when there is a single task. In particular, consider the increase in the marginal cost of effort provision, given an increase in the effort for task $\beth$ which will raise normalized effort by 1 unit (i.e., increase $\frac{H_{i\beth t}\tilde{e}_{i\beth t}}{\eta_{i,\beth,t}}$ by 1). Across different tasks, this might not be the same. Of course, with a single task, by construction this is only a single number. In order to overcome this issue, provide a tractable model, and generate clean results, we will proceed as follows. In particular, our approach will allow us to speak about "total effort" as simple sum across tasks of effort times hours. Each task has an associated difficulty $d_\beth$. We assume costs can be represented as a function of the output of a task, times the difficulty of the task, times the hours devoted to that task (as well as the worker type): $c(\sum_\beth d_\beth H_{i\beth t}\tilde{e}_{i\beth t}, \theta_i)$. Similarly, we assume non-pecuniary benefits are proportional to $a\sum_\beth d_\beth H_{i\beth t}\tilde{e}_{i\beth t}$. In order to simplify everything, we will define $e_{i\beth t} = d_\beth \tilde{e}_{i\beth t}$ as the difficulty adjusted effort.

The utility function is then (as usual we will suppress i and θ)

$$o + a\sum_\beth H_{\beth 1}e_{\beth 1} + w\hat{g}(\sum_\beth \tfrac{H_{\beth 1}e_{\beth 1}}{d_\beth \eta_{\beth,1}}) - c(\sum_\beth H_{\beth 1}e_{\beth 1})$$

$$+\delta(o + a\sum_\beth H_{\beth 2}e_{\beth 2} + w\hat{g}(\sum_\beth \tfrac{H_{\beth 2}e_{\beth 2}}{d_\beth(\varsigma e_{\beth 1}+(1-\varsigma)r_{1\beth})}) - c(\sum_\beth H_{\beth 2}e_{\beth 2}))$$

In this situation, we still find a ratchet effect, although with slightly more nuance as the next results demonstrates. In order to simplify our analysis, we will assume all optima for the decision-maker's problem generate the same $\sum_\beth H_{\beth 1}e_{\beth 1}$. This simply means that if the optimal effort allocation is not unique for the worker, all optimal allocations must generate the same total cost of effort.

Proposition 6. *Individuals for whom the Period-2 target rate depends on Period-1 effort, i.e., $\varsigma > 0$, compared to those for whom $\varsigma = 0$, either*

- *have a smaller total effort $\sum_\beth H_{\beth 1}e_{\beth 1}$ in Period 1, or*
- *earn less in Period 1.*

The proposition is more nuanced than previous propositions. It says that Treatment individuals either put in less effort (as measured by the cost of effort) or they earn less (i.e., they put in less effort in terms of the wage benefits of effort). As the proof makes clear, the first statement always holds, unless it is the case that Control workers have worked such they they are at or above the cap in Period 1. Then the second statement is true by construction. But it raises the question of why Control workers can be earning more, while paying a lower effort cost. Consider the vector space that defines effort combinations across these tasks. The “iso-effort cost” sets do not necessarily have the same gradient in this space as the “iso-effort-bonus” sets. Thus, it could be that Control can redistribute effort such that they pay a lower cost of effort in Period 1, but have a higher Period-1 bonus. This is because Treatment may exert effort differently, which lowers Period-1 bonuses, but increases Period-2 bonuses.

Our theoretical results from the model with multiple tasks raise the question to what extent can we measure “total effort” in Period 1. It can’t simply be the sum of outputs across tasks, since output is in different units across tasks. However, we know the firm initially set static target rates (which are also the rates used in Period 1) to reflect what they believed was relative difficulty across tasks. If we believe that this is true, then this would imply that a measure of total productivity that relied on the initial target rates used in Period 1 would measure total effort. In fact, our empirical analysis of INDIVIDUAL in the paper relies on this — we measure total output controlling for task fixed effects, to convert output on different tasks to the same units. Thus, under the assumption that Period-1 target rates reflect actual difficult differences, we are testing the prediction of the multi-task model.

Proof of Proposition 6:

Refer to Treatment as $\varsigma > 0$ and Control as $\varsigma = 0$. The structure of the proof is as follows: we will always attempt to show $\sum_{\beth} H_{\beth 1} e_{\beth 1}$ is smaller for Treatment than for Control or that Treatment must earn less than Control.

Utility for both Treatment and Control is piecewise linear because the cost function is differentiable, and $\hat{g}$ is piecewise differentiable. In particular, $\hat{g}$ is non-differentiable at two points, the quota and the cap. However, workers will never want to work at the quota unless the first-order condition also holds. Moreover, they will never want to put in infinite effort for any task. Thus, we will consider three kinds of optima in Period 1 for a worker in Treatment: (1) they choose 0 effort for all tasks, (2) they choose effort levels so that $\sum_{\beth} \frac{H_{\beth 1} e_{\beth 1}}{d_{\beth} \eta_{,\beth,1}}$ is at a differentiable incentive payment for Period 1 and the first order condition holds, or (3) they choose $\sum_{\beth} \frac{H_{\beth 1} e_{\beth 1}}{d_{\beth} \eta_{,\beth,1}}$ at the cap (and the first order condition may not hold). Denote a vector of efforts in Period 1 (one effort for each task) as e_1, with the corresponding vector of hours as H_1. Fix Period-2 behavior (a vector of efforts for each task) as e_2 for Treatment. We go over each of the three cases in turn.

1. If Treatment has $H_1 e_1 = 0$ at the optimum, then Control, by construction, must work weakly more. Thus the claim is true.

2. Next, suppose that an optimum effort for Treatment is at a differentiable point for Period-1 incentives. To begin with also assume that we are at a point where the Period-2 incentive payment is differentiable, and thus the first-order condition characterizes the optimum for Treatment. Consider the first-order condition with respect to task $\hat{\gimel}$.

$$\begin{gathered} aH_{\hat{\gimel}1} + \frac{H_{\hat{\gimel}1}}{d_{\gimel}\eta_{i,\hat{\gimel},1}} w\hat{g}'\left(\sum_{\gimel} \frac{H_{\gimel 1} e_{\gimel 1}}{d_{\gimel}\eta_{\gimel,1}}\right) - H_{\hat{\gimel}1} c'\left(\sum_{\gimel} H_{\gimel 1} e_{\gimel 1}\right) - \\ \delta\varsigma \frac{H_{\hat{\gimel}2} e_{\hat{\gimel}2}}{d_{\hat{\gimel}}^2(\varsigma e_{\hat{\gimel}1} + (1-\varsigma) r_{\hat{\gimel}})^2} w\hat{g}'\left(\sum_{\gimel} \frac{H_{\gimel 2} e_{\gimel 2}}{d_{\gimel}(\varsigma e_{i\gimel 1} + (1-\varsigma) r_{\gimel})}\right) = 0 \end{gathered}$$

or, rewritten

$$aH_{\hat{\gimel}1} + \frac{H_{\hat{\gimel}1}}{d_{\gimel}\eta_{\hat{\gimel},1}} w\hat{g}'\left(\sum_{\gimel} \frac{H_{\gimel 1} e_{\gimel 1}}{d_{\gimel}\eta_{\gimel,1}}\right) = H_{\hat{\gimel}1} c'\left(\sum_{\gimel} H_{\gimel 1} e_{\gimel 1}\right) + \delta\varsigma \frac{H_{\hat{\gimel}2} e_{\hat{\gimel}2}}{d_{\hat{\gimel}}^2(\varsigma e_{\hat{\gimel}1} + (1-\varsigma) r_{\hat{\gimel}})^2} w\hat{g}'\left(\sum_{\gimel} \frac{H_{\gimel 2} e_{\gimel 2}}{d_{\gimel}(\varsigma e_{\gimel 1} + (1-\varsigma) r_{\gimel})}\right)$$

We consider three sub-cases. The three cases vary by what assumption we make about the optimum for Control.

   (a) Suppose one of Control's ($\varsigma = 0$) optima is also characterized by the first-order condition. For Control the first-order condition is $aH_{\hat{\gimel}1} + \frac{H_{\hat{\gimel}1}}{d_{\gimel}\eta_{\hat{\gimel},1}} w\hat{g}'(\sum_{\gimel} \frac{H_{\gimel 1} e_{\gimel 1}}{d_{\gimel}\eta_{\gimel,1}}) = H_{\hat{\gimel}1} c'(\sum_{\gimel} H_{\gimel 1} e_{\gimel 1})$. Compare the first-order conditions across Treatment and Control. Notice that the left-hand side, which denotes the marginal benefit of an extra unit of effort, is the same. However, the right-hand side, which captures the marginal costs, is larger for any choice of e_1 when $\varsigma > 0$ compared to when $\varsigma = 0$, all else being equal. We now proceed by contradiction. Suppose that $H_1 e_1$ strictly decreased moving from Treatment to Control. This implies that for any task, the right-hand side of the first-order condition must have gone down because we dropped an additional term, and the argument of c' fell. This implies the left-hand side of the equal must have gone down as well. There are four sub-cases

i. Suppose that at the optimum for Treatment, H_1e_1 was either below the quota or above the cap, so that the marginal bonus was 0 (i.e. the derivative of $\hat{g}$ was equal to 0 for Treatment). In this case, observe that there is no way for the left-hand side of the first order condition to fall in order to ensure that the Control choice is an optimum. This leaves us with a contradiction.

ii. Suppose that at the optimum for Treatment, denoted e_1^T, effort generated positive marginal bonus. Note that H_1e_1 cannot go up when we move to Control (by assumption) which implies that for the left-hand side of Control's first-order condition to fall, it must be that effort either moves into the region below the quota for Control, where there is zero marginal bonus, or into the region above the cap. Assume for the moment, that effort moves into the region below the quota for control.

   The first-order condition for Control is then $a = c'(\sum_{j} H_{j1}e_{j1})$, which implies that the agent is indifferent between any allocation of efforts that provides the same H_1e_1, which must generate a marginal cost equal to a. Denote one of these optima as e_1^C. We assume that $e_1^C << e_1^T$. In this case, consider Control adjusting their effort from e_1^C to e_1^T. Observe that integrating along this path (by the Gradient Theorem which path doesn't matter) over the difference between marginal benefits and marginal costs for Control generates a negative number (by construction, since e_1^C is an optimum for Control). Now consider Treatment. Treatment has a marginal cost manifold that is shifted up everywhere relative to Control. This implies that integrating along the same path as before between e_1^C to e_1^T over the difference between marginal benefits and marginal costs must be negative for Treatment. Thus e_1^T cannot be an optimum for Treatment (since e_1^C generates a higher utility). This is a contradiction.

iii. We still assume that at the optimum for Treatment, denoted e_1^T, effort generated positive marginal bonus, and that Control is providing effort such that they are below the quota (just as in (ii)). However, we now assume that e_1^C is not strictly less than e_1^T. But, notice that we can consider the set of effort vectors that generate the same Period-1 effort costs for Control as e_1^C; i.e., we can consider the set of efforts e' such that $H_1 e' = H_1 e_1^C$. Because the iso-effort sets are parallel in the vector space of efforts, we can find an e' in this set such that $e' << e_1^T$. Moreover, since e_1^C is an optimum and generates no bonus, it must be the case that e' also generates no bonus (since if it did, Control could have higher bonuses and the same costs by choosing e' rather than e_1^C) and is an optimum. Now we can use our reasoning from (ii) and show that both Control and Treatment must prefer e' to e_1^T. This is a contradiction

iv. We again assume that at the optimum for Treatment, denoted e_1^T, effort generated positive marginal bonus, but now we suppose that Control is providing effort such that they are above the cap. Thus, Control must be earning more than Treatment (since Treatment is providing effort that leaves it below the cap).

This covers all the sub-cases where Control's ($\varsigma = 0$) optimum is also characterized by the first-order condition.

(b) Now suppose Control's optima all have $H_1 e_1$ equal to the cap. In this case Treatment workers must earn (weakly) less.

(c) Last, suppose Control's optimum is $e_1 H_1 = 0$. But observe that Treatment faces higher total costs from any level of effort compared to Control, and so Treatment's optimum must be $e_1 H_1 = 0$, which is a contradiction.

Continuing to assume that effort is at a differentiable point for today's incentives, now suppose that tomorrow Treatment chooses $e_1 H_1 = 0$. Then Treatment faces the same optimization problem as Control (since tomorrow they will not exert any effort), and so Treatment and Control choose the same amount of effort today.

Last, suppose that tomorrow Treatment chooses $e_2 H_2$ to be precisely at the cap (while still assuming that effort is at a differentiable point for today's incentive). If one of Control's optima is characterized by the first-order condition, then the arguments of 2(a) holds. If Control's optima are also at the cap, then both Treatment and Control are earning equal amounts. If Control's optimum is $e_1 H_1 = 0$, we know that Treatment faces higher total costs from any level of effort compared to Control, and so Treatment's optimum must be $e_1 H_1 = 0$, which is a contradiction.

3. Suppose that for Treatment all the optima are exactly at the cap. If Control's optima involve providing effort at or above the cap, then both Treatment and Control are earning equal amounts. If Control's optimum is $e_1 H_1 = 0$, we know that Treatment faces higher total costs from any level of effort compared to Control, and so Treatment's optimum must be $e_1 H_1 = 0$, which is a contradiction. Thus, we are left to consider what happens if at least one of Control's optima is characterized by the first order condition and is either below the quota or between the quota and cap (where positive marginal bonus is earned). Then we can repeat the arguments of 2(a).

Notice that these results hold for any e_2, and specifically for the e_2 that is chosen at optimum by Treatment. This proves the result. $\square$

B Structural calibration

B.1 Details of the structural calibration for INDIVIDUAL

B.1.1 Calibration and worker simulation

Here we discuss our calibration and simulation exercise for the field experiment in detail. We will first describe our modeling approach, then the particular values used, and finally, the results.

Our two period model, when reduced to a single period and a representative agent, implies that utility is $U = o + w\hat{g}(\frac{e}{\eta}) - c(e,\theta) + ae$, or in other words, $w\hat{g}(\frac{e}{\eta}) - c(e,\theta) + ae$ (dropping the fixed payment as well as the time and individual indicators).[56] Recall that $w\hat{g}$ takes on a piecewise linear structure, with no incentives being received if normalized effort is below some threshold $\underline{E}$, and the marginal incentive dropping to 0 once normalized effort is above some level $\bar{E}$. We can rewrite $w\hat{g}(\frac{e_i}{\eta})$ in order to make it more amenable to simulation: $w\hat{g}(\frac{e}{\eta}) = \max[\min[-\kappa + \frac{ew}{\eta}, M], 0]$, with $\frac{\kappa}{w}$ corresponding to $\underline{E}$ and $\frac{M+\kappa}{w}$ corresponding to $\bar{E}$. The maximum bonus obtainable is thus M. We observe κ and M in the data. These, along with our assumption about the cost function ($c(e,\theta) = \theta\frac{e^{\gamma+1}}{\gamma+1}$) implies that the worker's one-period maximization problem is:

$$\max_{e} \quad ae + \max[\min[-\kappa + \frac{ew}{\eta}, M], 0] - \theta\frac{e^{\gamma+1}}{\gamma+1}$$

The parameters we must calibrate (i.e. are unobserved to the researchers) are a, θ and γ. Once static incentives are introduced, the optimization problem is exactly described as above, but prior to the incentive scheme $w = 0$, and so the optimization problem simplifies to

[56] We assume a representative worker because we do not have adequate variation in wages for any given worker. Because there are many tasks, and the optimization problem for any given task is non-concave, solving for the optimization without a representative task becomes intractable. In the model, bonuses are calculated at a period level, while in reality they are calculated at the weekly level, but because target rates do not change over the three week period we can simply average them to obtain a representative bonus.

$$\max_{e} \quad ae - \theta \frac{e^{\gamma+1}}{\gamma+1}$$

The optimal effort before and after the introduction of incentives is thus $\left(\frac{a}{\theta}\right)^{\frac{1}{\gamma}}$ and $\left(\frac{a+\frac{w}{\eta}}{\theta}\right)^{\frac{1}{\gamma}}$, respectively.

We use two data points to calibrate our model. First, the response of workers to the introduction of static incentives (i.e. the ratio of effort after the introduction to effort prior to the introduction of static incentives), which we call R. Second, the level of effort when w is strictly positive, which we call E. Two key issues arise in making this choice of what values to assign to these moments, which also speak to why we need to assume a representative agent. First, the workers in the INDIVIDUAL trial were not at the warehouse during the change from no incentives to static incentives. Thus, in order to estimate responsiveness to the introduction of wages, we have to use a distinct set of workers from those who we observe participating in our experiment. Thus, we can only estimate aggregate, rather than individual responsiveness to the introduction of incentives, and assume that the workers in our experiment would respond in the same way as the workers who were actually present for the introduction of static incentives. As we discuss in more detail below, we are going to normalize all variables such that the target rate faced by workers in Period 1 of INDIVIDUAL is equal to 1. We will also assume that the R we calculate using data prior to INDIVIDUAL is precisely the same responsiveness that would have occurred had we moved from no incentives to static incentives in Period 1 of INDIVIDUAL. This implies that $\eta = 1$ for computing R.

We want to ensure that our calibration matches treatment worker behavior when they face only static incentives. Moreover, we would like any welfare losses to be driven solely by issues with the dynamic optimization problem the workers face. Thus, we want to ensure that if workers were to face the actual rates they faced in Period 2 (when they had only static incentives) there would be zero welfare loss from their actual behavior. Thus, we set E equal to effort in Period 2 of the treatment workers. This implies we need to normalize the wage by a distinct target rate $\eta = \eta_2$ (i.e. the observed target rate in Period 2 from the data) (because the target rate in Period 2 of the experiment is different than the target rate in Period 1). We discuss how we calculate η_2 below.

Given these considerations, the two moments we use are:

$$R = \left(\frac{a+w}{a}\right)^{\frac{1}{\gamma}} \qquad \text{and} \qquad E = \left(\frac{a+\frac{w}{\eta}}{\theta}\right)^{\frac{1}{\gamma}}$$

Because we have three parameters and two moments, we consider sets of allowable parameters by changing a from 10^{-6} to 10^{6} by powers of 10. We calculate for each a, the corresponding values of θ and γ that rationalize the observed data. Recall that we assume that workers optimize given a static contract.

Once we have calibrated a, θ and γ, we simulate behavior under dynamic incentives. We thus turn to our two-period model, which, given the discussion above, is

$$\begin{aligned} \max_{e_1,e_2} \quad & ae_1 + \max[\min[-\kappa + e_1 w, M], 0] - \theta \frac{e_1^{\gamma+1}}{\gamma+1} \\ + \quad & \delta(ae_2 + \max[\min[-\kappa + \frac{e_2 w}{\frac{1}{2}r + \frac{1}{2}\frac{e_1}{h}}, M], 0] - \theta \frac{e_2^{\gamma+1}}{\gamma+1}) \end{aligned}$$

The functional form has a couple of features that require explanation. We normalize the target rate to be equal to 1 in Period 1, in other words, we measure everything in terms of Period 1 SPH and the Period-1 bonus is thus $\frac{e_1 w}{1}$. Thus, all effort is measured in terms of Period-1 SPH (we discuss later how to normalize the measured effort in other periods to ensure they are expressed in terms of Period-1 SPH). The Period-2 target rate is an average of an exogenous rate r imposed by the firm, averaged with the per-hour normalized effort generated by the worker in Period 1, which is calculated by taking their total effort e_1 and dividing by the number of hours they worked, h. This means that a worker's rate is not affected by the effort of any other worker. We suppose that workers know what value the exogenous portion of their target rate will take on in Period 2 (i.e. they know r precisely). In reality, it is likely they had a good idea of what value r was likely to be, albeit not perfect knowledge.

Because we take a representative task, we need to measure aggregate effort across all tasks. In order to carefully describe our approach, for any given measured variable, we need to distinguish the value that the variable takes in the actual observed data, and the value that the variable takes on for the structural estimation. Because everything in the structural calibration is measured relative to Period 1 (i.e. we normalize the target rate in Period 1 to 1) the two are typically simply scalar multiples of one another. Thus, we will speak both of observed variable values, as well as structural variable values.

The first thing we do, as discussed, is set the structural target rate in Period 1 equal to 1. In the data, the observed target rate in Period 1 is 506.8. We calculate this number by (1) considering the set of individuals who work more than 20 hours per week (i.e. those who appear to be working full time) (2) calculating the time weighted average target rate that these workers face across all tasks they work in. We find the actual target rate of the workers in Period 2 similarly, and calculate it to be 542.2. Thus, the structural target rate that workers actually face in Period 2 of the experiment is the ratio of the actual target rate in Period 2 relative to the actual target rate in Period 1, and so the observed target rate in Period 2 (in Period 1 target rate units) is $\eta_2 = \frac{542.2}{506.8} = 1.06985$.

We rely on the difference-in-differences analysis of the introduction of static incentives (see column 6 in Table D.1) and set $R = 1.1052$. Recall we take E to be equal to the effort in Period 2. Because the target rate in Period 1 and Period 2 are different, we need to convert observed SPH (i.e. observed effort) in Period 2 to Period-1 units. Doing this gives us the structural value of E. In particular, we need to multiply the observed effort in Period 2 (30.8) by η_2: $E = 30.8 \times \frac{542.2}{506.8} = 32.95138$.

As shown in Figure C.3 in Appendix C, 93.1 percent of worker time is spent on the positive slope of the incentive scheme. The aggregate result that the representative agent works on the upward sloping portion of the scheme is thus also borne out at the individual level.

In order to match the actual contract, we set the parameters describing the shape of the bonus function to $\kappa = 78$ and $M = 54$. As discussed in Section 3.3, we set $\delta = 0.913$ which is the ratio of hours worked in Period 2 to Period 1 (including individuals who leave the firm). r is the exogenous normalization rate used in Period 2, which is averaged with the effort exerted by workers in Period 1. We turn observed r (which is 554.0) into structural r by normalizing by the Period-1 target rate, and so $r = \frac{554.0}{506.8} = 1.09313$. h is the number of hours worked by subjects in Period 1, $h = 31.2$. The actual effort in Period 1 is $e_1 = 33.6$ (recall we measure everything in terms of Period 1 and so no normalization is needed). Our structural e_2 is the actual effort in Period 2 in the data, 30.8, normalized back to Period 1 units $30.8 \times \eta_2 = 32.95138$ (observe that this is, as it should be, the same as E).

For each set of allowable parameter values (recall for each value of a we have corresponding values of θ and γ) we then find the values of e_1 and e_2, which maximize the utility function. We then calculate the utility obtained by the individual at the optimal and the actual effort levels.

Table B.1: Simulated optimal behavior in INDIVIDUAL

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
a	γ	θ	e_1 ***Predicted***	e_2 ***Predicted***	***Utility Predicted***	***Utility Actual***	e_1 ***Diff Ratio***	e_2 ***Diff Ratio***	***Utility Loss***
1.00E-06	148.97	2.16E-226	8.46	30.02	49.30	22.74	0.75	0.09	-26.56
1.00E-05	125.97	1.75E-191	8.52	30.06	49.30	25.46	0.75	0.09	-23.85
0.0001	102.97	1.42E-156	8.58	30.10	49.31	27.22	0.74	0.09	-22.09
0.001	79.97	1.14E-121	8.72	30.20	49.34	28.26	0.74	0.08	-21.08
0.01	57.00	8.39E-87	8.92	30.34	49.66	28.99	0.73	0.08	-20.66
0.1	34.30	2.51E-52	9.46	30.72	52.84	33.27	0.72	0.07	-19.58
1	13.85	3.65E-21	11.36	32.06	84.80	79.47	0.66	0.03	-5.34
10	2.62	1.35E-03	31.84	32.98	441.64	440.17	0.05	0.00	-1.47
100	0.30	36.63	31.76	33.00	1334.65	1333.16	0.05	0.00	-1.50
1000	0.03	903.24	31.76	33.00	1693.89	1692.39	0.05	0.00	-1.50
1.00E+04	0.00	9898.61	31.76	33.00	1740.97	1739.47	0.05	0.00	-1.50
1.00E+05	0.00	99898.13	31.76	33.00	1745.83	1744.33	0.05	0.00	-1.50
1.00E+06	0.00	999898.08	31.76	33.00	1746.32	1744.82	0.05	0.00	-1.50

Notes: Actual effort was 33.60 in Period 1 and 32.95 in Period 2.

Table B.1 reports the full results of the simulation separately for 13 representative values of a covering the entire range we considered. The first three columns of the table show allowable combinations of parameter values. The fourth and fifth columns show the predicted effort in Periods 1 and 2 for each allowable parameter combination, while the sixth shows the utility for the optimal choices of effort. The seventh column shows the actual utility, given the observed effort levels of 33.60 in Period 1 and 32.95 in Period 2.[57] Columns 8 and

[57]The reason we can find effort to be slightly higher in Period 1 than Period 2, despite the fact that units per hour were lower in Period 1 than Period 2 (see Figure 2), is that the exogenous portion of the target

9 show the difference between actual and predicted effort, divided by actual effort, while the tenth shows the utility difference between actual and predicted behavior. We see from the bottom rows of Table B.1, column 8, that treatment workers are predicted to reduce effort in Period 1 by at least 5 percent (for high values of a), far more than is observed.[58] Recall the observed treatment effect (-0.1 percent with 95 percent CI [-1.2, 1.0]) was over an order of magnitude smaller. It's also the case that the utility losses from mis-optimization are non-negligible. That said, our model predicts that for smaller values of a, the change in effort, and the utility losses are much larger. And, as discussed in the body of the paper, reasonable estimates of the elasticity of effort (which give an a of roughly 0.1) would indicate large (around 70 percent) changes in effort.

As seen in Table B.1, changing parameter values can sometimes cause dramatic changes in the optimal effort level, and thus in the difference between observed and optimal efforts (as well as utility losses). This is because nonlinearities in the optimization problem lead to multiplicity of local maxima. The multiplicity and the differences between optima occur for two reasons. First, the contract scheme features both a quota and a cap. Even in a one-period setting, this means that the marginal wage, as a function of SPH, goes from 0 to something positive and back to 0. Each of these regions has a potential optimum. If we have two periods, then the joint maximization over Period 1 and 2 efforts has up to 9 potential local optima. These local optima can differ wildly in terms of proscribed effort level. Second, because of the zero marginal wage below the quota, once effort falls below the quota it can fall to very low levels.

As we move down the table across allowable parameter combinations, there are both shifts of the location of the local optimum in each region, but also changes in the relative attractiveness of the different local optima, which can lead to jumps in optimal effort (this multiplicity of local optima also explains why we need to numerically estimate the optimal effort provisions).[59]

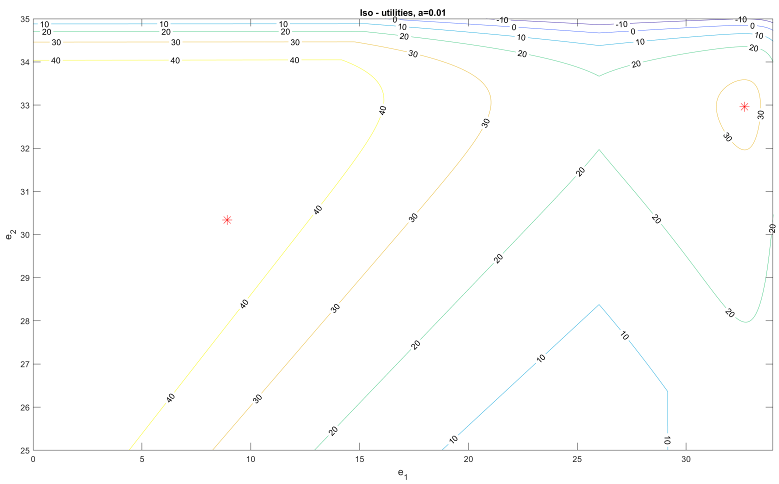
rate was easier in Period 1 (as discussed above, 506.8 versus 542.2). This reflected a policy of the firm to start new hires with easier rates and increase the exogenous portion of rates with tenure until reaching the rates faced by experienced workers. The firm calibrated this increase to correct for learning and make SPH reflect effort.

[58]As a comparison, this implies that the dynamic incentives should have eliminated around half of the effort increases from the introduction of static incentives.

[59]In Table B.1, the jump in predicted e_1 occurs between $a = 1$ and $a = 10$. If we zoom in further, the

In Figure B.1, we illustrate multiplicity by showing an indifference curve map for one set of parameter values involving a value for a in the empirically plausible range. There is a local optimum in the "North East" region, with effort in Period 1 roughly 5 percent less than in Period 2 (effort levels around 31 and 33 respectively). The contour line shows utility in this case is a little over 30. However, there is a second local optimum in the "North West" region, with optimal effort for Period 1 around 9 and Period-2 effort relatively high, entailing a ratchet effect in the range of 73 percent. As this gives a utility of around 50, this local maximum featuring a strong ratchet effect is the global optimum.

Figure B.1: Multiple local optima for e_1



Notes: The graph shows the worker's iso-utility curves depending on effort choices in Periods 1 and 2 for a=0.01. Red stars denote local maxima.

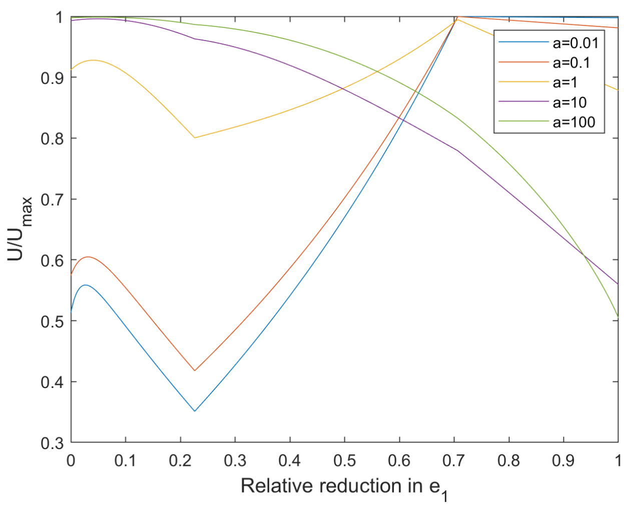
A natural question is, why is the low-effort local maximum so low? To illustrate the intuition, we will discuss in detail how the quota aspect of the scheme can lead to large ratchet effects. Remember that worker utility consists of monetary (extrinsic) motivation, intrinsic motivation and effort costs. When Period-1 effort drops below the quota, the extrinsic marginal benefit of effort drops discretely to zero. Thus, once workers reduce their effort down to the quota, they might as well continue to reduce it further. In particular, if intrinsic motivation a is lower than effort costs, our model predicts very large effort reductions.

To illustrate this more clearly, Figure B.2 focuses only on Period-1 effort. We fix Period-

jump is actually between $a = 1$ (with predicted $e_1 = 11.36$) and $a = 2$ (with predicted $e_1 = 32.04$).

2 effort at the observed level of effort in our field experiment. The horizontal axis shows reduction in Period-1 effort relative to Period-2 effort (so that 0 indicates no reduction, i.e., a Period-1 effort equal to Period-2 effort). The vertical axis shows the utility level for any given Period-1 effort, normalized by the utility achieved at the optimal Period-1 and Period-2 effort (we do this because as we change parameters the scale of utility changes, and we want to have a similar scale for all parameter combinations).[60] The different lines represent the relationship between effort and utility depending on the parameter values.

Figure B.2: Effect of quota on optimal effort



Notes: The graph shows worker utility (normalized by the utility achieved at the optimal Period-1 and Period-2 effort) for all Period-1 effort levels. Period-2 effort is fixed at the level we observe in the data. The horizontal axis shows reduction in Period-1 effort relative to Period-2 effort.

We can see in the figure that Period-1 effort can feature multiple local optima, even fixing Period-2 effort. Second, we can see precisely where the quota matters – this is where the lines feature a kink (this is at around a 20 percent reduction in effort). Third, we can also see that for low levels of a, the low local maximum is globally optimal and there will be very large ratchet effects.

The utility losses change between the rows for two different reasons. First, as just discussed, the location of the optimal effort combination in Period 1 and Period 2 changes (and sometimes dramatically). Second, as we move between rows, the utility function changes.

[60] Because we normalize by the maximum utility achieved for the optimal selection of Period-1 and Period-2 effort, and we fix Period-2 effort to be the observed effort, not all lines will achieve a maximum at 1.

For small a utility is primarily driven by the cost of effort provision and the pecuniary benefits that accrue from the incentive scheme. Thus, even small movements away from optimizing the amount of money earned can have large impacts on utility. In contrast, for large values of a, decision are primarily driven by intrinsic motivation and the costs of effort, and so failing to understand the dynamics of monetary earnings matters less.

