Incentive Complexity, Bounded Rationality and Effort Provision[*]

Johannes Abeler, David Huffman, Collin Raymond

May 16, 2025

Abstract

Using field and laboratory experiments, we demonstrate that the complexity of incentive schemes and worker bounded rationality can affect effort provision. This is because some attributes of the incentives become opaque, i.e., workers do not take them into account. In our setting, workers over-provide effort relative to a fully rational benchmark, improving efficiency. We identify contract features, and facets of worker cognitive ability, that matter for opacity. We find that even relatively small degrees of opacity can cause large shifts in behavior. Our results illustrate important implications of complexity and bounded rationality for designing and regulating workplace incentive contracts.

JEL Codes: D8, D9, J2, J3

Keywords: Complexity; Bounded Rationality; Ratchet effect; Dynamic Incentives; Field Experiments

1 Introduction

Contract theory has traditionally assumed that workers are fully rational and take into account all nuances of the incentives they face. Although there is a large empirical literature showing that workers react to incentives (e.g., Lazear (2000)) it is much less clear how varying degrees of bounded rationality on the part of workers may mediate such responses. This raises the possibility that not all workers will be equally aware of their incentives, and that varying the complexity of workplace incentive schemes might matter for effort provision. Recent theoretical advances have modeled (e.g., Gabaix and Laibson (2006)), and empirical has work documented (e.g., Chetty, Looney, and Kroft (2009); Abaluck and Gruber (2011); Grubb (2015)), how the fact that certain contract features are opaque (i.e. are not taken into account) may impact the response of individuals to contracts.[1]

In this paper we extend research on opacity to the setting of labor incentive contracts and answer two main research questions. First, can complexity and worker bounded rationality cause certain aspects of workplace incentives to be opaque, thereby affecting effort provision, efficiency, firm profit and worker utility? Our focus in answering this question is on whether dynamic incentives in the form of ratchet effects can be opaque. Second, we ask: What specific contract features, and facets of worker cognitive ability, matter for opacity?

Our study involves three complementary datasets. The first consists of personnel data from a firm, which employs workers to fulfill customer orders in a large warehouse. The firm agreed to alter its business practices to randomly assign some workers to have an incentive scheme that was potentially advantageous to the firm, but also had a perverse dynamic incentive to reduce effort. We conducted two large-scale field experiments to test whether these dynamic incentives are opaque to workers in a real work context over substantial periods of time. The second dataset is from online lab experiments, conducted with the same worker population, which provide more tightly controlled evidence on mechanisms. A third dataset is from experiments conducted with Amazon Mechanical Turk (AMT) workers. These allow us to replicate our results with another worker population, and extend our understanding of

---

[1] We use "complexity" to refer to a property of the contract; "bounded rationality" to denote a characteristic of the decision maker; and "opacity" for the combined effect of complexity and bounded rationality, namely that a decision maker does not take all aspects of the contract into account.

mechanisms.

The analysis yields four main sets of results. (1) Across our field and online experiments, the complexity of workplace incentives matters for effort provision, even in the presence of feedback and a substantial period of time for learning. (2) In the specific situation we consider, complexity actually increases effort by making dynamic incentives opaque. A welfare analysis based on a calibrated model of effort provision shows that, in our case, complexity increases efficiency and can benefit both workers and the firm in equilibrium. Our model estimates also show that, given sufficient nonlinearity of utility, even if feedback or learning leads to close to full transparency of dynamic incentives, there can remain large differences in behavior relative to the optimum. (3) A measure of cognitive ability, the Cognitive Reflection Test (CRT), predicts worker response to complex incentives, implying that CRT levels in the workforce will matter for whether incentives are opaque. (4) By exogenously varying different aspects of the contracts we identify key contractual features that matter for whether workers neglect dynamic incentives.

To derive these results, we study worker behavior in a widely-used class of dynamic incentive schemes that induce workers to reduce effort in the form of the so-called "ratchet effect" (e.g., Weitzman (1980); Laffont and Tirole (1988)) – where worker's productivity in future periods is normalized by past productivity, giving rational workers a motive to work slower, reducing efficiency.

We test the hypothesis that this type of dynamic incentives could be opaque, and explore mechanisms. Our test is based on whether workers act as though dynamic incentives are absent (full opacity), or in line with the rational benchmark (full transparency), or in-between. Regarding mechanisms, we hypothesize, and test, whether several specific features of the contract may be important contributors to opacity.[2] First, the framing of dynamic incentives in terms of target rates rather than money may make the monetary consequences implicit and thus less transparent; second, the fact that the incentives induce an impact of current effort on future marginal incentives, rather than just affecting the level of future

---

[2]There are various ideas of how to measure contract complexity and opacity (e.g., Oprea (2020)). Our metric is inspired by the recent literature on attention, see Section 2 for details. Although we identify features that lead to opacity, we do not take a specific stand on the precise psychological process that leads to opacity (e.g., not noticing a feature, or refusing to reason through it) — see Handel and Schwartzstein (2018) for a recent discussion.

earnings, may make the optimization problem relatively demanding in terms of need for contingent thinking and thus more opaque; and third, noise in parameters of such schemes, arising from how they are implemented in practice, may act as a distraction and contribute to opacity.[3] We also hypothesize that workers who are more boundedly rational, as proxied by lower cognitive ability, will be more likely to fail to take dynamic incentives into account, holding contract complexity constant. As cognitive ability is multidimensional, we measure several different facets.

In Section 2, we introduce a simple two-period model of effort provision. The model illustrates how the ratchet effect arises in our setting with fully rational workers. We also propose metrics for how much a boundedly rational worker mis-perceives dynamic incentives, e.g., a parameter, ψ, which ranges from 0 for complete opacity to 1 for complete transparency.

Section 3 describes our main field experiment, denoted INDIVIDUAL, which was designed to test whether workers respond rationally to dynamic incentives in a field setting. This experiment randomized new hires into treatment and control ($N = 1,294$). Treatment workers faced dynamic incentives to reduce effort. We find, however, only a very small and non-significant reduction in effort of 0.1 percent.[4] In contrast, we find a large increase in effort (12.5 percent) when the firm moved from paying only hourly wages to also paying incentives using fixed target rates. Using this substantial response, and another moment from our data, we calibrate a model of effort provision under the assumption that the contract is fully transparent, and find that the model predicts a much larger response to dynamic incentives than we observe. The model is only set-identified, due to three parameters but only two moments, but the model always predicts that rational workers reduce effort by at least 5.5 percent, and for some parameter values much more. Our metrics of mis-perception imply almost complete opacity of dynamic incentives (e.g., $\psi = 0$), but our model also shows that even if ψ were, e.g., 0.9 due to learning, the model still predicts large deviations from the rational optimum. The model thus offers an explanation for persistent behavioral effects of

---

[3]These hypotheses are inspired by previous findings in psychology and behavioral economics: Various ways in which costs or prices are implicit has been found to cause them to pass unnoticed (e.g., J. Brown, Hossain, and Morgan (2010)); people have trouble with relatively complicated contingent thinking in abstract settings (e.g., Esponda and Vespa (2019); and extraneous information can distract individuals who have limited attention (e.g., Hirshleifer, Lim, and Teoh (2009).

[4]We measure effort response by difference in worker output, controlling in various ways for other factors that could affect output besides effort. We discuss this in detail in Section 3.

opacity. The observed deviation from the model's optimum implies non-trivial utility losses for the workers. A detailed welfare analysis shows that the opacity of dynamic incentives implies higher overall surplus, higher firm profits, and, depending on parameters, higher worker utility, compared to a benchmark in which dynamic incentives are transparent and the firm optimizes the contract in equilibrium.

We conduct various robustness checks, including running a second field experiment, GROUP ($N = 1,447$), which allows for more learning opportunities because it takes place over the course of 10 months and incorporates potential motives for social pressure and learning. We again find that on average, treatment worker behavior is similar to behavior of control workers who face zero dynamic incentives. The response is somewhat larger in magnitude than in INDIVIDUAL, a reduction of 1.0 percent, but there is no significant time trend. Thus, the effect of opacity on behavior appears robust to learning opportunities.

Section 4 turns to our second data set, using online experiments with the same warehouse workers, which were designed to shed more light on underlying mechanisms. Participants work on a real-effort task. Workers were randomly assigned to treatments, some of which vary the way that the incentives are framed or structured to reduce complexity. This controlled setting rules out, by design, motives that might work against a response to dynamic incentives in the field, such as concerns about dismissal. In our main treatment, denoted COMPLEX, workers faced the same type of incentive scheme that was implemented for treatment workers in the INDIVIDUAL field experiment. In another treatment, SIMPLE, we made changes to eliminate the three contract factors we hypothesized as potentially contributing to opacity: The impact of current effort on future earnings was explained explicitly in terms of money rather than in terms of target rates; the structure of dynamic incentives was such that earnings in the current period were subtracted from the level of earnings in the future, rather than influencing future marginal incentives; and the future target rate was known with certainty so there was no noise in parameters. To capture cognitive ability we measured CRT, as well as tendency to bracket decisions broadly, and backwards induction ability, and asked what workers thought would be the optimal response to the incentives.

We find strong evidence that workers do not fully react to dynamic incentives because of complexity and worker bounded rationality. Workers respond little to dynamic incentives on

average in COMPLEX, despite feedback, but show a stronger reaction in SIMPLE. At the individual level, about twice as many workers show signs of a ratchet effect in SIMPLE compared to COMPLEX. Calibrating our (set-identified) representative-agent model of rational effort provision, using two moments of the data from the online experiment, we find that the average effort we observe in COMPLEX is substantially higher than the optimal effort, for all free parameter values. The observed average effort is, however, very close to the optimum in SIMPLE. Our mis-perception metrics also imply more opacity in COMPLEX, and workers also mention dynamic incentives less often in SIMPLE than COMPLEX. We find that lower scores on the CRT are associated with a weaker response to dynamic incentives and lower probability of mentioning dynamic incentives, in both COMPLEX and SIMPLE.

Since the same warehouse workers participated in the GROUP field experiment and the online experiment, we can link these two data sets. We find that workers who recognize dynamic incentives online also respond significantly more in the field experiment, suggesting that the same mechanisms play a role.

Section 5 discusses results from our third dataset, consisting of online experiments with AMT workers. The first set of experiments shows that the results found with warehouse workers replicate with AMT workers, although AMT workers do show a significantly stronger response to dynamic incentives, largely explainable by their higher average CRT score. For the AMT workers we have an additional treatment to point-identify and estimate our structural model. Observed average behavior of AMT workers is more than 400 clicks away from the optimum in COMPLEX, while it is only about 5 clicks away from the optimum in SIMPLE. A second set of experiments tests which of the three features of COMPLEX, which were eliminated in SIMPLE, contribute to opacity, and finds that each of these matters. A third set of experiments identifies contract features that can be changed without influencing opacity: nonlinearity of the piece rate schedule, and framing incentives in terms of target rate.

Section 6 concludes with a discussion and suggestions for future research.

Our findings contribute to the literature on dynamic incentives and specifically ratchet effects (e.g., Weitzman (1980); Laffont and Tirole (1988)).[5] There are anecdotal accounts

---

[5]This intuition has been extended to many other settings (e.g., Dillen and Lundholm (1996); Hart and

and case studies about ratchet effects from piece rate jobs in the late 1800's and early 1900's (see, e.g., Mathewson (1931); Clawson (1980); Roy (1952)). More recently, studies using observational data have found evidence interpreted as indicating ratchet effects, including historical data on serfs in Russia (Markevich and Zhuravskaya (2018)), teachers (Macartney (2016)), workers in sales occupations (e.g., Misra and Nair (2011)), and tree planters (Shearer (2022)). A literature from laboratory experiments, however, finds more mixed results: Chaudhuri (1998) finds no ratchet effect, while Cooper et al. (1999), Charness, Kuhn, and Villeval (2011) and Cardella and Depew (2018) do. Evidence is also mixed in the broader literature on whether consumers respond to various types of dynamic incentives in the field, e.g., in health insurance contracts some studies find a weak response (Keeler and Rolph (1988); Dalton, Gowrisankaran, and Town (2020)), but others find close to full understanding (Chevalier and Goolsbee (2009); Aron-Dine et al. (2015); Einav, Finkelstein, and Schrimpf (2015)). One challenge to explaining such mixed evidence is that there is little evidence on the mechanisms generating opacity (Chandra, Handel, and Schwartzstein (2019)). We contribute to this literature by (1) testing the response to dynamic incentives in the field, using large-scale RCTs, and (2) providing novel evidence on mechanisms – specific features of incentive structure and framing, and facets of cognitive ability– that predict when this classic incentive problem will be weak or strong.

Our findings also contribute to a literature documenting that important aspects of several different decision environments can be opaque, e.g., sales taxes when shopping, or out-of-pocket costs of health insurance.[6] (1) Our study is the first to extend research on opacity to the environment of workplace incentive systems, showing that complexity and bounded rationality are important factors left out of standard theories of optimal incentives. (2) In line with a key insight of papers like Chetty, Looney, and Kroft (2009), we show that opacity can be efficiency enhancing, but find that this can even extend to Pareto improvements that make both workers and firms better off in equilibrium, i.e., compared to a benchmark

---

Tirole (1988)), and remains an active area of research (e.g., Malcomson (2016); Gerardi and Maestri (2020)).

[6]Recent papers document unawareness or mis-estimation of aspects of taxation and government fees (e.g., Chetty, Looney, and Kroft (2009); Taubinsky and Rees-Jones (2018)), consumer contracts (e.g., Grubb (2015); J. Brown, Hossain, and Morgan (2010)), banking services (e.g., Alan et al. (2018)), and health insurance (for a survey see Chandra, Handel, and Schwartzstein (2019); see also Heiss et al. (2021); Z. Brown and Jeon (2023)). A related literature studies determinants of complexity in abstract laboratory settings (e.g., Herrnstein et al. (1993); Martínez-Marquina, Niederle, and Vespa (2019); Oprea (2020)).

in which dynamic incentives are transparent and the firm adjusts the structure of contract optimally.[7] (3) Unlike most evidence on opacity, we have exogenous variation in incentive structure, and micro-level measures of decision maker bounded rationality, providing evidence about the roles of both of these in creating opacity. (4) Our findings about aspects of workplace incentives that contribute to opacity have potential broader implications, suggesting analogous features of consumer contracts that may matter for complexity. (5) Our results suggest that CRT may be a trait that matters for heterogeneity in opacity in other decision environments as well.[8] (6) We provide converging evidence with other studies that find behavioral effects of opacity in settings with learning, such as taxation or cellular phone usage (Chetty, Looney, and Kroft (2009); Grubb and Osborne (2015)), but we are novel in offering an explanation for persistence. In our setting, the interaction of convex effort costs with mis-perception can lead worker behavior to deviate strongly from the optimum, even if perceptions are close to full transparency.