Of course, the assumptions we have made in generating these results are subject to concerns about their robustness. Thus, we conduct three robustness checks. One concern is that the non-pecuniary motives we model are linear in effort. But in reality they may be non-linear. For example, suppose non-pecuniary motives are driven by firing threats. Workers might believe that if they reduce effort by too much the chance of being fired increases quickly. The firm has actually told us that they do not fire workers due to low effort (and this indeed seems to be the case given the data we observe, see Appendix F.4 for details). However, workers still may believe that working slow could lead to firing. Thus, for the first check, we assume that workers will never reduce their Period-1 effort relative to the observed Period-2 level by more than 20 percent. This is based on the firm telling us that only if workers are 30 percent slower than average do they receive any extra attention (and this is extra training rather than firing). We are conservative and assume that workers want to avoid such attention, and might be miscalibrated and think it could occur already with a slowdown of 20 percent. This is meant to capture firing concerns that go beyond what we capture by the linear effect of a. In a sense, we assume that, first, the agent believes they will be fired for sure if they reduce effort by more than 20 percent, and second, that the agent has such a low utility from being fired, they will never choose an effort level such that firing would occur. Thus, Period-1 effort must be no lower than $0.8 \times 33.6 = 26.88$. The rest of the optimization problem is as before.

Table B.2: Simulated optimal behavior in INDIVIDUAL: Robustness to 20 percent maximum decline

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
a	γ	θ	e_1 ***Predicted***	e_2 ***Predicted***	***Utility Predicted***	***Utility Actual***	e_1 ***Diff Ratio***	e_2 ***Diff Ratio***	***Utility Loss***
1.00E-06	148.97	2.16E-226	32.82	32.96	32.49	22.74	0.02	0.00	-9.75
1.00E-05	125.97	1.75E-191	32.82	32.96	32.29	25.46	0.02	0.00	-6.84
0.0001	102.97	1.42E-156	32.82	32.96	32.00	27.22	0.02	0.00	-4.79
0.001	79.97	1.14E-121	32.72	32.96	31.60	28.26	0.03	0.00	-3.34
0.01	57.00	8.39E-87	32.72	32.96	31.33	28.99	0.03	0.00	-2.34
0.1	34.30	2.51E-52	32.52	32.96	34.96	33.27	0.03	0.00	-1.69
1	13.85	3.65E-21	32.12	32.96	80.90	79.47	0.04	0.00	-1.44
10	2.62	1.35E-03	31.82	32.98	441.64	440.17	0.05	0.00	-1.47
100	0.30	36.63268074	31.82	33.00	1334.65	1333.16	0.05	0.00	-1.50
1000	0.03	903.2357427	31.72	33.00	1693.89	1692.39	0.06	0.00	-1.50
1.00E+04	0.00	9898.608279	31.72	33.00	1740.97	1739.47	0.06	0.00	-1.50
1.00E+05	0.00	99898.12891	31.72	33.00	1745.83	1744.33	0.06	0.00	-1.50
1.00E+06	0.00	999898.0808	31.72	33.00	1746.32	1744.82	0.06	0.00	-1.50

Notes: Actual effort was 33.60 in Period 1 and 32.95 in Period 2.

Table B.2 reports the same variables as Table B.1, but under this restriction. We find that the bounds on behavior change: the robust bounds are that effort should decline by at least 2.3 percent. However, the bounds are still much larger than what we observe. For reasonable Frisch elasticities (values of which are discussed in the main text), the reduction is around 3 percent. In fact, the pattern of the relationship between the allowable parameters and the size of the reduction in Period 1 is reversed (we now see that smaller values of a lead to smaller reductions of e_1).

What drives these differences in results? In particular, why are there no parameter values for which the worker wants to set their effort at the boundary point? The explanation lies in the fact that the utility function for the worker has multiple local equilibria, as discussed before. In particular, there can be a "low effort" local optimum and a "high effort" local optimum. For small a, the former is globally optimal, while for large a, the latter is globally optimal. Both local optima feature effort levels which are smaller than the actual observed effort levels, but only the latter is within the 20 percent reduction of effort bound. In addition, decreasing effort between high and low effort optima causes utility to fall and then rise again. If a worker is only able to reduce effort by 20 percent, they are still in a region where the best response is to choose the high effort local optima (i.e. the constraint is in the "valley" of their payoff function and so it is better to choose the high effort local optima rather than reduce effort by 20 percent). For large a, the high effort local optima is also global, and so workers behave the same as if there was no bound on effort reduction. For a small a, the bound on reduction causes them to shift from the low effort (globally optimal) effort level to the high effort local optima. The optimal e_1 for this local optima is actually decreasing in a, and so we observe the tightest bounds for small a's now.

Interestingly, while we now observe the smallest reduction of effort for small a's, they still generate the largest utility losses from mis-optimization. Thus, workers would be highly incentivized to ensure that they reduced their effort, even if only by 2 percent. This is because when a is small, most utility benefits are generated because of incentive payments, and so mis-optimization is particularly costly (relative to a large a, where most utility benefits come from intrinsic motivation).

For the second robustness check we assume that individuals are present biased. We set $\delta = 0.6$ (which is below the lowest estimate of the short-run discount parameter from a recent meta-analysis by Cheung, Tymula, and Wang (2021)) and re-estimate the model. The results are displayed in Table B.3. We find that the bounds on effect sizes are smaller, but there is still at least a 4 percent reduction, and there are still large reductions (on the order of 70 percent) for reasonable Frisch elasticities. Thus, the overall conclusion remains the same. For any given allowable parameter combination, the bounds are somewhat tighter than under our standard parameterization due to the fact that with more discounting of the future, the worker is less concerned about giving up future benefits for current incentive payments.

Table B.3: Simulated optimal behavior in INDIVIDUAL: Robustness to $\delta = .6$

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
a	γ	θ	e_1 ***Predicted***	e_2 ***Predicted***	***Utility Predicted***	***Utility Actual***	e_1 ***Diff Ratio***	e_2 ***Diff Ratio***	***Utility Loss***
1.00E-06	148.97	2.16E-226	8.52	30.06	32.40	18.83	0.75	0.09	-13.57
1.00E-05	125.97	1.75E-191	8.58	30.10	32.40	21.58	0.74	0.09	-10.82
0.0001	102.97	1.42E-156	8.72	30.20	32.40	23.39	0.74	0.08	-9.01
0.001	79.97	1.14E-121	8.86	30.30	32.43	24.50	0.74	0.08	-7.93
0.01	57.00	8.39E-87	9.12	30.48	32.66	25.29	0.73	0.08	-7.38
0.1	34.30	2.51E-52	9.80	30.96	35.06	28.98	0.71	0.06	-6.08
1	13.85	3.65E-21	32.52	32.96	68.59	67.69	0.03	0.00	-0.90
10	2.62	1.35E-03	32.30	32.96	370.22	369.40	0.04	0.00	-0.82
100	0.30	36.63268074	32.26	32.96	1117.10	1116.28	0.04	0.00	-0.81
1000	0.03	903.2357427	32.26	32.96	1417.55	1416.74	0.04	0.00	-0.81
1.00E+04	0.00	9898.608279	32.24	32.98	1456.94	1456.12	0.04	0.00	-0.81
1.00E+05	0.00	99898.12891	32.24	32.98	1461.00	1460.18	0.04	0.00	-0.81
1.00E+06	0.00	999898.0808	32.24	32.98	1461.40	1460.59	0.04	0.00	-0.81

Notes: Actual effort was 33.60 in Period 1 and 32.95 in Period 2.

For the third check, we consider how robust our results are to the specification of effort costs. Thus, we assume a different cost function: we use an exponential cost function $\theta\frac{\exp(\gamma e)}{\gamma}$ (as, e.g., in DellaVigna and Pope (2018)) instead of $\theta\frac{e^{\gamma+1}}{\gamma+1}$. Our static labor supply equation becomes

$$\max_e ae + \max[\min[-\kappa + \frac{ew}{\eta}, M], 0] - \theta\frac{\exp(\gamma e)}{\gamma}$$

This implies that our two equations for calibration become

$$R = \frac{\log(\frac{a+w}{\theta})}{\log(\frac{a}{\theta})} \qquad \text{and} \qquad E = \frac{1}{\gamma}\log(\frac{a+\frac{w}{\eta}}{\theta})$$

The two period optimization problem is now:

$$\begin{aligned}\max_{e_1,e_2} \quad & ae_1 + \max[\min[-\kappa + e_1 w, M], 0] - \theta\frac{\exp(\gamma e_1)}{\gamma} \\ + \quad & \delta(ae_2 + \max[\min[-\kappa + \frac{e_2 w}{\frac{1}{2}r + \frac{1}{2}\frac{e_1}{h}}, M], 0] - \theta\frac{\exp(\gamma e_2)}{\gamma})\end{aligned}$$

Table B.4 reports the results of the calibration and simulation. Compared to the main specification the bounds change in a negligible fashion. In particular, for small a we find that the worker should engage in extremely large distortions, while for large a, the worker should reduce effort by about 5 percent.[61]

Table B.4: Simulated optimal behavior in INDIVIDUAL: Robustness to alternative cost

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
a	γ	θ	e_1 ***Predicted***	e_2 ***Predicted***	***Utility Predicted***	***Utility Actual***	e_1 ***Diff Ratio***	e_2 ***Diff Ratio***	***Utility Loss***
1.00E-06	4.75	3.08E-68	8.46	30.02	49.33	21.38	0.75	0.09	-27.95
1.00E-05	4.02	9.67E-58	8.52	30.06	49.32	24.69	0.75	0.09	-24.63
0.0001	3.28	3.03E-47	8.58	30.10	49.31	26.81	0.74	0.09	-22.50
0.001	2.55	9.50E-37	8.72	30.20	49.35	28.08	0.74	0.08	-21.27
0.01	1.82	2.90E-26	8.92	30.34	49.66	28.96	0.73	0.08	-20.70
0.1	1.09	6.87E-16	9.46	30.72	52.84	33.31	0.72	0.07	-19.54
1	0.44	1.92E-06	11.36	32.06	84.60	79.06	0.66	0.03	-5.54
10	0.08	8.28E-01	31.88	32.98	369.85	368.38	0.05	0.00	-1.47
100	0.01	7.55E+01	31.80	32.98	-1.47E+04	-1.47E+04	0.05	0.00	-1.49
1000	0.00	9.72E+02	31.80	32.98	-1.96E+06	-1.96E+06	0.05	0.00	-1.49
1.00E+04	0.00	9.97E+03	31.80	32.98	-2.01E+08	-2.01E+08	0.05	0.00	-1.49
1.00E+05	0.00	1.00E+05	31.80	32.98	-2.01E+10	-2.01E+10	0.05	0.00	-1.49
1.00E+06	0.00	1.00E+06	31.78	32.98	-2.01E+12	-2.01E+12	0.05	0.00	-1.49

Notes: Actual effort was 33.60 in Period 1 and 32.95 in Period 2.

B.1.2 Calibration of equilibrium outcomes with efficiency

We next turn to discussing in more detail how we conduct the analysis of equilibrium effects of transparency mentioned in Section 3.5. What is the structure of the optimal contract if the firm had to use a transparent contract, i.e., if workers took into account all aspects of the contract? And what would be the impact of those contracts (and the assumption of full transparency) on firm profits, worker utility, and overall efficiency? The second question is at the heart of the literature on the ratchet effect. As Laffont and Tirole (1988) illustrate in their early paper on dynamic incentives, if workers are fully rational, a firm facing ratchet effects actually does worse than the static second-best contract. Of course, given that our workers do not seem to fully respond to dynamic incentives in the contract, perhaps these

[61]Note that for low γ values the effort cost grows very large and overall utility can become negative.

dynamic contracts actually sidestep the efficiency losses that are implied by the analysis of Laffont and Tirole (1988).

In order to formalize the issues surrounding efficiency, the equilibrium effects of opaque contracts and recover parameters, we need to enrich our environment and make several assumptions.

1. The firm can sell one unit of effort by the worker for a price p, which is set in a competitive market for goods (recall that we focus on a single, representative worker). Moreover, given the worker's (recall we assume a representative worker) calibrated preference parameters, the firm is profit maximizing in Period 2. In other words, given the induced level of effort by workers in Period 2, the price for which the firm sells each unit of effort is equal to the marginal cost of effort.

2. In Period 1, the firm has perfect knowledge of the worker's preference parameters a and γ, but that they do not know θ.[62] Instead, they have an incorrect point belief about θ denoted $\hat{\theta}$. Thus, in the first period, the firm does not know the preference parameters of the workers, but rather faces uncertainty about them. Since effort is measurable and contractible, this implies that we are in a situation that features adverse selection but not moral hazard. Given this, in Period 1 the firm maximizes profit given their knowledge of the parameters of the model (including a, γ and $\hat{\theta}$).

3. The set of workers at the firm is fixed — there are no extensive margins of labor supply. Instead, there is only an intensive margin of labor supply determined by the worker's optimization problem.

4. The firm cannot adjust the hourly wage. We believe in our setting this is not unreasonable; as the firm felt it could not (and did not want to) alter the hourly wage in response to introducing static or dynamic incentives. Thus, the only available instrument for the firm to alter is the bonus scheme.

5. When simulating counterfactuals, in order to keep our model as close to reality as possible we will assume that the set of possible bonus schemes that the firm can utilize

[62] This assumption is similar to the many applied models, as well as empirical applications, which assume that the marginal cost of effort (which is θ in our setting) is the relevant unknown parameter.

must follow the general form set out in Section 2:

$$\hat{g}(x) = \begin{cases} 0 & \text{if } x \leq \bar{E} \\ [x - \underline{E}] & \text{if } \underline{E} \leq x \leq \bar{E} \\ [\bar{E} - \underline{E}] & \text{if } x \geq \bar{E} \end{cases}, \text{ where } x_1 = \frac{e_1}{\eta_1} \text{ and } x_2 = \frac{e_2}{0.5e_1 + 0.5r}.$$

We assume (as in reality) that r is simply a function of η_1, and so $r = 1.09313\eta_1$. Given this structure, we only allow the firm to adjust the target rate in Period 1. We call the target rates that the firm can choose counterfactually in Periods 1 and 2 ρ_1 and $\rho_2 = 1.09313\rho_1$, in order to distinguish them from the target rates in other parts of our analysis, which were given by the firm in the observed data (and which we called η_1 and η_2).[63] This implies that in observed data $\rho_1 = 1$. Our approach allows for a large set of possible contracts. The set contains contracts where the worker earns positive incentive pay in one, or both periods. It also allows for considering contracts where the firm pays no bonus. By setting the target rates in Period 1 and Period 2 to be extremely large (regardless of the effort put forth by the worker in Period 1), they effectively ensure that workers never earn any bonus – only the fixed hourly wage. Of course, one could consider other counterfactuals. For example, a distinct counterfactual is considering what would occur if the firm only used static contracts (in other words, setting $\varsigma = 0$). We can use our approach to speak to this scenario as well. In this case, workers would behave the same whether the contract was transparent or opaque. Moreover, it means, given our assumptions, that in Period 2 the firm mis-optimizes (as they have less information in Period 1 than in Period 2) and so this outcome would be worse than the status quo (for the firm) with the opaque contract. A distinct approach is where firm could change the length of time that passes before a new rate is set. The extreme points of such adjustments are easy to map to what we already do — if the rate is reset every period, then we simply have the setup we focus on in this section; if the rate never resets, this is equivalent to the firm refusing to update the contract. Last, instead of adjusting rates, the firm could adjust the wage per unit of effort. Our approach allows for "as if" wage adjustment through changes in target rate, subject to

[63] We focus on the firm controlling the Period-1 target rate, as this was the primary lever that the firm actually used to alter the incentive scheme.

the structure described above.

With these assumptions in hand, we then search for the profit maximizing target rate in Period 1, and with this target rate, solve out for the worker's actual behavior over the two periods and calculate counterfactual profits, utility and total surplus.

Given our assumptions, with a fully transparent contract there are three sources of inefficiency. The first is that the firm is a monopsonist purchaser of labor — they face an upwards sloping labor supply curve (observe that we suppose conditional on a worker being at the firm, they provide effort in accordance with the provision of incentives, which generates the labor supply curve for the firm). Second, the firm faces a potential asymmetric information problem — they do not know the preference parameters of the worker, which means that in the first period, the optimal contract will generically be incorrect. The third source is that in order to prevent the firm from learning the preference parameters fully, and thus eliminating second-period rents, workers will shirk in supplying effort in Period 1. In order to estimate the parameters of the model (in particular, the price, or marginal value, of a unit of effort, as well as the firm's mis-estimation of the preference parameter of the worker) we assume that in Period 2 the firm offers the first best contract, given the labor supply it faces, and in Period 1 it offers the constrained best optimum given its beliefs about the worker's preference parameters.[64]

Regardless of workers' rationality, the first source of inefficiency is always present. This implies that the firm will earn positive profits. We will examine the consequences of changing from a fully opaque contract (where the firm is optimally responding to workers ignoring the dynamic implications of the contract) to a situation where the contract is fully transparent and workers fully respond, and firms offer the profit maximizing contract in anticipation of this. How will this alter the payoffs of firms and workers, as well as overall efficiency (which

[64]Such an approach, while allowing us to consistently compute our counterfactuals, requires one caveat. Recall we assume that workers optimally choose labor supply in Period 2 of INDIVIDUAL. But the fact that observed effort is actually slightly *higher* in Period 1 than Period 2 implies that workers oversupply labor in Period 1 given the contract they face. Thus, firm profits in Period 1 tend to be higher than what the model would predict, above and beyond workers ignoring dynamic incentives. We could instead assume that labor supply is chosen optimally by completely myopic agents in Period 1, but this would lead to the same problem, but in reverse. This implies that we do not necessarily observe lower overall efficiency in Period 1 relative to Period 2, as we would expect given the assumption that firms have less information in Period 1. However, we conducted a robustness check where we assume that workers supply the same labor in Period 1 and Period 2 and find that it does not qualitatively affect the results we discuss below.

is simply the sum of the two)?

We now discuss the six steps we use in our simulation. First, we need to estimate the marginal revenue (output price) per unit effort, p. Denote $\frac{1}{2}r + \frac{1}{2}\frac{e_1}{h} = \rho_2$. Recall that the worker's problem (so long as they earn positive marginal wage, which they do in reality) implies that labor supply obeys the following equation: $\frac{w}{\rho_2} = \theta e_2^\gamma - a$. We assume that the firm is maximizing profits, in Period 2, given the actual preference parameters of the representative worker. Thus, the firm's problem is to maximize $pe_2 - \frac{w}{\rho_2}e_2$ (observe we can drop κ without altering the optimum). Substituting in, the profit maximization problem becomes $pe_2 - (\theta e_2^\gamma - a)e_2$.

The first-order condition implies that $p = (\gamma+1)\theta e_2^\gamma - a$. We substitute into the equation our value of e_2, denoted $e_{2,act}$ (i.e. the actual e_2) along with the rest of the parameters to solve out for p. This implies an output price of $p = (\gamma+1)\theta e_{2,act}^\gamma - a$.

In the second step, we solve out for payments to the worker (i.e. total wages), utility, profits for the firm, and total surplus in each period given observed behavior (we ignore the flat wages, and only focus on incentive payments).[65] We subscript these with "act", as well as a time indicator indicating which period it is relevant to. Denote total surplus as TS. Note that wages are simply transfers between firms and workers, and total surplus is simply the revenue of firms given the price, plus the non-pecuniary benefits of production, less the costs of producing effort.[66] Thus

- $TS_{1,act} = pe_{1,act} - \theta\frac{e_{1,act}^{\gamma+1}}{\gamma_1} + ae_{1,act}$ and $TS_{2,act} = pe_{2,act} - \theta\frac{e_{2,act}^{\gamma+1}}{\gamma_1} + ae_{2,act}$
- $Payments_{1,act} = \max[\min[-\kappa + e_{1,act}w, M], 0]$ and $Payments_{2,act} = \max[\min[-\kappa + \frac{e_{2,act}w}{\frac{1}{2}r+\frac{1}{2}\frac{e_{1,act}}{h}}, M], 0]$
- $Profits_{1,act} = pe_{1,act} - \max[\min[-\kappa + e_{1,act}w, M], 0]$ and $Profits_{2,act} = pe_{2,act} - \max[\min[-\kappa + \frac{e_{2,act}w}{\frac{1}{2}r+\frac{1}{2}\frac{e_{1,act}}{h}}, M], 0]$.

[65] We compute total surplus and firm profits without discounting, while we compute utility using our discount fact of 0.913 thus, the sum of utility and profits does equal surplus). We do this to make utility directly comparable to our previously "partial equilibrium" simulations.

[66] We assume that the non-pecuniary motives captured by a are intrinsic motivations, and so matter for agents' utility and total surplus, rather than concerns about firing threats, relational contracting considerations, etc. which might not enter those terms.

- $Utility_{1,act} = ae_{1,act} + \max[\min[-\kappa + e_{1,act}w, M], 0] - \theta\frac{e_{1,act}^{\gamma+1}}{\gamma+1}$, $Utility_{2,act} = \delta(ae_{2,act} + \max[\min[-\kappa + \frac{e_{2,act}w}{\frac{1}{2}r+\frac{1}{2}\frac{e_{1,act}}{h}}, M], 0] - \theta\frac{e_{2,act}^{\gamma+1}}{\gamma+1})$

Third, we estimate the firm's belief about the worker's θ in Period 1 (before they learn from effort in Period 1).[67] Given our limited data, we cannot estimate a distribution over potential values of θ that captures the firm's uncertainty in Period 1. Instead, we make a much simpler assumption, namely that the firm has degenerate beliefs on an incorrect value of θ. This means that between Periods 1 and 2 the firm must engage in non-Bayesian updating after they learn the true θ. Call $\hat{\theta}$ the firm's guess of θ in Period 1 (separate from the actual θ). We will assume that given $\hat{\theta}$ and the other parameters of the model, the firm was profit maximizing in Period 1. Using the firm's FOC, along with the fact that workers treat the first period as a static decision-problem (as the dynamic incentives are opaque), we can solve for $\hat{\theta}$: $\hat{\theta} = \frac{p+a}{(\gamma+1)e_{1,act}^{\gamma}}$

In the fourth step, we solve for the ρ_1 the firm would want to set in order to maximize profits if contracts were fully transparent to workers, given their beliefs about the parameters of the model (including $\hat{\theta}$) and the fact that $r = \rho_1 \times 1.09313$. First, the firm anticipates the worker's labor supply. They consider the solution to the worker's problem:

$$\begin{aligned}\max_{e_1,e_2} \quad & ae_1 + \max[\min[-\kappa + \tfrac{e_1 w}{\rho_1}, M], 0] - \hat{\theta}\tfrac{e_1^{\gamma+1}}{\gamma+1} \\ + \quad & \delta(ae_2 + \max[\min[-\kappa + \tfrac{e_2 w}{\frac{1}{2}r+\frac{1}{2}\frac{e_1}{h}}, M], 0] - \hat{\theta}\tfrac{e_2^{\gamma+1}}{\gamma+1})\end{aligned}$$

Denote the solution as $\hat{e}_1^*(\rho_1), \hat{e}_2^*(\rho_1)$. The firm then calculates total profits

$$\pi(\rho_1) = p\hat{e}_1^*(\rho_1) - \max[\min[-\kappa + \frac{\hat{e}_1^*(\rho_1)w}{\rho_1}, M], 0] + p\hat{e}_2^*(\rho_1) - \max[\min[-\kappa + \frac{\hat{e}_2^*(\rho_1)w}{\frac{1}{2}r + \frac{1}{2}\frac{\hat{e}_1^*(\rho_1)}{h}}, M], 0]$$

The firm then searches over all ρ_1 (we set the bounds between 0 and 200) to find the one that maximizes profits.[68] Denote this as ρ_1^*, with an associated r^*. This is the optimal

[67]Recall we assume in Period 2 the contract is set optimally given the parameters.

[68]If ρ_1 is large enough, e.g., 200, it implies that for any potential level of effort the worker can choose that

contract chosen by the firm given their beliefs and the fact that workers are rational.

Fifth, we compute how workers will actually behave. Notice this means we compute behavior using θ, rather than $\hat{\theta}$. Therefore, we find the solution to the following maximization

$$\begin{aligned} \max_{e_1,e_2} \quad & ae_1 + \max[\min[-\kappa + \tfrac{e_1 w}{\rho_1^*}, M], 0] - \theta \tfrac{e_1^{\gamma+1}}{\gamma+1} \\ + \quad & \delta(ae_2 + \max[\min[-\kappa + \tfrac{e_2 w}{\frac{1}{2}r^* + \frac{1}{2}\frac{e_1}{h}}, M], 0] - \theta \tfrac{e_2^{\gamma+1}}{\gamma+1}) \end{aligned}$$

Denote these solutions as $e_1^*(\rho_1^*), e_2^*(\rho_1^*)$.

Sixth, we then compute total surplus, given ρ_1^*, using all the actual parameters (including θ), as well as payments, profits and utilities. We subscript this with "pred" (to indicate that it is predicted as a counterfactual).

- $TS_{1,pred} = pe_1^*(\rho_1^*) - \theta \frac{e_1^*(\rho_1^*)^{\gamma+1}}{\gamma+1} + ae_1^*(\rho_1^*)$ and $TS_{2,pred} = pe_2^*(\rho_1^*) - \theta \frac{e_2^*(\rho_1^*)^{\gamma+1}}{\gamma+1} + ae_2^*(\rho_1^*)$
- $Payments_{1,pred} = \max[\min[-\kappa + \frac{e_1^*(\rho_1^*)w}{\rho_1^*}, M], 0]$ and $Payments_{2,pred} = \max[\min[-\kappa + \frac{e_2^*(\rho_1^*)w}{\frac{1}{2}r^* + \frac{1}{2}\frac{e_1^*(\rho_1^*)}{h}}, M], 0]$
- $Profits_{1,pred} = pe_1^*(\rho_1^*) - \max[\min[-\kappa + \frac{e_1^*(\rho_1^*)w}{\rho_1^*}, M], 0]$ and $Profits_{2,pred} = pe_2^*(\rho_1^*) - \max[\min[-\kappa + \frac{e_2^*(\rho_1^*)w}{\frac{1}{2}r^* + \frac{1}{2}\frac{e_1^*(\rho_1^*)}{h}}, M], 0]$
- $Utility_{1,pred} = ae_1^*(\rho_1^*) + \max[\min[-\kappa + \frac{e_1^*(\rho_1^*)w}{\rho_1^*}, M], 0] - \theta \frac{e_1^*(\rho_1^*)^{\gamma+1}}{\gamma+1}$, $Utility_{2,pred} = \delta(ae_2^*(\rho_1^*) + \max[\min[-\kappa + \frac{e_2^*(\rho_1^*)w}{\frac{1}{2}r^* + \frac{1}{2}\frac{e_1^*(\rho_1^*)}{h}}, M], 0] - \theta \frac{e_2^*(\rho_1^*)^{\gamma+1}}{\gamma+1})$

Notice that for solving out for the optimal ρ_1 we use $\hat{\theta}$, but to compute total surplus, ρ_1^*, we actually solve out the worker's problem using θ.

We do this for each allowable parameter combination. Table B.5 reports the results. In both panels, column 1 gives the value of a. In the top panel columns 2 and 3 provide p and $\hat{\theta}$. The fourth column gives ρ_1^*, and columns 5 and 6 give the induced effort levels by workers given the optimal ρ_1, $e_1^*(\rho_1^*), e_2^*(\rho_1^*)$. Columns 7 and 8 provide $TS_{1,obs}$ and $TS_{2,obs}$, while 9 and

would generate positive utility, the worker will earn no incentives. In other words, the workers will never earn any bonus. In situations where multiple ρ_1s generate the same value, we break ties by choosing the largest ρ_1.

10 give $TS_{1,pred}$ and $TS_{2,pred}$. Column 11 gives the difference between actual and predicted total surplus. In the bottom panel, columns 2–6 are analogous to 7–11 in the top panel but provide information on profits, while columns 7–11 in the bottom panel give information on utility. We find that depending on the set of allowable parameters we consider, the optimal Period-1 target rate ρ_1 varies, but is never equal to 1 (which is what it actually is).

A key result is that if we compare actual total surplus to predicted total surplus (column 11 in the top panel), we find that actual total surplus is always higher. In many ways, this should not be surprising – in order to estimate parameters in our model we had to assume that the firm was achieving either first best (in Period 2) or static second best (in Period 1). For some parameter estimates the gain in efficiency from a lack of understanding of the contract is relatively small (e.g. for very large values of a they are around a gain of 68 utils). However, for some parameterizations the efficiency gain is relatively large, on the order of thousands of utils. Not surprisingly, firm profits are always higher under the actual, opaque contract relative to the counterfactual, transparent contract (see column 6 in the bottom panel). This is because workers ignore any dynamic effects of opaque contracts which might cause them to shirk, while still allowing the firm to dynamically adjust the contract.

The results for workers are more ambiguous. Recall that workers, facing a fixed contract, will always be worse off if they disregard the dynamic impact of their effort today on pay tomorrow. The questions is – does such an intuition extend to situations where the firm can adjust the contract depending on how transparent the contract is to workers? Our approach shows that it varies. For some parameters, this intuition extends — workers are better off with a transparent contract, even though the firm offers a different contract than they would if the dynamic attributes were opaque. However, for other parameter values, workers are (like the firm) worse off in the presence of transparent contracts. In these contracts the optimal ρ_1 for the firm to maximize profit is larger than 1 (which is the value it is in reality). Thus, the firm sets a higher target rate than what we currently observe. As mentioned in the main text, such target rates can be optimal under transparency because they reduce the future benefit from shirking. In contrast, with a complex contract, they may mis-optimize, but have an easier target, leading to a larger bonus. The benefit of bonuses can thus outweigh the costs of mis-optimization.