2   Theoretical framework

Laffont and Tirole (1988) define the ratchet effect as arising when high effort today will make it more difficult to earn money in the future. In other words, when earnings in Period $t+1$ are g_{t+1} and effort in Period t is e_t, then the ratchet effect is that $\frac{\partial g_{t+1}}{\partial e_t} < 0$. In slight abuse of nomenclature we will often refer to both the effect of effort today on incentives tomorrow, as well as the induced behavioral response on effort today (a reduction in effort), as the ratchet effect. We will distinguish *static incentives*, where effort in period t only impacts bonuses in t, from *dynamic incentives*, where effort in period t may impact future bonuses.

Our model has two goals: we want to show that there should be a ratchet effect in the

---

[7]It has been shown that inertia or inattention to state of the world can benefit consumers (e.g., Handel (2013)). We show this can be the case for a distinct bias, opacity of dynamic incentives. Most studies on complex contracts conclude that opacity leads to inefficiency and exploitation of agents (e.g., Gabaix and Laibson (2006); Heidhues and Kőszegi (2010); Eliaz and Spiegler (2006)), implying that regulation should reduce opacity. Our findings suggest that in other cases a more nuanced regulatory approach may be needed.

[8]We find that opacity is pronounced for workers with a CRT score below a certain threshold, and in representative population samples about 64 percent have scores in this range (Brañas-Garza, Kujal, and Lenkei (2019)). Although heterogeneity in understanding has been shown to have important implications, e.g., for tax efficiency (Taubinsky and Rees-Jones (2018)), our findings document a measure of cognitive ability that can be used to predict such heterogeneity.

setting we study, if dynamic incentives are transparent. And we want to calibrate the model to the data to be able to make counterfactual predictions and to compute welfare. We will initially assume workers are perfectly rational (in line with our empirical null hypothesis), and later allow for boundedly rational understanding of contracts.

The model assumes two time periods, $t = 1, 2$, and in each period, the worker decides how much effort to exert. Effort affects utility in three ways: the worker pays a cost of effort, they receive non-monetary benefits, and they receive monetary benefits.

We suppose that there is a single worker, with a type θ_i, where i indexes a type drawn i.i.d. from cdf H. A higher type implies a higher marginal cost of effort.[9] The individual faces an effort cost $c(e_{i,t}, \theta_i)$. c is strictly increasing and convex in the first argument, differentiable, the cross partial is positive (so that higher types have higher marginal costs), $\frac{\partial c(0,\theta_i)}{\partial e_{i,t}} = 0$ and c grows without bound as $e_{i,t} \to \infty$.

Non-monetary benefits of effort in each period are $ae_{i,t}$, where $a \geq 0$. This can reflect intrinsic motivation, altruism by the worker towards the firm, or other non-monetary concerns, like concerns regarding dismissal.[10]

In each period the individual also receives monetary benefits of effort, g_t, in addition to a base income o. The incentive function g_t is the key part of the utility function, since it is through this that the ratchet effect operates. In Period 1, the incentive is simply a function of $e_{i,1}$: $g_1(e_{i,1})$ (unlike many papers in contract theory, we assume that effort is directly observable and contractible, and so given an effort level, there is no uncertainty about the incentive payment). In Period 2, the bonus can be a function of effort in both periods: $g_2(e_{i,1}, e_{i,2})$. This general form captures many incentive schemes that induce the ratchet effect, such as Weitzman (1980) (where past effort enters linearly into today's payoff) and Laffont and Tirole (1988) (where the incentives are determined as part of a Perfect Bayes Nash Equilibrium).[11]

---

[9]We include subscript i as a reminder that workers can have different types, which provides a rationale for the firm to adjust incentives based on worker behavior, as in Laffont and Tirole (1988). Indexing worker type is not necessary for the theoretical results in this section, as we take incentives as given.

[10]Although we model non-monetary motives as linear in effort, in line with the model we calibrate later, our theoretical results extend to more general concave functions of effort.

[11]In this section, we don't model *why* the firm would want an incentive system with dynamic incentives and rather explore what happens once workers face dynamic incentives. In the previous theoretical literature on the ratchet effect, e.g. Laffont and Tirole (1988), the firm's motive is that workers have private information

The specific form of g_t is modelled after the incentive contract implemented by the firm, which is an example of a widely used class of incentive schemes known as a "standard-hour plan." In standard-hour plans, bonuses are a function of "normalized effort". Specifically, Period-t effort is normalized using a standard rate of speed, denoted $\eta_{i,t}$, which is often referred to as a "target rate." Normalized effort is thus $\frac{e_{i,t}}{\eta_{i,t}}$ and is also referred to as Standard Productive Hours (SPH). It is converted into money using a bonus function $g_t(\frac{e_{i,t}}{\eta_{i,t}}, w) = w\hat{g}(\frac{e_{i,t}}{\eta_{i,t}})$, where $w > 0$ is the base marginal benefit, and $\hat{g}(\frac{e_{i,t}}{\eta_{i,t}})$ has a piecewise-linear form (see below). Dynamic incentives can arise in this class of incentives if Period-1 effort affects Period-2 target rates (i.e., $\eta_{i,2}$). $\eta_{i,1}$ is an exogenous number set by the firm, but the Period-2 target rate is $\eta_{i,2} = \varsigma e_{i,1} + (1 - \varsigma)r_1$, with $\varsigma \in [0, 1)$ (in our field experiment $\varsigma = \frac{1}{2}$), where r_1 is exogenously set by the firm. Thus, if $\varsigma > 0$, the Period-2 target rate depends on Period-1 effort. Such a scheme is similar to the original scheme considered in Weitzman (1980) but uses ratios of effort, rather than differences.

$\hat{g}$ has a piecewise-linear form. Such a form is commonly used and is known as a "cap-and-quota" form with quota $\underline{E}$ and cap $\bar{E}$:

$$
\hat{g}\left(\frac{e_{i,t}}{\eta_{i,t}}\right) = \begin{cases} 0 & \text{if } \frac{e_{i,t}}{\eta_{i,t}} \leq \bar{E} \\ [\frac{e_{i,t}}{\eta_{i,t}} - \underline{E}] & \text{if } \underline{E} \leq \frac{e_{i,t}}{\eta_{i,t}} \leq \bar{E} \\ [\bar{E} - \underline{E}] & \text{if } \frac{e_{i,t}}{\eta_{i,t}} \geq \bar{E} \end{cases}
$$

This means that marginal incentives are positive and constant if normalized effort is between quota $\underline{E}$ and cap $\bar{E}$. Elsewhere, marginal incentives are zero. We make no claims about the optimality for our particular firm, although researchers have found evidence of the usage of similar, nonlinear schemes across firms.[12]

---

about task difficulty. In Section 3.5, we add the firm to the model.

[12]In a survey of Fortune 500 firms, Joseph and Kalwani (1992) report that the vast majority of firms have incentive schemes featuring bonuses, which are often combined with commissions, implying kinks and non-linearities in compensation. Regarding the fact that workers earn no bonus for an effort below the quota, Misra and Nair (2011) note that "such quotas are ubiquitous in sales-force compensation and have been justified in the theory literature as a trade-off between the optimal provision of incentives versus the cost of implementing more complicated schemes (Raju and Srinivasan 1996), or as optimal under specific assumptions on agent preferences and the distribution of demand (Oyer 2000)." They go on to note that the cap in such schemes can be rationalized as a way to reduce potential windfall compensation.

Individuals discount the future by δ. Thus, total utility is

$$U_i = o + w\hat{g}\left(\frac{e_{i,1}}{\eta_{i,1}}\right) - c(e_{i,1}, \theta_i) + ae_{i,1} + \delta\left(o + w\hat{g}\left(\frac{e_{i,2}}{\varsigma e_{i,1} + (1-\varsigma)r_1}\right) - c(e_{i,2}, \theta_i) + ae_{i,2}\right)$$
(2.1)

The proposition shows that a ratchet effect arises when $\varsigma > 0$.

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

The proof is in Appendix A.1. The intuition for the proposition is fairly simple. When $\varsigma > 0$, the first-order condition with respect to e_1 has an extra negative term, which captures that working hard in Period 1 reduces pay in Period 2, all else being equal. Marginal costs have risen for all e_1, while marginal benefits have stayed the same, leading to a decrease in the optimal e_1.[13] As can be shown via counter-examples, the more general result that Period-1 effort is always decreasing in ς, does not hold.[14]

We would also like to have a formal definition, and measure, of the degree of opacity (or transparency) that a particular contract induces in a particular worker population. Let $\mathbb{P}$ be the optimization problem induced by the dynamic contract, with optimal effort in Period t as $e_t(\mathbb{P})$, while $\tilde{\mathbb{P}}$ is the optimization problem that occurs if the agent ignores the impact of changes in effort today on payment tomorrow (i.e., considers the static contract equivalent to $\mathbb{P}$) with optimal effort $e_t(\tilde{\mathbb{P}}) \geq e_t(\mathbb{P})$. Denoting observed effort as e_t^o, we can think of the degree of transparency as a function of $e_1(\mathbb{P})$, $e_1(\tilde{\mathbb{P}})$ and e_1^o, which is 1 if $e_1^o \leq e_1(\mathbb{P})$, 0 if $e_1^o \geq e_1(\tilde{\mathbb{P}})$, and decreasing in e_1^o. One simple example would be the ratio $\frac{e_1(\tilde{\mathbb{P}}) - e_1^o}{e_1(\tilde{\mathbb{P}}) - e_1(\mathbb{P})}$, for $e_1^o \in [e_1(\mathbb{P}), e_1(\tilde{\mathbb{P}})]$ so long as $e_1(\tilde{\mathbb{P}}) \neq e_1(\mathbb{P})$. This is the distance between the opaque

---

[13]The proof uses monotone comparative statics results and does not rely on first-order conditions. The model discussed here makes several simplifying assumptions relative to the reality that the workers in our field experiment faced. First, we assume that effort costs are additively separable across time periods, which we relax in Appendix A.4.1. Second, we assume that there is a single task with a single target rate in each period. In Appendix A.4.2, we generalize our model to allow for many tasks with an exogenous allocation of hours across them (in line with our empirical setup), where each task has a separate task-specific target rate. In Appendices A.2 and A.3, we also analyze two alternative contracts, one used in our GROUP field experiment (where $\eta_{i,t+1}$ is the average effort of several workers in t) and one used in our online "SIMPLE" treatment and show these also generate ratchet effects.

[14]Proposition 1 will also go through if we assume wages are a function not of effort, but rather of output, where output is effort plus mean-zero noise.

optimum and observed effort, normalized by the distance between the two optima. We will refer to this measure as the "ratio measure".

We can also measure the degree of transparency using a model-driven approach, inspired by the inattention framework developed in Gabaix (2019), in which individuals misperceive marginal incentives. We assume that individuals underestimate the marginal impact of their effort on their payments tomorrow. In particular, consider the first-order condition for an agent with respect to effort in Period 1 for the contract in Equation 2.1 (we can conduct similar exercises for other contracts considered in the paper). For simplicity assume that in both periods the agent is facing a positive marginal incentive, i.e., they are on the upwards sloping portion of $\hat{g}$. Then the first-order condition for Period 1 of the agent is $\frac{w}{\eta_{i,1}} - \frac{\partial c(e_{i,1}, \theta_i)}{\partial e_{i,t}} + a - \varsigma \delta w \frac{e_{i,2}}{(\varsigma e_{i,1} + (1-\varsigma) r_1)^2} = 0$. We assume that the agent underestimates the marginal impact of e_1 on Period-2 payments. In particular, they assume that the impact of e_1 on Period-2 payments is a fraction $\psi \in [0, 1]$ of its actual value. However, they understand the actual target rate that they will face in Period 2 (i.e., $\eta_{i,2}$).[15] Thus, the first-order condition that they solve is $\frac{w}{\eta_{i,1}} - \frac{\partial c(e_{i,1}, \theta_i)}{\partial e_{i,t}} + a - \psi \varsigma \delta w \frac{e_{i,2}}{(\varsigma e_{i,1} + (1-\varsigma) r_1)^2} = 0$.[16] If $e_1^o \in [e_1(\mathbb{P}), e_1(\tilde{\mathbb{P}})]$, ψ can be thought of as a simple measure of opacity, with 0 corresponding to complete opacity and 1 to full transparency (and equal to either 0 or 1 if effort is outside of these bounds).[17]

3 Field experiments

In this section we first describe the work context. We then explain the design of our main field experiment, INDIVIDUAL, present results, and contrast these with the predictions of calibrated version of our model of fully-optimal effort provision. The final part of the section discusses various robustness checks, including a second field experiment, GROUP.

---

[15]This assumption means that the worker does not encounter Period-2 earnings that are strongly at odds with their expectations about the level of Period-2 earnings. It corresponds to workers being "used to" experiencing high rates, without fully taking into account the role of their current effort in determining such rates.

[16]In essence, this is equivalent to having a higher discount rate specifically for the pecuniary benefits in Period 2, or by mis-estimating the target rate in Period 2 by a fraction $\frac{1}{\psi}$, but where these may differ by contract structure and framing.

[17]Notice that the ratio measure and ψ need not agree on the relative opacity revealed by two different effort levels, since the ratio is linear in observed effort, but ψ is typically not.

3.1 The work context

We collaborate with a firm that operates multiple warehouses in which workers fulfill the online orders of customers. The modal task involves workers collecting the desired products from storage and putting these into delivery containers. The workers scan each product before placing it into the container.[18] Work is done individually and a worker's output is independent of the effort of other workers, e.g., as there are lines of containers waiting at each station that serve as buffers between workers. Workers often work in different locations in the warehouse throughout their shift.

The firm has provided us with access to minute-by-minute data on the output of all workers, as captured by their scans, which we use to measure responses of worker effort to incentives (Anonymous Firm (2025), Abeler, Huffman, and Raymond (2023)). To compare effort across workers we will control for task fixed effects, to account for the fact that a scan can be more or less difficult across tasks and thus means something different in terms of effort. We will also control, as appropriate, for other factors besides effort that could influence output, such as time effects (e.g., machine breakdowns or demand shocks) and worker ability.

The warehouse we study was a relatively new one for the firm, and had the following timeline. Initially, the firm paid fixed wages at the warehouse. After about a year, the firm introduced the standard-hour plan incentive scheme described in Section 2, with fixed target rates (static incentives) but kept the base wage the same. Some months after that, we conducted field experiments testing the response of workers to including dynamic rate adjustment into the scheme (dynamic incentives). Figure C.1 in the appendix provides more details on the timeline.

The details of the static incentives introduced by the firm are as follows. In contrast to the simple theory described in the body of the text, which involves a single task (but in line with the theory described in Appendix A.4), workers typically work on several tasks in the warehouse over the course of a week. Different tasks involve different amounts of

---

[18]Other types of individually measured tasks in the warehouse include, for example, loading containers into frames that are put on delivery trucks or accepting deliveries of new products and placing these into storage.

exertion to do one scan, because of variation in the activity done and product characteristics. Accordingly, the firm divided the warehouse into 75 different tasks, each with a different target rate η. The initial target rates were based on the average output of all workers in each of the 75 tasks over a previous period of months, but were fixed in the sense that workers were explicitly told that the rates would not be changed without informing workers well ahead of time. The different rates were meant to capture different degrees of task difficulty across tasks.