Table B.5: Simulated counterfactuals with transparent contracts

(1) a	(2) p	(3) $\hat{\theta}$	(4) ρ_1^*	(5) $e_1^*(\rho_1^*)$	(6) $e_2^*(\rho_1^*)$	(7) TS_1 ***Actual***	(8) TS_2 ***Actual***	(9) TS_1 ***Predicted***	(10) TS_2 ***Predicted***	(11) TS ***Act-Pred***
1.00E-06	420.53	1.184E-227	0.9	12.1	30.1	14118.28	13856.36	5088.38	12657.89	10228.36
1.00E-05	356.04	1.505E-192	0.9	12.1	30.1	11954.20	11731.18	4308.04	10716.7	8660.61
0.0001	291.55	1.911E-157	0.9	12.1	30.1	9789.50	9606.24	3527.81	8775.80	7092.12
0.001	227.14	2.405E-122	0.85	32.80	33.00	7626.42	7483.49	7449.46	7494.39	166.07
0.01	163.22	2.761E-87	0.80	32.80	33.00	5479.66	5377.13	5352.80	5384.93	119.06
0.1	102.42	1.288E-52	0.85	32.70	33.00	3439.12	3375.31	3350.18	3380.15	84.10
1	55.48	2.789E-21	0.90	32.30	33.10	1886.42	1852.62	1818.00	1860.43	60.61
10	36.36	1.281E-03	1.10	31.70	32.90	1432.59	1411.04	1368.27	1409.32	66.04
100	33.16	3.642E+01	1.15	31.50	32.80	1791.81	1772.32	1727.28	1767.71	69.13
1000	32.81	9.027E+02	1.15	31.50	32.80	1967.90	1948.63	1904.11	1944.08	68.34
10000	32.77	9.898E+03	1.15	31.50	32.80	1991.33	1972.08	1927.62	1967.54	68.26
1.00E+05	32.77	9.990E+04	1.15	31.50	32.80	1993.75	1974.51	1930.04	1969.96	68.26
1.00E+06	32.77	9.999E+05	1.15	31.50	32.80	1994.00	1974.75	1930.29	1970.20	68.26

(1) a	(2) $Profit_1$ ***Actual***	(3) $Profit_2$ ***Actual***	(4) $Profit_1$ ***Predicted***	(5) $Profit_2$ ***Predicted***	(6) ***Profit*** ***Act-Pred***	(7) $Utility_1$ ***Actual***	(8) $Utility_2$ ***Actual***	(9) $Utility_1$ ***Predicted***	(10) $Utility_2$ ***Predicted***	(11) ***Utility*** ***Act-Pred***
1.00E-06	14106.94	13843.87	5088.38	12603.89	10258.53	11.34	11.40	0.00	49.3	-26.55
1.00E-05	11940.04	11718.80	4308.04	10662.71	8688.08	14.15	11.30	0.00	49.30	-23.84
0.0001	9773.44	9594.02	3527.81	8721.79	7117.85	16.06	11.16	0.00	49.30	-22.08
0.001	7609.12	7471.49	7412.45	7473.48	194.68	17.30	10.96	37.01	19.08	-27.84
0.01	5461.47	5365.29	5308.69	5361.35	156.73	18.18	10.81	44.10	21.53	-36.64
0.1	3418.35	3361.62	3311.57	3357.39	111.02	20.77	12.50	38.61	20.78	-26.13
1	1841.30	1815.00	1762.31	1815.81	78.18	45.12	34.34	55.69	40.74	-16.96
10	1198.84	1184.95	1144.11	1185.03	54.66	233.75	206.42	224.16	204.78	11.23
100	1091.25	1079.44	1040.25	1078.52	51.92	700.56	632.60	687.03	629.23	16.89
1000	1079.55	1067.97	1029.28	1067.10	51.14	888.34	804.04	874.82	800.68	16.89
10000	1078.37	1066.81	1028.18	1065.95	51.06	912.96	826.52	899.44	823.15	16.89
1.00E+05	1078.26	1066.69	1028.07	1065.84	51.05	915.50	828.83	901.98	825.47	16.89
1.00E+06	1078.24	1066.68	1028.06	1065.82	51.05	915.75	829.07	902.23	825.70	16.89

B.2 Details of model calibration for the online experiments with warehouse workers

The calibration and simulation for the online experiments is quite similar to the calibration for the INDIVIDUAL field experiment. Again, we assume workers optimally respond to static incentives. The data show that the representative agent chooses effort on the upward-sloping portion of the incentive scheme (this is also true at the individual level, see Figure G.2). For each allowable parameter combination we then predict workers' response to the two different dynamic contract schemes. In COMPLEX, workers choose e_1 and e_2 to find the optimal solution to the utility function $U = o + w\hat{g}(\frac{e_1}{\eta_1}) - \theta\frac{e_1^{\gamma+1}}{\gamma+1} + ae_1 + \delta(o + w\hat{g}(\frac{e_2}{\eta_2}) - \theta\frac{e_2^{\gamma+1}}{\gamma+1} + ae_2)$, with the functional forms of $\hat{g}$ and η_2 discussed in Section 2. Of course the particular parameters we use here differ from those in the field study. The second scheme is the SIMPLE contract. Here workers maximize $U = o + w\hat{g}(\frac{e_1}{\eta_1}) - \theta\frac{e_1^{\gamma+1}}{\gamma+1} + ae_1 + \delta(o + w\hat{g}(\frac{e_2}{\eta_2}) - w\hat{g}(\frac{e_1}{\eta_1}) - \theta\frac{e_2^{\gamma+1}}{\gamma+1} + ae_2)$, where $\hat{g}$ satisfies the functional form described previously, and $\eta_2 = \eta_1$. As previously, for each allowable parameter combination we consider, we can compare the model prediction under SIMPLE and COMPLEX to the observed behavior, and calculate the predicted utility loss.[69]

We now take actual (observed) values as given, and do not engage in the normalization exercise we did with the field data. As before, the one-period worker problem is (with the exogenous target rate being η)[70]

$$\max_{e} \quad ae + \max[\min[-\kappa + \tfrac{ew}{\eta}, M], 0] - \theta\tfrac{e^{\gamma+1}}{\gamma+1}$$

In order to identify the three preference parameters θ, γ and a, we leverage two moments in the data. First, we use the ratio of effort when workers receive a positive bonus to when they receive no bonus for effort. In particular, we average the effort in Periods 3 and 4 of STATIC and STATIC_ZERO and take the ratio between them. Second, we use the average effort level in STATIC in Periods 3 and 4.[71] We consider a wide range of values of a, and

[69]We take the average behavior over Periods 1 and 3 in the experiment to be the observed e_1, and average behavior over Periods 2 and 4 in the experiment to the be e_2. In the experiment the exogenous part of the target rate in Period 2 was randomly drawn from three numbers which had a mean of 300. To simplify, we assume that workers optimize against the expected value.

[70]Although workers may not have been completely sure of what the exogenous portion of their rate would be in Period 2 (recall it could take on one of three closely spaced value), because all three potential values are quite close to one another, we assume that it takes on the expected value, which is equal to η (from Period 1).

[71]We use Periods 3 and 4 so that the amount of experience workers have is the same in the data from STATIC and STATIC_ZERO.

for each of them find a combination of θ and γ that rationalize the data (as before, we refer to these as allowable preference combinations). Thus, the two moments are[72]

$$R = \left(\frac{a + \frac{w}{\eta}}{a}\right)^{\frac{1}{\gamma}} \qquad \text{and} \qquad E = \left(\frac{a + \frac{w}{\eta}}{\theta}\right)^{\frac{1}{\gamma}}$$

As before, we consider sets of allowable parameters by changing a from 10^{-6} to 10^{6} by powers of 10. When we simulate optimal effort we assume no discounting, i.e., $\delta = 1$, since effort in all periods is paid out at the same time, and experimental periods are separated by a very short period of time.

Thus, the dynamic problem workers face in COMPLEX is

$$\begin{aligned} \max_{e_1,e_2} \quad & ae_1 + \max[\min[-\kappa + \tfrac{e_1 w}{\eta}, M], 0] - \theta \tfrac{e_1^{\gamma+1}}{\gamma+1} \\ + \quad & ae_2 + \max[\min[-\kappa + \tfrac{e_2 w}{0.5e_1+0.5r}, M], 0] - \theta \tfrac{e_2^{\gamma+1}}{\gamma+1} \end{aligned}$$

In the experiment, $r = \eta$.

In SIMPLE the optimization problem becomes

$$\begin{aligned} \max_{e_1,e_2} \quad & ae_1 + g(e_1) - \theta \tfrac{e_1^{\gamma+1}}{\gamma+1} \\ + \quad & ae_2 + [g(e_2) - g(e_1)] - \theta \tfrac{e_2^{\gamma+1}}{\gamma+1} \\ = \max_{e_1,e_2} \quad & ae_1 - \theta \tfrac{e_1^{\gamma+1}}{\gamma+1} + ae_2 + g(e_2) - \theta \tfrac{e_2^{\gamma+1}}{\gamma+1} \end{aligned}$$

where $g(e_i) = \max[\min[-\kappa + \frac{e_i w}{\eta}, M], 0]$.

For parameters, we have $R = \frac{472.15}{347.05}$, which is the ratio of average effort in the last two periods of STATIC relative to average effort in the last two periods of STATIC_ZERO, and $E = 472.15$, the numerator of R. $w = 1.25$, $\kappa = 0.125$, $M = 3.625$ and $\eta = 300$. The observed efforts in COMPLEX were 440.85 in Period 1 (here we average across the two periods where the workers faced dynamic incentives, i.e., Periods 1 and 3 of the actual

[72]Here, η appears in the moments because, unlike in the field data, we do not normalize the Period-1 target rate to 1. In the lab, we control the target rate and set them equal to η for all periods.

experiment), and 462.2 in Period 2 (here we average across the two periods after the workers faced dynamic incentives, i.e., Periods 2 and 4 of the actual experiment); while in SIMPLE they were 337.65 in Period 1 and 446.25 in Period 2 (where we again calculate these as averages across the relevant periods in the experiment).

Table B.6: Simulated behavior for warehouse lab experiment, COMPLEX contract

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
a	γ	θ	e_1 *Predicted*	e_2 *Predicted*	*Utility Predicted*	*Utility Actual*	e_1 *Diff Ratio*	e_2 *Diff Ratio*	*Utility Loss*
1.00E-06	27.08	1.66E-75	0	450	3.61	3.10	1.00	0.03	-0.51
1.00E-05	19.60	1.59E-55	0	450	3.59	3.07	1.00	0.03	-0.52
0.0001	12.19	1.06E-35	0	450	3.59	3.06	1.00	0.03	-0.53
0.001	5.33	2.81E-17	0	450	3.79	3.46	1.00	0.03	-0.33
0.01	1.13	1.34E-05	404	454	6.49	6.47	0.08	0.02	-0.02
0.1	0.13	4.60E-02	398.4	452.4	10.91	10.88	0.10	0.02	-0.03
1	0.01	9.24E-01	397.8	452.2	12.03	12.00	0.10	0.02	-0.03
10	0.00	9.92E+00	397.7	452.2	12.16	12.13	0.10	0.02	-0.03
100	0.00	99.92086	397.7	452.2	12.18	12.15	0.10	0.02	-0.03
1000	1.35E-05	999.9208	397.7	452.2	12.18	12.15	0.10	0.02	-0.03
1.00E+04	1.35E-06	9999.921	397.7	452.2	12.18	12.15	0.10	0.02	-0.03
1.00E+05	1.35E-07	99999.92	397.7	452.2	12.18	12.15	0.10	0.02	-0.03
1.00E+06	1.35E-08	999999.9	397.7	452.2	12.18	12.15	0.10	0.02	-0.03

Notes: Actual effort was 440.85 in Period 1 and 462.2 in Period 2.

Table B.7: Simulated behavior for warehouse lab experiment, SIMPLE contract

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
a	γ	θ	e_1 *Predicted*	e_2 *Predicted*	*Utility Predicted*	*Utility Actual*	e_1 *Diff Ratio*	e_2 *Diff Ratio*	*Utility Loss*
1.00E-06	27.08	1.66E-75	347.04	472.15	1.77	1.72	-0.03	-0.06	-0.05
1.00E-05	19.60	1.59E-55	347.04	472.15	1.75	1.71	-0.03	-0.06	-0.04
0.0001	12.19	1.06E-35	347.04	472.15	1.77	1.74	-0.03	-0.06	-0.03
0.001	5.33	2.81E-17	347.04	472.15	2.22	2.20	-0.03	-0.06	-0.02
0.01	1.13	1.34E-05	347.04	472.15	5.27	5.25	-0.03	-0.06	-0.01
0.1	0.13	4.60E-02	347.04	472.15	9.70	9.68	-0.03	-0.06	-0.01
1	0.01	9.24E-01	347.04	472.15	10.82	10.81	-0.03	-0.06	-0.01
10	0.00	9.92E+00	347.04	472.15	10.95	10.94	-0.03	-0.06	-0.01
100	0.00	99.92086	347.04	472.15	10.96	10.95	-0.03	-0.06	-0.01
1000	1.35E-05	999.9208	347.04	472.15	10.96	10.95	-0.03	-0.06	-0.01
1.00E+04	1.35E-06	9999.921	347.04	472.15	10.96	10.95	-0.03	-0.06	-0.01
1.00E+05	1.35E-07	99999.92	347.03	472.15	10.96	10.95	-0.03	-0.06	-0.01
1.00E+06	1.35E-08	999999.9	347.02	472.09	10.96	10.95	-0.03	-0.06	-0.01

Notes: Actual effort was 337.65 in Period 1 and 446.25 in Period 2.

Tables B.6 and B.7 show the results of the calibration and simulation for the COMPLEX and SIMPLE contracts respectively. The first three columns show the allowable parameter estimates. Columns 4 and 5 show the optimal effort levels in Periods 1 and 2, while columns 6 and 7 show the utility at the optimum and the utility resulting from actual effort choices. Columns 8 and 9 show the difference between actual and predicted effort, divided by actual effort, while column 10 shows the utility difference between actual and optimum efforts.

We first turn to COMPLEX. As can be seen, in COMPLEX workers fail to choose the predicted effort, with the minimum difference between actual and predicted effort being at least 10 percent. Just as in INDIVIDUAL, we see a wide range of possible responses – from 10 percent reduction up to a 100 percent reduction. Just as in INDIVIDUAL this is driven by the fact there are multiple local maxima, and changes in the parameters both move the location of each local maximum as well as which local maximum is global. Despite the fact that the minimum reduction is rather small (only 10 percent), our best guess is that the actual reduction is much larger. In the AMT experiments (discussed in Section B.3), where we can point identify the optimal effort, we estimate that subjects should put in zero effort in Period 1. We might thus expect the actual optimum to be close to 0 here as well.

In contrast, for SIMPLE, as Table B.7 shows, workers actually work too little relative to the prediction of the model, and the absolute difference, regardless of parameters, is always about 3 percent. This is because SIMPLE is equivalent to having a zero wage in Period 1. The model is parameterized to match data which pins down the effort at a wage of 0, and so all parameters make the same prediction. Notice that these results imply that workers fully take into account that SIMPLE implies a wage of 0 in Period 1.

B.3 Details of estimation for the experiments with AMT workers

The details of estimating and simulation for the AMT experiments are similar to those of the online experiments with warehouse workers. We again incorporate the fact that the representative worker chooses clicks in-between the quota and cap values, so that the relevant first order conditions are for the interior (it is also true at the individual level that almost every worker is in the interior, see Figure H.2). One key difference is that we now have three wage levels for the static contract across the treatments STATIC_ZERO, STATIC_LOW and STATIC and so the model is point identified. Figure H.3 shows the average effort levels for the three different piece rate levels. We use the individual-level data under these three piece rates to estimate parameter values for our model, using bootstrapping for the confidence intervals. Denote the three wage levels $w_0 = 0$, $w' > 0$ and $w'' > w'$, for the three treatments respectively. Given that we have three moments and three parameters (a, θ, γ), we can directly solve for parameters as a function of the data.

The three first-order conditions are[73]

$$e_0 = \left(\frac{a}{\theta}\right)^{\frac{1}{\gamma}}$$

$$e' = \left(\frac{a + \frac{w'}{\eta}}{\theta}\right)^{\frac{1}{\gamma}}$$

[73]Just as in the warehouse workers' lab experiments, and unlike the calibrations involving field data, η appears here because we do not normalize effort to Period 1 SPM.

$$e'' = \left(\frac{a + \frac{w''}{\eta}}{\theta}\right)^{\frac{1}{\gamma}}$$

The first equation can be rewritten $\theta e_0^\gamma = a$. Substitution into the latter two equations gives

$$e' = \left(\frac{\theta e_0^\gamma + \frac{w'}{\eta}}{\theta}\right)^{\frac{1}{\gamma}} = (e_0^\gamma + \frac{w'}{\eta\theta})^{\frac{1}{\gamma}}$$

and

$$e'' = \left(\frac{\theta e_0^\gamma + \frac{w''}{\eta}}{\theta}\right)^{\frac{1}{\gamma}} = (e_0^\gamma + \frac{w''}{\eta\theta})^{\frac{1}{\gamma}}$$

We can rewrite these as

$$e'^\gamma = e_0^\gamma + \frac{w'}{\eta\theta}$$

and

$$e''^\gamma = e_0^\gamma + \frac{w''}{\eta\theta}$$

Solving out the first of these two equations for θ gives $\theta = \frac{w'}{\eta(e'^\gamma - e_0^\gamma)}$. Substituting into the final equation we obtain $e''^\gamma = e_0^\gamma + \frac{w''}{w'}(e'^\gamma - e_0^\gamma)$. Thus, we first solve for γ using $e''^\gamma = e_0^\gamma + \frac{w''}{w'}(e'^\gamma - e_0^\gamma)$; then given γ solve for θ using $\theta = \frac{w'}{\eta(e'^\gamma - e_0^\gamma)}$, and then for a using $a = \theta e_0^\gamma$.

Otherwise we proceed just as before. The parameters are $w_0 = 0$, $w' = 0.01$, $w'' = 0.5$, $\eta = 400$, $\kappa = 0.05$ and $M = 1.45$. The wage in the non-STATIC treatments is $w = 0.5$. We observe efforts of $e_0 = 205.45$, $e' = 387.75$ and $e'' = 491.3$ using (the average of the last two periods of) STATIC_ZERO, STATIC_LOW and STATIC, respectively.

For simulating, the optimization problems for COMPLEX and SIMPLE are exactly the same as before. The only difference is that we now only consider a single allowable (point identified) parameter combination. The observed efforts were 425.2 in Period 1 and 497.4 in Period 2 for COMPLEX and 211.0 in Period 1 and 489.6 in Period 2 for SIMPLE (again these are actually the average across Periods 1 and 3 of the experiment and Periods 2 and 4 of the experiment for the former and latter numbers).

To obtain confidence intervals around the model predictions, we use a bootstrap procedure, following the methodology of DellaVigna and Pope (2018). We have three exogenous wage treatments (STATIC_ZERO STATIC_LOW and STATIC) with sample size of $N = 73$, 118 and 82 respectively. Denote N_i as the sample size for treatment i. For each participant in the sample we take the average effort across the final two periods of the experiment (recall that all participants faced the same baseline wage in the first three periods).

To conduct our bootstrap, we draw N_i average efforts from the respective treatment, with replacement. We then compute the average effort in that bootstrapped sample for each treatment. We then use those three average efforts, along with our first-order conditions, to solve for the parameters that rationalize the average efforts. Each sample thus yields a vector of estimated parameters a, γ and θ. Given those parameters, we then predict effort and utility for Period 1 and 2 of the dynamic incentive scheme. We conduct this procedure 100 times, leading to 100 parameter vectors and the corresponding predictions of effort and utility. For each parameter or prediction, the 5th and 95th percentile yield the lower and upper bounds of the 95% confidence interval.

Table B.8 reports the results. The top panel gives the results for the COMPLEX contract, showing the point estimate, i.e., the prediction from the original data, as well as the bounds of the confidence intervals. The bottom panel gives the results for SIMPLE. The first three columns report the parameter estimates. Columns 4 and 5 show the optimal effort levels in Periods 1 and 2, while the sixth column shows the utility at the optimum. Columns 7–9 show the actual effort levels and the induced utility. Columns 10 and 11 show the difference between actual and predicted effort, divided by actual effort, while column 12 shows the utility difference between actual and optimum efforts.

The most apparent feature is that in COMPLEX workers should work 0 in Period 1 (with 95% CI [0,0]), but they actually work 425, while in SIMPLE they should work 204 (with CI [200, 210], and they actually work 211. Thus, our predictions here are in line with the predictions we observe for the warehouse workers in the same experiment, when a is less than 0.001 (observe that we estimate among the AMT workers a is 6.9×10^{-10} which is consistent with this bound on a among warehouse workers).

Thus, while workers actual effort is relatively far from predicted effort in COMPLEX, workers work in SIMPLE very close to the level that is predicted by our model, even if just outside the confidence interval. Because the model is designed specifically to match the observed effort when wages are 0, this implies that in SIMPLE the subjects understand that they earn a net zero wage by working in Period 1.

Table B.8: Simulated behavior for AMT lab experiment

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	a	γ	θ	e_1 ***Predicted***	e_2 ***Predicted***	***Utility Predicted***	e_1 ***Actual***	e_2 ***Actual***	***Utility Actual***	e_1***Diff Ratio***	e_2***Diff Ratio***	***Utility Loss***
COMPLEX												
Point estimate	6.23E-10	16.52	4.32E-48	0.00	512.36	1.16	425.20	497.45	0.99	1.00	-0.03	-0.17
95% CI lower bound	6.87E-11	15.69	-7.19E-47	0.00	508.81	1.15	425.20	497.45	0.98	1.00	-0.02	-0.17
95% CI upper bound	9.85E-10	17.00	8.62E-48	0.00	516.10	1.16	425.20	497.45	0.99	1.00	-0.04	-0.17
SIMPLE												
Point estimate	6.23E-10	16.52	4.32E-48	204.10	491.31	0.53	210.95	489.60	0.53	0.03	0.00	-6.04E-04
95% CI lower bound	6.87E-11	15.69	-7.19E-47	199.61	488.19	0.53	210.95	489.60	0.53	0.05	0.00	0.0004
95% CI upper bound	9.85E-10	17.00	8.62E-48	210.23	494.07	0.53	210.95	489.60	0.53	0.00	-0.01	-1.00E-04

B.4 Details of calibrating ψ

As discussed in the body of the paper, we assume that the individual underestimates how much a change in their effort will impact their pay tomorrow, while still understanding correctly the level of pay. In order to simplify our approach, while also remaining consistent with the data, we will assume that workers are exerting effort in a region where they face a strictly positive marginal bonus. We will first discuss our model in the context of the field data, and then extend it to the online experiment. For each data set and contract scheme, we can take the given and calibrated parameters, along with the observed e_1 and e_2 and find the ψ that rationalizes the data.

Recall the first-order condition for our worker in INDIVIDUAL (conditional on earning a positive marginal wage) is:

$$a + w - \theta e_1^{\gamma} - \delta \frac{1}{2h(\frac{1}{2}r + \frac{1}{2}\psi\frac{e_1}{h})} \frac{e_2 w}{\frac{1}{2}r + \frac{1}{2}\frac{e_1}{h}} = 0$$

We want to capture the worker misperceiving the marginal impact of their effort in Period 1 on their payment in Period 2, which is captured by the term $\delta \frac{1}{2h(\frac{1}{2}r+\frac{1}{2}\psi\frac{e_1}{h})} \frac{e_2 w}{\frac{1}{2}r+\frac{1}{2}\frac{e_1}{h}}$. We thus multiply this by ψ, where $\psi \in [0, 1]$ captures the degree to which the worker underperceives the impact of their effort in Period 1 on Period 2 payoffs (with $\psi = 1$ indicating full transparency). Notice that the worker still correctly perceives the level of payoff that they will receive tomorrow, which is the second fraction, $\frac{e_2 w}{\frac{1}{2}r+\frac{1}{2}\frac{e_1}{h}}$.

Thus, we focus on a worker with the first-order condition

$$a + w - \theta e_1^{\gamma} - \delta \frac{\psi}{2h(\frac{1}{2}r + \frac{1}{2}\psi\frac{e_1}{h})} \frac{e_2 w}{\frac{1}{2}r + \frac{1}{2}\frac{e_1}{h}} = 0$$

The simplest way to find ψ would be to take all the known parameters, and observed effort levels, and solve it for ψ.

However, for any ψ that solves the first-order condition (and note it will be unique, since the first-order condition is linear in ψ) we need to check the second-order condition. If that does not hold, given the ψ we have solved, we might not actually have found a point where the worker is maximizing utility. Thus, we need to check whether the second-order condition,

$$-\gamma\theta e_1^{\gamma-1} + \delta\frac{\psi}{2h^2}\frac{e_2 w}{(\frac{1}{2}r + \frac{1}{2}\frac{e_1}{h})^3}$$

is negative over the entire range of relevant e_1's. Unfortunately, this is almost never true. If $\psi > 0$, for small enough e_1 this is always positive (so long as e_2 is positive). Given this, we cannot just use the first-order condition and need to pursue a different approach. What we do is take the first-order condition and integrate up, and then substitute in ϵ for e_1. This gives us an objective function:

$$a\epsilon - \kappa + \epsilon w - \theta\frac{\epsilon^{\gamma+1}}{\gamma+1} + \delta\psi\frac{e_2 w}{\frac{1}{2}r + \frac{1}{2}\frac{\epsilon}{h}}$$

Notice that the derivative of this corresponds to our first-order condition where $e_1 = \epsilon$. For each value of $\psi \in [0,1]$, we compute the ϵ that maximizes this equation. Call this $\epsilon^*(\psi)$. Our goal is then to find the value of ψ that minimizes the distance between e_1 and $\epsilon^*(\psi)$, which we then report as our value of ψ. In other words, we find the ψ such that the predicted first-period effort is as close to the observed first-period effort as possible.

Table B.9: Estimates of ψ in warehouse

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
INDIVIDUAL				Warehouse Workers in Lab				
a	γ	θ	ψ	a	γ	θ	ψ ***COMPLEX***	ψ ***SIMPLE***
1.00E-06	148.97	2.16E-226	0	1.00E-06	27.08	1.66E-75	0.8	1
1.00E-05	125.97	1.75E-191	0	1.00E-05	19.60	1.59E-55	0.8	1
0.0001	102.97	1.42E-156	0	0.0001	12.19	1.06E-35	0.79	1
0.001	79.97	1.14E-121	0	0.001	5.33	2.81E-17	0.75	1
0.01	57.00	8.39E-87	0	0.01	1.13	1.34E-05	0.51	1
0.1	34.30	2.51E-52	0	0.1	0.13	4.60E-02	0.45	1
1	13.85	3.65E-21	0	1	0.01	9.24E-01	0.44	1
10	2.62	1.35E-03	0	10	0.00	9.92E+00	0.44	1
100	0.30	3.66E+01	0	100	0.00	99.92086	0.44	1
1000	0.03	9.03E+02	0	1000	1.35E-05	999.9208	0.44	1
1.00E+04	0.00	9.90E+03	0	1.00E+04	1.35E-06	9999.921	0.44	1
1.00E+05	0.00	9.99E+04	0	1.00E+05	1.35E-07	99999.92	0.44	1
1.00E+06	0.00	1.00E+06	0	1.00E+06	1.35E-08	999999.9	0.44	1

Table B.9 provides the results of the analysis. Columns 1–3 provide the allowable parameter combinations we consider. The next column provides the resulting value of ψ. Note that for all parameter combinations we find ψ to be 0. This should come as little surprise — the model is calibrated to match Period 2 effort, and we know from the data that workers are working harder in Period 1 than in Period 2, the opposite of the predictions for a rational agent.[74]

The process for finding ψ for the COMPLEX contract in the lab studies are exactly the same. Now the formula that we maximize over ϵ, for a given ψ, is[75]

$$a\epsilon - \kappa + \frac{\epsilon w}{\eta} - \theta \frac{\epsilon^{\gamma+1}}{\gamma+1} + \psi \frac{e_2 w}{\frac{1}{2}\eta + \frac{1}{2}\epsilon}$$

Table B.9 provides the resulting values of ψ for COMPLEX for our online experiments with the warehouse workers. Columns 5–7 give allowable parameter combinations, while column 8 provides our ψ for the COMPLEX contract. The value of ψ vary between about

[74] In fact, if we had allowed ψ to be negative, we would have found a negative ψ.

[75] Recall that in this setting $r = \eta$.

0.44 and 0.8 depending on the exact parameter combinations we consider. Why do we observe a much larger ψ here than in the field data? In online experiments the model is designed to predict that if ψ is 0, then effort should be around 472. We observe effort in Period 1 in complex as being around 440. Thus, we do observe some reduction in effort based on the point estimates. Because the mapping between effort and ψ is non-linear, we observe that individuals are inattentive ($\psi < 1$), but not fully so. In particular, even this relatively small reduction in observed effort relative to the prediction of $\psi = 0$ leads to a relatively large increase in the value ψ. This value of ψ is in contrast to the field, where workers in Period 1 worked too hard relative to what the model would predict, leading to $\psi = 0$.

Table B.10: Estimates of ψ for AMT

(1)	(2)	(3)	(4)
a	γ	θ	ψ
COMPLEX			
6.90E-10	16.52	4.30E-48	0.85
NOISE_MARGINAL			
6.90E-10	16.52	4.30E-48	0.87
NOISE			
6.90E-10	16.52	4.30E-48	1
SIMPLE			
6.90E-10	16.52	4.30E-48	1

In the AMT experiments, we use the same approach as in the previous paragraph to find ψ for COMPLEX and NOISE_MARGINAL. Table B.10 provides values of ψ for these two treatments. The columns again correspond to the allowable parameters and the value of ψ. Different rows correspond to different treatments (because each treatment is point identified in terms of parameters). Although the rational optimum for both of these treatments is to exert 0 effort, observed average efforts in periods with dynamic incentives are 425 and 362 for COMPLEX and NOISE_MARGINAL, respectively. However, these changes translate into only small differences in the ψ value, which range from 0.85 to 0.87. Again, finding $\psi > 0$ should not be a surprise: the model is estimated so that an agent with $\psi = 0$ puts forth an effort of 491. All of the observed efforts are below that level. Moreover, utility is a non-concave function of e_1. There is both an interior local maximum, as well as a corner local maximum at $e_1 = 0$. As ψ increases, it affects two things: it shifts the interior local maximum downwards, and it also causes the non-interior local maximum to become more attractive, compared to the interior maximum. Around ψ's of 0.85 the interior maximum becomes non-globally optimal, causing a dramatic shift in optimal effort to $e_1 = 0$. This generates an upper bound on the value of ψ. Moreover, just as in COMPLEX among the warehouse workers, the non-linearity of the objective function of the agent leads to a highly non-linear mapping between observed effort and ψ, so that small decreases in observed effort from the fully opaque optimum lead to dramatic increases in the value of ψ.

We next turn to considering our SIMPLE contracts in both the online experiments with warehouse workers and in our AMT experiments. We do a similar exercise. The first-order condition for a rational worker (who is earning a positive marginal bonus) is

$$a + \frac{w}{\eta} - \theta e_1^{\gamma} - \frac{w}{\eta} = 0$$

We again assume that the agent underestimates the impact of their effort today on their payment tomorrow by a degree $\psi \in [0, 1]$

$$a + \frac{w}{\eta} - \theta e_1^{\gamma} - \frac{\psi w}{\eta} = 0$$

The second-order condition is

$$-\gamma \theta e_1^{\gamma - 1}$$

which is always negative, and so we just need to solve the first-order condition for ψ.

The ninth column of Table B.9 provides the values of ψ for our SIMPLE treatment with the warehouse workers. It is 1 regardless of the allowable parameter combinations we consider. Full rationality implies that the individual should treat the SIMPLE contract exactly the same as a contract that pays 0 in Period 1. In the data we observe that individuals worked 347 when given a contract that paid 0. We see that workers actually work 337 in the SIMPLE treatment. Thus, workers are actually working less than they should, leading to our robust finding of $\psi = 1$.[76]

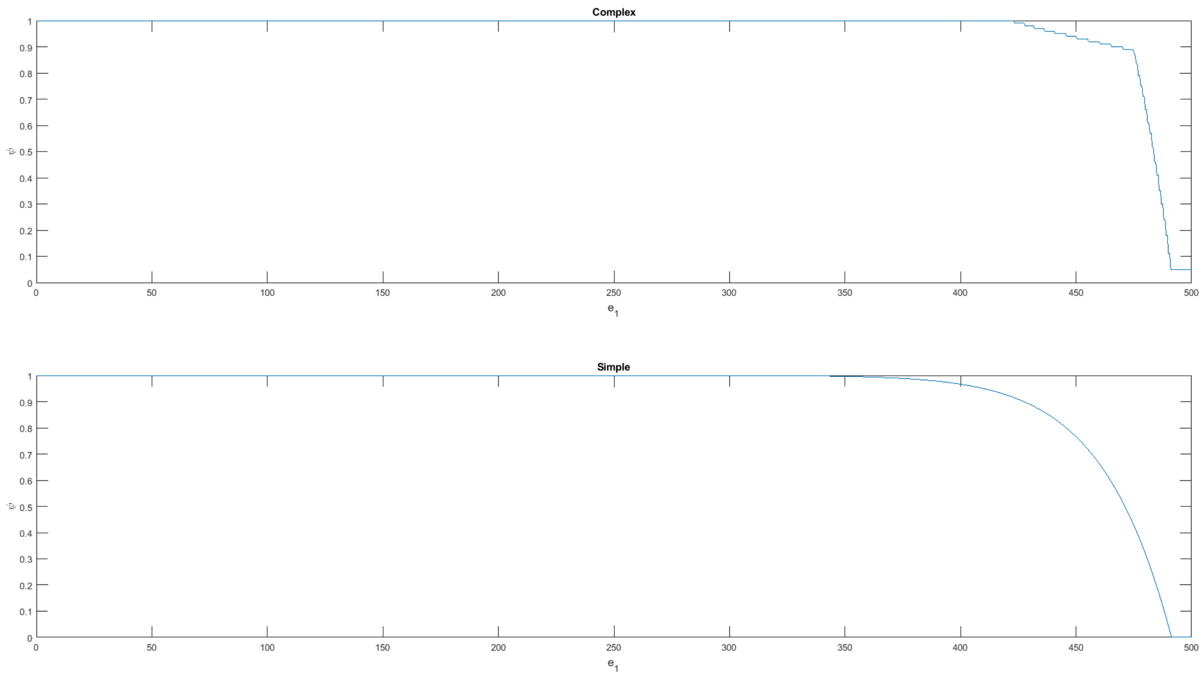
The lower rows of Table B.10 provide results for SIMPLE and NOISE, both of which are contracts that feature payments in Period 1 being subtracted from payments in Period 2.[77] Although there is minor variation in the value of ψ, due to differences in observed effort levels, both are extremely close to 1. As mentioned, perfectly rational workers should exert effort in SIMPLE as if they were facing a one-period problem, which means that they should provide an effort of 204. We observe average efforts in periods with dynamic incentives of 247 and 211, for NOISE and SIMPLE, respectively. However, these differences in effort level only translate to small changes in our ψ value. This mirrors the reasons provided for COMPLEX: starting from the fact that an effort of around 491 implies a ψ of 0, small decreases in effort lead to large increases in ψ (due to the fact that the model is highly non-linear). A decrease of 20 units of effort to 470 corresponds to an increase in ψ from 0 to around 0.5, while a decrease of effort to around 450 corresponds to ψ rising to 0.75. Thus, the marginal impact of even large additional decreases in effort from 450 to 200 cause only minor changes in ψ.