With the implementation of static incentives workers began to earn a weekly bonus in addition to the base wage. This involves normalizing a worker's total weekly scans from each task by the corresponding target rate, to calculate Standard Productive Hours (SPH) for that task, and adding up across tasks to determine total SPH. SPH are then remunerated according to a cap-and-quota piece rate scheme. The average bonus has typically been about 10 percent of weekly salary, with the maximal bonus corresponding to about 38 percent of weekly salary.

While the static incentives were not implemented as an experiment, we can use a second warehouse of the firm, where incentives were constant over time, as a "control" and estimate the impact of introducing these performance bonuses on worker productivity using a difference-in-difference analysis. Figure 1 shows the main result. Appendix D describes the setup and analyses in detail. The figure shows that introducing static incentives strongly increased workers' effort at the warehouse we study. The corresponding regressions in Table D.1 show that the effect is about 12.5 percent. We will use this observed response, and related results, to calibrate our structural model of effort provision in Section 3.3.

The firm was interested in adding dynamic rate adjustment to the incentive scheme, because both absolute and relative task difficulties rarely stay constant over time. Changing task difficulties are due to changes in capital (e.g., machinery or software) that affect worker productivity and due to changes in the composition of products across tasks. Using past worker output to calibrate target rates on an on-going basis was seen by the firm as a simple and efficient way to keep incentives well-calibrated. Such a rate-setting scheme, however, induces dynamic incentives, in particular a ratchet effect, which might reduce workers' effort. A key question facing the firm was thus whether, and how, workers would respond to the

13

Figure 1: Response to static incentives in the warehouse
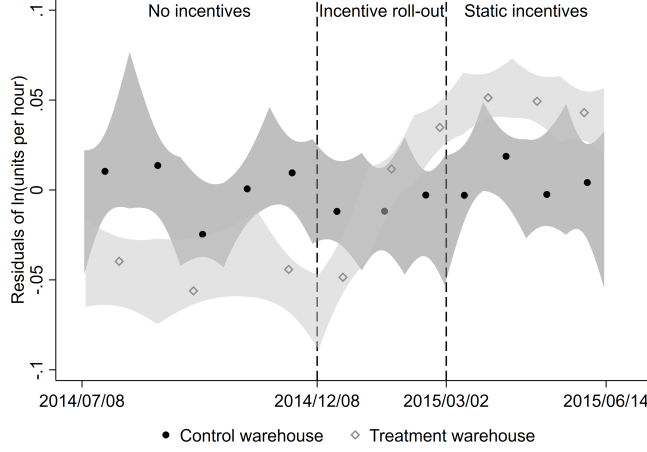
Notes: Binscatter graph of ln(units per hour) in the treated and the control warehouse, residualized for control variables. The treated warehouse initially just paid an hourly wage. It then gradually phased in an additional weekly performance bonus, task-by-task, between 8 December 2014 and 2 March 2015 (time period denoted by shaded gray region). After that the performance bonus system was in place for all tasks, with unchanged target rates until 14 June 2015. The control warehouse had a similar bonus scheme, which did not change across the studied period. The graph corresponds to column 2 in Table D.1. The dependent variable is thus residualized for task fixed effects and warehouse fixed effects, as well as controls for the total time worked in a given shift and warehouse, and controls for average worker tenure in a given shift and warehouse. The shaded areas show 95-confidence bands (Cattaneo et al. 2024).

introduction of dynamic rate adjustment. To shed light on this question, we conducted field experiments within the warehouse.

3.2 Experiment on dynamic incentives: INDIVIDUAL trial

About six months after the introduction of the incentive pay system with static target rates, the firm agreed to randomly assign workers to face different forms of incentive contracts, within the context of its daily business operations. Our main experiment is the INDIVIDUAL trial, so-named because treatment workers could substantially determine their own individual future target rates. The dynamic rate adjustment could thus potentially tailor incentives to differences in individual worker ability. Notably, the design rules out any potential motives related to own effort affecting target rates of others. Also, own rates are not affected by the effort of other participants in the trial. Such motives are present in our second field experiment, the GROUP trial, which serves as a robustness check.

Participants in the INDIVIDUAL trial were newly hired workers in the warehouse. Each

week, on average about 32 workers joined the firm, and for a period of 40 weeks, all workers in each new cohort were randomly divided into treatment and control groups. We thus have a sequence of 40 experiments. The random allocation of workers to treatments was done by us. Workers were extensively informed about all the details outlined below, except for the fact that the trial was designed together with university researchers.[19]

Table 1 summarizes the experimental design. For each cohort, the experiment lasted 12 weeks. The first three weeks were a baseline period in which all workers faced exogenously given target rates, calibrated to the warehouse rates calculated using more experienced (more than 13 weeks of tenure) workers. In Week 3, the workers learned about their assignment to the treatment (or control) condition and how their target rates would be determined in the weeks going forward.

The period of interest is weeks 4 to 6. In these weeks, treatment workers and control workers faced the same contemporaneous, fixed rates, but treatment workers knew that their individual target rates in weeks 7 to 9 would be the average of their output in weeks 4 to 6, and the warehouse average of experienced workers (workers with more than 13 weeks of experience) for the same activity and time period. In terms of the model, $\varsigma = 0.5$ and r_1 is the experienced worker average hourly output. Thus, the rate was not completely determined by an individual treatment worker's output, as it depended partly on the warehouse average, but worker output had a substantial weight of 0.5. Control workers, by contrast, knew they would again face fixed rates in weeks 7 to 9, and thus lacked dynamic incentives in weeks 4 to 6. Note that treatment workers did not influence anyone's rate but their own. To maintain fairness, and to avoid a Hawthorne effect, our design ensured that control workers also had similar dynamic incentives, but later on, in weeks 7 to 9.[20] From week 10, the rates of treatment workers reverted to the overall warehouse rates. Thus, our test for a causal effect of dynamic incentives focuses on weeks 4 to 6 of the trial, comparing the behavior of

---

[19]Treatment spillovers are unlikely, including because treatment and control workers rarely met outside the induction meetings, as they worked across the warehouse among a daily workforce of about 900 workers.

[20]In weeks 7 to 9, control workers knew their individual output would determine their target rates for weeks 10 to 12 (recall that treatment workers have rates revert to warehouse averages in weeks 10 to 12). This period is, however, not a clean comparison for measuring their response to dynamic incentives because static incentives also differ at that time, due to treated workers endogenously determining their rates for weeks 7 to 9.

treated to control workers.[21]

Appendix Table C.1 contains summary statistics and randomization checks (all $p > 0.34$). In total, 1294 workers started the treatment period, which began four weeks after joining the firm. Appendix Figure C.2 and Table C.2 show that there is no differential attrition before or during the trial between treatment and control group.

Table 1: Design of the INDIVIDUAL trial

Baseline 3 weeks	Rates = fixed fraction of site rates	
Condition assigned	Treatment workers ($N = 631$)	Control workers ($N = 663$)
Weeks 4 to 6	Rates = fixed fraction of site rates	
Weeks 7 to 9	Rates = (individual speed in weeks 4 to 6 + site rate)/2	Rates = fixed fraction of site rate
Weeks 10 to 12	Rates = fixed fraction of site rate	Rates = (individual speed in weeks 7 to 9 + site rate)/2

Finding 1. *The INDIVIDUAL trial yields a very small ratchet effect that is not significantly different from zero.*

Figure 2 illustrates the finding graphically. The figure shows a time series plot of worker output, measured as units (=scans) per hour, separately for treatment and control workers, pooling all cohorts of the experiment. Output is binned by week. We control for task fixed effects, to account for task difficulty, and for cohort fixed effects (and their interaction with task FE), to account for the possibility that cohorts starting in different calendar weeks could differ. The figure shows that output is lower in weeks 4–6 than 7–9 for treatment workers, but there is also the same time trend for control workers, indicating that this is a general increase in ability, i.e., a learning curve. We see that the treatment difference at

---

[21]We did not pre-register our analyses as we started the data collection when pre-registration was less common. This might raise concerns about p-hacking. Our main results from the field are not, or only marginally, significant, potentially reducing concerns about p-hacking, but the treatment differences in the online experiment with warehouse workers are significant (see Section 4). Also, our results on cognitive ability stem from a sub-group analysis and are only significant for one of three aspects of cognitive ability we measured. Addressing these concerns is a reason why we conducted a replication with AMT workers (see Section 5). More broadly, all our main results – the main ratchet-effect result, the treatment differences in the online experiments and the results on cognitive ability – obtain across all our three datasets and in various robustness checks.

Figure 2: Opacity of dynamic incentives in the INDIVIDUAL trial
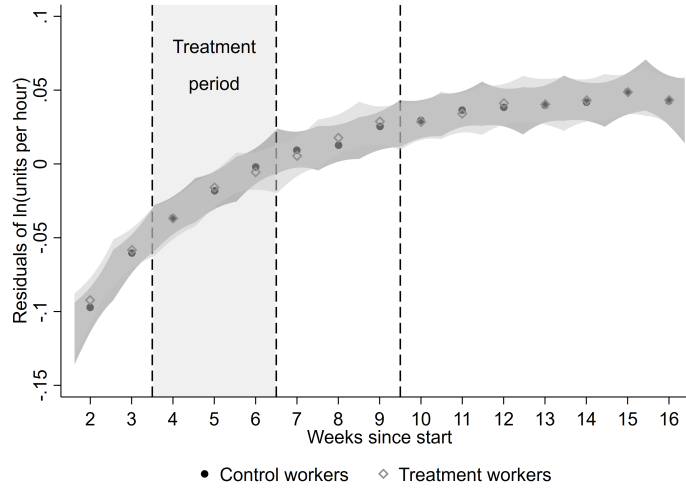
Notes: Time series plot of ln(units per hour) (UPH) for treatment and control workers. UPH are residualized for task and cohort fixed effects and their interactions. Both groups of workers faced the same static incentives. Treatment workers additionally faced a dynamic (ratchet) incentive to reduce effort in weeks 4–6. The shaded areas show 95-confidence bands (Cattaneo et al. 2024).

each point on the learning curve is approximately zero. This very small reaction of workers to the introduction of dynamic incentives is in stark contrast to the reaction of workers to the introduction of static incentives shown in Figure 1.

We begin the statistical analysis with the simplest approach, treating each of the 40 cohorts as a separate experiment that provides a noisy but unbiased estimate of the treatment effect. For each cohort, we regress worker output in weeks 4–6 on a treatment dummy and task fixed effects. In all our regressions, we will use standard errors that are two-way clustered on individual workers and on shifts.[22] Taking the average of the treatment effects, we find that treatment workers had lower output, but only by -0.2 percent and not significantly different from zero (with a 95 percent confidence interval of [-1.6, 1.1]).

We can also pool data from all cohorts, if we account for potential differences across cohorts starting at different dates. In addition to task fixed effects, we thus include cohort fixed effects, task × cohort FE, shift fixed effects (each shift corresponds to a day or night shift on a particular calendar date) and shift × cohort FE. These control for, e.g., the possibility that different cohorts might have different time shocks, or ability levels, or learning curves.

---

[22]Alternatively, we can just cluster on workers as the unit of randomization. Across all tables in the paper, this changes p-values only in the third digit, if at all, and leaves all results unchanged.

Table C.3 in Appendix C shows the detailed results. The main regression includes the full sample of weeks 4–6. Figure 2 is closest to this specification. The point estimate implies a change of effort of -0.1 percent (CI: [-1.2, 1.0]). We can also restrict the sample to only week 6, in case the response increases over time, or further restrict it to only those workers who kept working for the firm until at least the end of week 9. These latter workers enjoy the full benefit of reducing effort in weeks 4–6 and they thus face the strongest ratchet incentives. The (absolute) point estimates are only slightly larger, implying a change of -0.2 percent (CI: [-1.4, 1.1]) or -0.5 percent (CI: [-1.8, 0.9]).[23]

3.3   Comparison to fully transparent benchmark

Our theoretical model predicts that treatment workers should reduce effort relative to control workers, if the contract is fully transparent to workers. But what size of a response should we expect, and is it larger than the response we observe? In order to derive a benchmark for behavior, we revisit our model of rational effort provision from Section 2, which allows for several considerations that will affect the size of the effort reduction, such as time discounting, the elasticity of work effort, and intrinsic motivation or fear of being dismissed.

In order to calibrate our model, we must make some assumptions (for more details on the calibration see Appendix B.1). We focus on a single, representative worker who works on a single representative task over the course of two periods, as in the framework of Section 2. Period 1 corresponds to weeks 4–6 of INDIVIDUAL, and Period 2 to weeks 7–9. In line with other papers studying labor supply we suppose the cost of effort is a power function: $c(e, \theta) = \theta \frac{e^{\gamma+1}}{\gamma+1}$. We set the discount factor $\delta = 0.913$.[24]   As we show in Appendix B.1, the results reported below are robust to assuming an exponential cost function for $c(e, \theta)$,

---

[23]As Figure 2 shows, there is a similarly small opposite ratchet effect in weeks 7–9, when the control group faces incentives to ratchet and the treatment group does not. The corresponding point estimates for the control group in weeks 7–9 are very similar, e.g., -0.1 percent in the full sample (as in column 1 of Table C.3). Since the spot incentives are different, this estimate is potentially biased, as mentioned before.

[24]Because the time elapsed between the end of the two periods is small (3 weeks in reality), there is unlikely to be significant true time discounting. In contrast, there is some potential "as if" discounting because workers might leave the firm, or be scheduled to work less, in Period 2 than in Period 1. This reduces the future benefits of sacrificing bonus in Period 1, and thus implies a weaker ratchet effect, similar to greater time discounting. We calculate this as-if discounting by calculating the ratio of the total time workers spent working in Period 2 relative to Period 1 (including workers who leave the firm permanently), and find that it is 0.913.

non-linear non-pecuniary preferences or stronger time discounting.

To identify the parameters of the model we use data on workers' observed behavior and assume that workers rationally choose their effort level when facing static incentives, i.e., when they face only an hourly wage or when they receive incentive pay with exogenous target rates. We consider the single period (i.e. only Period 1) equivalent of Equation 2.1.

We match two features of the data. First, our estimated reaction of the workers to the introduction of static incentives, mentioned in Section 3.1 (details are provided in Appendix D), corresponding to marginal incentives rate w moving from zero to strictly positive. In the group of workers most similar to participants in the INDIVIDUAL trial, the estimated response to static incentives is 10.5 percent. We will assess whether, given this substantial response, the model can explain the much smaller observed response to dynamic incentives in INDIVIDUAL of only -0.1 percent. The second moment we will use for the calibration is the observed effort level in Period 2 of the INDIVIDUAL trial (weeks 7–9), when treatment workers only face static incentives.[25] Given the parameter values we derive from these two moments, we then simulate what workers' responses should be in Period 1 of the two-period model (corresponding to weeks 4–6). This implies that any utility losses our model generates must occur solely because of issues that arise due to mis-optimization in Period 1 against the dynamic incentives of the contract.[26]

We only have two moments of data, while we have three parameters to identify: the two parameters of the cost function θ and γ, as well as non-monetary motivation a. Our approach therefore focuses on set identification. We establish bounds in terms of the behavioral response and utility loss across the entire set of possible parameter combinations that match the observed moments. In particular, for any positive value of a we can find corresponding values of θ and γ that rationalize the behavior under static incentives. We call these triple-tons "allowable". Since the model features many local optima, we solve numerically. In order to establish bounds, we consider a large range of potential values of a (and corresponding θ

---

[25]As discussed in detail in Appendix 2 we use the prevailing target rates during this period, weighted by the share of time each task was done by the representative worker, to account for task difficulty and convert observed output levels into representative agent effort.

[26]One could instead calibrate the model to match the Period 1 behavior of workers in control. But this would imply that we would predict utility losses in Period 2 that are distinct from any issues that arise because of mis-understanding of the dynamic implications of contracts, and so we do not pursue this direction.

and γ). Given an allowable triple (θ, γ, a), we then numerically simulate workers' response to the 2-period dynamic maximization problem, conditional on them taking all aspects of the contract into account, and compare this to the observed behavior. Looking over the range of allowable triples leads to the following result.

Finding 2. *Given their observed response to static incentives, workers should have reduced effort by at least 5.5 percent in Period 1 of the INDIVIDUAL trial.*

Table B.1 in Appendix B.1 summarizes the results of this exercise. Looking across all parameter combinations that rationalize static behavior (i.e., with perfect perception of dynamic incentives), the lower bound is an order of magnitude larger than the observed reaction. Indeed, the ratio metric and ψ metrics of mis-perception are both zero when estimated based on actual behavior, implying full opacity (see Appendix B.4 for more details). We can also calculate the utility loss that treatment workers experienced given their actual effort, compared to what would have happened had they behaved optimally.[27] Again, the size of the utility loss varies by the parameter values, but never falls below 1.5 utils (about 6 percent of the average weekly bonus). These values, of course, are the most conservative estimates, and for other allowable parameter combinations the treatment effects should be much larger – up to 75 percent reduction in effort, and losses of around 26 utils. The reason for this range of possible responses, and potentially very large predicted reductions, arises from the way that various nonlinearities in the optimization problem, e.g., the cap and quota, lead to multiple local maxima. For example, one maximum involves relatively high efforts in both periods, albeit with at least 5 percent lower effort in Period 1, and corresponds to our conservative estimates. Another local maximum, however, involves very low effort in Period 1 and high effort in Period 2, corresponding to a large predicted ratchet effect. As parameter values change, the predicted response to dynamic incentives can change continuously up to a point, but then increase dramatically because the local maximum with low Period-1 effort switches to being the global maximum. For example, when a worker finds it optimal to work below the quota in Period 1, the discrete reduction in financial incentives can make it optimal to reduce Period-1 effort to a very low level. We provide a more detailed discussion

---

[27]In related work, Copeland and Monnet (2009) study the impact of bonuses on welfare in an environment where workers are paid a daily bonus, under the assumption of full transparency.

in Appendix B.1.

The wide variation in these results across allowable parameters combinations raises the question of which allowable combinations are most reasonable. One way of answering this is to consider literature estimates of the intensive elasticity of labor supply. Although this elasticity is subject to dispute, two recent studies using natural experiments of tax holidays, Stefansson (2019) and Martinez, Saez, and Siegenthaler (2021), find intensive-margin elasticities of 0.07 and 0.025, respectively. In our model, the elasticity of effort with respect to the value of effort is $\frac{1}{\gamma}$ and so these estimates of intensive labor supply imply that γ is around 10 to 50. In an online experiment using a real-effort tasks, DellaVigna and Pope (2018) find estimates of γ around 20. These parameter values imply utility losses and effort reductions dramatically larger than our lower bounds.

With the version of our calibrated model that incorporates the mis-perception parameter ψ, we can also ask the question of how sensitive are predictions to degree of mis-perception. The model shows mis-perception can be relatively minor, e.g., with ψ equal to 0.9, and predicted behavior can diverge strongly from predictions with full transparency, i.e., there can be very little in the way of response to dynamic incentives. This reflects the way that mis-perception enters the model and interacts with nonlinearity in the optimization problem. This suggests that even if factors such as learning were to reduce mis-perception substantially, there could still be strong deviations of behavior from the fully transparent optimum, as long as there remains some opacity. We discuss this more rigorously in Appendix B.4.

In sum, our results imply that workers should have reduced effort significantly more strongly than what we observe in INDIVIDUAL, if dynamic incentives were fully transparent. Our results also suggest that the rate-setting scheme would not be advantageous for the firm if all workers were rational. We explore this in more detail when the firm can alter its contracts in Section 3.5.

3.4 Robustness checks

Learning, experience and social pressure: Our findings from INDIVIDUAL are consistent with dynamic incentives induced by contracts being opaque. Our second field experiment, denoted the GROUP trial, investigates whether ratchet effects can remain opaque for

Figure 3: Opacity of dynamic incentives over time in the GROUP trial
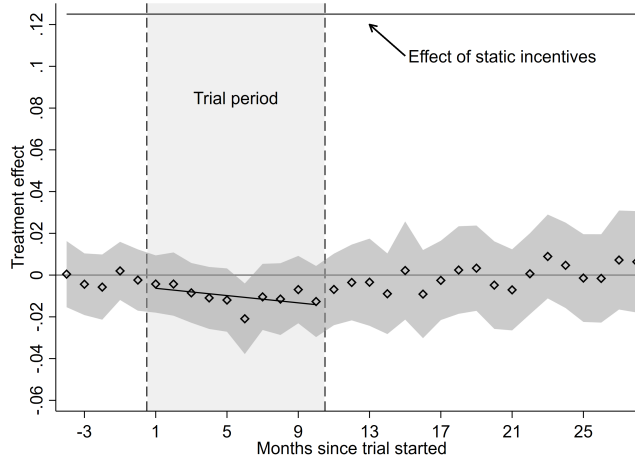
Notes: Event study graph of the treatment difference on ln(units per hour), i.e., the ratchet effect. The vertical lines depict the start and the end of the trial. "Months" counts the four-week rate-setting periods since the start of the trial. The graph is restricted to the first randomization cohort for whom the trial lasted for 10 periods. Point estimates are from regressions as in Table E.4, column 1, separately for each month. The shaded area is the 95 percent confidence interval.

a longer time period and with more opportunities and motives for learning.

The GROUP trial randomized all workers at the warehouse into two conditions, treated workers and control workers. During the 10-month trial period, next month's rate for a given task was set equal to the average speed of the treated workers in that task during the current month, but these rates applied to all workers equally, including those in the control group. The contemporaneous static incentives were thus identical for treatment and control workers, but treated workers faced an additional dynamic incentive to hold back effort. Appendix E describes the design and analysis of the GROUP trial in more detail. Section A.4 shows theoretically that the treatment should have induced a ratchet effect.

Because GROUP lasted for 10 months, with feedback each time[28], and because treatment workers influenced the rates of others, there were substantial opportunities for individual learning, and potential motives for social pressure and social learning.[29]

Figure 3 plots the treatment effect in the GROUP trial over time. The overall effect is slightly larger than what we find in INDIVIDUAL, about -1.0 percent, but still quite small

---

[28]Each month, workers were explicitly informed about the new rates and workers were reminded that "you set the rates", i.e., rates are set as a function of workers' past speed.

[29]The sociological literature on ratchet effects (e.g., Mathewson (1931); Roy (1952)) reports instances of social pressure and learning playing a role in workers colluding and holding back effort.

relative to, e.g., the impact of the introduction of static incentives, which is shown in the figure for comparison. The point estimate for the ratchet effect starts out close to zero in the first month, similar to INDIVIDUAL, and grows over time, reaching -1.3 percent by the end of 10 months, but this time trend is not statistically significant (p=0.475). This suggests that opacity is still present to some degree after 10 months.[30]

Additional robustness checks: In Appendix F we discuss the robustness of our conclusions to various potential concerns. These include our simplifying assumption of a single task; potential for time discounting to attenuate ratchet effects; whether the firm could commit to its stated policy of adjusting rates; whether dismissal concerns could explain the weak response to dynamic incentives. We provide evidence and arguments for why these do not appear to explain our findings, and we also note that all of these are eliminated in the context of our online experiments, discussed below.

3.5 Equilibrium efficiency and surplus with transparency

Our results so far suggest that complexity is good for the firm, as it increases worker effort. The effect on workers, fixing the contract, are negative. However, the firm would likely not want to offer the same contract to workers who fully took into account all the contractual features compared to those for whom some are opaque. In this case, it seems less clear what the equilibrium effects on workers and firms might be.

Appendix B.1.2 describes a full-equilibrium welfare analysis where we allow the firm to choose which contract it offers, under the assumption that the contract is fully transparent. Although complete transparency may be too strong an assumption if certain payment schemes are intrinsically complex (see Section 5.2 for which features determine transparency

---

[30]An alternative explanation would be that the workers come to fully recognize the dynamic incentives due to the opportunities for learning, but for some reason find it optimal to respond only very weakly. It is difficult to assess the plausibility of this with a formal benchmark for rational effort provision, which we do not have because the repeated game aspect of GROUP can induce complicated equilibrium play and makes point predictions indeterminant. To provide further evidence on this, we analyze the behavior of workers in our online experiments, which took place after the end of GROUP, who previously participated in GROUP. In the COMPLEX treatment of the online experiments, workers face a very similar contract as in GROUP but potential motives to not respond are removed by design. A response to dynamic incentives should thus emerge strongly in the online experiment if most workers are in fact aware of them after their experience with GROUP. Section 4 shows that this is generally not the case, except for a sub-sample of workers with high cognitive ability.

in reality) we believe it forms a good baseline for comparison. We compare outcomes with optimally chosen transparent contracts to the actual observed situation, where dynamic incentives are opaque.

In order to formalize the issues surrounding efficiency, we need to enrich our environment by making several assumptions. Specifically, we assume: The output market is competitive; the firm has imperfect knowledge about the worker's cost function parameter θ; the set of workers at the firm is fixed; the firm cannot adjust the hourly wage; the set of bonus schemes available to the firm follows the general form described in Section 2 (this still allows the firm considerable flexibility, including being able to vary the initial target rate, to pay only static incentives, or not pay any bonus at all). For more details on these assumptions see Appendix B.1.2.

We find that firm profits and total surplus are always higher with opacity. This is in line with the intuition in the literature on ratchet effects (e.g., Laffont and Tirole (1988)). If dynamic contracts are transparent, then workers will tend to work less than they would if they were simply maximizing one-period utility, given incentives. We refer to this effort reduction as shirking. Anticipating that workers will shirk, the firm no longer wants to produce at the first best, reducing efficiency. Opacity reduces shirking and so improves efficiency, i.e., total surplus. Without ex-post transfers between the firm and workers, the effects on the utility of the workers are more nuanced. In partial equilibrium, increased transparency (fixing the contract), always benefits workers. For some parameters, this intuition carries over into full equilibrium reasoning. However, for many parameters, the optimal transparent contract involves the firm setting higher target rates. Such target rates can be optimal under transparency because they reduce the future benefit from shirking and thus reduce the incentive to shirk. In this case, workers can be worse off when the contract is transparent, compared to our actual workers, i.e., if the optimal transparent contract involves no bonuses then workers may be better off by mis-optimizing against a scheme with some positive bonuses.[31]

---

[31]There are also distributional effects between workers, if workers differ in their sophistication. Fixing a contract, sophisticates are always better off than naifs. For example, for $a = 0.1$, Table B.1 shows that sophisticates earn 53 utils and naifs 33 utils. However, our results imply that both sophisticated and naive workers may be better off when the firm provides a contract that is designed to maximize profits against a population of naifs (i.e. the observed contract), relative to a counterfactual where the firm chooses a contract

4 Online experiments with warehouse workers

Our field experiments showed only a weak response to dynamic incentives in a real work setting, and our structural analysis casts doubt on motives that might make this a rational response. While this is indicative of complexity and bounded rationality playing a role, we turn to analyzing data from online experiments with the same warehouse workers to provide sharper tests for these potential mechanisms (Abeler, Huffman, and Raymond (2025)).

The controlled setting of online experiments has several advantages. First, the online experiments allow us to randomly allocate workers to a broader range of incentives schemes, including ones that seek to make dynamic incentives more transparent. Unsurprisingly, the firm would not allow us to conduct similarly transparent treatments in the warehouse, out of concern that this might lead to a large and damaging ratchet effect. Second, the setting rules out, by design, motives related to discounting, firing threats, or social preferences towards the firm, because experiments took place over a short time period, were known to be conducted by outside, academic researchers rather than the firm, involved a task of no intrinsic value, and promised confidentiality of responses. Third, we can measure cognitive ability and test directly whether lack of response to dynamic incentives is related to bounded rationality. Fourth, because the experiment was run with the same warehouse workers after the end of the GROUP trial, we can link behavior in the lab to that in the field.

4.1 Design of experiments

Participants in the online experiments worked on a real-effort task, for an incentive scheme that was very similar to the one in the warehouse. The task was clicking a button on the screen, either with a finger (if using a smart phone or tablet), or with a mouse (if using a computer). During the experiment, workers could work on the task for multiple periods of 90 seconds each. The number of clicks in any given period is our measure of effort. It was divided by a target rate η_t to give "Standard Productive Minutes" (SPM). There was a cap-and-quota payment schedule. Complete instructions for the online experiment are provided in Appendix I.

---

optimized against a population of sophisticates.

Workers were recruited via e-mail invitations, which specified that the study was being done by outside researchers, and promised confidentiality of individual responses from the firm. Compared to the warehouse population we do not see any differential selection into the online experiment according to gender, age, tenure, nationality or speed in the warehouse (both speed in the entire period before the online experiment or speed only during the period without dynamic incentives; all $p > 0.204$).

Workers were randomized into one of four treatments, which varied the nature of the incentive scheme. In all we have 430 warehouse workers in the online experiments.[32] Table G.1 in the appendix summarizes the design of the conditions.

In a condition that we denote COMPLEX, workers faced incentives with a similar degree of complexity, and the same type of dynamic rate adjustment, as in the INDIVIDUAL trial. Workers first had a baseline period in which they could do the task for static incentives, with a fixed target rate of 300. Subsequently they learned the rules for working in Periods 1 and 2: Period 1 would again have the exogenous target rate of 300, but in Period 2, the target rate would equal the number of clicks done in Period 1, averaged with a randomly drawn number X, which would be uniformly drawn from a narrow range of values centered around the target rate (285, 300, or 315) at the beginning of Period 2. The implementation of X mimics the fact that the general warehouse rate, r_1, which helped determine the target rate in INDIVIDUAL in weeks 7–9, was not known in advance. After completing Periods 1 and 2, workers learned about Periods 3 and 4, which had the same structure as Periods 1 and 2. Given the dynamic incentives, we would expect workers to reduce effort in Period 1 and 3.