Figure B.3 shows that the relationship of Period-1 effort and ψ is highly concave. This means that if workers start out with very little understanding that Period-1 effort impacts Period-2 payments (low ψ), then a small increase in that understanding will change e_1 very little. Conversely, if a worker has a high degree of understanding (e.g., $\psi = 0.95$) then a little more understanding can change behavior to a far larger degree (large reduction in e_1).

[76] In fact, if ψ could be larger than 1 we would find it to be indeed being larger than 1.

[77] These all have the same first-order condition, because we observe the representative individual in all treatments being in the region where they are earning positive marginal wage even with non-linear transformation of normalized effort into money, and because we have abstracted away from the random noise in the target rate.

Figure B.3: Relationship of Period-1 effort and ψ



Notes: The graph shows the relationship between Period-1 effort on the x-axis and ψ on the y-axis, fixing Period-2 effort at the level we observe in the data. The top panel shows the relationship for the COMPLEX treatment and the bottom panel the one for the SIMPLE treatment.

C Additional empirical results for the INDIVIDUAL trial

In this appendix, we show a timeline of all the changes to the incentive scheme at the warehouse we study (Figure C.1). We then provide summary statistics and randomization checks for the INDIVIDUAL trial (Table C.1). Note that we have age, gender and nationality for only about 40 percent of workers (but who account for about 80 percent of the time worked during the trial). Missing demographic information is not related to treatment status ($p = 0.377$).

As can be seen in Figure C.2 and Table C.2, there is no differential attrition before and during the INDIVIDUAL trial.

In Table C.3, we provides the regression table underpinning Figure 2, including additional robustness checks.[78]

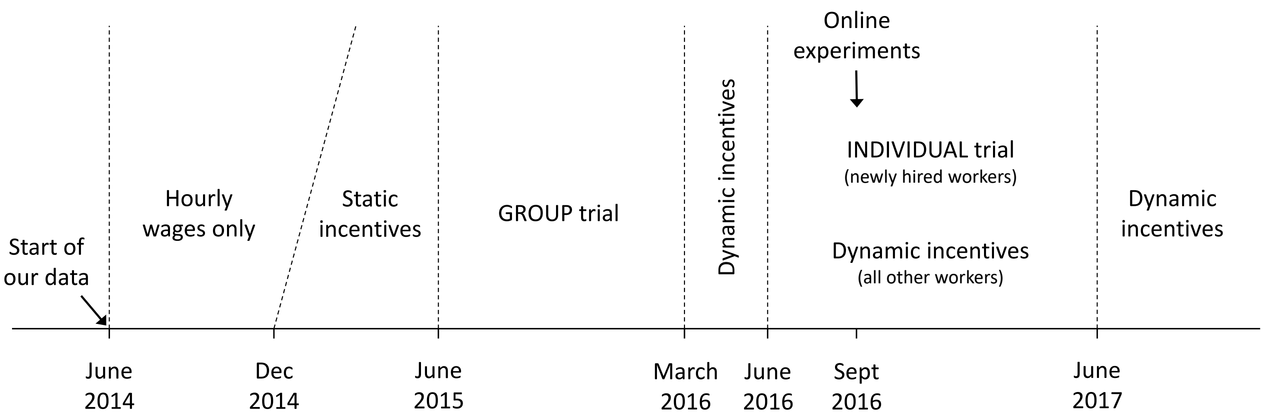
Figure C.3 shows that workers almost never worked below the quota or above the cap in the INDIVIDUAL trial. The figure shows the distribution of SPH per week during the INDIVIDUAL trial compared to the quota (26 SPH) and cap (44 SPH). We find that 93.1 percent of worker-weeks are on the positive slope of the incentive scheme. Only 2.2 percent are below the quota and 4.7 percent are above the cap, when marginal incentives are zero, and these worker-weeks are not far off the kink. There is no difference between treatment and control workers. The aggregate data indicate that the representative worker is on the positive slope, an empirical fact that we incorporate into our structural calibration. Figure C.3 shows that this holds at the individual level as well, for the vast majority of worker-weeks.

Figure C.4 shows the treatment effect in the INDIVIDUAL trial for each task separately; we see that for almost every task the treatment effect is less than the most conservative prediction of the rational model. The size of each bubble is proportional to the time spent in that task. The line is the pooled treatment effect as in Table C.3. The gray-shaded area depicts the range of treatment effects that is in line with our rational model (see Section 3.3). Since the sample size becomes relatively small if we consider the 75 tasks separately,

[78] If we repeat the regressions in columns 2 and 3, but restrict the sample to week 4 or 5 (instead of to week 6), the treatment effects are -0.5 percent and -0.7 percent (week 4) and 0.1 percent and -0.2 percent (week 5).

we find some variation between tasks, but there is only one task where the ratchet effect is in the predicted range of the rational model.[79]

Figure C.1: Timeline of changes to the incentive scheme



Notes: A timeline of the changes to the incentive scheme in the treated warehouse. Our data start in June 2014. Static incentives were gradually introduced starting in December 2014, with roll-out completed by March 2015.

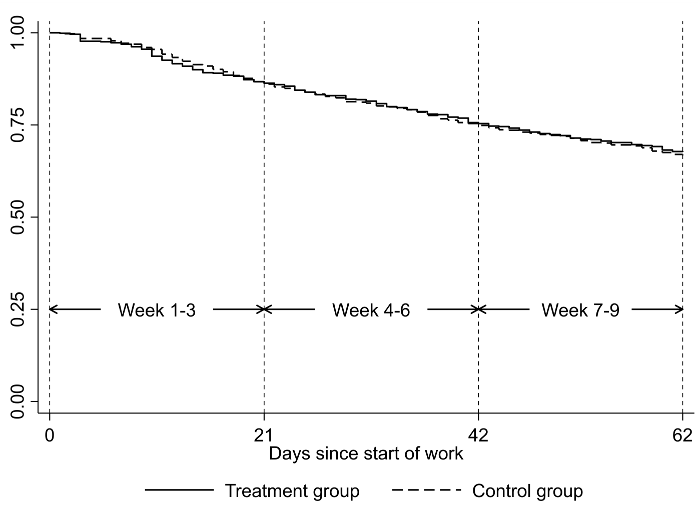
[79]We can also calculate an overall treatment effect by aggregating the treatment effect for each task. The resulting treatment effect is -0.2 percent with CI [-1.3, 0.8], almost identical to the pooled regression estimates in Table C.3. The one outlier task represents a trivial portion of workers or worker time. The share of workers who ever worked in this task is 5.9 percent. Of these workers, 60 percent spent less than a tenth of their time on the task. The share of minutes worked in the task was only 1.2 percent.

Table C.1: Summary statistics and randomization checks for the INDIVIDUAL trial

	Mean		p-value
	Control	Treatment	
Age at randomization	30.61	30.47	0.864
1 if female	0.29	0.30	0.910
1 if non-native	0.65	0.68	0.591
1 if temp/agency worker	0.34	0.32	0.342
# Workers	776	739	

Notes: Summary statistics of the workers randomized in the INDIVIDUAL trial. In total, we randomized 1515 workers into treatment and control conditions of which 1294 worked in the treatment period (weeks 4–6). The number of workers shown in the table includes workers with missing demographic information. P-values are from t-tests.

Figure C.2: Attrition in the INDIVIDUAL trial



Notes: Kaplan-Meier survival estimates for the INDIVIDUAL trial. The vertical lines show the start and end of the treatment period (weeks 4–6). Corresponding regressions are in Table C.2.

Table C.2: Attrition in the INDIVIDUAL trial

Dependent variable: Worker left firm						
	Week 1-9		Week 1-3		Week 4-9	
	(1)	(2)	(3)	(4)	(5)	(6)
1 if treated	0.9730	0.9796	1.0089	1.0361	0.9599	0.9665
	(0.087)	(0.089)	(0.142)	(0.150)	(0.113)	(0.115)
Cohort FE	No	Yes	No	Yes	No	Yes
# Workers	1515	1515	1515	1515	1306	1306

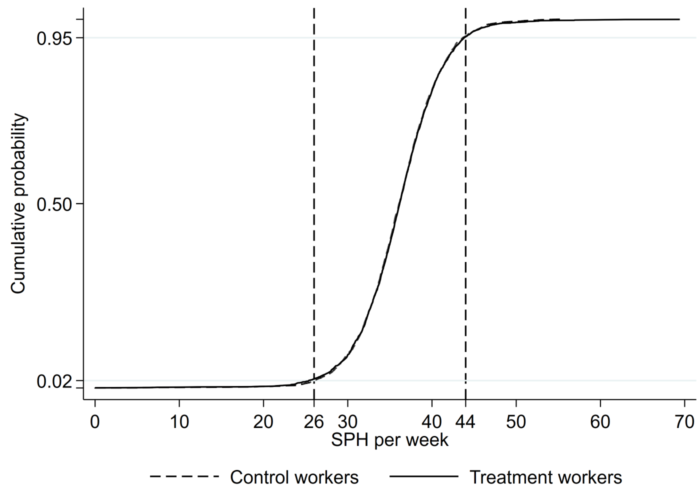
Notes: Hazard ratios from Cox proportional hazard models for the INDIVIDUAL trial. Robust standard errors in parentheses. The treatment dummy is 1 for the workers who faced ratchet incentives in weeks 4–6. The control workers did not face such incentives during that time. Since we have little pre-trial data for this trial, we only control for cohort fixed effects.

Table C.3: Ratchet effect in INDIVIDUAL trial

Dependent variable: ln(units per hour)			
	(1)	(2)	(3)
1 if treated	-0.001	-0.002	-0.005
	(0.005)	(0.006)	(0.007)
Sample	Weeks 4–6	Week 6	Week 6
			Attrited after week 9
Task FE	Yes	Yes	Yes
Shift FE	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes
all FE's × cohort	Yes	Yes	Yes
# Workers	1294	1147	969
# Shifts	607	550	549

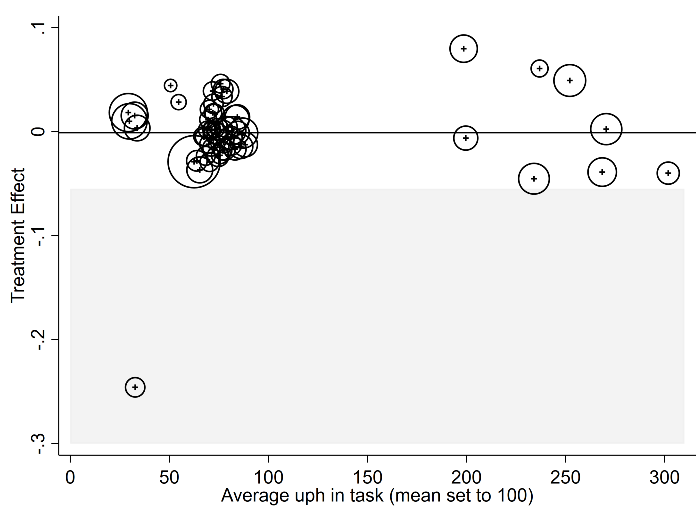
Notes: OLS regressions. Robust standard errors, using two-way clusters on individual workers and on shifts, are in parentheses. The sample is restricted to weeks 4–6, when the treatment workers ('1 if treated') faced a ratchet incentive to work more slowly, while the control workers did not face such an incentive. The spot incentives were identical for both groups. Specification 1 is the main regression using the full sample. Specification 2 restricts the sample to only week 6 to allow for some learning. Specification 3 further restricts the sample to only include workers who kept working for the firm until at least the end of week 9. These workers enjoy the full benefit of reducing effort in weeks 4–6, as the individualized rates were in effect for weeks 7–9. Participants are workers who had just started working for the firm. Within each starting week, workers were randomized into treatment and control. Cohort fixed effects control for this weekly cohort. All other fixed effects are also interacted with cohort.

Figure C.3: Distribution of effort in the INDIVIDUAL trial



Notes: Distribution of effort, measured as the sum of normalized effort (SPH) per week, in the INDIVIDUAL trial. Vertical dashed lines indicate the quota and cap of 26 SPH and 44 SPH respectively. Marginal incentives were positive between the quota and cap, and zero elsewhere. Horizontal light gray lines indicate the fractions of observations below the quota, or below the cap.

Figure C.4: Treatment effect in the INDIVIDUAL trial by task



Notes: The graph plots the treatment effect in the INDIVIDUAL trial for each task separately. The size of each bubble is proportional to the total time spent in that task. Tasks are sorted by the average number of units per hour (uph) achieved by workers. The solid line marks the overall treatment effect (see column 1 in Table C.3). The gray-shaded area depicts the range of treatment effects that is in line with our rational model (see Section 3.3).

D Analysis of the introduction of static incentives

This appendix presents the empirical analysis of the introduction of static incentives. This provides useful context for assessing the results of our field experiments on the response to dynamic incentives. We also use the estimated response to static incentives to estimate our structural model of effort provision, as discussed in Section 3.3.

At the warehouse we study, the firm initially just paid workers an hourly wage, but after about a year the firm rolled out an incentive pay scheme (see Figure C.1 for a timeline). The scheme left the base wage unchanged but added a weekly performance bonus. The performance bonus was implemented in the form of a standard-hour plan, with output being normalized by target rates into "standard-productive hours," as described in Section 2. When incentives were first rolled out, target rates were based on the average effort of all workers in each task over a previous period of months. Workers were explicitly told that the rates were static in the sense that they would remain in place until further notice and not be changed without informing the workers well ahead of time. The incentive system thus only introduced static incentives, i.e., their effort in period t did not affect their potential incentive pay in period $t + 1$.

Since incentives were not randomly allocated, we use a difference-in-differences estimation with the other main warehouse of the firm as control, adding either warehouse or worker fixed effects. Both warehouses serve the same purpose of receiving goods and fulfilling customer orders. Both warehouse thus contain the same types of jobs, use similar machines, have similar size and face the same seasonal and weekly demand shocks. They just serve different geographical areas. The control warehouse had an incentive system in place that did not change across the studied period. The data set runs from July 2014 to June 2015 (when the GROUP trial started). Between December 2014 and March 2015, the treatment warehouse gradually rolled out incentives, task by task. From then on, all tasks were incentivized. We thus have effort data for about five months before the roll-out of incentives and three months after the roll-out was completed. During the entire time target rates remained static.

Finding 9. *The introduction of static incentives increases worker effort by 12.5 percent.*

Figure 1 in Section 3.1 of the main text plots average worker effort for each warehouse by week, measured as residuals of ln(units per hour) residualized for the control variables in column 2 of Table D.1 (see below for details). The figure shows that effort in the treated warehouse is stable, and parallel to the control warehouse, before the introduction of incentives, then slowly increases while incentives are rolled out, and is then relatively stable again at a higher level. By contrast, effort in the control warehouse does not change much across the entire period.

The corresponding difference-in-differences regressions are shown in Table D.1. The regressions control for any time-invariant differences between warehouses by using warehouse fixed effects (columns 1 and 2) or worker fixed effects (columns 3 and 4). To control for time-varying differences, the regressions in columns 2 and 4 add total time worked per shift and average tenure per shift. Since the treated warehouse was newer, its workforce was still growing. The time profile of tenure and total time worked is thus different between the two warehouses. The two control variables correct for these different time profiles. To avoid issues with two-way fixed-effect regressions in staggered diff-in-diff analyses (e.g., Goodman-Bacon 2021), all specifications exclude the roll-out period. We thus only have one pre- and one post-period.[80]

[80]When we include the roll-out period (which would be valid under the assumption of time-invariant treatment effects), the point estimates become slightly larger.

Table D.1: Diff-in-Diff analysis of introduction of static incentive on effort

Dependent variable: ln(units per hour)						
	(1)	(2)	(3)	(4)	(5)	(6)
1 if static incentives	0.1276	0.1252	0.1359	0.1319	0.1038	0.1052
	(0.008)	(0.009)	(0.008)	(0.008)	(0.009)	(0.009)
Total time worked per WH & shift		-0.0372		-0.0324		-0.0399
		(0.014)		(0.013)		(0.010)
Average tenure per WH & shift		0.0615		0.0568		1.5663
		(0.028)		(0.025)		(0.315)
Sample	Full	Full	Full	Full	Restricted	Restricted
Task FE	Yes	Yes	Yes	Yes	Yes	Yes
Shift FE	Yes	Yes	Yes	Yes	No	No
Warehouse FE	Yes	Yes	No	No	No	No
Worker FE	No	No	Yes	Yes	No	No
# Workers	4580	4580	3534	3534	1263	1263
# Shifts	514	514	514	514	443	443

Notes: OLS regressions. Robust standard errors using two-way clusters on workers and shifts are in parentheses. '1 if static incentives' is an indicator for periods when workers were paid static incentives. It is always 1 in the control warehouse. 'Full sample' includes workers in treated and control warehouse and excludes the period when incentives were gradually rolled-out across activities. 'Restricted sample' includes only workers similar to the sample of the INDIVIDUAL trial, i.e., workers in the treated warehouse, if they worked for at least 20 hours per week on average and only during their first 13 weeks in the warehouse. The restricted sample again excludes the roll-out period. In the specifications with worker fixed effects, the number of workers only includes those workers in the treated warehouse who were present both in the before- and the after-period, and whose effort is thus not absorbed by the fixed effects.

Column 2 is the specification that corresponds to Figure 1 and is our preferred specification. It shows that the introduction of static incentives lead to a 12.5 percent increase in worker effort. The specifications in columns 1, 3, and 4 yield very similar results (Figure D.1 shows the corresponding event study graph for column 1). This suggests that workers are in fact motivated by the static incentives that are present in the firm's performance pay system. This makes the very small ratchet effect we find in our two field experiments particularly striking. The introduction of the incentive scheme increased overall worker pay by about 10 percent on average. The per-unit labor cost thus did not change by much. The firm was, however, still pleased about the outcome, as it increased machine utilization and thus the capacity of the warehouse.

Columns 5 and 6 restrict the sample to the workers most similar to the participants in the INDIVIDUAL trial, i.e., only workers during their first 13 weeks in the warehouse and who work at least 20 hours per week on average. As we have very few such workers in the control warehouse, columns 5 and 6 only use data from the treated warehouse, so this is just a before-after comparison. Since effort in the control warehouse does not change over the time period, this should not affect results much. The estimates are quite similar to the estimates in columns 1–4, and we use the estimate in column 6 (10.5 percent) for the structural calibration in Section 3.3.[81]

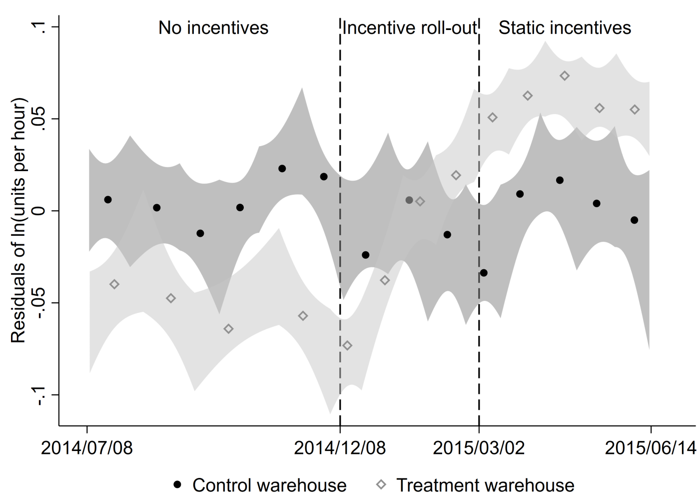
Figure D.2 and Table D.2 analyze differential attrition between the two warehouses in the time before and after the roll-out of incentives. We separately analyze attrition for the time before the incentive roll-out (July to December 2014), for the time during and after the incentive roll-out (December 2014 to June 2015) and for the time after the incentive roll-out (March to June 2015). Since the treatments were not randomly allocated, it is not surprising that attrition is different between the warehouses. In particular, the treated warehouse has a higher attrition than the control warehouse. This is mostly driven by the differences in worker tenure. Turnover is particularly high for new hires and once a workers has been in the firm for about a year, turnover is very low. We are particularly concerned about potential differential attrition with respect to worker effort, as this would bias the results in Table D.1. Column 2 of D.2 shows that faster workers (as measured by their pre-incentive-rollout speed) are *more* likely to leave in the treated warehouse compared to the control warehouse in the time before the incentive roll-out. This works against the effect in Table D.1, where we find that workers in the treated warehouse become faster on average, whereas differential attrition will create a slower work force in the treated warehouse over time. Columns 5 and 8 show that this differential attrition is not significant for the time during and after the

[81] The sample of workers in columns 5 and 6 of Table D.1 and the participants in the INDIVIDUAL trial are similar on many observable characteristics. The average age is 30.5 and 31.0 ($p = 0.387$), respectively. The share female is 0.29 and 0.27 ($p = 0.390$). The share of non-native workers is 0.66 and 0.65 ($p = 0.676$). Only the share of agency workers differs: 0.33 and 0.20 (p=0.000). For worker speed, we find mixed results depending on how we measure speed. In terms of units per hour, the INDIVIDUAL sample is 8.5 percent faster ($p = 0.003$). But the INDIVIDUAL trial happened two years later and the warehouse had become more productive, i.e., average units per hour had gone up or, put differently, task difficulty had decreased. When we take the speed of older workers as benchmark for task difficulty and correct for task difficulty, then the INDIVIDUAL sample is 2.7 percent slower ($p = 0.000$). However, the learning rate, i.e., how quickly new hires improve their speed in the first 13 weeks, is not different between the two samples ($p = 0.387$).

incentive roll-out.[82]

Finally, Figure D.3 shows a scatter plot of the effect of static incentives (measured during the roll-out of incentives) and the observed ratchet effect (measured during the INDIVIDUAL trial). Each bubble corresponds to one task. The size of the bubble is proportional to the amount worked in this task during the INDIVIDUAL trial. The lines mark the overall treatment effects. We find no relationship between the two effects: the size of the ratchet effect is uncorrelated with the effect of static incentives ($p = 0.784$). This is in line with our boundedly rational model, described in Section 2, where workers do not realize the incentive to ratchet, regardless of how elastic the task is. Appendix A.4.2 discusses a rational model that allows for multiple tasks with interdependent effort choices across tasks.

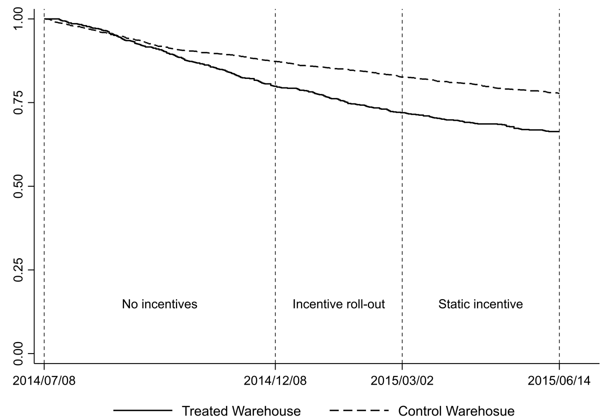
Figure D.1: Visual Diff-in-Diff of introduction of static incentives (without additional controls)



Note: Binscatter graph of the residuals of ln(units per hour) in the treated and the control warehouse, binned by week. The incentives were rolled out, task by task, between 8 December 2014 and 2 March 2015, for the treated warehouse and were always present in the control warehouse. The graph corresponds to column 1 in Table D.1. The dependent variable is thus residualized for task fixed effects and warehouse fixed effects. Target rates were static for the treated warehouse for the entire period shown in the graph. Target rates in the control warehouse were set according to the previous month's average effort in that warehouse. This rate setting rule was unchanged during the period shown in the graph. The shaded areas show 95-confidence bands (Cattaneo et al. 2024).

[82]The analysis considers workers who were employed on 8 July 2014. We find the same results regarding differential attrition if we consider the set of workers employed on 8 December 2014 or on 2 March 2015.

Figure D.2: Attrition during the introduction of static incentives (workers employed in July 2014)



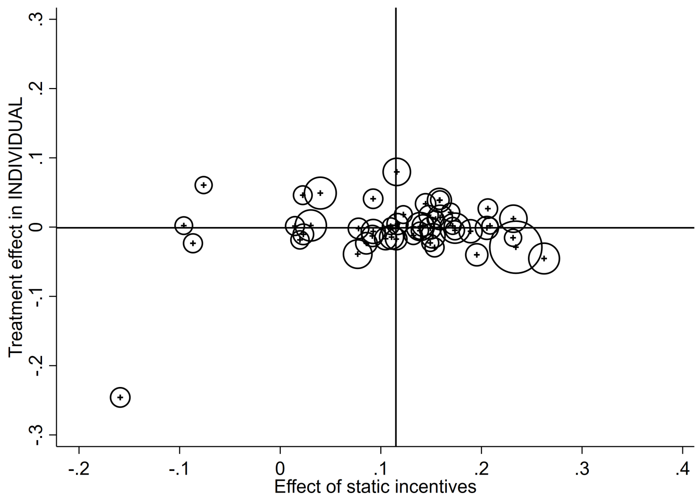
Notes: Kaplan-Meier survival estimates for the introduction of static incentives for workers who were employed on 8 July 2014. The vertical lines show the start and the end of the roll-out of static incentives in the treated warehouse. Corresponding regressions are in Table D.2.

Table D.2: Attrition during the introduction of static incentives (workers employed in July 2014)

Dependent variable: Worker left firm	Jul 2014 - Dec 2014			Dec 2014 - Jun 2015			Mar 2015 - Jun 2015		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1 if treated warehouse	1.6195	0.2068	0.3580	1.6083	1.0009	0.9781	1.3270	0.7162	0.6879
	(0.172)	(0.072)	(0.108)	(0.203)	(0.191)	(0.218)	(0.248)	(0.225)	(0.246)
Tenure at start of baseline period		0.4334	0.4498		0.6381	0.7633		0.5974	0.7147
		(0.055)	(0.068)		(0.059)	(0.085)		(0.073)	(0.108)
Tenure × treated WH		0.0725	0.7688		1.6367	1.4967		1.6203	1.5833
		(0.052)	(0.487)		(0.483)	(0.457)		(0.944)	(0.931)
Pre-roll-out speed		0.6001	0.6317		0.8554	0.9044		0.8074	0.8550
		(0.052)	(0.059)		(0.099)	(0.099)		(0.136)	(0.131)
Pre-roll-out speed × treated WH		1.5239	1.5454		1.1523	1.1253		1.2496	1.1659
		(0.181)	(0.199)		(0.154)	(0.148)		(0.254)	(0.215)
1 if female			0.6674			0.4616			0.3575
			(0.136)			(0.119)			(0.141)
1 if female × treated WH			1.0597			1.6955			1.9368
			(0.378)			(0.648)			(1.210)
Age at start of baseline period			0.7527			0.6297			0.6958
			(0.093)			(0.085)			(0.136)
Age × treated WH			1.2657			1.7575			1.3667
			(0.201)			(0.298)			(0.347)
# Workers	2315	1903	1694	1959	1609	1468	1824	1499	1371

Notes: Hazard ratios from Cox proportional hazard models for workers who were employed by the firm on 8 July 2014. Robust standard errors in parentheses. In the treated warehouse, static incentives were rolled out from December 2014. The control warehouse always had incentives in place. Columns 1–3 analyze the time before the roll-out of incentives. Columns 4–6 analyze the time during and after the roll-out. Columns 7–9 analyze the time after the roll-out. A worker's pre-roll-out speed is their average units per hour in the period before the incentive roll-out, controlling for task fixed effects, i.e., correcting for the fact that a unit is harder or easier in different tasks. This is calculated for all workers who worked for at least 16 hours before the incentive roll-out. Pre-roll-out speed, tenure and age are normalised.

Figure D.3: Effect of static incentives and observed ratchet effect by task



Notes: Scatter plot of the effect of static incentives (measured during the roll-out of incentives) and the observed ratchet effect (measured during the INDIVIDUAL trial). Each bubble corresponds to one task. The size of the bubble is proportional to the amount worked in this task during the INDIVIDUAL trial. The lines mark the overall treatment effects.

E Analysis of the GROUP trial

This appendix presents more details about the design of the GROUP trial, and results from the the empirical analysis.

E.1 Design

Table E.1 summarizes the design of GROUP. We randomized all workers into two conditions, treated workers (40 percent of workers) and control workers (60 percent of workers), and workers kept the same roles throughout the trial.

Table E.1: Design of the GROUP trial

Baseline period	Fixed rates	
Condition assigned	Treatment workers ($N = 573$)	Control workers ($N = 874$)
Month 1	Fixed rates	
Month 2	Rates = average speed of **treatment workers** in previous month	
...		
Month 10	Rates = average speed of **treatment workers** in previous month	
Month 11+	Rates = average speed of **all workers** in previous month	

Workers were extensively informed about all the details outlined below, except for the fact that the trial was designed together with university researchers. In the baseline period, before the trial, all workers faced incentive pay with exogenous target rates. During and after the trial, rates were changed every four weeks. For simplicity, we refer to a 4-week rate-setting period as a "month". In Month 1 of the trial, all workers faced the same target rates, but workers in the treatment group knew that their effort in that month would determine the target rates for all workers (treatment and control) for the second month. Specifically, in Month 2, the rate for each task would be the average output per hour in Month 1 in that task, with the average calculated across the group of all treated workers who worked at some point in that task. Control workers knew that rates were determined by the treated workers, and that their own efforts would have no impact on anyone's rates. Thus, treated workers faced dynamic incentives in Month 1 whereas control workers did not. In Month 2, both groups faced the same rates (determined by treated workers' effort in Month 1). Treated workers again faced dynamic incentives, because their effort determined rates in Month 3, while control workers did not influence rates. This continued for 10 months. After this point, rates were set as the average speed of *all* workers in the previous month.

In June of 2015, we randomized all workers into treatment and control. 1075 workers started the trial. In September of 2015 (i.e., Month 4 of the first randomization cohort), we randomized workers who had been hired since June. This added 263 workers to the sample and gives a second cohort of treated and control workers. The trial period for the second cohort was thus shorter, lasting from Month 4 to Month 11.[83] The random allocation of workers to treatments was done by us, stratifying the randomization on above median pre-trial speed, being a temp/agency worker, working mostly on the night shift, and working mostly in the modal warehouse activity. Table E.2 contains summary statistics and randomization checks for the GROUP trial. Treatment and control group are not significantly different, including in terms of characteristics on which we did not stratify. Figure E.1 and Table E.3 show that there is no differential attrition between treated and control workers.

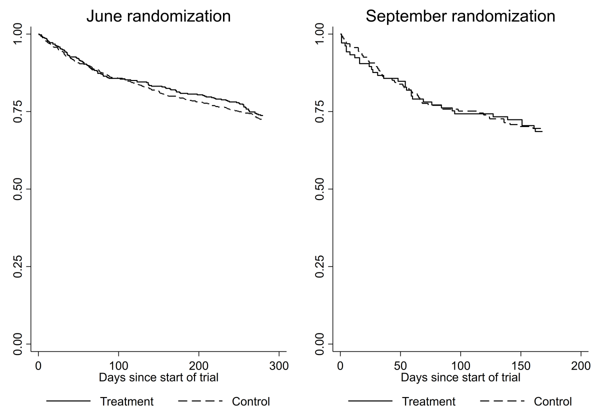
[83]During the baseline period for the second cohort, rates were the rates used for all workers, determined by the treated workers of the first randomization cohort.

Table E.2: Summary statistics and randomization checks in the GROUP trial

	Mean		p-value
	Control	Treatment	
Pre-trial speed	0.95	0.94	0.447
1 if temp/agency worker	0.07	0.08	0.657
1 if mostly working at night	0.69	0.70	0.772
1 if mostly working in modal task	0.51	0.49	0.372
Tenure at start of trial	263.01	263.74	0.956
Experience at start of trial	379.08	367.54	0.431
Age at start of trial	33.19	32.50	0.340
1 if female	0.26	0.22	0.139
1 if non-native	0.56	0.58	0.687
# Workers	874	573	

Notes: Summary statistics of the workers randomized in the GROUP trial. Across the two randomization cohorts, we randomized 1447 workers into treatment and control conditions of which 1338 started the treatment period. The p-values are from t-tests. A worker's pre-trial speed is their average units per hour in the period before the start of the trial, controlling for task fixed effects, i.e., correcting for the fact that a unit is harder or easier in different tasks. This is calculated for all workers who worked for at least 16 hours before the start of the trial. Tenure at start of trial is the number of days between the first day a worker starts working in the firm and the start of the trial. Experience at start of trial is the total time worked in hours between the first day of work and the start of trial.

Figure E.1: Attrition in the GROUP trial



Notes: Kaplan-Meier survival estimates for the GROUP trial, shown separately for the two randomization cohorts.

Table E.3: Attrition in the GROUP trial

Dependent variable: Worker left firm			
	(1)	(2)	(3)
1 if treated	0.9565	0.9294	0.8817
	(0.101)	(0.102)	(0.209)
Tenure at start of trial		0.8293	1.2228
		(0.095)	(0.080)
Tenure × treated		0.7888	0.9030
		(0.128)	(0.097)
Pre-trial speed		0.9029	0.9634
		(0.056)	(0.095)
Pre-trial speed × treated		0.9609	0.9768
		(0.101)	(0.188)
1 if female			0.6825
			(0.208)
1 if female × treated			0.6491
			(0.374)
Age at start of trial			1.0359
			(0.132)
Age × treated			1.1167
			(0.216)
Randomization cohort FE	Yes	Yes	Yes
# Workers	1359	1331	792

Notes: Hazard ratios from Cox proportional hazard models for the full sample of the GROUP trial. Robust standard errors in parentheses. Treated workers faced ratchet incentives during the trial, while control workers did not. A worker's pre-trial speed is their average units per hour in the period before the start of the trial, controlling for task fixed effects, i.e., correcting for the fact that a unit is harder or easier in different tasks. This is calculated for all workers who worked for at least 16 hours before the start of the trial. Tenure at start of trial, pre-trial speed and age at start of trial are normalised.

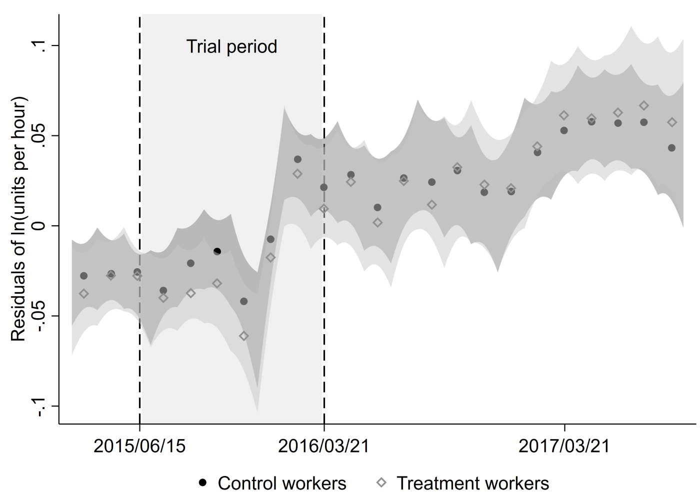
E.2 Results

Figure E.2 depicts the evolution of effort before, during and after the GROUP trial for treated and control workers (similar to Figure 2 for the INDIVIDUAL trial). For simplicity, the graph is restricted to the first randomization cohort. The figure shows a small treatment effect in the expected direction. Figure 3 in the main text plots the *treatment effect* over time. That figure shows more clearly that the treatment effect grows slowly over time.

We analyse the treatment effect using regression analyses. Table E.4 mirrors Table C.3

for the INDIVIDUAL trial. It shows results from OLS regressions, again using ln(units per hour), our measure of workers' effort, as dependent variable. Column 1 shows results from the contemporaneous comparison of treatment and control group, i.e., comparing the effort of treated to control workers during the trial. The fixed effects on cohort, tasks, shift and cohort interacted with all other fixed effects are like in Table C.3. We thus flexibly control for differences between the two randomization cohorts. Treated workers are on average slower by -1.0 percent, with the difference marginally statistically significant (95 percent confidence interval: [-2.1, 0.2]). The effect is still small, relative to the benchmark of response to static incentives, but larger than in the INDIVIDUAL trial.

Figure E.2: Effort in GROUP trial by treatment



Notes: The figure shows a time series plot of worker effort for treatment and control workers. Both groups of workers face the same static incentives and treatment workers additionally face a dynamic (ratchet) incentive to reduce effort during the trial period. Effort is residualized for task fixed effects. The graph is restricted to the first randomization cohort. The shaded areas show 95-confidence bands (Cattaneo et al. 2024).

Table E.4: Ratchet effect in GROUP trial

Dependent variable: ln(units per hour)			
	(1)	(2)	(3)
1 if treated	-0.0096	-0.0124	-0.0124
	(0.006)	(0.006)	(0.006)
Sample	During trial	During trial, periods 3+	During trial, periods 3+ Working entire next period
Task FE	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes
Shift FE	Yes	Yes	Yes
all FE's × cohort	Yes	Yes	Yes
# Workers	1338	1165	1073
# Shifts	556	444	444

Notes: OLS regressions. Robust standard errors, using two-way clusters on individual workers and on shifts, are in parentheses. The sample is restricted to the time during the GROUP trial, when the treated workers faced a ratchet incentive to work more slowly, while the control workers did not face such an incentive.

GROUP gives workers more time to learn and notice the dynamic incentives, and it also gives workers a potential motive to put pressure on, or teach, treated workers to slow down. If there were indeed learning over time, then the point estimate in column 1 of Table E.4 would underestimate the long-term ratchet effect. In column 2, we thus drop the first two months of the trial. The point estimate grows slightly to -1.2 percent (CI: [-2.5, -0.0]) but is still small.[84]

Column 3 further restricts the sample to only those workers who kept working for the firm until at least the end of the following rate-setting period. These workers enjoy the full benefit of reducing effort in the current period and they thus face the strongest ratchet incentives. The point estimate is unchanged compared to column 2 (-1.2 percent, CI: [-2.5, 0.0]). Across the two trials, INDIVIDUAL and GROUP, we can thus reject that ratchet incentives reduce effort by more than 2.5 percent.

Because rates in GROUP are based on the average speeds of groups of treated workers, with a group involving N individuals, each individual treated worker's impact on the rate is scaled by $\frac{1}{N}$. Note that in INDIVIDUAL, since own speed was averaged with one other

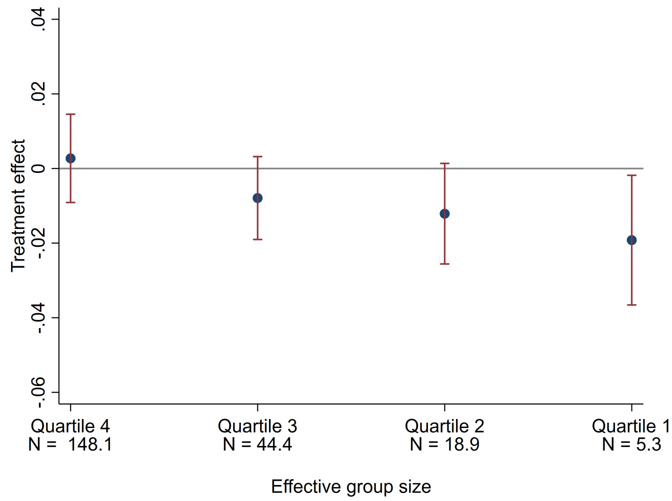
[84] Excluding the first three or four months yields very similar results.

number, individual impact was scaled by $\frac{1}{2}$, i.e., we effectively had a group size of 2. This raises the question whether treated workers in GROUP might come to notice the dynamic incentives over time due to learning, but still find it not worthwhile to respond because they see their impact on the rates as being too small (but note that social pressure motives could actually make the utility benefits of slowing down greater in GROUP than INDIVIDUAL). Since we know how many workers work on a particular task, we know how many workers affect the corresponding target rate. We can thus study the effect of naturally occurring variation in group size on the ratchet effect. To explore the effect of group size, we calculate each treated worker's share of the time worked by all treated workers in a given task, for each rate-setting period. 1 divided by this share is the effective group size of treated workers. When we split the effective group sizes into quartiles, the average group size per quartile is 148.1, 44.4, 18.9 and 5.3 workers, respectively.

Figure E.3 plots the ratchet effect, i.e., the difference between treatment and control workers during the trial, for the four quartiles. We find that smaller groups do show a larger ratchet effect (Quartile 1 vs. 4, $p = 0.014$). This is in line with the hypothesis that individuals respond more strongly when they have a bigger individual impact. However, even in the smallest groups, the ratchet effect is only about -2 percent, and also groups consisting of around 40 workers show a ratchet effect of about -0.8 percent. Thus, even large variations in group size are having a relatively minor impact on responses.

One way to shed light on whether non-response in GROUP could reflect workers understanding, but not finding it to worthwhile to respond, is to check how these same workers respond in the online experiments, which took place after the GROUP trial and where we have an effective group size of 2. If workers were fully aware of dynamic incentives based on learning in GROUP, but did not respond due to group size being large, we would expect a strong response in the online environment, because workers are fully aware and group size is only 2. But this is not what we find (see Section 4 in the main text and Appendix G).

Figure E.3: Ratchet effect in GROUP trial by workers' effective group size



Notes: The graph plots the treatment difference on ln(units per hour), i.e., the ratchet effect, by effective group size. We calculate each treated worker's share of the time worked by all treated workers in a given task, for each rate-setting period. 1 divided by this share is the effective group size of treated workers. Point estimates are from regressions as in Table E.4, column 1, separately for each group size quartile. Error bars show 95 percent confidence intervals.

The ratchet effect essentially results from a trade-off between reduced earnings now and reduced effort costs in the future. The ratchet effect could thus also be small because workers put too little value on the future. This could be because they are liquidity constraint or generally present-biased or because they put a small likelihood on still working for the firm in the next month.

We measure the value workers should or do put on the future in the firm in three ways. First, we can assume that workers have at least some foresight about whether they will work at the firm in the following rate-setting period. We can then compare the ratchet effect among those workers who ended up working in the firm for the entire next rate-setting period to those workers who ended up not working for the firm. The workers who do not work for the entire next rate-setting period do not enjoy the full benefit of reducing effort in the current period. They thus face weaker ratchet incentives and should reduce effort less (this is similar to comparing columns 2 and 3 in Table E.4). Table E.5 shows this comparison. The coefficient of interest is on the interaction of not working the entire next month × treated. We find no significant difference between the two groups. The point estimate goes in the opposite direction compared to what a rational model would predict.

Table E.5: Ratchet effect in GROUP trial for workers who will vs. won't work the entire next month

Dependent variable: ln(units per hour)		
	(1)	(2)
1 if treated	-0.0091	-0.0124
	(0.006)	(0.006)
1 if not working entire next month × treated	-0.0136	-0.0048
	(0.011)	(0.013)
Sample	During trial	During trial, periods 3+
Task FE	Yes	Yes
Cohort FE	Yes	Yes
Shift FE	Yes	Yes
all FE's × cohort	Yes	Yes
all FE's × not working next month	Yes	Yes
# Workers	1338	1165
# Shifts	556	444

Notes: OLS regressions. Robust standard errors, using two-way clusters on individual workers and on shifts, are in parentheses. This table replicates Table E.4 (columns 1 and 2) but adds interactions of the treatment dummy with a dummy for the observations when the worker is not working for the entire next rate-setting period.

Second, the majority of workers in our sample have a permanent contract with the firm. However, a sizable minority of workers are employed by an agency and are drafted into the warehouse on a more ad-hoc basis. A third group of workers started out as temp/agency workers and then became permanent. The permanent workers should have a higher expectation to stay in the firm than the first-agency-then-permanent workers who in turn should have a higher expectation to stay than the agency workers. Table E.6 compares the ratchet effect across these three groups. We find no significant differences between the groups.

Table E.6: Ratchet effect in GROUP trial for permanent vs. agency workers

Dependent variable: ln(units per hour)			
	(1)	(2)	(3)
1 if treated	-0.0081	-0.0112	-0.0105
	(0.009)	(0.009)	(0.009)
1 if temp/agency worker × treated	-0.0116	-0.0108	-0.0497
	(0.019)	(0.026)	(0.035)
1 if permanent worker × treated	-0.0031	-0.0026	-0.0041
	(0.012)	(0.012)	(0.012)
Sample	During trial	During trial, periods 3+	During trial, periods 3+ Working entire next period
Task FE	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes
Shift FE	Yes	Yes	Yes
all FE's × cohort	Yes	Yes	Yes
all FE's × agency and permanent	Yes	Yes	Yes
# Workers	1338	1165	1073
# Shifts	556	444	444

Notes: OLS regressions. Robust standard errors, using two-way clusters on individual workers and on shifts, are in parentheses. This table replicates Table E.4 but adds interactions of the treatment dummy with being a temp/agency worker or a permanent worker. The omitted category are workers who start out as agency workers and then become permanent.

Third, we directly measure workers' time discounting for the sample of workers participating in the online experiments (see Section 4). Workers had to choose between receiving $15 in the next paycheck or receiving a larger amount in the following paycheck, four weeks later. Workers made five of these choices and one of the five choices was randomly chosen to be paid out for 1 in 10 workers. The five choices were determined in a staircase method (Falk et al. (2023), see Appendix I for the full instructions). We calculate workers' discount rate from their choices and split workers at the median. Again, workers with large or small discount rates do not show differential ratchet effects (Table E.7).

Table E.7: Ratchet effect in GROUP trial by time preferences

Dependent variable: ln(units per hour)			
	(1)	(2)	(3)
1 if treated	0.0068	0.0050	0.0050
	(0.019)	(0.020)	(0.020)
1 if patient × treated	-0.0048	-0.0110	-0.0110
	(0.026)	(0.026)	(0.026)
Sample	Online exp. During trial	Online exp. During trial, periods 3+	Online exp. During trial, periods 3+ Working entire next period
Task FE	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes
Shift FE	Yes	Yes	Yes
all FE's × cohort	Yes	Yes	Yes
all FE's × low disount rate	Yes	Yes	Yes
# Workers	247	244	244
# Shifts	555	443	443

Notes: OLS regressions. Robust standard errors, using two-way clusters on individual workers and on shifts, are in parentheses. This table replicates Table E.4 but adds interactions of the treatment dummy with being patient, i.e., preferring larger-later payments over smaller-sooner payments in the online experiment. The sample is restricted to the workers who participated in the GROUP trial and in the online experiment.

In Section 4.3 in the main text, we depict graphically the correlation between behavior in the online experiment and behavior in the Group trial. Table E.8 contains the underlying regression results. The table replicates Table E.4, but adds an interaction of treatment with the variable "showed RE online" (see Section 4.3 for a definition of the variable). Columns 1–3 have this variable as a median split, columns 4–6 have the continuous variable. All specifications show that workers who exhibited a ratchet effect in the online experiment also showed a stronger ratchet effect in the warehouse.[85]

Workers with higher scores in the CRT also show a stronger ratchet effect in the warehouse, but this effect is not significant. We can, however, include CRT as a fourth variable in the principal component analysis and this increases the point estimate in regressions with the same specifications as those in Table E.8.

[85] We find similar point estimates if we use each of the three ingredients in the PCA separately as interaction variable, although only one of them is individually significant.

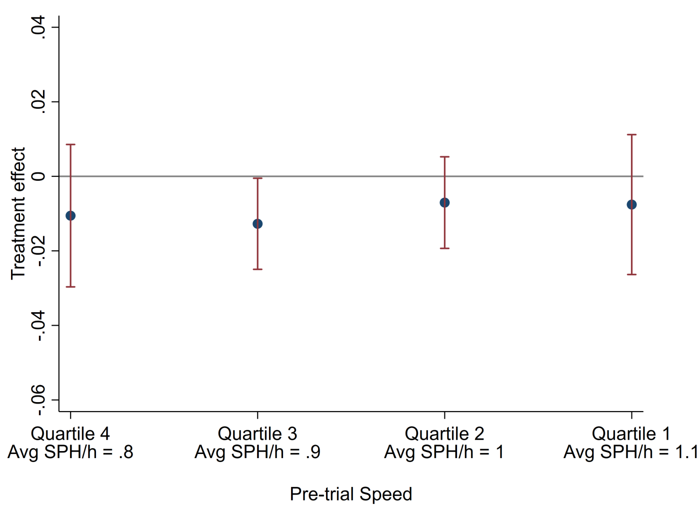
Table E.8: Correlation of online and field experiment behavior

Dependent variable: ln(units per hour)	(1)	(2)	(3)	(4)	(5)	(6)
1 if treated in warehouse	0.0064	0.0067	0.0067	-0.0327	-0.0375	-0.0375
	(0.019)	(0.018)	(0.018)	(0.023)	(0.025)	(0.025)
Showed RE online (dummy) × treated in warehouse	-0.0651	-0.0876	-0.0876			
	(0.032)	(0.029)	(0.029)			
Showed RE online (PCA) × treated in warehouse				-0.0635	-0.0658	-0.0658
				(0.021)	(0.023)	(0.023)
Sample	COMPLEX During trial	COMPLEX During trial, Period 3+	COMPLEX During trial, Period 3+, Working entire next period	COMPLEX During trial	COMPLEX During trial, Period 3+	COMPLEX During trial, Period 3+, Working entire next period
Task FE	Yes	Yes	Yes	Yes	Yes	Yes
Shift FE	Yes	Yes	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
all FE's × cohort	Yes	Yes	Yes	Yes	Yes	Yes
all FE's × showed RE	Yes	Yes	Yes	Yes	Yes	Yes
# Workers	79	79	79	79	79	79
# Shifts	555	443	443	555	443	443

Notes: OLS regressions. Robust standard errors, using two-way clusters on individual workers and on shifts, are in parentheses. The dependent variable is worker speed in the warehouse. The sample is restricted to workers who participated in the GROUP trial in the warehouse and also in the online experiment and were there randomly allocated to the COMPLEX treatment. The sample is also restricted to the time during the GROUP trial, when the treatment workers ('1 if treated in warehouse') faced a ratchet incentive to work more slowly, while the control workers did not face such an incentive. The treatment dummy is interacted with a variable for having showed a ratchet effect in the online experiment. 'Showed RE online' is the standardized first principal component of a PCA of reducing effort from baseline period to Period 1; reducing effort from baseline period to Period 3; and mentioning arguments relating to ratchet effects or dynamic incentives in an open-ended question. Columns 4–6 use this variable directly, while columns 1–3 use a median-split dummy of it. Specifications 1 and 4 are the main regressions for the full sample. Specifications 2 and 5 restrict the sample to rate-setting periods 3 and later to allow for some learning. A period lasts 4 weeks, and the trial lasted for 10 periods. Specifications 3 and 6 further restrict the sample to only include workers who kept working for the firm until at least the end of the following rate-setting period. These workers enjoy the full benefit of reducing effort in the current period (since the online experiment took place after the GROUP trial, this restriction does not drop any workers). All warehouse workers employed by the firm in June 2015 were randomized into treatment and control. Workers starting after this date entered the trial in September and were randomized then. Cohort fixed effects control for these two randomization cohorts. All other fixed effects are also interacted with cohort and with the variable for showing a ratchet effect in the online experiment.

In our online experiments, we find significant differences in behavior due to worker cognitive ability, we thus further explore the heterogeneity of the ratchet effect between workers. Figure E.4 shows the ratchet effect separately for fast and slow workers, measured by their pre-trial speed. As can be seen from the figure, the ratchet effect does not vary with pre-trial speed. The ratchet effect is slightly stronger for men than for women, but not significantly so (p=0.494, in a regression akin to Table E.5, column 1). It is also not different by nationality (p=0.674), age (p=0.109) or tenure (p=0.193), even if older and more experienced workers show a marginally stronger ratchet effect.

Figure E.4: Ratchet effect in GROUP trial by workers' pre-trial speed



Notes: The graph plots the treatment difference on ln(units per hour), i.e., the ratchet effect, by workers' pre-trial speed. We calculate each worker's speed in the period between the roll-out of static incentives and the start of the trial and split workers into quartiles. Point estimates are from regressions as in Table E.4, column 1, separately for each pre-trial speed quartile. Error bars show 95 percent confidence intervals. The graph also shows the average number of Standard Productive Hours (SPH) workers in this quartile achieve per hour. SPH are units per hour corrected for the fact that a unit is harder or easier in different tasks.

F Robustness checks for the warehouse field experiments

In this section we discuss robustness of our conclusions from the field experiments to a variety of potential concerns.

F.1 Multiple tasks and rational effort provision

Our simulation of how far workers deviate from the rational optimum in INDIVIDUAL uses the simplifying assumption of a representative task, whereas in reality there are many tasks in the warehouse. In our online experiments with warehouse workers, however, we actually have only a single task, and we again find that the deviation from the rational optimum is large in magnitude, showing that our conclusion that the incentive scheme is complex does not hinge on the simplifying assumption. Also, as discussed in detail in Appendix A.4.2, we can allow for multiple tasks in our model. The model with multiple tasks does not allow simulating a quantitative prediction for the rational optimum, due to computational intractability (more on this below), but under empirically plausible assumptions it makes similar qualitative predictions to the single task model, in terms of total effort across tasks. Specifically, total effort will be lower for treatment workers in INDIVIDUAL. Thus, our focus on aggregate effort in analyzing INDIVIDUAL is not inconsistent with having multiple tasks in the model. We do not find support for this qualitative prediction, as we cannot reject a response of zero. Another indication that assuming a single task is problematic would be if we observed strong heterogeneity across tasks in response to dynamic incentives. As shown in Appendix Figure C.4, however, responses are very small for almost all of the 75 tasks, except for one outlier task that only occupies a trivial portion of worker time. Moreover, if workers were responding rationally to dynamic incentives across tasks, one might also expect a correlation between how much workers respond to static incentives for a task, and how much they adjust in response to dynamic incentives. Instead, as shown in Appendix Figure D.3, there is no correlation.

F.2 Time discounting

One explanation for a weak response to our treatments could be that workers put only a small value on the future. This could be because of time preferences, or liquidity constraints, or perceiving a small likelihood of still being employed by the firm in the next month. Our structural model allows for discounting (which we calibrate using the data) and results are robust to assuming much stronger time discounting (see Appendix B.1). We can also investigate these concerns directly: (1) We compare workers who have (or expect to have) a longer lasting relationship with the firm to those who don't. (2) We control for experimentally measured discount rates of individual workers. We find that these factors do not matter for observed behavior (see Appendix E). (3) We find similar patterns of behavior in the online experiments, where discounting is irrelevant by design.

F.3 Firm's ability to commit

Can the firm credibly commit to not using the data from control workers for setting future rates, in particular in GROUP? If control workers doubted this, we should see a ratchet effect in both treatment and control. First, the firm did in fact not use the data. This is likely due to the firm wanting to retain workers, and keeping a good working relationship with worker representatives. Second, the firm had stuck to a similar incentive system for several years in another warehouse, thus building a reputation for trustworthiness among workers. Third, the firm explicitly stated that it would not use the data. Any deviation would have been a clear breach of workers' trust. Fourth, such concerns are reduced for INDIVIDUAL and non-existent for the online experiments, where we find similar results.[86]

[86] A different worry could be that workers were unaware of their treatment assignment in the study, or of the nature of the incentive system. The firm worked hard to ensure that workers knew both. The treatment status, for example, was transmitted in text or in person to workers. Moreover, all workers in GROUP were reminded that treatment workers determined future target rates. It is unlikely that forgetting status is a major concern, as we already see a weak treatment effect at the beginning of GROUP. Moreover, we observe similar effects in the online experiments, where there is negligible delay between learning about the incentive system, and making effort choices.

F.4 Concerns about dismissal or promotion

Maybe the weak ratchet effect is due to fears of being dismissed if working too slowly? Our structural model already allows for non-pecuniary concerns, including concerns about dismissals and promotions, so to some extent this is already ruled out. We also show in online experiments with the warehouse workers, where firing threat is ruled out by design, that there is only a weak response to dynamic incentives. We can also directly study these concerns, however, using data and information from the warehouse.

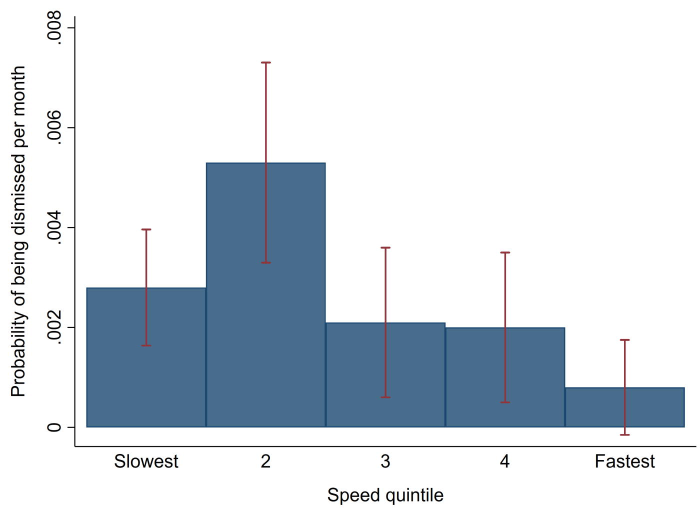
A first observation is that very few workers are dismissed by the firm to begin with. The vast majority of turnover comes from workers deciding to leave the firm. We have access to a 6-month sample of dismissal data after the end of the two field experiments. In this sample, the likelihood per month of being dismissed is 0.2 percent. There are also few promotions. We have a 19-month sample of secondment data and the likelihood per month of being seconded (which often leads to a permanent promotion) is 1.1 percent.

Second, we know that the firm does not dismiss anybody because of low effort, at least not in the short-run. The human-resources policy of the firm is that, if a worker works more than 30 percent slower than the average worker over a longer period of time, they receive additional training. This means that a treated worker in the INDIVIDUAL trial could have slowed down dramatically during the 3-week trial period, i.e., showed a large ratchet effect, and would not have been fired. Instead, we find (and the firm tells us) that dismissals are mostly about attendance or sometimes gross misconduct.

Third, some workers are dismissed for unspecific reasons (e.g., "Other substantial reason"), so we cannot exclude, on basis of the recorded reason, that these dismissals might be effort related, despite the stated HR policies of the firm. However, we know the effort of the dismissed workers and can correlate effort and dismissal probability. Figure F.1 shows the likelihood per month of being dismissed for an unspecific reason, split by worker speed. Unspecific-reason dismissals happen across the speed distribution. Low-speed workers are very slightly more likely to be dismissed but this difference is not significant ($p=0.276$).

Finally, we saw in the analysis of the introduction of static incentives that effort provision is quite elastic. It seems that workers before the introduction of incentives were fine with working at a slower pace. Put more formally, our model in Section 3.3 and Appendix B.1.1 actually incorporates the fear of being fired and the hope of being promoted for workers in the INDIVIDUAL trial, as these motives are part of parameter a. We show that even with levels of a that match the observed behavior of workers before and after the introduction of static incentives, i.e., with levels of workers' actual beliefs about dismissals and promotions, the ratchet effect should be much larger than what we observe in the data. Appendix B.1.1 adds a robustness check, which assumes that workers never want to reduce effort by more than 20 percent, e.g., they believe they will be fired for sure if they reduce effort by more than 20 percent and they don't want to be fired. Even under this strong assumption, the lower bound on rational effort reduction is still much larger than what we observe.

Figure F.1: Probability per month of being dismissed for unspecific reasons



Notes: The graph shows the probability of being dismissed per month, split by worker speed. The graph only contains dismissals for unspecific, and thus potentially speed-related, reasons. All workers are divided into five quintiles based on their average speed in the last 26 weeks before being dismissed. A placebo leave date that is distributed equally to the actual leave dates is assigned to workers who are not dismissed to create the control group. A worker's speed is their average units per hour, controlling for task fixed effects, i.e., correcting for the fact that a unit is easier or harder in different tasks. Error bars show 95 percent confidence intervals.

G Appendix for online experiments with warehouse workers

Table G.1: Design of online experiments, warehouse workers

Introductory phase	Consent, device type, educational attainment			
Condition assigned	COMPLEX ($N = 141$)	SIMPLE ($N = 140$)	STATIC ($N = 75$)	STATIC_ZERO ($N = 74$)
Baseline work period	Rate is 300			
Preferences	Time discounting and risk aversion measures			
Period 1 work	Rate is 300			
Period 2 work	Rate is average of Period-1 clicks and random number X	Rate is 300, Period-1 earnings subtracted	Rate is 300	Rate is 300
Cognitive ability	CRT, narrow choice bracketing measure, backwards induction ability measure			
Period 3 work	Rate is 300	Rate is 300	Rate is 300	Rate is 300, piece rate reduced to 0
Period 4 work	Rate is average of Period-3 clicks and random number X	Rate is 300, Period-3 earnings subtracted	Rate is 300	Rate is 300, piece rate reduced to 0
Questionnaire	Open-ended question about the best strategy for Periods 3 and 4			

G.1 Details on the rationale for experimental design of SIMPLE

There were several factors that determined our experimental design for SIMPLE, and led to a design in which the rational optimum was different from in COMPLEX. We knew intrinsic motivation would be positive, but it was difficult to predict the exact level ex ante and piloting with the warehouse workers was not an option. Moreover, we had to start from the warehouse incentive scheme, since the online experiments try to explain behavior observed in the field. For this scheme, the most natural simplifications affected the optimal effort for rational workers. While we planned to use our model to study deviations from the optima in SIMPLE versus COMPLEX, we were also reassured by the fact that the difference in optima for dynamic incentive periods predicts lower effort in COMPLEX than SIMPLE, which works against the prediction of complexity, that effort will be higher in COMPLEX.

We considered whether there is a way to combine all three simplifications, and have a “simple” contract which also had an optimum at zero as in COMPLEX, but we found that this would require making other changes to the incentive scheme, such as making the piece rate schedule fully linear, which would require shifting the entire level of payments. Such a change diverged from our aim to understand what is making the dynamic incentives complex within the class of incentive schemes implemented by the firm and to only simplify the dynamic incentive portion of the scheme.

G.2 Additional results for online experiments with warehouse workers

Table G.2: Diff-in-Diff of clicks relative to baseline period and STATIC

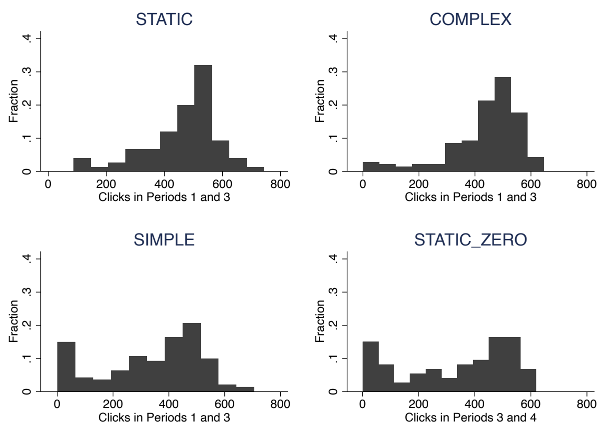
	(1)	(2)
Period1*COMPLEX	-29.77	-56.76
	(19.13)	(14.13)
Period2*COMPLEX	1.77	3.43
	(18.52)	(10.59)
Period3*COMPLEX	-17.64	-115.11
	(19.00)	(21.00)
Period4*COMPLEX	-16.46	-17.00
	(16.33)	(14.54)
Period1*SIMPLE	-88.47	-267.00
	(22.18)	(21.63)
Period2*SIMPLE	-7.01	-25.02
	(18.11)	(13.25)
Period3*SIMPLE	-150.27	-358.51
	(24.17)	(21.26)
Period4*SIMPLE	-29.84	-29.35
	(16.49)	(16.02)
Period1*STATIC_ZERO	12.51	4.46
	(19.16)	(10.19)
Period2*STATIC_ZERO	16.38	10.39
	(19.41)	(10.79)
Period3*STATIC_ZERO	-80.57	-269.12
	(28.28)	(29.00)
Period4*STATIC_ZERO	-110.73	-275.74
	(28.09)	(27.45)
Additional coefficients suppressed	Yes	Yes
# Workers	436	449

Notes: OLS regressions. Fully interacted difference-in-differences model with STATIC and the baseline period as omitted category. Only the coefficients for the interaction of period with treatment are shown. Negative coefficients mean that individuals in that treatment and period have a larger drop relative to baseline than individuals in STATIC. Robust standard errors in parentheses, clustering on worker.

Figure G.1 shows the individual-level effort choices in our four treatments. We see that at the individual level, behavior in COMPLEX in periods with dynamic incentives looks very similar to behavior in corresponding periods in STATIC, where dynamic incentives are absent. This is suggestive of most workers not taking dynamic incentives into account in

COMPLEX. We also see that clicks in SIMPLE in periods with dynamic incentives are bi-modal at the individual level, with modes at 0 and around 450, and this is very similar to the distribution observed in STATIC_ZERO in periods where the piece rate is zero. This is consistent with many workers recognizing dynamic incentives in SIMPLE, with some having low intrinsic motivation and others having high intrinsic motivation. We cannot conclude, however, that every worker recognized dynamic incentives in SIMPLE, and indeed, it is plausible that at least some of those who click around 450 did not recognize the dynamic incentives. Our main conclusion, however, is that in relative terms, more workers in SIMPLE than COMPLEX recognized dynamic incentives.