Another condition, denoted SIMPLE, was very similar in set-up to COMPLEX. However, the incentive scheme created dynamic incentives to reduce effort in Period 1 in a form that was intended to be more transparent. Specifically, the scheme eliminated the three aspects of COMPLEX that we hypothesized might contribute to opacity. These changes were intended to (1) describe the dynamic incentives explicitly in terms of money; (2) make dynamic incentives about the level of future earnings, rather than future marginal incentives; (3) eliminate noise in parameters in the form of the random variable X.[33] The target rate was

---

[32] There is no significantly different attrition in any of the treatments compared to COMPLEX as baseline treatment.

[33] We consider X to be a distractor, rather than a meaningful element of the incentive scheme, since in our

fixed at 300 for both periods, but any earnings in Period 1 would be subtracted from earnings in Period 2. After completing Periods 1 and 2, workers learned about Periods 3 and 4, which had the same structure as Periods 1 and 2. In such a scheme, it is easy to see that one should reduce effort in Periods 1 and 3, a ratchet effect (for a formal proof see Appendix A.3).[34]

We randomized the remaining workers into one of two additional treatments. In STATIC, workers faced an exogenous target rate of 300 in all five work periods. Any changes in effort over time thus reflect other potential factors like learning by doing or fatigue. This treatment provides a benchmark of behavior when dynamic incentives are truly absent rather than opaque. In STATIC_ZERO, workers also faced a static target rate of 300 in all five periods, but for Periods 3 and 4, the piece rate was reduced to zero. Clicking in these latter periods therefore reveals the extent of intrinsic motivation. Notably, the comparison of behavior with zero piece rate to behavior with positive piece rate is analogous to the data moment we observe in the warehouse, of effort under fixed wages versus effort under static incentives.

As summarized in Table G.1, we also collected various other types of measures about participating workers at several points during the experiment (for the exact wording see Appendix I). This included incentivized measures of time preferences and risk aversion, workers' educational attainments, and three facets of cognitive ability. One aspect of cognitive ability was the CRT, a test consisting of three questions, each with seemingly obvious but incorrect answers (Frederick (2005)). We used the CRT because it is a measure of a tendency to think deeply, and avoid superficially plausible but incorrect answers. It thus focuses on an aspect of cognitive ability that could be particularly important for noticing opaque attributes of contracts. Other questions included hypothetical lottery questions, designed to identify narrow versus broad bracketing of decision making (Rabin and Weizsäcker 2009). We included this measure in case recognizing dynamic incentives across periods is related to broad bracketing. The workers also played a simple game as part of the questionnaire, called Hit 7, designed to measure ability at backwards induction (Burks et al. 2009). We implemented this measure in case difficulties in backwards induction might make it harder

---

model of effort provision noise in Period-2 target rates is irrelevant for the optimal Period-1 effort.

[34]Once we calibrate our model of rational effort provision (see Appendix B.2), we find that the optimal effort reduction in SIMPLE is actually smaller than in COMPLEX, which works against the result we will find.

to reason backwards from incentives in Period 2 to the optimal choice in Period 1.

Finally, after Period 4, workers were asked an open-ended question about what they would recommend to another person as the best strategy for work in Periods 3 and 4: "If someone were trying to get the most money, total, from [Period 3 and Period 4], what do you think would be the best approach?" The question wording was chosen so that it could prompt potential comments on dynamic incentives in both COMPLEX and SIMPLE. As it is difficult to ask about utility, the question focused on money.[35] The reason why the question asked them to explain a strategy that someone else might use, was in case workers were reluctant to describe themselves in such a role due to a heuristic or habit of caring about reputation (although there were no actual reputational consequences since responses were confidential).

4.2 Results

Panel (a) of Figure 4 shows average clicks by period and treatment. In STATIC, where dynamic incentives are absent, clicks are largely constant over time, although there is a slight increasing trend, potentially reflecting some learning by doing.[36] The time profile of clicks in STATIC_ZERO is very similar to STATIC initially, but then there is a sharp drop for Periods 3 and 4 when the piece rate is reduced to zero. This latter difference indicates that workers recognize the static incentives in the scheme. It also points to an important role of intrinsic motivation, because clicks are around 350 on average even when the piece rate is zero. Such a level of intrinsic motivation is in line with previous studies using button-clicking tasks (e.g., DellaVigna and Pope (2018)).

In COMPLEX and SIMPLE, dynamic incentives were present in Periods 1 and 3. Panel (a) shows that warehouse workers in COMPLEX are largely unresponsive to these, with

---

[35]In our experiment, if participants hold back effort in line with the ratchet effect, they can earn more with less effort in the future. So long as the worker faces a contemporaneous trade-off of providing effort in Period 1 (i.e. Period-1 earnings vs. Period-1 costs) then the advice "reduce effort in Period 1 compared to Period 2" is optimal both for maximizing money and for maximizing utility. For COMPLEX, we thus need to assume that workers are not able to reach the full bonus. This turns out to be a weak assumption since only 1 out of 430 participating workers was able to reach the full bonus in any of the periods. No additional assumption is needed for SIMPLE.

[36]The total increase is about 11 percent over the five periods. The linear time trend in STATIC is statistically significant in an OLS regression of clicks on period ($p < 0.001$).

Figure 4: Opacity of dynamic incentives, warehouse workers online
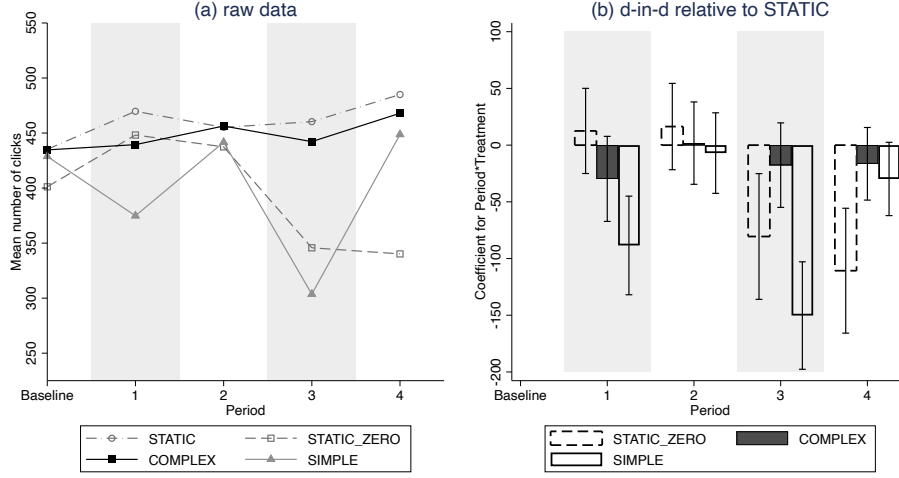
Notes: Panel (a) shows average clicks in a given work period. Shaded gray bars denote periods with dynamic incentives in COMPLEX and SIMPLE. Panel (b) plots coefficients of interaction terms, Period*Treatment, from a difference–in–differences regression relative to baseline period and the treatment STATIC (see column 1 of Table G.2 in the appendix for all coefficients). Error bars show 95 percent confidence intervals.

behavior very similar to STATIC. Factors such as reputation concerns or time discounting are absent and thus cannot explain the lack of response. This further strengthens the case that, for most warehouse workers, the dynamic incentives are opaque in this type of incentive contract. Comparing behavior in COMPLEX and SIMPLE demonstrates the direct effect of complexity. Panel (a) shows the zig-zag pattern that is consistent with workers recognizing dynamic incentives in SIMPLE, with workers strongly reducing effort in Periods 1 and 3. Since workers were randomly allocated to treatments, this treatment difference measures the causal effect of changing the incentive scheme to reduce complexity.

Finding 3. *In COMPLEX, warehouse workers behave on average as though dynamic incentives are absent, and are far from the optimum predicted by our representative agent model of fully-rational effort provision. Workers in SIMPLE exhibit a significantly stronger response, and average effort is closer to the corresponding rational optimum. Individual-level measures also indicate that fewer workers respond to dynamic incentives in COMPLEX than in SIMPLE.*

To test statistically between treatments, we use difference-in-differences regressions, regressing effort in each period on treatment and period dummies and their interactions. Such

a regression guards against any randomization failure between treatments by controlling for effort in the baseline period, and controls for any time trends arising for reasons other than dynamic incentives, by comparing to STATIC. Panel (b) plots the interaction terms of Period*Treatment from the regression. The full regression is in column 1 of Table G.2 in the appendix. To assess joint significance of effort differences across Periods 1 and 3, we use F-tests (see footnotes for p-values). We find that effort in SIMPLE is on average significantly lower than in COMPLEX or STATIC in Periods 1 and 3, whereas COMPLEX and STATIC are not different.[37]

Just as with our INDIVIDUAL field experiment, we can compare results on aggregate worker behavior to the predictions of our representative-agent model of fully-rational effort provision. We proceed in a way that mirrors our approach in Section 3.3 (all details are in Appendix B.2). As before, we have only two moments to match, and three preference parameters, so we use set-identification of parameters and construct bounds on the behavior induced by COMPLEX and SIMPLE. The model shows that the optimum in COMPLEX is lower than in SIMPLE for many parameters, including the most reasonable parameter values (e.g., those that correspond to the point estimated values from our AMT data, see Section 5). Participants also incur larger loss from ignoring dynamic incentives in COMPLEX compared to SIMPLE. These features work against our empirical findings, since we see that workers reduce effort less in COMPLEX despite the lower optimum and the larger utility loss.[38]

In COMPLEX, the model predicts effort levels from 0 up to 397 in periods with dynamic incentives, while the observed average effort in Period 1 and 3 is 441. Thus, observed effort is *at least* 10 percent higher than predicted for rational workers. The ratio metric of transparency for COMPLEX ranges from 0 to an upper bound of at most 0.42, and the ψ

---

[37]P-values of F-tests of joint significance for Periods 1 and 3: STATIC vs. COMPLEX $p = 0.296$, STATIC vs. SIMPLE $p < 0.001$, COMPLEX vs. SIMPLE $p < 0.001$. Periods 2 and 4: STATIC vs. COMPLEX $p = 0.484$, STATIC vs. SIMPLE $p = 0.169$, COMPLEX vs. SIMPLE $p = 0.696$. STATIC vs. STATIC_ZERO Periods 1 and 2 $p = 0.688$, Periods 3 and 4: $p < 0.001$

[38]Intuitively, the reason why the optimum in SIMPLE is higher than in COMPLEX is intrinsic motivation. Dynamic incentives reduce monetary incentives to zero in SIMPLE, and so workers should provide effort to the extent that they are concerned with their intrinsic motivation (and effort costs). In contrast, dynamic incentives can push Period-1 monetary incentives below zero in COMPLEX. In other words, reducing effort in COMPLEX can increase future earnings by more than the loss of earnings in this period. If workers have no intrinsic motivation, both treatments have the same optimum of zero effort. With positive intrinsic motivation, rational workers in SIMPLE provide positive effort in line with their intrinsic motivation. In COMPLEX, for many levels of intrinsic motivation, it is still optimal to provide zero effort.

that best fits the data ranges from 0.45-0.85, depending on parameter assumptions, so both metrics indicate substantial opacity.[39]

In SIMPLE, the model predicts effort, regardless of the parameter combination, of 347 in periods with dynamic incentives, while observed average effort is 339. Thus, the model prediction is only around 3 percent different from observed effort in SIMPLE (workers work slightly less hard than predicted) and the ratio metric and best-fitting ψ are both at 1, indicating full transparency.[40] Clearly workers are closer to their optimum in SIMPLE than COMPLEX. Moreover, for any allowable parameter combination, workers are losing more utility compared to the optimum in COMPLEX than in SIMPLE.

We also see signs, at the individual level, that fewer workers respond to dynamic incentives in COMPLEX than SIMPLE. In COMPLEX, the fractions of workers reducing clicks in Periods 1 and 3, relative to their baseline clicks, were 0.30 and 0.33, respectively. These are not significantly different from the fractions observed in STATIC, 0.23 and 0.28 (Wilcoxon tests relative to COMPLEX: p=0.26, p=0.49). By contrast, in SIMPLE, the fractions were significantly larger than in COMPLEX, with the majority of workers exhibiting a decline by the second dynamic incentive period: 0.41 and 0.56 in Periods 1 and 3, respectively (Wilcoxon tests relative to COMPLEX: p=0.04; p<0.001).

Another indication that dynamic incentives are more opaque in COMPLEX is that workers mention dynamic incentives much less often in COMPLEX than SIMPLE, when asked about optimal work strategies. Three evaluators, who were unaware of this paper's research question or hypotheses, independently coded responses to the open-ended question about optimal work strategies. The main focus was to find any indication that the worker rec-

---

[39]The non-linearity of the ψ metric implies that it can deliver different conclusions about the quantitative magnitudes of opacity underlying observed behavior in a given environment, compared to our linear (e.g., ratio) metrics. The two metrics deliver robust qualitative conclusions, however, about how varying the environment (e.g., COMPLEX vs. SIMPLE) affects opacity.

[40]The finding of full transparency in SIMPLE is a result for the average or representative worker, and thus does not necessarily imply that every individual worker recognizes dynamic incentives in SIMPLE. Data limitations prevent us from calculating a personal rational optimum for each individual. Figure G.1 shows, however, that the distribution of individual behavior in SIMPLE, in periods with dynamic incentives, is very similar to that in STATIC_ZERO, in periods with zero piece rate. This is suggestive that many workers in SIMPLE recognize that dynamic incentives are present and effectively imply a piece rate of zero, although it does not exclude that some individual workers did not notice dynamic incentives. By contrast, the distribution of individual level behavior in COMPLEX is very similar to STATIC, consistent with most workers not recognizing dynamic incentives in COMPLEX.

ognized a reason to click less in Period 3 than Period 4, which is a lenient classification of noticing dynamic incentives. A worker was coded as showing awareness if at least two evaluators agreed (evaluators almost always agreed, with an average Spearman correlation of 0.93 between rater evaluations).

Finding 4. *Few workers mention dynamic incentives in COMPLEX in the open-ended question, while many do in SIMPLE.*

Only 19 percent of workers in COMPLEX mention dynamic incentives vs 44 percent in SIMPLE.[41] The difference is highly statistically significant (Wilcoxon test; $p < 0.001$). Table G.3 in the appendix lists the categorizations of all responses. The modal response in COMPLEX is to say things like "Click fast!" or "Do your best in both periods," i.e., strategies focused on going fast without reference to dynamic incentives, while the modal response in SIMPLE is about dynamic incentives.

If a lack of response to dynamic incentives reflects opacity, one might also expect the response to depend on the degree of worker bounded rationality, as captured by cognitive ability. We find this to be the case.