Figure G.1: Distributions of clicks by treatment and relevant periods, warehouse workers online



Notes: Each panel of the figure shows the distribution of clicks observed in the corresponding treatment for the specified periods. One outlier observation in STATIC is not shown in the graph.

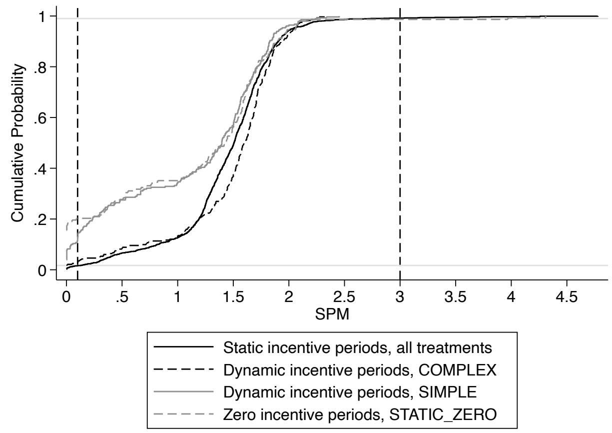
Table G.3: Categorization of open-ended responses about optimal work strategies, warehouse workers

	COMPLEX	SIMPLE
	(percent of responses)	
Response focused on dynamic incentives	19.15%	43.57%
Response focused on working fast or constantly	34.75%	12.14%
Response said no idea	3.55%	5.71%
Response mentioned reverse dynamic incentives	2.13%	3.57%
Response missing or nonsense	7.09%	4.29%
None of the above	33.33%	30.71%
Total	100%	100%

Notes: The open ended question asked workers what they would recommend to someone else as the best way to approach working in Periods 3 and 4 of the online experiment. Responses were assigned to the first category if at least two out of three independent evaluators categorized the response as focused on dynamic incentives. All other responses were assigned to one of the other mutually exclusive categories by a member of the research team.

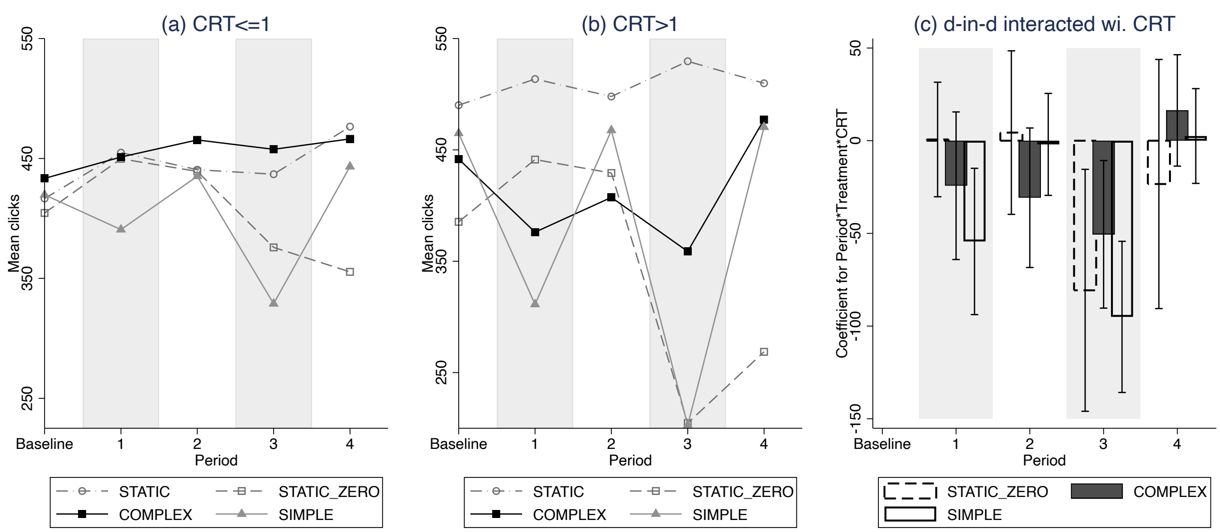
Our structural calibration for warehouse workers online incorporates the observation from the data that the representative worker chooses effort between the quota and cap when facing static incentives. Figure G.2 shows this is true for the median worker but also in almost every case at the individual level. We see that in periods without dynamic incentives, the quota was low enough that it was almost always attained, and workers almost never reached the cap. Specifically, 1.5 percent of observations were below the quota, and roughly 1 percent were above the cap. Thus, to a close approximation, workers were always working in the "interior." The figure also shows that in periods with dynamic incentives, workers in COMPLEX were also essentially always in the interior (behaving as though incentives were static). In SIMPLE and STATIC_ZERO, by contrast, we see lower effort levels, and more workers going below the quota, in periods with dynamic incentives, or zero piece rate, respectively. This is consistent with noticing dynamic incentives or responding to the cut in the piece rate, respectively.

Figure G.2: Cumulative distributions of SPM and the quota and cap values, warehouse workers online



Notes: Vertical dashed lines indicate the quota and cap of 0.1 SPM and 3 SPM, respectively. Horizontal light gray lines indicate the fractions of observations below the quota, or below the cap, 0.015 and 0.99, respectively, for static incentive periods.

Figure G.3: Opacity of dynamic incentives and CRT, warehouse workers online



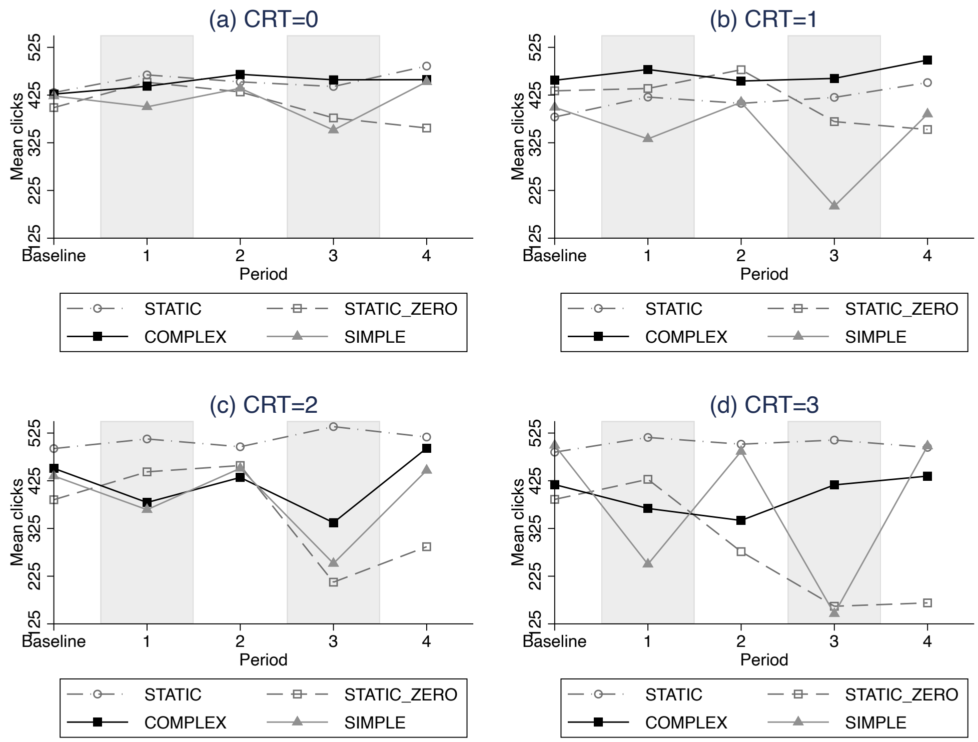
Notes: Panels (a) and (b) show average clicks in a given work period for workers with CRT ≤ 1 and CRT > 1, respectively. Shaded gray bars denote periods with dynamic incentives in COMPLEX and SIMPLE. Panel (c) plots coefficients of interaction terms, Period*Treatment*CRT, from a difference-in-differences regression relative to baseline period and the treatment STATIC (see column 1 of Table G.4 for all coefficients; CRT score enters the interaction term linearly). Error bars show 95 percent confidence intervals.

Table G.4: Diff-in-Diff of clicks relative to baseline period and STATIC, interacted with cognitive ability, warehouse workers

	(1)	(2)	(3)	(4)
Period1*COMPLEX*Cog. Ability	-24.28	-5.56	-12.59	-66.59
	(20.32)	(5.53)	(40.61)	(38.62)
Period2*COMPLEX*Cog. Ability	-30.76	-0.43	45.77	-90.53
	(19.20)	(5.82)	(35.54)	(37.71)
Period3*COMPLEX*Cog. Ability	-50.52	1.11	-30.29	-6.70
	(20.31)	(6.33)	(44.16)	(38.33)
Period4*COMPLEX*Cog. Ability	16.35	2.56	46.86	-23.75
	(15.34)	(5.25)	(38.47)	(35.20)
Period1*SIMPLE*Cog. Ability	-54.35	-17.86	-15.72	-40.26
	(20.14)	(7.35)	(45.64)	(45.64)
Period2*SIMPLE*Cog. Ability	-1.98	-2.98	16.75	-66.75
	(14.03)	(5.86)	(33.29)	(37.33)
Period3*SIMPLE*Cog. Ability	-95.04	-14.47	-58.88	20.07
	(20.80)	(8.46)	(53.11)	(49.31)
Period4*SIMPLE*Cog. Ability	2.53	-1.61	21.90	-0.04
	(13.04)	(5.34)	(37.06)	(34.39)
Period1*STATIC_ZERO*Cog. Ability	0.66	-1.61	70.30	-74.67
	(15.76)	(4.84)	(37.83)	(39.03)
Period2*STATIC_ZERO*Cog. Ability	4.39	0.29	39.56	-53.78
	(22.50)	(6.64)	(37.40)	(39.49)
Period3*STATIC_ZERO*Cog. Ability	-80.71	4.14	-110.85	57.37
	(33.30)	(10.96)	(69.28)	(58.58)
Period4*STATIC_ZERO*Cog. Ability	-23.36	2.46	-15.13	26.79
	(34.28)	(9.37)	(67.35)	(57.97)
Cognitive ability measure	CRT	Education	Back. induction	Broad bracketing
Additional coefficients suppressed	Yes	Yes	Yes	Yes
# Workers	436	436	436	436

Notes: OLS regressions. Fully interacted difference-in-differences model with STATIC and the baseline period as omitted category. Only the coefficients for the triple interaction of period with treatment and the different cognitive ability measures are shown. The measures for cognitive ability in columns 1 to 4 are CRT, years of schooling, ability to do backwards induction and ability to do broad bracketing, respectively. CRT is the linear CRT score (0–3). Backwards induction ability is an indicator for having won the Hit 7 game against the computer. Broad bracketing is an indicator for not violating dominance in a set of paired lottery choices. Negative coefficients mean that individuals with higher cognitive ability in a given treatment and period have a larger drop in clicks relative to baseline and STATIC than individuals with lower cognitive ability. Robust standard errors in parentheses, clustering on worker.

Figure G.4: Average clicks by value of CRT, warehouse workers



Notes: Each panel shows the average number of clicks in a given work period for workers with a given CRT score. The vertical shaded bars denote periods with dynamic incentives to reduce effort in COMPLEX and SIMPLE. Note that we have very few observations for panel (d): 6 workers in STATIC, 2 in STATIC_ZERO, 6 in COMPLEX and 13 in SIMPLE.

H Additional results for online experiments with AMT workers

H.1 Replicating experiments with warehouse workers and results on cognitive ability

In this appendix, we describe the replication treatments among AMT workers. In August of 2019, we conducted the same four treatments (COMPLEX, SIMPLE, STATIC, STATIC_ZERO) as with the warehouse workers. We added one treatment, STATIC_LOW, that implements a low but non-zero level of piece rate. In all, we had $N = 571$ AMT workers participate in these five treatments. An overview of all treatments and complete instructions are provided in Appendix J.[87] One notable difference relative to the online experiments with warehouse workers is that we adjusted the parameters slightly, to account for the typical wages of AMT workers, and to allow for the fact that AMT workers almost exclusively use computers rather than smartphones, which tends to increase speed of clicking. Specifically, the baseline target rate was increased to 400, and the piece rate was $0.50 rather than the value of $1.25 used with the warehouse workers.

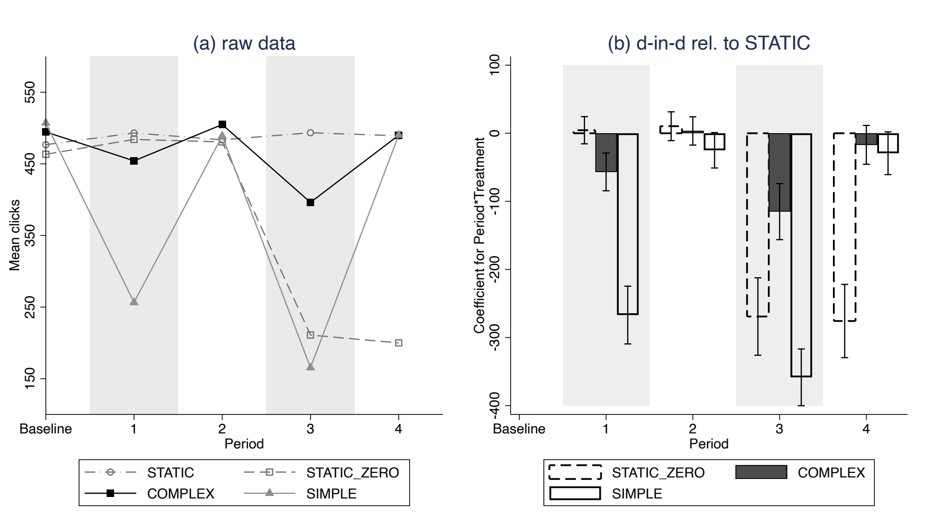
AMT workers are an interesting worker population to study because they have on average higher cognitive ability than the warehouse workers. Average CRT score is 2 for AMT workers, versus 0.6 for warehouse workers. Moreover, the typical educational attainment is a college degree among AMT workers as opposed to high school among warehouse workers. The AMT subject pool thus allows us to test whether our results hold in a similar, but not identical, group of participants and allows us to further explore the role of cognitive ability in the reaction to dynamic incentives.

Overall, we find that AMT workers respond to treatments in a very similar way to warehouse workers. Panel (a) of Figure H.1 shows effort across treatments and periods.

[87] There is no significantly different attrition in any of the AMT treatments compared to COMPLEX as baseline treatment. In a very low fraction of observations, AMT workers achieved extremely high number of clicks (up to 6000 clicks per 90-second period), which indicates the use of an auto clicker. Including these observations, in which workers always click high could over-estimate the inattention to ratchet incentives, because it could just reflect the auto clicker inducing very low effort cost. We thus exclude the 1.0 percent of observations, which have more than 900 clicks per 90-second period. However, all treatment differences remain virtually unchanged if we include them.

AMT workers respond much less to dynamic incentives in COMPLEX than in SIMPLE. Panel (b) shows the coefficients for the interactions of period with each treatment, from our difference-in-difference regression analysis (the full regression is in column 2 of Table G.2). All treatment differences are statistically significant relative to STATIC, including a modest but significant decrease in effort in COMPLEX.[88]

Figure H.1: Replication with AMT workers



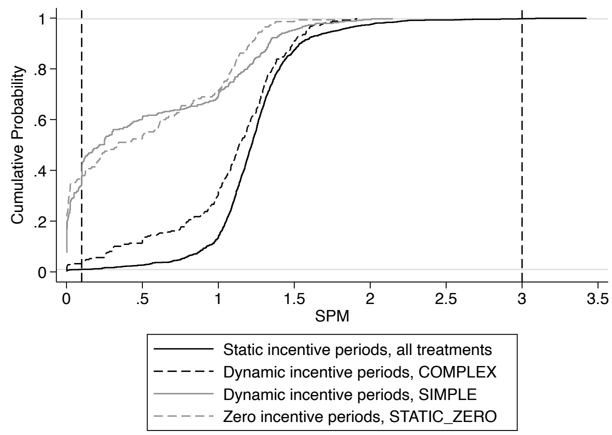
Notes: Panel (a) shows average number of clicks in a given work period. Panel (b) plots coefficients of interaction terms, Period*Treatment, from a difference-in-differences regression relative to baseline period and the treatment STATIC (see column 2 of Table G.2 for all coefficients). The vertical shaded bars in both panels denote periods with dynamic incentives to reduce effort in COMPLEX and SIMPLE. The piece rate was reduced to 0 in Periods 3 and 4 in the treatment STATIC_ZERO.

Our structural estimation for AMT workers incorporates the observation from the data that the representative worker chooses effort between the quota and cap when facing static incentives. Figure H.2 shows this is true for the median worker but also in almost every case at the individual level. We see that in periods without dynamic incentives, the quota was low enough that it was almost always attained, and workers almost never reached the cap: In periods with only static incentives, only 0.9 percent of SPM were below the quota and only 0.2 percent were above the cap. The figure shows some response to dynamic incentives in COMPLEX, with lower efforts in periods with dynamic incentives, and a much stronger reduction in SIMPLE and STATIC_ZERO, in periods with dynamic incentives and zero

[88]The p-values of the F-tests for the interactions with Periods 1 and 3 are: COMPLEX vs. SIMPLE: $p < 0.001$, STATIC vs. COMPLEX: $p < 0.001$, STATIC vs. SIMPLE: $p < 0.001$.

piece rate, respectively.

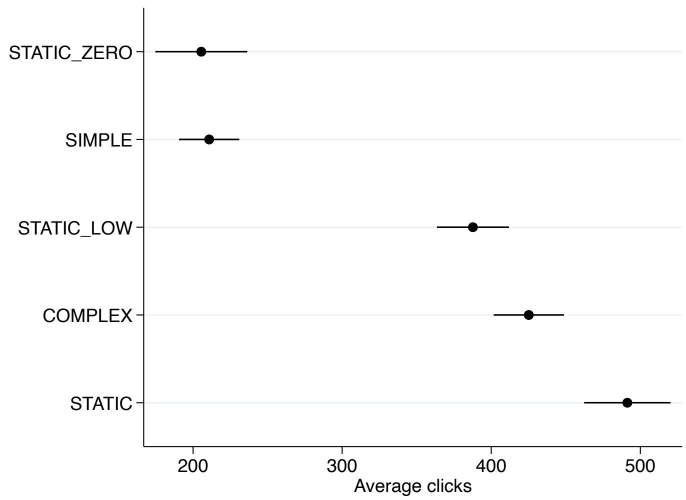
Figure H.2: Cumulative distributions of SPM and the quota and cap values, AMT workers



Notes: Vertical dashed lines indicate the quota and cap of 0.1 SPM and 3 SPM, respectively. Horizontal light gray lines indicate the fractions of observations below the quota, or below the cap, 0.009 and 0.997, respectively, for static incentive periods.

To estimate our structural model for AMT workers we use data on clicks across our three treatments that vary the piece rate, STATIC, STATIC_LOW, and STATIC_ZERO. Figure H.3 shows average clicks in these treatments, which illustrate that AMT workers on average increase effort as the piece rate increases. We also see that clicks in SIMPLE are very similar to STATIC_ZERO, as predicted if workers realize that the effective piece rate in SIMPLE in periods with dynamic incentives is zero. Clicks in COMPLEX are much higher than in SIMPLE, indicating greater opacity of dynamic incentives in COMPLEX.

Figure H.3: Response of average effort to piece rate variation, bench-marked by responses to dynamic incentives, AMT workers



Notes: The figure shows average clicks and 95% C.I.'s. For STATIC_ZERO, STATIC_LOW, and STATIC, the averages are from periods with across-treatment variation in the piece rate (Periods 3 and 4), where piece rates are zero, \$0.01, or \$0.50, respectively. For comparison, the figure shows results from COMPLEX and SIMPLE in periods with dynamic incentives (Periods 1 and 3).

Just as for warehouse workers, we also find that AMT workers are substantially less likely to mention dynamic incentives in COMPLEX than in SIMPLE. The corresponding fractions based on the three independent evaluators are 40 percent in COMPLEX versus 80 percent in SIMPLE (Wilcoxon test; $p < 0.001$). Thus, the majority of AMT workers do not seem to recognize the dynamic incentives in COMPLEX, while the vast majority do in SIMPLE.

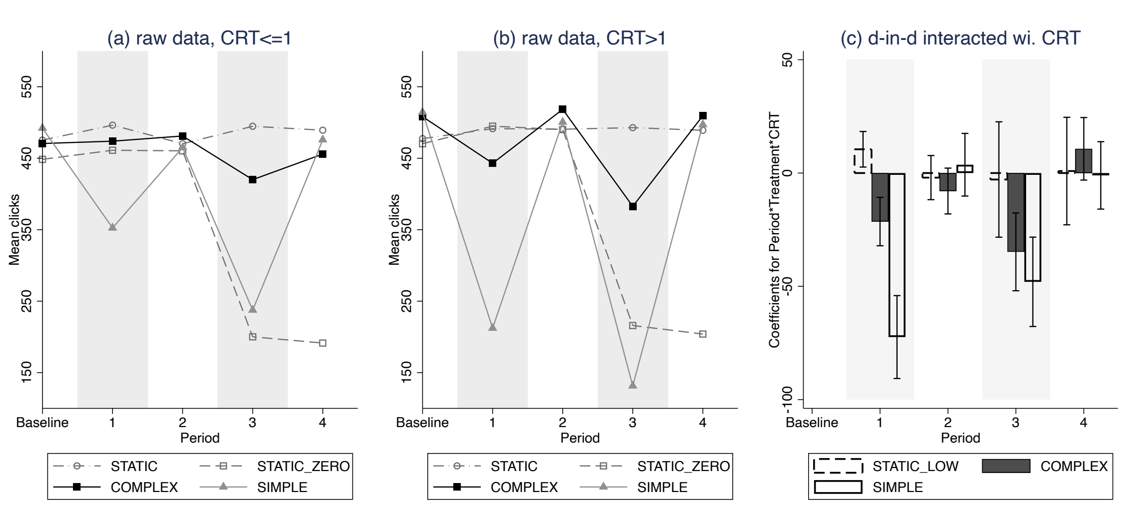
We also replicate with AMT workers that bounded rationality, as captured by CRT, matters for opacity of dynamic incentives (see Figure H.4 and column 1 of Table H.1).[89] AMT workers with higher CRT scores exhibit significantly greater responses to dynamic incentives in both COMPLEX and in SIMPLE, compared to STATIC.[90] Higher CRT is also significantly positively correlated with mentioning dynamic incentives, in both COMPLEX and SIMPLE (Spearman correlations; $\rho = 0.27$, $\rho = 0.17$, $p = 0.024$, $p = 0.025$). As was the case for warehouse workers, our other measures of cognitive ability have limited explanatory

[89]Figure H.5 shows results by each value of CRT separately, and as for warehouse workers, shows that transparency increases strongly when CRT surpasses 1.

[90]P-values for the F-tests for interactions with CRT in Periods 1 and 3: COMPLEX vs. STATIC: $p = 0.061$, SIMPLE vs. STATIC: $p < 0.001$, COMPLEX vs. SIMPLE: $p = 0.023$. P-value for the F-tests for interactions with CRT in Periods 3 and 4: STATIC_ZERO vs. STATIC: $p = 0.935$.

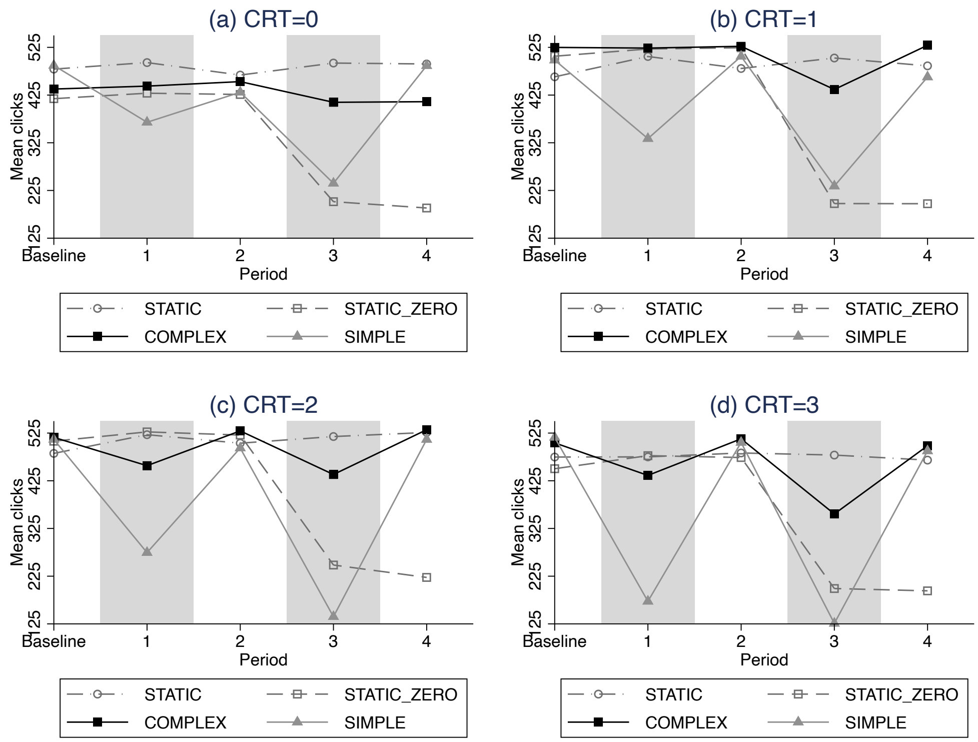
power for responses to dynamic incentives (see columns 2–4 in Table H.1).

Figure H.4: Opacity of dynamic incentives and CRT, AMT workers



Notes: Panels (a) and (b) show the average number of clicks in a given work period for workers with CRT ≤ 1 and CRT > 1, respectively. Panel (c) plots coefficients of interaction terms, Period*Treatment*CRT, from a difference-in-differences regression relative to baseline period and the treatment STATIC (see column 1 of Table H.1 for all coefficients; CRT score enters the interaction term linearly). The vertical shaded bars in all panels denote periods with dynamic incentives to reduce effort in COMPLEX and SIMPLE.

Figure H.5: Average clicks by value of CRT, AMT workers



Notes: Each panel shows the average number of clicks in a given work period for workers with a given CRT score. The vertical shaded bars denote periods with dynamic incentives to reduce effort in COMPLEX and SIMPLE.

Table H.1: Diff-in-Diff of clicks relative to baseline period and STATIC, interacted with cognitive ability, AMT workers

	(1)	(2)	(3)	(4)
Period1*COMPLEX*Cog. Ability	-21.42	-4.27	-46.44	72.56
	(10.69)	(9.36)	(30.73)	(29.21)
Period2*COMPLEX*Cog. Ability	-7.94	1.23	-25.67	20.79
	(10.11)	(7.13)	(23.22)	(21.06)
Period3*COMPLEX*Cog. Ability	-34.77	-16.39	-53.27	84.60
	(17.17)	(16.65)	(43.64)	(45.07)
Period4*COMPLEX*Cog. Ability	10.71	10.81	-11.02	54.52
	(13.80)	(10.59)	(26.14)	(31.18)
Period1*SIMPLE*Cog. Ability	-72.39	-44.42	-162.85	11.12
	(18.30)	(16.35)	(43.99)	(46.20)
Period2*SIMPLE*Cog. Ability	3.66	0.45	6.99	50.66
	(13.82)	(9.09)	(29.19)	(26.63)
Period3*SIMPLE*Cog. Ability	-48.00	-25.93	-41.84	76.59
	(19.72)	(18.09)	(42.43)	(45.72)
Period4*SIMPLE*Cog. Ability	-1.04	18.59	11.86	37.60
	(14.87)	(13.13)	(31.38)	(35.99)
Period1*STATIC_ZERO*Cog. Ability	10.49	4.64	-12.51	18.03
	(7.85)	(7.72)	(23.22)	(23.27)
Period2*STATIC_ZERO*Cog. Ability	-2.02	-5.87	-8.88	16.14
	(9.75)	(8.15)	(23.28)	(21.58)
Period3*STATIC_ZERO*Cog. Ability	-2.86	-1.28	-38.14	20.20
	(25.46)	(24.79)	(59.47)	(60.81)
Period4*STATIC_ZERO*Cog. Ability	0.88	19.83	-51.56	59.21
	(23.72)	(21.83)	(56.74)	(57.49)
Cognitive ability measure	CRT	Education	Back. induction	Broad bracketing
Additional coefficients suppressed	Yes	Yes	Yes	Yes
# Workers	449	449	449	449

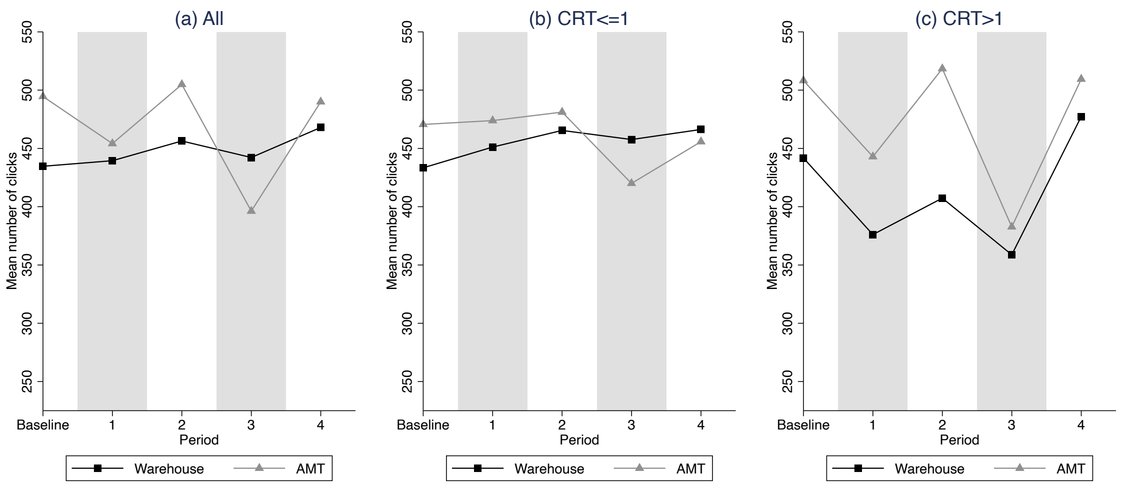
Notes: OLS regressions. Fully interacted difference-in-differences model with STATIC and the baseline period as omitted category. Only the coefficients for the triple interaction of period with treatment and the different cognitive ability measures are shown. The measures for cognitive ability in columns 1 to 4 are CRT, education, ability to do backwards induction and ability to do broad bracketing, respectively. CRT is the linear CRT score (0–3). Education is measured by six educational attainment categories (some high school; high school degree; some college; 2 year college degree; 4 year college degree; graduate or professional degree). Backwards induction ability is an indicator for having won the Hit 7 game against the computer. Broad bracketing is an indicator for not violating dominance in a set of paired lottery choices. Negative coefficients mean that individuals with higher cognitive ability in a given treatment and period have a larger drop in clicks relative to baseline and STATIC than individuals with lower cognitive ability. Robust standard errors in parentheses, clustering on worker.

While dynamic incentives are substantially opaque in COMPLEX for AMT workers, AMT workers do show signs of a greater relative awareness compared to warehouse workers. AMT workers in COMPLEX have a statistically significant difference relative to STATIC in Periods 1 and 3, unlike warehouse workers (see Table G.2). While far from the rational optimum, AMT workers are closer than warehouse workers. The percentage of AMT workers mentioning dynamic incentives is also higher than what we observed for warehouse workers, in both COMPLEX and SIMPLE.

One explanation for these differences is that AMT workers have higher CRT on average, an aspect of cognitive ability that we have shown matters for noticing opaque contract attributes.[91] Indeed, behavior of warehouse and AMT workers is more similar if we condition on CRT. Figure H.6 shows that behavior in COMPLEX becomes more similar for warehouse and AMT workers, if we compare within categories of CRT ≤ 1 and CRT > 1. Table H.2 presents regressions using the pooled sample of warehouse and AMT workers and shows that AMT workers have significantly stronger responses to dynamic incentives than warehouse workers in Periods 1 and 3, in both COMPLEX and SIMPLE. These differences are cut by about half, however, if the regressions are run separately for samples of high and low CRT workers. Differences in other facets of cognitive ability that we do not measure, but which might affect noticing opaque attributes, could be a reason for the remaining discrepancies in behavior of warehouse and AMT workers. Our findings illustrate how responses to the same incentive scheme can vary across worker populations according to differences in average cognitive ability and how this affects noticing opaque attributes.

[91] Comparing ability at backwards induction, as measured by the HIT 7 game, about 27 percent of warehouse workers win, versus 33 percent of AMT workers. Regarding choice bracketing, about 61 percent of warehouse workers bracket narrowly, compared to 40 percent for AMT workers.

Figure H.6: Comparing behavior in COMPLEX, warehouse versus AMT workers



Notes: Panel (a) shows the average number of clicks in a given work period for all warehouse and AMT workers in COMPLEX. Panels (b) and (c) compare warehouse and AMT workers who have CRT ≤ 1, and CRT > 1, respectively.

Table H.2: Diff-in-Diff of clicks relative to baseline period and STATIC, warehouse versus AMT workers

	(1) All workers	(2) CRT<=1	(3) CRT>1
Period2*COMPLEX*AMT	-26.99	2.90	9.86
	(23.77)	(27.74)	(45.56)
Period3*COMPLEX*AMT	1.66	7.57	39.17
	(21.32)	(31.21)	(37.88)
Period4*COMPLEX*AMT	-97.48	-73.86	-18.57
	(28.31)	(38.97)	(52.68)
Period5*COMPLEX*AMT	-0.54	-1.57	-26.25
	(21.86)	(32.31)	(34.10)
Period2*SIMPLE*AMT	-178.53	-93.24	-138.88
	(30.96)	(43.65)	(48.26)
Period3*SIMPLE*AMT	-18.00	-13.57	-21.70
	(22.43)	(36.87)	(22.97)
Period4*SIMPLE*AMT	-208.24	-162.81	-96.47
	(32.17)	(49.77)	(50.46)
Period5*SIMPLE*AMT	0.49	6.32	-14.68
	(22.98)	(34.55)	(29.58)
Period2*STATIC_ZERO*AMT	-8.04	-15.01	-21.84
	(21.68)	(27.27)	(24.21)
Period3*STATIC_ZERO*AMT	-5.99	6.30	-29.46
	(22.20)	(31.17)	(43.60)
Period4*STATIC_ZERO*AMT	-188.55	-218.57	-49.66
	(40.49)	(62.21)	(84.46)
Period5*STATIC_ZERO*AMT	-165.00	-161.83	-141.56
	(39.26)	(59.27)	(86.08)
Additional coefficients suppressed	Yes	Yes	Yes
# Workers	879	499	380

Notes: OLS regressions. The sample for column 1 includes all warehouse and AMT workers participating in the four treatments. Samples for columns 2 and 3 are warehouse and AMT workers with CRT scores ≤ 1 and > 1, respectively. Fully interacted difference-in-differences model with STATIC and the baseline period as omitted category. Only the coefficients for the triple interaction of period*treatment*AMT are shown. AMT is an indicator variable for AMT worker. Negative coefficients mean that AMT workers in that treatment and period have a larger drop relative to baseline and STATIC than warehouse workers. Robust standard errors in parentheses, clustering on worker.

H.2 Additional results on contract features contributing to opacity

Table H.3: Diff-in-Diff of clicks relative to baseline period and COMPLEX, contract features contributing to opacity

	(1)
Period1*SIMPLE	-210.24
	(22.70)
Period2*SIMPLE	-28.44
	(11.63)
Period3*SIMPLE	-243.40
	(24.13)
Period4*SIMPLE	-12.36
	(14.58)
Period1*NOISE	-124.23
	(23.54)
Period2*NOISE	13.67
	(11.17)
Period3*NOISE	-128.15
	(27.06)
Period4*NOISE	21.44
	(13.97)
Period1*NOISE_MARGINAL	-47.70
	(18.43)
Period2*NOISE_MARGINAL	11.39
	(9.14)
Period3*NOISE_MARGINAL	-25.49
	(24.49)
Period4*NOISE_MARGINAL	19.03
	(13.60)
Additional coefficients suppressed	Yes
# Workers	531

Notes: OLS regressions. Fully interacted difference-in-differences model with COMPLEX as the benchmark treatment. Only the coefficients for the interaction of period with treatment are shown. Negative coefficients mean that individuals in that treatment and period have a larger drop relative to baseline than individuals in COMPLEX. Robust standard errors in parentheses, clustering on worker.

Table H.4: Word count, reading grade level, and ease of reading scores for experiment instructions, from online experiments with warehouse and AMT workers

	Word count	Reading grade level	Ease of reading score
Main treatments:			
STATIC	475	7	76.3
STATIC_ZERO	421	6.3	79.9
COMPLEX	785	7.1	75.5
SIMPLE	704	8	73.1
Contract features contributing to opacity:			
NOISE	991	9.3	67.1
NOISE_MARGINAL	985	6.5	78.2
Robustness of opacity:			
LINEAR	747	6.2	79.8
NOSPM	1154	10.6	66
LINEAR_NOSPM	776	7.2	72.1
Additional treatments:			
STATIC_LOW	484	6.8	76.9
SIMPLE_NOLOSS	755	5.9	80.6
Firm's actual communication materials:			
Static incentives	824	6.9	72.5
INDIVIDUAL Trial	633	7.3	75.6
GROUP Trial	612	7.4	73.2

Notes: Statistics are calculated from instructions for each treatment. The first four treatments were conducted with both warehouse and AMT workers, and had the same instructions for both groups except for slightly different parameter values for target rate and piece rate. Note that instructions for periods 3 and 4 were essentially identical to periods 1 and 2 for all treatments, except for STATIC, STATIC_ZERO, and STATIC_LOW; excluding period 3 and 4 instructions does not change the qualitative rankings of treatments in terms of difficulty. We measure reading grade level, and the related ease of reading score, using the Flesch-Kincaid Grade Level and Flesch Ease of Reading tests as implemented in Microsoft Word.

H.3 Additional results on robustness of opacity

We implement three variations on COMPLEX, the online treatment most similar to the actual incentive scheme in the warehouse. Treatment LINEAR eliminates the quota and cap, i.e., it pays for SPM starting right at zero, rather than 0.1, and without a cap at 3 SPM. It is plausible that firms might want to try such a perturbation, and indeed, discussions with managers at our firm suggest that this is a change they may consider. We also implement

NOSPM, which eliminates the construct of SPM from the instructions all together, and explains everything in monetary terms directly, e.g., we speak of a wage per click.[92] Lastly, we implement a treatment LINEAR_NOSPM, which makes the piece rate linear and eliminates SPM. In all, we had 369 AMT workers participate in these three treatments.

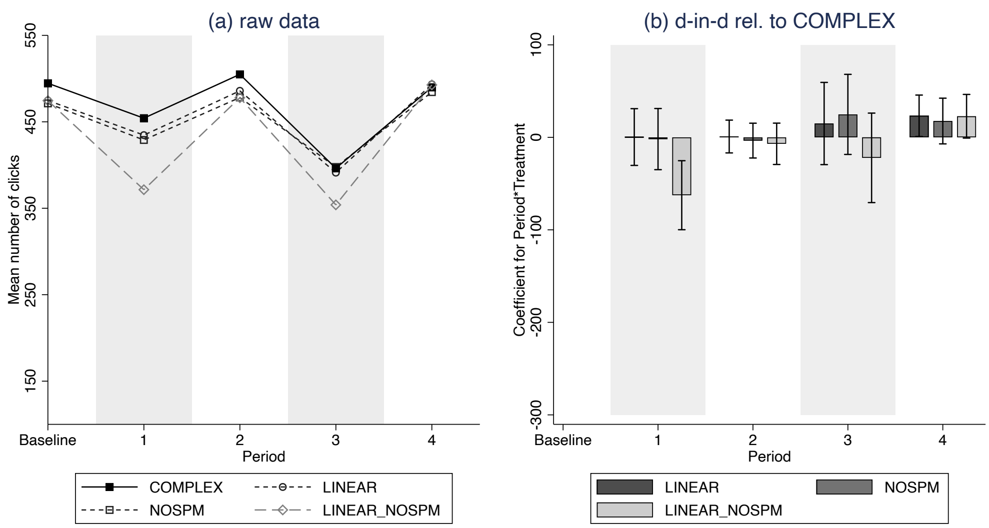
Finding 10. *Opacity of ratchet incentives is robust to making the scheme linear, or making monetary consequences more salient by eliminating SPM. There is a stronger response to dynamic incentives when we combine both, but the response is still modest and far smaller than for SIMPLE.*

We find that AMT workers in LINEAR and NOSPM behave almost exactly the same as workers in COMPLEX (see Figure H.7 and Table H.5). Workers in LINEAR_NOSPM react more strongly to dynamic incentives than workers in COMPLEX, or workers in LINEAR or NOSPM, but the differences are modest in size, and much smaller than the response observed in SIMPLE.[93]

[92] Treatment SIMPLE describes the dynamic incentives without reference to SPM, but still uses SPM for the rest of the incentive scheme.

[93] F-tests on Periods 1 and 3: LINEAR vs. COMPLEX $p = 0.706$, NOSPM vs. COMPLEX $p = 0.352$, COMPLEX vs. LINEAR_NOSPM $p = 0.002$. The response in LINEAR_NOSPM is also stronger compared to LINEAR or NOSPM: LINEAR_NOSPM vs. LINEAR $p = 0.003$, LINEAR_NOSPM vs. NOSPM $p = 0.010$.

Figure H.7: Robustness of opacity, AMT workers



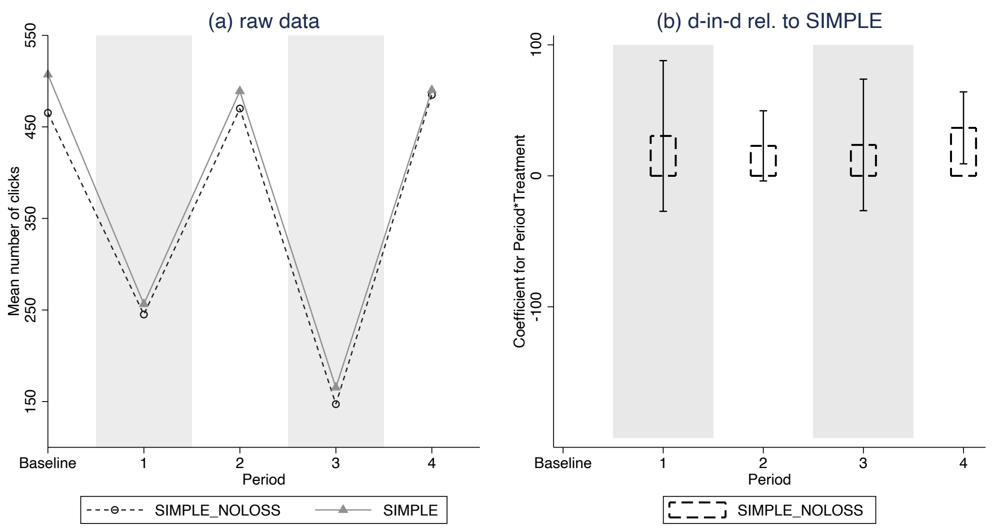
Notes: Panel (a) shows average number of clicks in a given work period. Panel (b) plots coefficients of interaction terms, Period*Treatment, from a difference-in-differences regression relative to baseline period and the treatment COMPLEX (see column 1 of Table H.5 for all coefficients). The vertical shaded bars in both panels denote periods with dynamic incentives to reduce effort in all treatments.

Table H.5: Diff-in-Diff of clicks relative to baseline period and COMPLEX, robustness of opacity

	(1)
Period1*LINEAR	0.34
	(15.65)
Period2*LINEAR	0.89
	(9.05)
Period3*LINEAR	14.93
	(22.67)
Period4*LINEAR	23.31
	(11.43)
Period1*NOSPM	-1.89
	(16.87)
Period2*NOSPM	-3.51
	(9.60)
Period3*NOSPM	24.81
	(22.08)
Period4*NOSPM	17.64
	(12.61)
Period1*LINEAR_NOSPM	-62.57
	(19.04)
Period2*LINEAR_NOSPM	-6.98
	(11.39)
Period3*LINEAR_NOSPM	-22.17
	(24.68)
Period4*LINEAR_NOSPM	22.79
	(12.04)
Additional coefficients suppressed	Yes
# Workers	493

Notes: OLS regressions. Fully interacted difference-in-differences model with COMPLEX and the baseline period as omitted category. Only the coefficients for the interaction of period with treatment are shown (interactions of SIMPLE with period are also suppressed). Negative coefficients mean that individuals in that treatment and period have a larger drop relative to baseline than individuals in COMPLEX. Robust standard errors in parentheses, clustering on worker.

Figure H.8: Effect of potential losses, AMT workers



Notes: Panel (a) shows average number of clicks in a given work period, comparing treatment SIMPLE_NOLOSS to SIMPLE. Panel (b) plots coefficients of interaction terms, Period*Treatment, from a difference-in-differences regression relative to baseline period and the treatment SIMPLE (see column 1 of Table H.6 for all coefficients). The vertical shaded bars denote periods with dynamic incentives to reduce effort in all treatments.

Table H.6: Diff-in-Diff of clicks relative to baseline period and SIMPLE, additional treatment SIMPLE_NOLOSS, AMT workers

	(1)
Period1*SIMPLE_NOLOSS	30.45
	(29.34)
Period2*SIMPLE_NOLOSS	22.88
	(13.66)
Period3*SIMPLE_NOLOSS	23.61
	(25.62)
Period4*SIMPLE_NOLOSS	36.62
	(14.01)
Additional coefficients suppressed	Yes
# Workers	289

Notes: OLS regressions. Fully interacted difference-in-differences model with SIMPLE and the baseline period as omitted category. Only the coefficients for the interaction of period with treatment are shown. Negative coefficients mean that individuals in that treatment and period have a larger drop relative to baseline than individuals in SIMPLE. Robust standard errors in parentheses, clustering on worker.

I Instructions for online experiments with warehouse workers

This document provides the instruction text for the online experiments with warehouse workers. A few words and phrases have been modified to preserve confidentiality of the firm, but without changing the sense of the instructions.

Decision Study

Welcome to the study!

This is research by economists at the University of Oxford, University of Pittsburgh, and Amherst College. Current [Firm] employees are being invited to participate.

The purpose is to understand what employees like and how they make decisions. You must be at least 18 years old to participate. Participating is completely voluntary.

Your individual decisions will be completely confidential, although we will share the general results with [Firm].

You will be paid for participating

- You get $10.50 as a "thank you" for participating
- You can also get more money, based on your choices; on average you could get $13.5 more, for a total of $24**.**
- The money does not come from [the Firm], but from the universities.
- You will get the money through [the Firm's] payroll.
- [The Firm] will not know how choices in the study influence money, so the money you earn will not tell [the Firm] what you chose.
- [The Firm] will know if you participate or not.
- You must complete the entire study to receive payment

The study involves questions, tasks, and games. It should take less than 25 minutes to do the study.

At this time, only [Firm] employees who are eligible for the [warehouse] incentive scheme are eligible to participate and be paid for the study. This means that [other types of workers] are not eligible and will not be paid for participating.

There are no risks, beyond those usually involved in online activities, and no personal benefits. The study may help increase scientific knowledge.

The researchers will not learn your name or any other identifiable information. De-identified data may be shared with others interested in similar research. The researchers will get productivity data from the warehouse, and match this to your survey responses.

If you have any questions about this study now, you should feel free to contact the Principal Investigator, Dr. David Huffman, at huffmand@pitt.edu. If you have questions later, or wish to withdraw, please contact the Principal Investigator.

If you have questions pertaining to your rights as a research participant, or want to report concerns about this study, you should contact the Human Subjects Protection Advocate at the University of Pittsburgh IRB Office (866-212-2668). You may want to write this phone number down in case you want it later.

You may discontinue participation at any time during the research activity. There is no penalty for choosing not to participate or choosing to stop/discontinue participation

You need a smart phone, tablet, or computer to participate. Please do the study by yourself, in a quiet place.

Are you 18 years or older?

- Yes
- No

I have read and understand the information above.

- Yes
- No

I want to participate in this research and continue with the study.

- Yes
- No

A Few Study Tips

Please do not use the “back” button on your browser or smartphone during the study! On a few types of devices using the "back" button can cause errors and you might be unable to complete the study.

If you are using a smartphone we suggest that you set it on a flat surface so you don't press the "back" button accidentally while holding the phone.

A final tip: The session will expire if you are inactive for more than 20 minutes and it will be considered incomplete, so it's best to not take long breaks once you've started the study. Please remember that only completed studies will be paid.

Thank you for participating!

Part 1: Basic Demographics

In what year were you born?

______

Are you male or female?

- o Male
- o Female

For how many years did you go to school or university in total? (add up all the years you spent in primary school, secondary school, university, or doing an apprenticeship.)

______

Is English your first language (mother toungue)?

- o Yes
- o No

What type of device are you using for the study?

- o Smartphone
- o Tablet
- o Computer
- o Other

Please enter your user ID. You need to enter your correct ID so that you can be paid.

______

Part 2: Instructions

<u>How does it work?</u>

1. In this part you can do a task, as much or as little as you like.
2. Doing the task gives you Standard Productive Minutes (SPM)
3. You can earn real money.

<u>What is the task?</u>

The task is very simple: You can click a button on the screen.

The task lasts 1 minute and 30 seconds.

If you use a touch screen:
You can tap the button with your finger.
Please do not tap the button with **two fingers at the exact same time**! This can cause the task to **end right away** on a few types of phones. You will be able to continue with the study, but will only get the score from the time the task stopped.

If you use a computer:
You must use the **mouse or trackpad** to click the button. Using the keyboard is not allowed, and pressing keys can **take away** from your score.

<u>How do you get SPM?</u>

Clicking the button gets you Standard Productive Minutes (SPM), just like SPH in the CFC.

The amount of clicks it takes to get SPM depends on the "**target rate**."

The target rate is 300.

We calculate SPM by dividing your total clicks by the target rate of 300.

Example:
You do 150 clicks
Total SPM is 150/300 = 0.5 SPM

<u>How do you get money?</u>

Once your SPM gets larger than 0.1, you start to earn money

You get $1.50 times the number of SPM units above 0.1.

You do not earn money for SPM units above 3.

Example:
Suppose you get 1 SPM.
This is 1-0.1 = 0.9 SPM units above 0.1 so you earn $1.50*0.9 = $1.35.

Please click "Next" to start with the task.

Part 3: Sooner or Later?

How does it work?

In this part you will make choices between different pairs of options.

One option will be: Get $15.00 added to your next paycheck.
The other option will be: Get a larger amount added to the paycheck **after** next.

So, if you want the larger amount, you have to wait longer, for one more paycheck.

Example:

Which option do you like better?
Option 1: $15.00 added to your next paycheck.
or
Option 2: $45.00 added to your paycheck after next.

You will make choices for 5 pairs of options.

How do you get money?

For this part, **only 1 out of every 10 participants is paid**.

At the end of the study, the computer randomly determines if you are someone who will be paid for this part.

If you are selected to be paid for this part:

- The computer randomly selects **one** of the 5 pairs of options.
- You get whatever option you chose for that pair.

Please click "Next" to continue.

[The following table summarizes how the 5 questions were selected based on subject responses]

				Skip logic: Go to Q x	
				If choose	If choose
Iteration	Q. number	Early payment	Delayed payment	Early	Late
5	5	15	61.5	stop	stop
4	4	15	60	to 5	to 6
5	6	15	58.5	stop	stop
3	3	15	57	to 4	to 7
5	9	15	55.5	stop	stop
4	7	15	54	to 9	to 8
5	8	15	52.5	stop	stop
2	2	15	51	to 3	to 10
5	12	15	49.5	stop	stop
4	11	15	48	to 12	to 13
5	13	15	46.5	stop	stop
3	10	15	45	to 11	to 14
5	15	15	43.5	stop	stop
4	14	15	42	to 15	to 16
5	16	15	40.5	stop	stop
1	1	15	39	to 2	to 17
5	27	15	37.5	stop	stop
4	26	15	36	to 27	to 28
5	28	15	34.5	stop	stop
3	25	15	33	to 26	to 29
5	30	15	31.5	stop	stop
4	29	15	30	to 30	to 31
5	31	15	28.5	stop	stop
2	17	15	27	to 25	to 18
5	21	15	25.5	stop	stop
4	19	15	24	to 21	to 20
5	20	15	22.5	stop	stop
3	18	15	21	to 19	to 22
5	24	15	19.5	stop	stop
4	22	15	18	to 24	to 23
5	23	15	16.5	stop	stop

Part 4: Decisions About Risk

How does it work?

In this part you again make choices between different pairs of options.
This time, either option is paid out with the next paycheck. However, one option is a lottery and the other is a sure payment.

Option 1 is always to play this lottery:

The computer “flips a coin”

- If heads, you win $45,
- If tails, you win $0

Option 2 is to get a payment for sure.

- The amount of the sure payment is different for different pairs.

Example:
Which option do you like better?

Option 1: Lottery with equal chance to win $45 or win £0
or
Option 2: Sure payment of $15

You will make choices for 5 pairs of options.

How do you get money?

For this part, **only 1 out of every 10 participants is paid**.

At the end of the study, the computer randomly determines if you are someone who will be paid.

If you are selected to be paid for this part:

- The computer randomly selects **one** of the 5 pairs of options.
- You get whatever option you chose for that pair.
-

Please click "Next" to continue.
You will make choices for 5 pairs of options.

[The following table summarizes how the 5 questions were selected based on subject responses]

Iteration	Q. number	Lottery EV	Sure	Skip logic: Go to Q x If choose Lottery	If choose Sure
5	5	22.5	46.5	stop	stop
4	4	22.5	45	to 5	to 6
5	6	22.5	43.5	stop	stop
3	3	22.5	42	to 4	to 7
5	9	22.5	40.5	stop	stop
4	7	22.5	39	to 9	to 8
5	8	22.5	37.5	stop	stop
2	2	22.5	36	to 3	to 10
5	12	22.5	34.5	stop	stop
4	11	22.5	33	to 12	to 13
5	13	22.5	31.5	stop	stop
3	10	22.5	30	to 11	to 14
5	15	22.5	28.5	stop	stop
4	14	22.5	27	to 15	to 16
5	16	22.5	25.5	stop	stop
1	1	22.5	24	to 2	to 17
5	27	22.5	22.5	stop	stop
4	26	22.5	21	to 27	to 28
5	28	22.5	19.5	stop	stop
3	25	22.5	18	to 26	to 29
5	30	22.5	16.5	stop	stop
4	29	22.5	15	to 30	to 31
5	31	22.5	13.5	stop	stop
2	17	22.5	12	to 25	to 18
5	21	22.5	10.5	stop	stop
4	19	22.5	9	to 21	to 20
5	20	22.5	7.5	stop	stop
3	18	22.5	6	to 19	to 22
5	24	22.5	4.5	stop	stop
4	22	22.5	3	to 24	to 23
5	23	22.5	1.5	stop	stop

Part 5: Instructions

[*COMPLEX* version]

<u>How does it work?</u>

1. You can do the clicking task again, as much or as little as you like.
2. This time there will be 2 rounds of 1 minute and 30 seconds each.
3. In round 1, your earnings depend on clicks and the round 1 target rate.
4. In round 2, your earnings depend on clicks and the round 2 target rate.
5. **But there is also a special rule in round 2:**
   - The round 2 target rate depends partly on how many clicks you do in round 1.
   - It is higher if you do more clicks in round 1, and lower if you do fewer clicks in round 1.
6. There is no round 3.

*<u>How do you earn money in **round 1**?</u>*

We take the total clicks you did in round 1, and divide by a target rate of 300**.**

You get $1.50 times the number of SPM units above 0.1.

But, as before, you only earn for SPM units between 0.1 and 3

*<u>How do you earn money in **round 2**?</u>*

We take the total clicks you did in round 2, and divide by the round 2 target rate**.**

You get $1.50 times the number of SPM units above 0.1.

But you only earn for SPM units between 0.1 and 3

<u>How do we set the target rate in round 2?</u>

We calculate the round 2 target rate as the **average** of:

The total clicks you do in round 1

and

A number X that is 285, 300, or 315, with equal chance of being 285, 300, or 315.

You do not find out the number X until the beginning of round 2.

Example 1:
You do 300 total clicks total in round 1; X turns out to be 300.
The new target rate for round 2 is (300+300)/2 = 300
In round 2 you will need 300 clicks to get 1 SPM.

Example 2:
You do 100 clicks in round 1; X turns out to be 300.
The new target is (100+300)/2 = 200
In round 2 you will need 200 clicks to get 1 SPM.

Please click "Next" to start with the task.

[*SIMPLE* version]

How does it work?

1. You can do the clicking task again, as much or as little as you like.
2. This time, though, there will be 2 rounds of 1 minute and 30 seconds each.
3. In round 1, your earnings depend on clicks and the round 1 target rate.
4. In round 2, your earnings depend on clicks and the round 2 target rate.
5. **But there is also a special rule in round 2:**
    - We **subtract** any money you earn in round 1 from your money in round 2.
    - If you earned more money in round 1 than round 2, we subtract the difference from your total earnings for the study.
6. There is no round 3.

*How do you earn money in **round 1**?*

We take the total clicks you did in round 1, and divide by a target rate of 300**.**

You get $1.50 times the number of SPM units above 0.1.

But, you only earn for SPM units between 0.1 and 3

*How do you earn money in **round 2**?*

We take the total clicks you did in round 2, and divide by a target rate of 300**.**

You get $1.50 times the number of SPM units above 0.1.

But, you only earn for SPM units between 0.1 and 3

Special rule for round 2

This is how we calculate your final earnings for round 2:

We take away any money you earned in round 1, from the money you earn in round 2.

If you earned more money in round 1 than round 2, we take away the difference from your total earnings at the end of the study.

Example:
You did 330 clicks and earned $1.50 in round 1
You did 330 clicks and earned $1.50 in round 2
But, since you earned $1.50 in round 1, we take away $1.50 from round 2 and you end up with $0.00 for round 2.

Example:
You did 90 clicks and earned $0.30 in round 1
You did 330 clicks and earned $1.50 in round 2
Since you earned $0.30 in round 1, we take away $0.30 from round 2 and you end up with $1.20 for round 2.

Please click "Next" to start with the task.

[*STATIC* and *STATIC_ZERO* version]

How does it work?

1. You can do the clicking task again, as much or as little as you like.
2. This time there will be 2 rounds of 1 minute and 30 seconds each.
3. In round 1, your earnings depend on clicks and the round 1 target rate.
4. In round 2, your earnings depend on clicks and the round 2 target rate.
5. There is no round 3.

How do you earn money in ***round 1****?*

We take the total clicks you did in round 1, and divide by a target rate of 300**.**

You get $1.50 times the number of SPM units above 0.1

You do not earn for SPM units above 3.

*How do you earn money in **round 2**?*

We take the total clicks you did in round 2, and divide by a target rate of 300**.**

You get $1.50 times the number of SPM units above 0.1

You do not earn for SPM units above 3.

Please click "Next" to start with the task.

Part 6: Hit 7 Game

How does it work?

- In this part you play a game with the computer, called “hit 7.”
- You can choose a number; either 1, 2, or 3.
- Then, the computer chooses a number; either 1, 2, 3.
- We add the computer’s number to your number.
- You and the computer keep taking turns choosing until someone wins.
- The winner is the first one to choose the number that makes the sum add to 7.

Example:

- You choose a number
- The computer chooses a number, but the sum is not yet 7
- You choose another number, and the sum is 7, so you win.

How do you earn money?

If you win you earn $1.50, if you lose you win zero.

Part 7: Additional Questions

A bat and a ball cost $1.10. The bat costs $1.00 more than the ball. How much does the ball cost? Please provide answer in **cents**.

______

Suppose you were given the following two choices, and suppose the gains and losses from both choices would be added to your final payment. What would you choose?

- A sure gain of $3.60
- A 25% chance to gain $15 and a 75% chance to gain zero

Suppose you were given the following two choices, and suppose the gains and losses from both choices would be added to your final payment. What would you choose?

- A sure loss of $11.25
- A 25% chance to lose $15 and a 75% chance to lose zero

On a scale from 0 to 10, with 0 being "Completely **unwilling** to take risks" and 10 being "Completely **willing** to take risks". How do you see yourself: Are you a person who is generally fully prepared to take risks, or do you try to avoid risks?

0 = completely unwilling, 10 = completely willing

On a scale from 0 to 10, with 0 being "Completely **unwilling** to give it up" and 10 being "Completely **willing** to give it up". How willing are you to give up something that is beneficial for you today in order to benefit more from that in the future?

0 = completely unwilling, 10 = completely willing

If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?

______

In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?

______

Part 8: Instructions

[*COMPLEX* version]

How does it work?

1. You can do the clicking task again, as much or as little as you like.
2. There will be 2 rounds of 1 minute and 30 seconds each with the same rules as before.
3. In round 1, your earnings depend on clicks and the round 1 target rate.
4. In round 2, your earnings depend on clicks and the round 2 target rate.
5. **But there is also a special rule in round 2:**
   - The round 2 target rate depends partly on how many clicks you do in round 1.
   - It is higher if you do more clicks in round 1, and lower if you do fewer clicks in round 1.
6. There is no round 3.

*How do you earn money in **round 1**?*

We take the total clicks you did in round 1, and divide by a target rate of 300**.**

You get $1.50 times the number of SPM units above 0.1.

But, as before, you only earn for SPM units between 0.1 and 3

*How do you earn money in **round 2**?*

We take the total clicks you did in round 2, and divide by the round 2 target rate**.**

You get $1.50 times the number of SPM units above 0.1

But you only earn for SPM units between 0.1 and 3

How do we set the target rate in round 2?

We calculate the round 2 target rate as the **average** of:

The total clicks you do in round 1

and

A number X that is 285, 300, or 315, with equal chance of being 285, 300, or 315.

You do not find out the number X until the beginning of round 2.

Example 1:
You do 300 total clicks total in round 1; X turns out to be 300.
The new target rate for round 2 is (300+300)/2 = 300
In round 2 you will need 300 clicks to get 1 SPM.

Example 2:
You do 100 clicks in round 1; X turns out to be 300.
The new target is (100+300)/2 = 200
In round 2 you will need 200 clicks to get 1 SPM.

Please click "Next" to start with the task.

[*SIMPLE* version]

How does it work?

1. You can do the clicking task again, as much or as little as you like.
2. There will be 2 rounds of 1 minute and 30 seconds each with the same rules as before.
3. In round 1, your earnings depend on clicks and the round 1 target rate.
4. In round 2, your earnings depend on clicks and the round 2 target rate.
5. **But there is also a special rule in round 2:**
   - We **subtract** any money you earn in round 1 from your money in round 2.
   - If you earned more money in round 1 than round 2, we subtract the difference from your total earnings for the study.
6. There is no round 3.

*How do you earn money in **round 1**?*

We take the total clicks you did in round 1, and divide by a target rate of 300**.**

You get $1.50 times the number of SPM units above 0.1.

But, you only earn for SPM units between 0.1 and 3

*How do you earn money in **round 2**?*

We take the total clicks you did in round 2, and divide by a target rate of 300**.**

You get $1.50 times the number of SPM units above 0.1.

But, you only earn for SPM units between 0.1 and 3

Special rule for round 2

This is how we calculate your final earnings for round 2:

We take away any money you earned in round 1, from the money you earn in round 2.

If you earned more money in round 1 than round 2, we take away the difference from your total earnings at the end of the study.

Example:
You did 330 clicks and earned $1.50 in round 1
You did 330 clicks and earned $1.50 in round 2
But, since you earned $1.50 in round 1, we take away $1.50 from round 2 and you end up with $0.00 for round 2.

Example:
You did 90 clicks and earned $0.30 in round 1
You did 330 clicks and earned $1.50 in round 2
Since you earned $0.30 in round 1, we take away $0.30 from round 2 and you end up with $1.20 for round 2.

Please click "Next" to start with the task.

[*STATIC* version]

How does it work?
1. You can do the clicking task again, as much or as little as you like.
2. There will be 2 rounds of 1 minute and 30 seconds each with the same rules as before.
3. In round 1, your earnings depend on clicks and the round 1 target rate.
4. In round 2, your earnings depend on clicks and the round 2 target rate.
5. There is no round 3.

*How do you earn money in **round 1**?*

We take the total clicks you did in round 1, and divide by a target rate of 300**.**

You get $1.50 times the number of SPM units above 0.1

You do not earn for SPM units above 3.

How do you earn money in ***round 2****?*

We take the total clicks you did in round 2, and divide by a target rate of 300**.**

You get $1.50 times the number of SPM units above 0.1

You do not earn for SPM units above 3.