Finding 5. *Warehouse workers with higher CRT scores show a significantly stronger response to dynamic incentives. Other aspects of cognitive ability have limited explanatory power.*

Panels (a) and (b) of Figure G.3 show behavior of workers according to CRT scores. We see that those workers in COMPLEX who have relatively low CRT scores exhibit essentially no response to dynamic incentives, whereas workers who score higher do show signs of the zig-zag pattern characteristic of recognizing the dynamic incentives. In SIMPLE, even low CRT workers exhibit a response to dynamic incentives, and the response is much larger for workers with higher CRT scores.[42] To test statistically for the effect of CRT, we again use

---

[41] These fractions are lower than the corresponding fractions exhibiting a behavioral response to dynamic incentives. This may be because some individual workers who understand dynamic incentives intuitively are unable to describe this, or unwilling to go to the trouble to do so. It could also reflect some of the behaviors being classified as response to dynamic incentives as instead reflecting some other motive.

[42] In Figure G.4 in the appendix we show graphs for each level of CRT separately, which supports our binarization of the CRT score for Figure G.3. We only binarize CRT for the graphs and use the linear CRT score in all regressions.

difference-in-differences regression, now adding triple interactions of Period*Treatment*CRT score. The regression is shown in column 1 of Table G.4 in the appendix, and Panel (c) of Figure G.3 plots the coefficients of these triple interactions. We find that in COMPLEX, a higher CRT is associated with a significantly stronger drop in effort in Period 1 and 3, compared to STATIC, and the same is true in SIMPLE.[43] Higher CRT is also significantly positively correlated with the tendency to mention dynamic incentives in the open-ended question, in both COMPLEX and SIMPLE (Spearman correlations; $\rho = 0.21$, $\rho = 0.32$, $p = 0.012$, $p < 0.001$).

We also explore whether measures of other aspects of cognitive ability and worker traits matter for opacity (see columns 2–4 in Table G.4). Higher educational attainment, measured by years of schooling, is associated with significantly stronger responses to dynamic incentives in SIMPLE, but not in COMPLEX. Our indicators for the ability to do broad choice bracketing or to do backwards induction, are not significantly related to responses to dynamic incentives in any systematic way. Time preference is unrelated to responses to dynamic incentives in the online experiments, as expected given the short time frame, as is risk aversion.

4.3   Combining field and online experiments

We can combine the evidence from field and online experiments, since many of the workers in the online experiment also participated in the GROUP trial.[44] We can thus correlate their behavior in the field and in the online experiment.

Finding 6. *Workers who show a ratchet effect in the online experiment also show a significant ratchet effect in the warehouse.*

To classify which workers showed a ratchet effect in the online experiment (and to reduce measurement error), we conduct a principal component analysis of three variables: (1) a

---

[43]P-values of F-tests for joint significance of interactions with CRT in Periods 1 and 3: STATIC vs. COMPLEX $p = 0.045$, STATIC vs. SIMPLE $p < 0.001$, COMPLEX vs. SIMPLE $p = 0.19$. P-values of F-tests of interactions with CRT in Periods 2 and 4: STATIC vs. COMPLEX $p = 0.035$, STATIC vs. SIMPLE $p = 0.942$, COMPLEX vs. SIMPLE $p = 0.037$. P-values of F-tests of interactions with CRT in periods 3 and 4: STATIC vs. STATIC_ZERO: $p = 0.028$

[44]Only a handful participated in the INDIVIDUAL trial, as the vast majority of INDIVIDUAL workers joined the firm after the online experiments had been conducted, see Figure C.1.

dummy indicating whether the worker reduced effort in Period 1 relative to the baseline period; (2) a dummy indicating whether the worker reduced effort in Period 3 relative to the baseline period; and (3) a dummy indicating that the worker mentioned any arguments relating to ratchet effects or dynamic incentives in the open-ended question. The variable "showed RE online" is the standardized first principal component of these three variables.

Figure 5 plots the correlation between online and field behavior. Workers who show a ratchet effect online (median splitting the sample) also show a ratchet effect in the warehouse. Workers who do not show a ratchet effect online, by contrast, do not show a ratchet effect in the warehouse. The second panel depicts the estimated slope coefficients. The full regressions are in Table E.8 in the appendix. This table replicates Table E.4, but adds an interaction of treatment with a median split or the continuous version of the variable "showed RE online". All specifications demonstrate that showing a ratchet effect in the online experiment is significantly correlated with showing a stronger ratchet effect in the warehouse.[45]

The online experiments actually occurred after the GROUP trial, so most participants online had experience with dynamic incentives in the warehouse. The fact that most of these workers do not respond to dynamic incentives in the arguably simpler environment of the online experiments supports the conclusion that full transparency did not occur in the warehouse despite learning opportunities. We also find that recognizing dynamic incentives online is not significantly related to length of tenure in the warehouse ($p = 0.665$).[46]

5 Online experiments with AMT workers

We conducted additional experiments with AMT workers, to replicate and extend our findings (Abeler, Huffman, and Raymond (2025); all instructions and a summary of all treatments are available in Appendix J).

---

[45]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. Figure 5 and Table E.8 restrict the sample to participants in the COMPLEX treatment, since the design of COMPLEX is as close to the warehouse incentive scheme as possible. Many participants might understand incentives in SIMPLE, but not understand the incentives in the warehouse. However, we can also include participants in SIMPLE and we find the correlation between online and field behavior to be of similar size and equally significant.

[46]This result is based on an OLS regression of the indicator for recognizing dynamic incentives online on tenure in the warehouse.

Figure 5: Correlation of behavior in online and field experiment
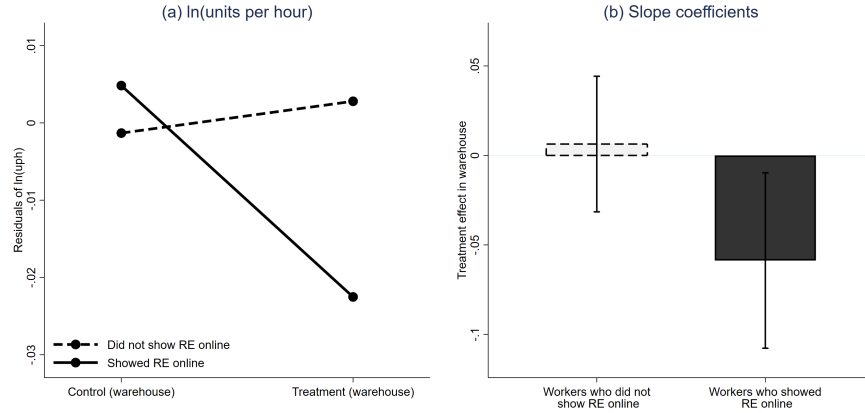
Notes: Panel (a) depicts ln(units per hour) (UPH) in the warehouse during the GROUP trial, separately for workers who showed (or did not show) a ratchet effect in the online experiments (median split), and separately for workers who were in the treatment or control group in the GROUP trial. The sample consists of the workers who participated in the GROUP trial and in the COMPLEX treatment in the online experiment. UPH are residualized for task, shift and cohort (and their interactions) fixed effects. Panel (b) plots the corresponding slope coefficients from the joint regression (see column 1 of Table E.8 in the appendix). Error bars show 95 percent confidence intervals.

5.1 Replication

We conducted the same four treatments as with the warehouse workers (albeit with slightly different payment and target rate parameters to account for differing wage expectations and ability among AMT workers). We added one treatment, STATIC_LOW, which implements a low but non-zero level of piece rate and which allows for point identification of our structural model (N = 571 across the five treatments). Notably, AMT workers have, on average, higher cognitive abilities than the warehouse workers, as captured by higher CRT scores, and higher educational attainment.

Finding 7. *AMT workers respond similarly to the treatments compared to warehouse workers. Average behavior is far from the rational optimum for COMPLEX, but close to the optimum for SIMPLE, based on a point-identified and estimated version of our model. The relationship of CRT to noticing dynamic incentives is also replicated. AMT workers do respond more strongly to dynamic incentives than warehouse workers, with a substantial portion of the difference explainable by higher CRT levels.*

Appendix H describes all of these results in detail. We report parameter estimates as well as the associated confidence intervals from a bootstrapping procedure, following the

methodology of DellaVigna and Pope (2018). Observed behavior is more than 400 clicks away from the optimum in COMPLEX and far outside the 95% CI, allowing us to reject statistically that behavior matches the optimum ($p < 0.001$). Observed behavior is very close, only 5 clicks away, from the optimum in SIMPLE, although it is just slightly outside of the CI (so we can reject that it is exactly optimal, $p < 0.001$). AMT workers also lose more utility in COMPLEX than in SIMPLE.[47] Thus, our conclusions from the warehouse, and online experiments with warehouse workers, are confirmed in online experiments with AMT workers, where our model is point identified and we can test statistically for differences from the rational benchmarks.

The findings imply that dynamic incentives can also be opaque for other worker populations besides warehouse workers, although opacity may be attenuated in more sophisticated workforces, holding complexity of the contract constant.

5.2 Identifying contract features that contribute to opacity

SIMPLE was designed to remove three features of COMPLEX that potentially contribute to opacity. In this section, we present two treatments ($N = 238$) that test which of these contract features contribute to the opacity of dynamic incentives in COMPLEX.[48]

The first treatment, NOISE, is the same as SIMPLE except that it adds one element found in COMPLEX: Noise in the Period 2 (Period 4) target rate. Specifically, while in SIMPLE the target rate is fixed at 400, in NOISE it is the average of 400 and a random variable $X \in \{380, 400, 420\}$. Importantly, our model makes the same prediction for Period 1 (Period 3) effort in SIMPLE and NOISE; in both cases, optimal effort is determined by intrinsic motivation, and is unaffected by noise in future target rates. Comparing the two treatments thus allows testing for an impact of noise in parameters that should be irrelevant if incentives are fully transparent, but which could matter if it contributes to complexity.

---

[47]The optimal effort levels are 0 and 204 in COMPLEX and SIMPLE, respectively. The observed effort levels are 425 and 211, the ratio metrics are 0.17 and 0.94, and the best-fitting ψ are 0.85 and 1, respectively. The utility losses relative to optimum are 15 percent and approximately 0 percent, respectively.

[48]One minor difference between COMPLEX and SIMPLE is that subjects could incur losses in SIMPLE, which were taken out of the show-up fee. In COMPLEX no such losses were possible. We conducted an additional treatment SIMPLE_NOLOSS which is identical to SIMPLE, except that no losses are possible. We find that this does not affect effort (F-test $p = 0.551$, see Figure H.8 and Table H.6).

The second treatment, NOISE_MARGINAL, builds on NOISE, but adds another element found in COMPLEX: Working fast in Period 1 now affects the *target rate* in Period 2 and thus the marginal incentives (like in COMPLEX), rather than directly affecting the level of earnings (like in SIMPLE). One implication is that responding optimally to dynamic incentives now involves a more complicated, non-separable optimization problem that potentially requires more complex contingent thinking.[49] Our model predicts optimal effort of about 200 in NOISE (the same as in SIMPLE), but 0 in NOISE_MARGINAL (the same as in COMPLEX). The difference in optima thus goes against the prediction of complexity, which is lower effort in NOISE than NOISE_MARGINAL in periods with dynamic incentives.

Turning to the comparison of NOISE_MARGINAL to COMPLEX, the only difference between these conditions is that COMPLEX frames dynamic incentives in terms of SPM rather than framing them directly in monetary terms. Our model thus predicts the same optimal effort in periods with dynamic incentives, 0, for the two conditions, since incentives are identical. This treatment comparison allows us to estimate the effect of the implicit, non-monetary framing of dynamic incentives, which should be irrelevant if incentives are transparent but could matter if this increases complexity and leads to opacity. We find that all three contract features matter.

Finding 8. *Noise in parameters, having dynamic incentives that involve future marginal incentives, and making financial consequences of dynamic incentives implicit, are all aspects of incentive contracts that can cause workers to neglect dynamic incentives.*

Panel (a) of Figure 6 shows effort across treatments and periods. The reaction to dynamic incentives is strongest in SIMPLE, and decreases monotonically going to NOISE, NOISE_MARGINAL and finally to COMPLEX. We again use difference-in-differences regressions to test across treatments. Panel (b) shows the coefficients of the relevant interaction terms. The full regression is in column 1 of Table H.3 in the appendix. Comparing responses to dynamic incentives across treatments, as captured by interaction terms for Periods 1 and 3, all treatment differences are statistically significant and magnitudes of the differential

---

[49]SIMPLE decouples the decision on current effort from thinking about how current effort would effect incentives for future effort. With n different effort levels in each period, the worker needs to search up to n^2 potential combination of effort levels in COMPLEX and NOISE_MARGINAL, while only $2n$ combinations in NOISE and SIMPLE (see Camara (2021)).

Figure 6: Contract features contributing to opacity, AMT workers
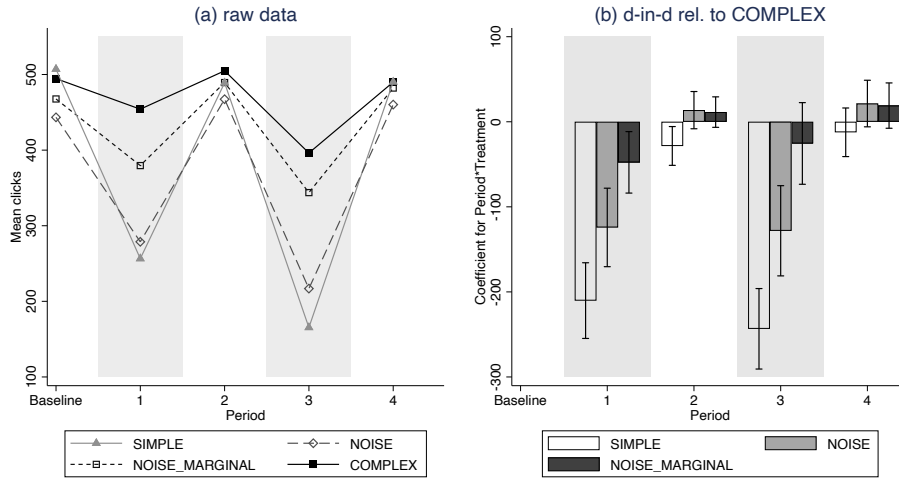
Notes: Panel (a) shows average clicks in a given work period. Shaded gray bars denote periods with dynamic incentives in all treatments. Panel (b) plots coefficients of interaction terms, Period*Treatment, from a difference–in-differences regression relative to baseline period and the treatment COMPLEX (see column 1 of Table H.3 in the appendix for all coefficients). Error bars show 95 percent confidence intervals.

responses to dynamic incentives are substantial.[50] The ratio metrics and best-fitting ψ for the treatments decrease (weakly) monotonically, as does the propensity to mention dynamic incentives, going from SIMPLE, to NOISE, to NOISE_MARGINAL, to COMPLEX.[51]

Appendix H.3 discusses additional treatments ($N = 369$) that extend our results on which contract features affect or do not affect opacity. We implement variations of our scheme that (1) eliminate the piece-wise linear nature of the incentives scheme as a function of SPM, (2) frame everything in terms of money (eliminating the need for SPM) and (3) do both of

---

[50]P-values for F-tests of interactions with Periods 1 and 3: SIMPLE vs. NOISE: $p < 0.001$, NOISE vs. NOISE_MARGINAL: $p < 0.001$, NOISE_MARGINAL vs. COMPLEX: $p = 0.028$. The coefficient estimates imply that response to dynamic incentives (drop in clicks relative to baseline, benchmarked against the drop in STATIC) is 30 percent less in COMPLEX than NOISE_MARGINAL (averaging across Periods 1 and 3). The corresponding response in NOISE is 32 percent less than in SIMPLE.