Please click "Next" to start with the task.

[*STATIC_ZERO* version]

How does it work?

1. You can do the clicking task again, as much or as little as you like.
2. There will be 2 rounds of 1 minute and 30 seconds each
3. This time there is no payment for clicking
4. In round 1, you earn nothing from clicking.
5. In round 2, you earn nothing from clicking.
6. There is no round 3.

How do you earn money in ***round 1****?*

You do not earn money for clicking in round 1.

How do you earn money in ***round 2****?*

You do not earn money for clicking in round 2.

Please click "Next" to start with the task.

Please tell us how you think about the following:
If someone were trying to get the most money, total, from round 1 and round 2 of this last part, what do you think would be the best approach?

J Overview of treatments and instructions for online experiments with AMT workers

Table J.1: Descriptions of treatments with AMT workers

Main treatments (replication):	
STATIC	Control treatment with no dynamic incentives, normal piece rate for all 5 periods
STATIC_ZERO	Control treatment with no dynamic incentives, normal piece rate for first 3 periods, zero for last two periods
COMPLEX	Complex dynamic incentives, similar to warehouse
SIMPLE	Nature of dynamic incentives and explanation changed to make dynamic incentives more transparent
Contract features contributing to opacity:	
NOISE	Same as SIMPLE, but adds noise to the target rate
NOISE_MARGINAL	Same as NOISE, but dynamic incentives affect the slope of future earnings, requiring complex contingent thinking
Robustness of opacity:	
LINEAR	Dynamic incentives like COMPLEX, but with linear piece rate schedule
NOSPM	Dynamic incentives like COMPLEX, but whole incentive scheme explained without SPM
LINEAR_NOSPM	Dynamic incentives like COMPLEX, but with linear piece rate schedule and whole scheme explained without SPM
Additional treatments:	
STATIC_LOW	Control treatment with no dynamic incentives, normal piece rate for first 3 periods, lower but non-zero for last two periods
SIMPLE_NOLOSS	Like SIMPLE but losses are not possible

Table J.2: Summary of characteristics for treatments with AMT workers

	Dynamic incentives	Dynamic incentives affect marginal bonus	Explain dynam. incentives using SPM	Noise in target rate	Entire scheme explained without SPM	Piece rate is linear	Dynam. incent. can lead to negative earnings
Main treatments (replication)							
STATIC							
STATIC_ZERO							
SIMPLE	Yes						Yes
COMPLEX	Yes	Yes	Yes	Yes			
Contract features contributing to opacity:							
NOISE	Yes			Yes			Yes
NOISE_MARGINAL	Yes	Yes		Yes			
Robustness of opacity:							
LINEAR	Yes	Yes	Yes	Yes		Yes	
NOSPM	Yes	Yes	Yes	Yes	Yes		
LINEAR_NOSPM	Yes	Yes	Yes	Yes	Yes	Yes	
Additional treatments							
STATIC_LOW							
SIMPLE_NOLOSS	Yes						

Instructions for Online Experiments with AMT Workers

Replication treatments:

1. COMPLEX
2. SIMPLE
3. STATIC
4. STATIC_ZERO

Contract features contributing to opacity:

5. NOISE
6. NOISE_MARGINAL

Robustness of opacity:

7. LINEAR
8. NOSPM
9. LINEAR_NOSPM

Additional treatments:

10. STATIC_LOW
11. SIMPLE_NOLOSS

1. STATIC

Period 1

How does it work?

1. In this part you can do a task,
2. Doing the task gives you Standard Productive Minutes (SPM)
3. You can earn real money.

What is the task?

The task is very simple: You can click a button on the screen.

The task lasts 1 minute and 30 seconds.

If you use a touch screen you can tap the button with your finger.

If you use a computer, you must use the **mouse or trackpad** to click the button. Using the keyboard is not allowed, and pressing keys can **take away** from your score.

How do you get SPM?

Clicking the button gets you Standard Productive Minutes (SPM).

The amount of clicks it takes to get SPM depends on the "**target rate**."

The target rate is 400.

We calculate SPM by dividing your total clicks by the target rate of 400.

Example:
You do 300 clicks
Total SPM is 300/400 = 0.75 SPM

How do you get money?

Once your SPM gets larger than 0.1, you start to earn money

You get $0.50 times the number of SPM units above 0.1.

You do not earn money for SPM units above 3.

Example:

Suppose you get 1 SPM.
This is 1-0.1 = 0.9 SPM units above 0.1 so you earn $0.50*0.9 = $0.45.

Please click "Next" to start with the task.

Period 2

How does it work?

1. You can do the clicking task again.
2. This time there will be 2 rounds of 1 minute and 30 seconds each.
3. In round 1, your earnings depend on clicks and the round 1 target rate.
4. In round 2, your earnings depend on clicks and the round 2 target rate.
5. There is no round 3.

*How do you earn money in **round 1**?*

We take the total clicks you did in round 1, and divide by a target rate of 400**.**

You get $0.50 times the number of SPM units above 0.1

You do not earn for SPM units above 3.

*How do you earn money in **round 2**?*

We take the total clicks you did in round 2, and divide by a target rate of 400**.**

You get $0.50 times the number of SPM units above 0.1

You do not earn for SPM units above 3.

Please click "Next" to start with the task.

Period 4

How does it work?

1. You can do the clicking task again.
2. There will be 2 rounds of 1 minute and 30 seconds each with the same rules as before.
3. In round 1, your earnings depend on clicks and the round 1 target rate.
4. In round 2, your earnings depend on clicks and the round 2 target rate.
5. There is no round 3.

*<u>How do you earn money in **round 1**?</u>*

We take the total clicks you did in round 1, and divide by a target rate of 400**.**

You get $0.50 times the number of SPM units above 0.1

You do not earn for SPM units above 3.

*<u>How do you earn money in **round 2**?</u>*

We take the total clicks you did in round 2, and divide by a target rate of 400**.**

You get $0.50 times the number of SPM units above 0.1

You do not earn for SPM units above 3.

Please click "Next" to start with the task.

2. STATIC_ZERO

Period 1
How does it work?

5. In this part you can do a task,
5. Doing the task gives you Standard Productive Minutes (SPM)
5. You can earn real money.

What is the task?

The task is very simple: You can click a button on the screen.

The task lasts 1 minute and 30 seconds.

If you use a touch screen you can tap the button with your finger.

If you use a computer, you must use the **mouse or trackpad** to click the button. Using the keyboard is not allowed, and pressing keys can **take away** from your score.

How do you get SPM?

Clicking the button gets you Standard Productive Minutes (SPM).

The amount of clicks it takes to get SPM depends on the "**target rate**."

The target rate is 400.

We calculate SPM by dividing your total clicks by the target rate of 400.

Example:
You do 300 clicks
Total SPM is 300/400 = 0.75 SPM

How do you get money?

Once your SPM gets larger than 0.1, you start to earn money

You get $0.50 times the number of SPM units above 0.1.

You do not earn money for SPM units above 3.

Example:

Suppose you get 1 SPM.
This is 1-0.1 = 0.9 SPM units above 0.1 so you earn $0.50*0.9 = $0.45.

Please click "Next" to start with the task.

Period 2

How does it work?

1. You can do the clicking task again.
2. This time there will be 2 rounds of 1 minute and 30 seconds each.
3. In round 1, your earnings depend on clicks and the round 1 target rate.
4. In round 2, your earnings depend on clicks and the round 2 target rate.
5. There is no round 3.

*How do you earn money in **round 1**?*

We take the total clicks you did in round 1, and divide by a target rate of 400**.**

You get $0.50 times the number of SPM units above 0.1

You do not earn for SPM units above 3.

*How do you earn money in **round 2**?*

We take the total clicks you did in round 2, and divide by a target rate of 400**.**

You get $0.50 times the number of SPM units above 0.1

You do not earn for SPM units above 3.

Please click "Next" to start with the task.

Period 4

How does it work?

6. You can do the clicking task again.
6. There will be 2 rounds of 1 minute and 30 seconds each
6. This time there is no payment for clicking
6. In round 1, you earn nothing from clicking.
6. In round 2, you earn nothing from clicking.
6. There is no round 3.

*How do you earn money in **round 1**?*

You do not earn money for clicking in round 1.

*How do you earn money in **round 2**?*

You do not earn money for clicking in round 2.

Please click "Next" to start with the task.

3. COMPLEX

Period 1

How does it work?

1. In this part you can do a task,
2. Doing the task gives you Standard Productive Minutes (SPM)
3. You can earn real money.

What is the task?

The task is very simple: You can click a button on the screen.

The task lasts 1 minute and 30 seconds.

If you use a touch screen you can tap the button with your finger.

If you use a computer, you must use the **mouse or trackpad** to click the button. Using the keyboard is not allowed, and pressing keys can **take away** from your score.

How do you get SPM?

Clicking the button gets you Standard Productive Minutes (SPM).

The amount of clicks it takes to get SPM depends on the "**target rate**."

The target rate is 400.

We calculate SPM by dividing your total clicks by the target rate of 400.

Example:
You do 300 clicks
Total SPM is 300/400 = 0.75 SPM

How do you get money?

Once your SPM gets larger than 0.1, you start to earn money

You get $0.50 times the number of SPM units above 0.1.

You do not earn money for SPM units above 3.

Example:

Suppose you get 1 SPM.
This is 1-0.1 = 0.9 SPM units above 0.1 so you earn $0.50*0.9 = $0.45.

Please click "Next" to start with the task.

Period 2

How does it work?

1. You can do the clicking task again.
2. This time there will be 2 rounds of 1 minute and 30 seconds each.
3. In round 1, your earnings depend on clicks and the round 1 target rate.
4. In round 2, your earnings depend on clicks and the round 2 target rate.
5. **But there is also a special rule in round 2:**
    - The round 2 target rate depends partly on how many clicks you do in round 1.
    - It is higher if you do more clicks in round 1, and lower if you do fewer clicks in round 1.
6. There is no round 3.

*How do you earn money in **round 1**?*

We take the total clicks you did in round 1, and divide by a target rate of 400**.**

You get $0.50 times the number of SPM.

But, as before, you only earn for SPM units between 0.1 and 3

*How do you earn money in **round 2**?*

We take the total clicks you did in round 2, and divide by the round 2 target rate**.**

You get $0.50 times the number of SPM units above 0.1

But you only earn for SPM units between 0.1 and 3

How do we set the target rate in round 2?

We calculate the round 2 target rate as the **average** of:

The total clicks you do in round 1

and

A number X that is 380, 400, or 420, with equal chance of being 380, 400, or 420.

You do not find out the number X until the beginning of round 2.

Example 1:
You do 400 total clicks total in round 1; X turns out to be 400.
The new target rate for round 2 is (400+400)/2 = 400
In round 2 you will need 400 clicks to get 1 SPM.

Example 2:
You do 100 clicks in round 1; X turns out to be 400.
The new target is (100+400)/2 = 250
In round 2 you will need 250 clicks to get 1 SPM.

Please click "Next" to start with the task.

4. SIMPLE

Period 1

How does it work?

1. In this part you can do a task,
2. Doing the task gives you Standard Productive Minutes (SPM)
3. You can earn real money.

What is the task?

The task is very simple: You can click a button on the screen.

The task lasts 1 minute and 30 seconds.

If you use a touch screen you can tap the button with your finger.

If you use a computer, you must use the **mouse or trackpad** to click the button. Using the keyboard is not allowed, and pressing keys can **take away** from your score.

How do you get SPM?

Clicking the button gets you Standard Productive Minutes (SPM).

The amount of clicks it takes to get SPM depends on the "**target rate**."

The target rate is 400.

We calculate SPM by dividing your total clicks by the target rate of 400.

Example:
You do 300 clicks
Total SPM is 300/400 = 0.75 SPM

How do you get money?

Once your SPM gets larger than 0.1, you start to earn money

You get $0.50 times the number of SPM units above 0.1.

You do not earn money for SPM units above 3.

Example:

Suppose you get 1 SPM.
This is 1-0.1 = 0.9 SPM units above 0.1 so you earn $0.50*0.9 = $0.45.

Please click "Next" to start with the task.

Period 2

How does it work?

1. You can do the clicking task again.
2. This time there will be 2 rounds of 1 minute and 30 seconds each.
3. In round 1, your earnings depend on clicks and the round 1 target rate.
4. In round 2, your earnings depend on clicks and the round 2 target rate.
5. **But there is also a special rule in round 2:**
    - We **subtract** any money you earn in round 1 from your money in round 2.
    - If you earned more money in round 1 than round 2, we subtract the difference from your total earnings for the study.
6. There is no round 3.

How do you earn money in ***round 1****?*

We take the total clicks you did in round 1, and divide by a target rate of 400**.**

You get $0.50 times the number of SPM.

But, as before, you only earn for SPM units between 0.1 and 3

How do you earn money in ***round 2****?*

We take the total clicks you did in round 2, and divide by the round 2 target rate**.**

You get $0.50 times the number of SPM units above 0.1

But you only earn for SPM units between 0.1 and 3

Special rule for round 2

This is how we calculate your final earnings for round 2:

We take away any money you earned in round 1, from the money you earn in round 2

If you earned more money in round 1 than round 2, we take away the difference from your total earnings at the end of the study.

Example:
You did 440 clicks and earned $0.50 in round 1
You did 440 clicks and earned $0.50 in round 2
But, since you earned $0.50 in round 1, we take away $0.50 from round 2 and you end up with $0.00 for round 2.

Example:
You did 160 clicks and earned $0.15 in round 1
You did 440 clicks and earned $0.50 in round 2
Since you earned $0.15 in round 1, we take away $0.15 from round 2 and you end up with $0.35 for round 2.

Please click "Next" to start with the task.

5. NOISE

Period 1
How does it work?

11. In this part you can do a task,
11. Doing the task gives you Standard Productive Minutes (SPM)
11. You can earn real money.

What is the task?

The task is very simple: You can click a button on the screen.

The task lasts 1 minute and 30 seconds.

If you use a touch screen you can tap the button with your finger.

If you use a computer, you must use the **mouse or trackpad** to click the button. Using the keyboard is not allowed, and pressing keys can **take away** from your score.

How do you get SPM?

Clicking the button gets you Standard Productive Minutes (SPM).

The amount of clicks it takes to get SPM depends on the "**target rate**."

The target rate is 400.

We calculate SPM by dividing your total clicks by the target rate of 400.

Example:
You do 300 clicks
Total SPM is 300/400 = 0.75 SPM

How do you get money?

Once your SPM gets larger than 0.1, you start to earn money

You get $0.50 times the number of SPM units above 0.1.

You do not earn money for SPM units above 3.

Example:

Suppose you get 1 SPM.
This is 1-0.1 = 0.9 SPM units above 0.1 so you earn $0.50*0.9 = $0.45.

Please click "Next" to start with the task.

Period 2

How does it work?

11. You can do the clicking task again.
11. This time there will be 2 rounds of 1 minute and 30 seconds each.
11. In round 1, your earnings depend on clicks and the round 1 target rate.
11. In round 2, your earnings depend on clicks and the round 2 target rate.
11. **But there is also a special rule in round 2:**
    a. We **subtract** any money you earn in round 1 from your money in round 2.
    b. This means we subtract more in round 2 if your round 1 earnings are higher, we subtract less in round 2 if your round 1 earnings are lower.
11. There is no round 3.

How do you earn money in ***round 1****?*

We take the total clicks you did in round 1, and divide by a target rate of 400**.**

You get $0.50 times the number of SPM.

But, as before, you only earn for SPM units between 0.1 and 3

How do you earn money in ***round 2****?*

We take the total clicks you did in round 2, and divide by the round 2 target rate**.**

You get $0.50 times the number of SPM

But you only earn for SPM units between 0.1 and 3

Then we subtract any money you earned in round 1 from your money in round 2.

If earnings from round 1 are greater than earnings from round 2 then we subtract the difference from your total earnings for the study.

How do we set the target rate in round 2?

We calculate the round 2 target rate as the **average** of:

400

and

A number X that is 380, 400, or 420, with equal chance of being 380, 400, or 420.

You do not find out the number X until the beginning of round 2.

Example 1:
You do 350 clicks total in round 1 and earn $0.39
X turns out to be 400.
The target rate for round 2 is (400+400)/2 = 400
You do 400 clicks in round 2.
SPM from round 2 clicks are 400/400 = 1 so you earn $0.45
We subtract round 1 earnings of $0.39
Final earnings in round 2 are $0.06

Example 2:
You do 100 clicks in round 1 and earn $0.08
X turns out to be 420.
The target rate for round 2 is (400+420)/2 = 410
You do 451 clicks in round 2
SPM from round 2 clicks are 451/410 = 1.1 and you earn $0.50
We subtract round 1 earnings of $0.08
Final earnings in round 2 are $0.42

Please click "Next" to start with the task.

6. NOISE_MARGINAL

Period 1
How does it work?
1. In this part you can do a task,
2. Doing the task gives you Standard Productive Minutes (SPM)
3. You can earn real money.

What is the task?

The task is very simple: You can click a button on the screen.

The task lasts 1 minute and 30 seconds.

If you use a touch screen you can tap the button with your finger.

If you use a computer, you must use the **mouse or trackpad** to click the button. Using the keyboard is not allowed, and pressing keys can **take away** from your score.

How do you get SPM?

Clicking the button gets you Standard Productive Minutes (SPM).

The amount of clicks it takes to get SPM depends on the "**target rate**."

The target rate is 400.

We calculate SPM by dividing your total clicks by the target rate of 400.

Example:
You do 300 clicks
Total SPM is 300/400 = 0.75 SPM

How do you get money?

Once your SPM gets larger than 0.1, you start to earn money

You get $0.50 times the number of SPM units above 0.1.

You do not earn money for SPM units above 3.

Example:
Suppose you get 1 SPM.
This is 1-0.1 = 0.9 SPM units above 0.1 so you earn $0.50*0.9 = $0.45.

Please click "Next" to start with the task.

Period 2

How does it work?

4. You can do the clicking task again.
5. This time there will be 2 rounds of 1 minute and 30 seconds each.
6. In round 1, your earnings depend on clicks and the round 1 target rate.
7. In round 2, your earnings depend on clicks and the round 2 target rate.
8. **But there is also a special rule in round 2:**
    a. How much you earn per click in round 2 depends partly on how many clicks you do in round 1.
    b. You earn less per click in round 2 if you click more in round 1, and you earn more per click in round 2 if you click less in round 1.
9. There is no round 3.

*How do you earn money in **round 1**?*

We take the total clicks you did in round 1, and divide by a target rate of 400**.**

You get $0.50 times the number of SPM.

But, as before, you only earn for SPM units between 0.1 and 3

*How do you earn money in **round 2**?*

We take the total clicks you did in round 2, and divide by the round 2 target rate**.**

You get $0.50 times the number of SPM units above 0.1

But you only earn for SPM units between 0.1 and 3

How do we set the target rate in round 2?

We calculate the round 2 target rate as the **average** of:

The total clicks you do in round 1

and

A number X that is 380, 400, or 420, with equal chance of being 380, 400, or 420.

You do not find out the number X until the beginning of round 2.

If you do more clicks in round 1, this increases the round 2 target rate. This means you earn less SPM, and less money, per click in round 2.

Example 1:
You do 400 total clicks total in round 1; X turns out to be 400.
The new target rate for round 2 is (400+400)/2 = 400
For each 100 clicks you do (between 0.1 and 3 SPH) in round 2 you get $0.13.

Example 2:
You do 100 clicks in round 1; X turns out to be 400.
The new target rate for round 2 is (100+400)/2 = 250
For each 100 clicks you do (between 0.1 and 3 SPH) in round 2 you get $0.20.

Please click "Next" to start with the task.

7. LINEAR

Period 1

How does it work?

1. In this part you can do a task,
2. Doing the task gives you Standard Productive Minutes (SPM)
3. You can earn real money.

What is the task?

The task is very simple: You can click a button on the screen.

The task lasts 1 minute and 30 seconds.

If you use a touch screen you can tap the button with your finger.

If you use a computer, you must use the **mouse or trackpad** to click the button. Using the keyboard is not allowed, and pressing keys can **take away** from your score.

How do you get SPM?

Clicking the button gets you Standard Productive Minutes (SPM).

The amount of clicks it takes to get SPM depends on the "**target rate**."

The target rate is 400.

We calculate SPM by dividing your total clicks by the target rate of 400.

Example:
You do 300 clicks
Total SPM is 300/400 = 0.75 SPM

How do you get money?

You get $0.50 times the number of SPM

Example:
Suppose you get 0.9 SPM.
You earn $0.50*0.9 = $.45

Please click "Next" to start with the task.

Period 2

How does it work?

1. You can do the clicking task again.
2. This time there will be 2 rounds of 1 minute and 30 seconds each.
3. In round 1, your earnings depend on clicks and the round 1 target rate.

10. In round 2, your earnings depend on clicks and the round 2 target rate.
11. **But there is also a special rule in round 2:**
    a. The round 2 target rate depends partly on how many clicks you do in round 1.
    b. It is higher if you do more clicks in round 1, and lower if you do fewer clicks in round 1.
12. There is no round 3.

*How do you earn money in **round 1**?*

We take the total clicks you did in round 1, and divide by a target rate of 400**.**

You get $0.50 times the number of SPM.

*How do you earn money in **round 2**?*

We take the total clicks you did in round 2, and divide by the round 2 target rate**.**

You get $0.50 times the number of SPM

How do we set the target rate in round 2?

We calculate the round 2 target rate as the **average** of:

The total clicks you do in round 1

and

A number X that is 380, 400, or 420, with equal chance of being 380, 400, or 420.

You do not find out the number X until the beginning of round 2.

Example 1:
You do 400 total clicks total in round 1; X turns out to be 400.
The new target rate for round 2 is (400+400)/2 = 400
In round 2 you will need 400 clicks to get 1 SPM.

Example 2:
You do 100 clicks in round 1; X turns out to be 400.
The new target is (100+400)/2 = 250
In round 2 you will need 250 clicks to get 1 SPM.

Please click "Next" to start with the task.

8. NOSPM

Period 1

How does it work?

1. In this part you can do a task,
2. Doing the task gives you real money

What is the task?

The task is very simple: You can click a button on the screen.

The task lasts 1 minute and 30 seconds.

If you use a touch screen you can tap the button with your finger.

If you use a computer, you must use the **mouse or trackpad** to click the button. Using the keyboard is not allowed, and pressing keys can **take away** from your score.

How do you get money?

Once your clicks get larger than 40 clicks you start to earn money.

The **wage per click** is $0.00125 for clicks above 40. In other words, you get $0.12 for every 100 clicks above 40.

You do not earn money for clicks above 1200.

Example:
Suppose you do 400 clicks.
This is 400 - 40 = 360 clicks above 40 so you earn $0.00125*360 = $0.45.

Please click "Next" to start with the task.

Period 2

How does it work?

9. You can do the clicking task again.
9. This time there will be 2 rounds of 1 minute and 30 seconds each.
9. In round 1, your earnings depend on clicks and the round 1 wage per click.
9. In round 2, your earnings depend on clicks and the round 2 wage per click.
9. **But there is also a special rule in round 2:**

- The round 2 wage per click depends partly on how many clicks you do in round 1.
- It is lower if you click more in round 1, and higher if you click less in round 1.

9. There is no round 3.

*<u>How do you earn money in **round 1**?</u>*

We take the total clicks you did in round 1 and multiply by the round 1 wage per click of $0.00125**.**

But, as before, you only earn for clicks between 40 and 1200

*<u>How do you earn money in **round 2**?</u>*

We take the total clicks you did in round 2, and multiply by the round 2 wage per click

But, as before, you only earn for clicks above a certain level and below a certain level.

<u>How do we set the wage per click in round 2?</u>

We calculate the round 2 wage per click as $1 divided by the **sum of**:

The total clicks you do in round 1

and

A number X that is 380, 400, or 420, with equal chance of being 380, 400, or 420.

You do not find out the number X until the beginning of round 2.

In other words, the round 2 wage per click is given by the formula:

Wage per click in round 2 = $1 / (X + Clicks in round 1)

The wage per click in round 2 is smaller, if you do more clicks in round 1.

You are only paid, however, for clicks above a certain level, **Y**, and below a certain level, **Z**. Both Y and Z depend on how many clicks you do in round 1.

Y is given by the formula:

Y = (X + Clicks in round 1) / 20.

Z is given by the formula:

Z = (X + Clicks in round 1) x 1.5.

Y and Z are both larger if you do more clicks in round 1.

Example 1:
You do 400 total clicks total in round 1; X turns out to be 400.
Y is (400 + 400) / 20 = 40.
The round 2 wage per click is $1 / (400 + 400) = $0.00125.
In other words, for each 80 clicks you do above 40 in round 2 you get $0.10.
Z is (400 + 400) x 1.5 = 1200 so you do not earn for clicks above 1200.

Example 2:
You do 100 clicks in round 1; X turns out to be 400.
Y is (400 + 100) / 20 = 25.
The round 2 wage per click is $1 / (400 + 100) = $0.002.
In other words, for each 80 clicks you do above 25 in round 2 you get $0.16.
Z is (400 + 100) x 1.5 = 750 so you do not earn for clicks above 750.

Please click "Next" to start with the task.

9. LINEAR_NOSPM

Period 1
How does it work?

8. In this part you can do a task,
8. Doing the task gives you real money

What is the task?

The task is very simple: You can click a button on the screen.

The task lasts 1 minute and 30 seconds.

If you use a touch screen you can tap the button with your finger.

If you use a computer, you must use the **mouse or trackpad** to click the button. Using the keyboard is not allowed, and pressing keys can **take away** from your score.

How do you get money?

The **wage per click** is $0.00125 . In other words, you get $0.25 for every 200 clicks.

Example:
Suppose you do 360 clicks.
So you earn $0.00125*360 = $0.45.

Please click "Next" to start with the task.

Period 2

How does it work?

8. You can do the clicking task again.
8. This time there will be 2 rounds of 1 minute and 30 seconds each.
8. In round 1, your earnings depend on clicks and the round 1 wage per click.
8. In round 2, your earnings depend on clicks and the round 2 wage per click.
8. **But there is also a special rule in round 2:**
   a. The round 2 wage per click depends partly on how many clicks you do in round 1.
   b. It is lower if you click more in round 1, and higher if you click less in round 1.
8. There is no round 3.

*How do you earn money in **round 1**?*

We take the total clicks you did in round 1, and multiply by the round 1 wage per click of $0.00125**.**

*How do you earn money in **round 2**?*

We take the total clicks you did in round 2, and multiply by the round 2 wage per click

How do we set the round 2 wage per click?

We calculate the round 2 wage per click as $0.50 divided by the **average of:**

The total clicks you do in round 1

and

A number X that is 380, 400, or 420, with equal chance of being 380, 400, or 420.

You do not find out the number X until the beginning of round 2.

In other words, the round 2 wage per click is given by the formula:

Wage per click in round 2 = $1 / (X + Clicks in round 1)

Example 1:
You do 400 total clicks total in round 1; X turns out to be 400.
In round 2 the wage per click is $1 / (400 + 400) = $0.00125.
In other words, doing 80 clicks in round 2 gives you $0.10.

Example 2:
You do 100 clicks in round 1; X turns out to be 400.
In round 2 the wage per click is $1 / (400 + 100) = $0.002.
In other words, doing 80 clicks in round 2 gives you $0.16.

Please click "Next" to start with the task.

10. STATIC_LOW

Period 1

<u>How does it work?</u>

1. In this part you can do a task,
2. Doing the task gives you Standard Productive Minutes (SPM)
3. You can earn real money.

<u>What is the task?</u>

The task is very simple: You can click a button on the screen.

The task lasts 1 minute and 30 seconds.

If you use a touch screen you can tap the button with your finger.

If you use a computer, you must use the **mouse or trackpad** to click the button. Using the keyboard is not allowed, and pressing keys can **take away** from your score.

<u>How do you get SPM?</u>

Clicking the button gets you Standard Productive Minutes (SPM).

The amount of clicks it takes to get SPM depends on the "**target rate**."

The target rate is 400.

We calculate SPM by dividing your total clicks by the target rate of 400.

Example:
You do 300 clicks
Total SPM is 300/400 = 0.75 SPM

<u>How do you get money?</u>

Once your SPM gets larger than 0.1, you start to earn money

You get $0.50 times the number of SPM units above 0.1.

You do not earn money for SPM units above 3.

Example:
Suppose you get 1 SPM.
This is 1-0.1 = 0.9 SPM units above 0.1 so you earn $0.50*0.9 = $0.45.

Please click "Next" to start with the task.

Period 2

How does it work?

1. You can do the clicking task again.
2. This time there will be 2 rounds of 1 minute and 30 seconds each.
3. In round 1, your earnings depend on clicks and the round 1 target rate.
4. In round 2, your earnings depend on clicks and the round 2 target rate.
5. There is no round 3.

*How do you earn money in **round 1**?*

We take the total clicks you did in round 1, and divide by a target rate of 400**.**

You get $0.50 times the number of SPM units above 0.1

You do not earn for SPM units above 3.

*How do you earn money in **round 2**?*

We take the total clicks you did in round 2, and divide by a target rate of 400**.**

You get $0.50 times the number of SPM units above 0.1

You do not earn for SPM units above 3.

Please click "Next" to start with the task.

Period 4

How does it work?

1. You can do the clicking task again.
2. This time there will be 2 rounds of 1 minute and 30 seconds each.
3. **This time you get only 1 cent per SPM instead of 50 cents.**

4. In round 1, your earnings depend on clicks and the round 1 target rate.
5. In round 2, your earnings depend on clicks and the round 2 target rate.
6. There is no round 3.

*How do you earn money in **round 1**?*

We take the total clicks you did in round 1, and divide by a target rate of 400**.**

You get $0.01 times the number of SPM units above 0.1

You do not earn for SPM units above 3.

*How do you earn money in **round 2**?*

We take the total clicks you did in round 2, and divide by a target rate of 400**.**

You get $0.01 times the number of SPM units above 0.1

You do not earn for SPM units above 3.

Please click "Next" to start with the task.

11. SIMPLE_NOLOSS

Period 1

How does it work?

1. In this part you can do a task,
2. Doing the task gives you Standard Productive Minutes (SPM)
3. You can earn real money.

What is the task?

The task is very simple: You can click a button on the screen.

The task lasts 1 minute and 30 seconds.

If you use a touch screen you can tap the button with your finger.

If you use a computer, you must use the **mouse or trackpad** to click the button. Using the keyboard is not allowed, and pressing keys can **take away** from your score.

How do you get SPM?

Clicking the button gets you Standard Productive Minutes (SPM).

The amount of clicks it takes to get SPM depends on the "**target rate**."

The target rate is 400.

We calculate SPM by dividing your total clicks by the target rate of 400.

Example:
You do 300 clicks
Total SPM is 300/400 = 0.75 SPM

How do you get money?

Once your SPM gets larger than 0.1, you start to earn money

You get $0.50 times the number of SPM units above 0.1.

You do not earn money for SPM units above 3.

Example:

Suppose you get 1 SPM.
This is 1-0.1 = 0.9 SPM units above 0.1 so you earn $0.50*0.9 = $0.45.

Please click "Next" to start with the task.

Period 2

<u>*How does it work?*</u>

4. You can do the clicking task again.
5. This time there will be 2 rounds of 1 minute and 30 seconds each.
6. In round 1, your earnings depend on clicks and the round 1 target rate.
7. In round 2, your earnings depend on clicks and the round 2 target rate.
8. **But there is also a special rule in round 2:**
   a. We **subtract** any money you earn in round 1 from your money in round 2.
   b. This means we subtract more in round 2 if your round 1 earnings are higher, we subtract less in round 2 if your round 1 earnings are lower.
9. There is no round 3.

<u>*How do you earn money in **round 1**?*</u>

We take the total clicks you did in round 1, and divide by a target rate of 400**.**

You get $0.50 times the number of SPM.

But, as before, you only earn for SPM units between 0.1 and 3

<u>*How do you earn money in **round 2**?*</u>

We take the total clicks you did in round 2, and divide by a target rate of 380**.**

You get $0.50 times the number of SPM

But you only earn for SPM units between 0.1 and 3

Then we subtract any money you earned in round 1 from your money in round 2.

You cannot have negative earnings for round 2.

Example 1:
You do 350 clicks total in round 1 and earn $0.39
You do 400 clicks in round 2 and earn $0.48
We subtract round 1 earnings of $0.39

Final earnings in round 2 are $0.09

Example 2:
You do 100 clicks in round 1 and earn $0.08
You do 451 clicks in round 2 and earn $0.54
We subtract round 1 earnings of $0.08
Final earnings in round 2 are $0.46

Please click "Next" to start with the task.