[51]Ratio metrics are 0.98, 0.85, 0.26, and 0.13, ψ values are 1, 1, 0.87, and 085, and fractions mentioning dynamic incentives are 0.80, 0.71, 0.58, and 0.40, respectively. We also investigated whether a contributor to greater or lesser opacity across the treatments could be word count, reading grade level, or ease of reading score for the corresponding instructions (see Table H.4 in the appendix for these statistics for all treatments). The reading level for COMPLEX is quite comparable to the reading level that we calculate for the actual communication materials that the firm used to explain the static incentives, and the two field experiments (as shown in Table H.4, all are roughly at reading grade level 7). It turns out that while SIMPLE has slightly more words than COMPLEX, it actually requires a higher reading grade level, and has a lower ease of reading score. Moreover, across all four treatments we do not see any systematic pattern between difficulty of reading, and behavior in our experiments, leading us to conclude that general ease of reading of instructions does not explain our treatment differences.

the preceding at once. We find that (1) and (2) have negligible effects on behavior, while under (3) workers do reduce their effort, but by a far smaller extent than in SIMPLE (or relative to the predicted optimum). These findings have important practical implications, as they show how various plausible changes to a widely used class of incentive schemes affect opacity of perverse dynamic incentives. Opacity is relatively robust, in the sense that it does not depend on one particular formulation of the scheme, although combining perturbations starts to lead to more transparency.

6   Conclusion

This paper provides empirical evidence for the importance of contract complexity, and heterogeneity in worker bounded rationality, for understanding optimal incentives within organizations. We show that dynamic incentives in the form of the so-called ratchet effect can be made opaque by contract complexity. We document specific contract features, and aspects of cognitive ability, which affect or do not affect the opacity of dynamic incentives. In our setting, opacity enhances efficiency and can lead to Pareto improvements.

More generally, our results indicate that incentive design should be sensitive to details that would not matter under standard theory, but have implications for complexity and opacity. This includes the specific way in which incentives are communicated, the details of the structure of incentives, the distribution of bounded rationality in the workforce, and potential opportunities and motives for learning about opaque attributes. Firms may use complexity to try and mitigate the emergence of ratchet effects, as well as other undesirable side-effects of contracts, or may regularly cycle through different contracts as workers learn about the details of the contract (e.g., as in Li, Mukherjee, and Vasconcelos (2021)). Obviously, complexity need not always be beneficial for firms, or efficiency.[52] Only effort-reducing aspects of contracts should be complex, while effort-enhancing aspects should stay simple.

Our findings suggest various directions for future research investigating the effect of complexity on other contract features beyond dynamic incentives, e.g., multi-tasking problems

---

[52]Anecdotally, it appeared from the initial discussions that some managers at the firm had not considered that dynamic contracts could induce the ratchet effect. This raises the possibility that both principal and agents may not fully take into account all aspects of the contracts they agree to.

(Holmstrom and Milgrom 1991). Understanding the optimal design of workplace incentive contracts when particular features induce opacity is an important area of research (e.g., Jehiel (2015); Ederer, Holden, and Meyer (2018)). Our structural estimation suggests that the relationship between the degree of opacity and the size of behavioral responses can be sensitive to the structure of the utility function of agents. Understanding how this plays out in other environments, such as taxation and consumer contracts, could lead to additional insights about the welfare and behavioral impacts of even relatively small degrees of opacity, and potentially explain persistent deviations from optimality.

References

Abaluck, Jason and Jonathan Gruber (2011). "Choice inconsistencies among the elderly: Evidence from plan choice in the Medicare Part D program". In: *American Economic Review* 101.4, pp. 1180–1210.

Abeler, Johannes, David Huffman, and Collin Raymond (June 2023). *Incentive Complexity, Bounded Rationality, and Effort Provision.* AEA RCT Registry. DOI: 10.1257/rct.11427-1.0.

— (2025). *Data for: Incentive Complexity, Bounded Rationality, and Effort Provision.* OpenICPSR number 217061.

Alan, Sule et al. (2018). "Unshrouding: Evidence from bank overdrafts in Turkey". In: *The Journal of Finance* 73.2, pp. 481–522.

Anonymous Firm (2025). *Data on the productivity of warehouse workers.* Proprietary dataset, unavailable to the public.

Aron-Dine, Aviva et al. (2015). "Moral hazard in health insurance: do dynamic incentives matter?" In: *Review of Economics and Statistics* 97.4, pp. 725–741.

Brañas-Garza, Pablo, Praveen Kujal, and Balint Lenkei (2019). "Cognitive reflection test: Whom, how, when". In: *Journal of Behavioral and Experimental Economics* 82, p. 101455.

Brown, Jennifer, Tanjim Hossain, and John Morgan (2010). "Shrouded attributes and information suppression". In: *The Quarterly Journal of Economics* 125.2, pp. 859–876.

Brown, Zach and Jihye Jeon (2023). "Endogenous Information and Simplifying Insurance Choice". In: *mimeo.*

Burks, S.V. et al. (2009). "Cognitive skills affect economic preferences, strategic behavior, and job attachment". In: *Proceedings of the National Academy of Sciences* 106.19, pp. 7745–7750.

Camara, Modibo (2021). "Computationally tractable choice". In: *Mimeo, Stanford University.*

Cardella, Eric and Briggs Depew (2018). "Output restriction and the ratchet effect: Evidence from a real-effort work task". In: *Games and Economic Behavior* 107, pp. 182–202.

Cattaneo, Matias D et al. (2024). "On binscatter". In: *American Economic Review* 114.5, pp. 1488–1514.

Chandra, Amitabh, Benjamin Handel, and Joshua Schwartzstein (2019). "Behavioral economics and health-care markets". In: *Handbook of Behavioral Economics.* Vol. 2. Elsevier, pp. 459–502.

Charness, Gary, Peter Kuhn, and Marie Claire Villeval (2011). "Competition and the ratchet effect". In: *Journal of Labor Economics* 29.3, pp. 513–547.

Chaudhuri, Ananish (1998). "The ratchet principle in a principal agent game with unknown costs". In: *Journal of Economic Behavior & Organization* 37.3, pp. 291–304.

Chetty, Raj, Adam Looney, and Kory Kroft (2009). "Salience and taxation: Theory and evidence". In: *American Economic Review* 99.4, pp. 1145–77.

Cheung, Stephen, Agnieszka Tymula, and Xueting Wang (2021). "Quasi-Hyperbolic Present Bias: A Meta-Analysis". In: *IZA Discussion Paper.*

Chevalier, Judith and Austan Goolsbee (2009). "Are durable goods consumers forward-looking? Evidence from college textbooks". In: *The Quarterly Journal of Economics* 124.4, pp. 1853–1884.

Clawson, Dan (1980). *Bureaucracy and the labor process: The transformation of US industry, 1860-1920.* NYU Press.

Cooper, David et al. (1999). "Gaming against managers in incentive systems: Experimental results with Chinese students and Chinese managers". In: *American Economic Review* 89.4, pp. 781–804.

Copeland, Adam and Cyril Monnet (2009). "The welfare effects of incentive schemes". In: *The Review of Economic Studies* 76.1, pp. 93–113.

Dalton, Christina M, Gautam Gowrisankaran, and Robert J Town (2020). "Salience, myopia, and complex dynamic incentives: Evidence from Medicare Part D". In: *The Review of Economic Studies* 87.2, pp. 822–869.

DellaVigna, Stefano and Devin Pope (2018). "What motivates effort? Evidence and expert forecasts". In: *The Review of Economic Studies* 85.2, pp. 1029–1069.

Dillen, Mats and Michael Lundholm (1996). "Dynamic income taxation, redistribution, and the ratchet effect". In: *Journal of Public Economics* 59.1, pp. 69–93.

Ederer, Florian, Richard Holden, and Margaret Meyer (2018). "Gaming and strategic opacity in incentive provision". In: *The RAND Journal of Economics* 49.4, pp. 819–854.

Einav, Liran, Amy Finkelstein, and Paul Schrimpf (2015). "The response of drug expenditure to nonlinear contract design: Evidence from Medicare Part D". In: *The Quarterly Journal of Economics* 130.2, pp. 841–899.

Eliaz, Kfir and Ran Spiegler (2006). "Contracting with diversely naive agents". In: *The Review of Economic Studies* 73.3, pp. 689–714.

Esponda, Ignacio and Emanuel Vespa (2019). "Contingent thinking and the sure-thing principle: Revisiting classic anomalies in the laboratory". In: *UCSB Working paper.*

Falk, Armin et al. (2023). "The preference survey module: A validated instrument for measuring risk, time, and social preferences". In: *Management Science* 69.4, pp. 1935–1950.

Frederick, Shane (2005). "Cognitive reflection and decision making". In: *Journal of Economic Perspectives* 19.4, pp. 25–42.

Gabaix, Xavier (2019). "Behavioral inattention". In: *Handbook of behavioral economics: Applications and foundations 1.* Vol. 2. Elsevier, pp. 261–343.

Gabaix, Xavier and David Laibson (2006). "Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets". In: *The Quarterly Journal of Economics*, pp. 505–540.

Gerardi, Dino and Lucas Maestri (2020). "Dynamic contracting with limited commitment and the ratchet effect". In: *Theoretical Economics* 15.2, pp. 583–623.

Goodman-Bacon, Andrew (2021). "Difference-in-differences with variation in treatment timing". In: *Journal of Econometrics* 225.2, pp. 254–277.

Grubb, Michael (2015). "Consumer inattention and bill-shock regulation". In: *The Review of Economic Studies* 82.1, pp. 219–257.

Grubb, Michael and Matthew Osborne (2015). "Cellular service demand: Biased beliefs, learning, and bill shock". In: *American Economic Review* 105.1, pp. 234–271.

Handel, Benjamin (2013). "Adverse selection and inertia in health insurance markets: When nudging hurts". In: *American Economic Review* 103.7, pp. 2643–82.

Handel, Benjamin and Joshua Schwartzstein (2018). "Frictions or mental gaps: What's behind the information we (don't) use and when do we care?" In: *Journal of Economic Perspectives* 32.1, pp. 155–178.

Hart, Oliver and Jean Tirole (1988). "Contract renegotiation and Coasian dynamics". In: *The Review of Economic Studies* 55.4, pp. 509–540.

Heidhues, Paul and Botond Kőszegi (2010). "Exploiting naivete about self-control in the credit market". In: *American Economic Review* 100.5, pp. 2279–2303.

Heiss, Florian et al. (2021). "Inattention and switching costs as sources of inertia in medicare part D". In: *American Economic Review* 111.9, pp. 2737–2781.

Herrnstein, Richard et al. (1993). "Utility maximization and melioration: Internalities in individual choice". In: *Journal of Behavioral Decision Making* 6.3, pp. 149–185.

Hirshleifer, David, Sonya Lim, and Siew Teoh (2009). "Driven to distraction: Extraneous events and underreaction to earnings news". In: *The Journal of Finance* 64.5, pp. 2289–2325.

Holmstrom, Bengt and Paul Milgrom (1991). "Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design". In: *Journal of Law, Economics, & Organization* 7, pp. 24–52.

Jehiel, Philippe (2015). "On transparency in organizations". In: *The Review of Economic Studies* 82.2, pp. 736–761.

Joseph, Kissan and Manohar U Kalwani (1992). "Do bonus payments help enhance salesforce retention?" In: *Marketing Letters* 3, pp. 331–341.

Keeler, Emmett B and John E Rolph (1988). "The demand for episodes of treatment in the health insurance experiment". In: *Journal of Health Economics* 7.4, pp. 337–367.

Laffont, Jean-Jacques and Jean Tirole (1988). "The dynamics of incentive contracts". In: *Econometrica* 56.5, pp. 1153–1175.

Lazear, E.P. (2000). "Performance Pay and Productivity". In: *The American Economic Review* 90.5, pp. 1346–1361.

Li, Jin, Arijit Mukherjee, and Luis Vasconcelos (2021). "Learning to game the system". In: *The Review of Economic Studies* 88.4, pp. 2014–2041.

Macartney, Hugh (2016). "The dynamic effects of educational accountability". In: *Journal of Labor Economics* 34.1, pp. 1–28.

Malcomson, James (2016). "Relational incentive contracts with persistent private information". In: *Econometrica* 84.1, pp. 317–346.

Markevich, Andrei and Ekaterina Zhuravskaya (2018). "The economic effects of the abolition of serfdom". In: *American Economic Review* 108.4-5, pp. 1074–1117.

Martinez, Isabel, Emmanuel Saez, and Michael Siegenthaler (2021). "Intertemporal labor supply substitution? Evidence from the Swiss income tax holidays". In: *American Economic Review* 111.2, pp. 506–546.

Martínez-Marquina, Alejandro, Muriel Niederle, and Emanuel Vespa (2019). "Failures in contingent reasoning: The role of uncertainty". In: *American Economic Review* 109.10, pp. 3437–74.

Mathewson, Stanley (1931). *Restriction of Output among Unorganized Workers.* The Viking Press New York.

Misra, Sanjog and Harikesh S Nair (2011). "A structural model of sales-force compensation dynamics". In: *Quantitative Marketing and Economics* 9, pp. 211–257.

Oprea, Ryan (2020). "What makes a rule complex?" In: *American Economic Review* 110.12, pp. 3913–51.

Oyer, Paul (2000). "A theory of sales quotas with limited liability and rent sharing". In: *Journal of Labor Economics* 18.3, pp. 405–426.

Rabin, Matthew and Georg Weizsäcker (2009). "Narrow bracketing and dominated choices". In: *American Economic Review* 99.4, pp. 1508–1543.

Raju, Jagmohan and V Srinivasan (1996). "Quota-based compensation plans for multiterritory heterogeneous salesforces". In: *Management Science* 42.10, pp. 1454–1462.

Roy, Donald (1952). "Quota restriction and goldbricking in a machine shop". In: *American Journal of Sociology* 57.5, pp. 427–442.

Shearer, Bruce (2022). "Piece-rate cuts and ratchet effects". In: *Canadian Journal of Economics/Revue canadienne d'économique* 55.3, pp. 1371–1403.

Stefansson, Arnaldur (2019). "Labor supply response to a tax holiday: The take-home from a large and salient shock". In: *PhD diss. Uppsala University.*

Taubinsky, Dmitry and Alex Rees-Jones (2018). "Attention variation and welfare: theory and evidence from a tax salience experiment". In: *The Review of Economic Studies* 85.4, pp. 2462–2496.

Weitzman, Martin L (1980). "The "ratchet principle" and performance incentives". In: *The Bell Journal of Economics* 11.1, pp. 302–308.

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

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

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

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

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

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 η. This allows for the existence of an optimal policy.

4

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. $\square$

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. $\square$

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.

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

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 $\beth = 1, ..., J$ over time periods $t = 1, 2$. The worker spends hours $H_{i\beth t}$ on task $\beth$ out of a total of H hours. $H_{i\beth t}$ 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 $\tilde{e}_{i\beth t}$ to produce output $\tilde{e}_{i\beth t}H_{i\beth t}$. $\eta_{i\beth t}$ 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 $H_{i\beth t}$ 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)

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

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

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 $\hat{\jmath}$.

$$aH_{\hat{\jmath}1} + \frac{H_{\hat{\jmath}1}}{d_{\jmath}\eta_{i,\hat{\jmath},1}}w\hat{g}'(\sum_{\jmath}\frac{H_{\jmath1}e_{\jmath1}}{d_{\jmath}\eta_{\jmath,1}}) - H_{\hat{\jmath}1}c'(\sum_{\jmath}H_{\jmath1}e_{\jmath1}) - $$
$$\delta\varsigma\frac{H_{\hat{\jmath}2}e_{\hat{\jmath}2}}{d_{\hat{\jmath}}^2(\varsigma e_{\hat{\jmath}1}+(1-\varsigma)r_{\hat{\jmath}})^2}w\hat{g}'(\sum_{\jmath}\frac{H_{\jmath2}e_{\jmath2}}{d_{\jmath}(\varsigma e_{i\jmath1}+(1-\varsigma)r_{\jmath})}) = 0$$

or, rewritten

$$aH_{\hat{\jmath}1} + \frac{H_{\hat{\jmath}1}}{d_{\jmath}\eta_{\hat{\jmath},1}}w\hat{g}'(\sum_{\jmath}\frac{H_{\jmath1}e_{\jmath1}}{d_{\jmath}\eta_{\jmath,1}}) = H_{\hat{\jmath}1}c'(\sum_{\jmath}H_{\jmath1}e_{\jmath1}) + \delta\varsigma\frac{H_{\hat{\jmath}2}e_{\hat{\jmath}2}}{d_{\hat{\jmath}}^2(\varsigma e_{\hat{\jmath}1}+(1-\varsigma)r_{\hat{\jmath}})^2}w\hat{g}'(\sum_{\jmath}\frac{H_{\jmath2}e_{\jmath2}}{d_{\jmath}(\varsigma e_{\jmath1}+(1-\varsigma)r_{\jmath})})$$

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 $aH_{\hat{\jmath}1} + \frac{H_{\hat{\jmath}1}}{d_{\jmath}\eta_{\hat{\jmath},1}}w\hat{g}'(\sum_{\jmath}\frac{H_{\jmath1}e_{\jmath1}}{d_{\jmath}\eta_{\jmath,1}}) = H_{\hat{\jmath}1}c'(\sum_{\jmath}H_{\jmath1}e_{\jmath1})$. 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

12

i. Suppose that at the optimum for Treatment, $H_1 e_1$ 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 $H_1 e_1$ 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 $a = c'(\sum_\beth H_{\beth 1} e_{\beth 1})$, which implies that the agent is indifferent between any allocation of efforts that provides the same $H_1 e_1$, 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

13

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

14

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:

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

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

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

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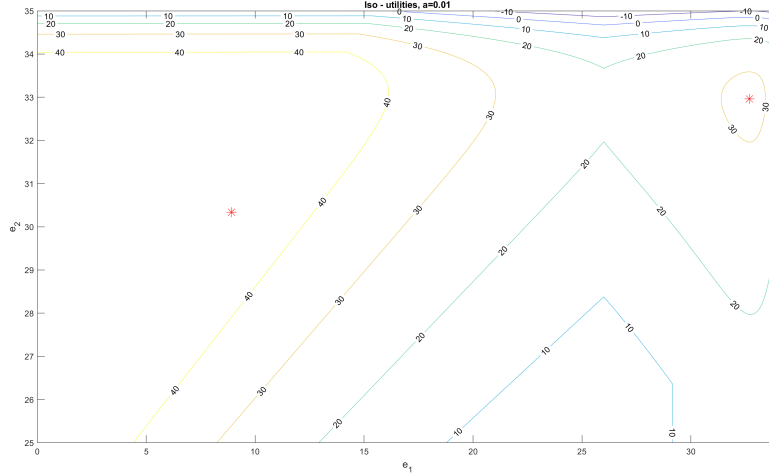
Notes: The graph shows the worker's iso-utility curves depending on effort choices in Periods 1 and 2 for a=0.01. Red stars denote local maxima.

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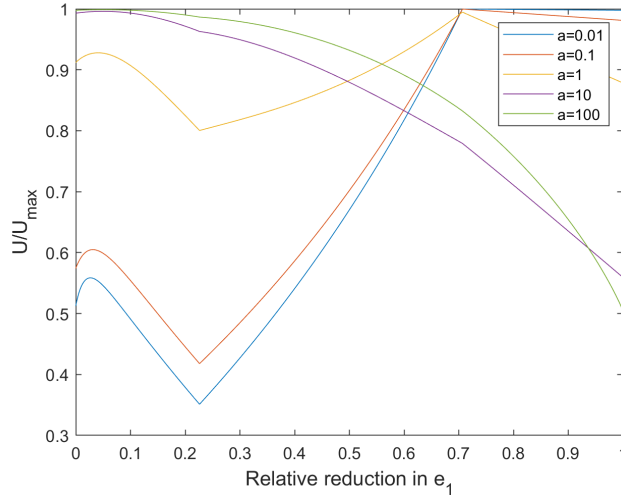
Notes: The graph shows worker utility (normalized by the utility achieved at the optimal Period-1 and Period-2 effort) for all Period-1 effort levels. Period-2 effort is fixed at the level we observe in the data. The horizontal axis shows reduction in Period-1 effort relative to Period-2 effort.

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

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

The two period optimization problem is now:

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

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:

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

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:

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

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

33

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

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

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)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
a	p	$\hat{\theta}$	ρ_1^*	$e_1^*(\rho_1^*)$	$e_2^*(\rho_1^*)$	TS_1 Actual	TS_2 Actual	TS_1 Predicted	TS_2 Predicted	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)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
a	$Profit_1$ Actual	$Profit_2$ Actual	$Profit_1$ Predicted	$Profit_2$ Predicted	$Profit$ Act-Pred	$Utility_1$ Actual	$Utility_2$ Actual	$Utility_1$ Predicted	$Utility_2$ Predicted	$Utility$ 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]

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

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]

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

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

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

In the experiment, $r = \eta$.

In SIMPLE the optimization problem becomes

$$\max_{e_1, e_2} \quad ae_1 + g\left(e_1\right) - \theta \frac{e_1^{\gamma+1}}{\gamma+1}$$
$$+ \quad ae_2 + [g\left(e_2\right) - g\left(e_1\right)] - \theta \frac{e_2^{\gamma+1}}{\gamma+1}$$
$$= \max_{e_1, e_2} \quad ae_1 - \theta \frac{e_1^{\gamma+1}}{\gamma+1} + ae_2 + g\left(e_2\right) - \theta \frac{e_2^{\gamma+1}}{\gamma+1}$$

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) a	(2) γ	(3) θ	(4) e_1 Predicted	(5) e_2 Predicted	(6) Utility Predicted	(7) e_1 Actual	(8) e_2 Actual	(9) Utility Actual	(10) e_1 Diff Ratio	(11) e_2 Diff Ratio	(12) 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 under-perceives 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
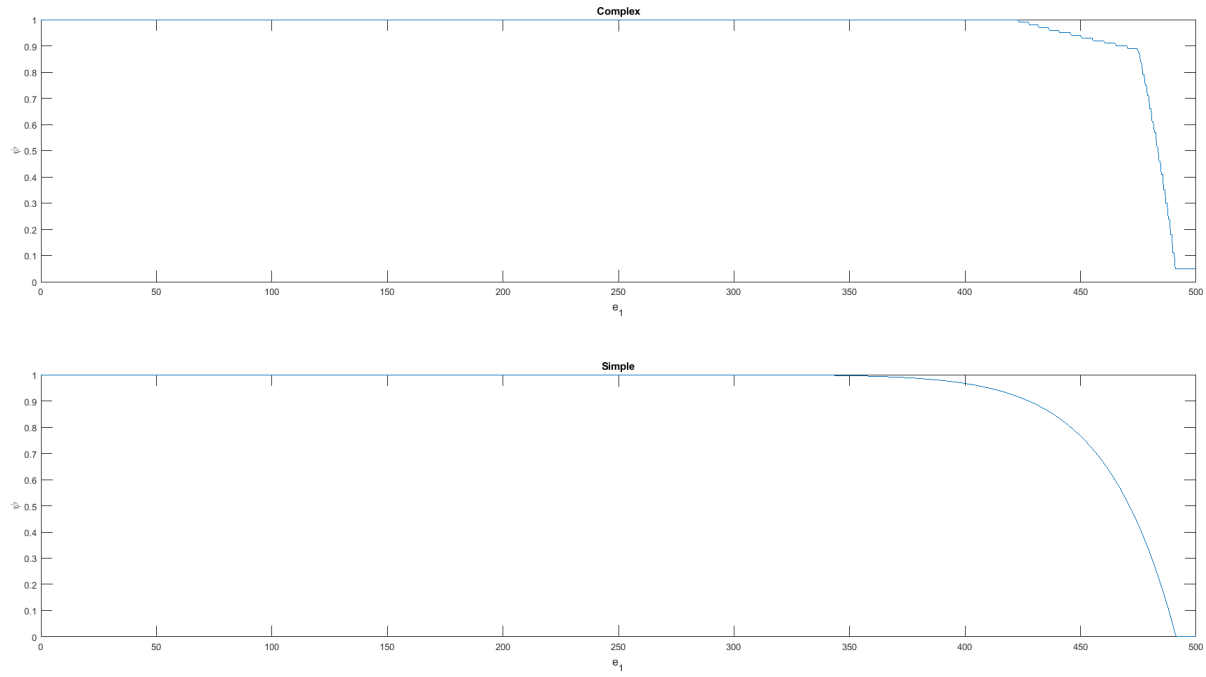
Notes: The graph shows the relationship between Period-1 effort on the x-axis and ψ on the y-axis, fixing Period-2 effort at the level we observe in the data. The top panel shows the relationship for the COMPLEX treatment and the bottom panel the one for the SIMPLE treatment.

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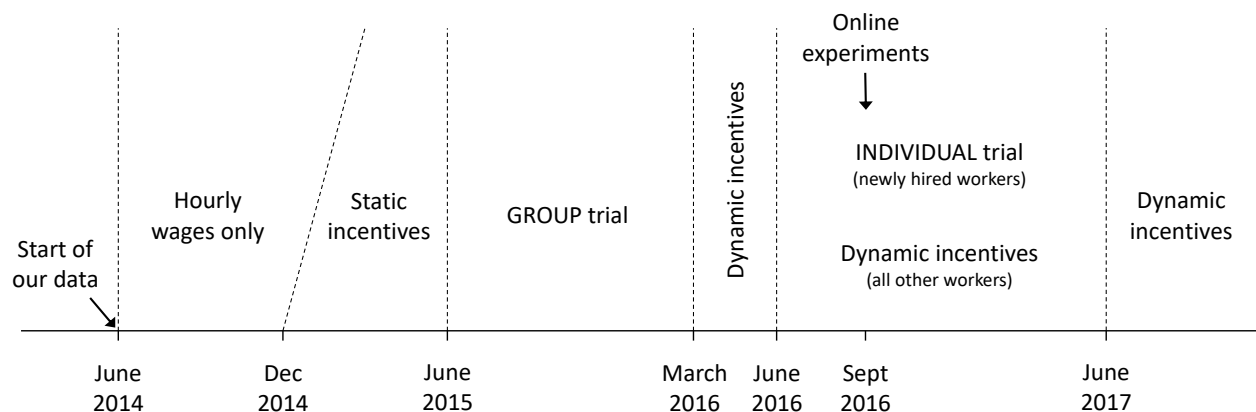
Notes: A timeline of the changes to the incentive scheme in the treated warehouse. Our data start in June 2014. Static incentives were gradually introduced starting in December 2014, with roll-out completed by March 2015.

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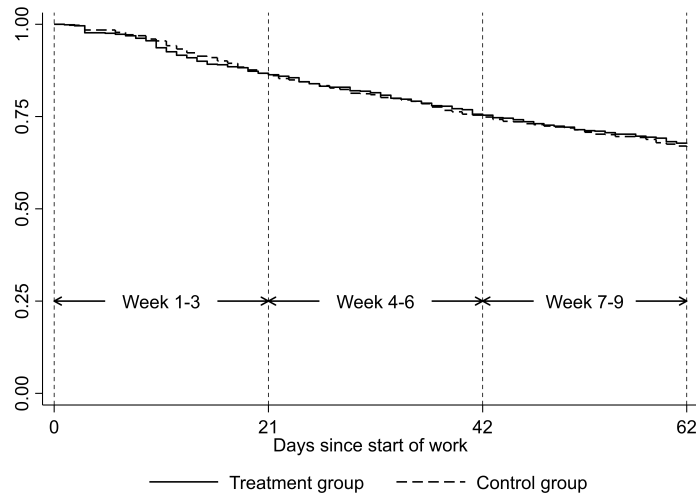
Notes: Kaplan-Meier survival estimates for the INDIVIDUAL trial. The vertical lines show the start and end of the treatment period (weeks 4–6). Corresponding regressions are in Table C.2.

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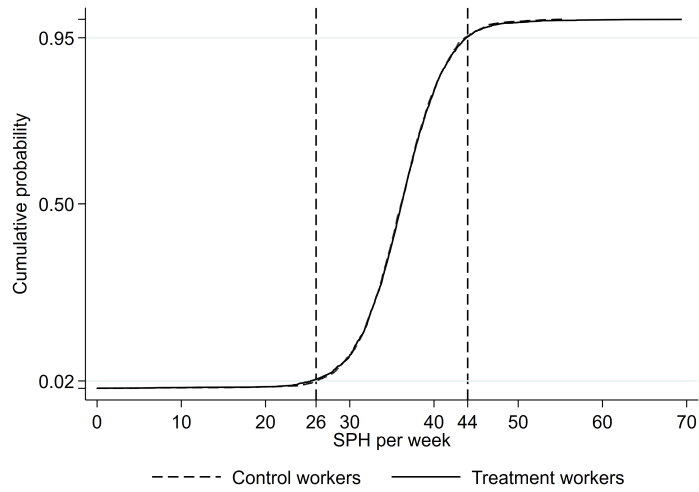
Notes: Distribution of effort, measured as the sum of normalized effort (SPH) per week, in the INDIVIDUAL trial. Vertical dashed lines indicate the quota and cap of 26 SPH and 44 SPH respectively. Marginal incentives were positive between the quota and cap, and zero elsewhere. Horizontal light gray lines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

Figure C.4: Treatment effect in the INDIVIDUAL trial by task
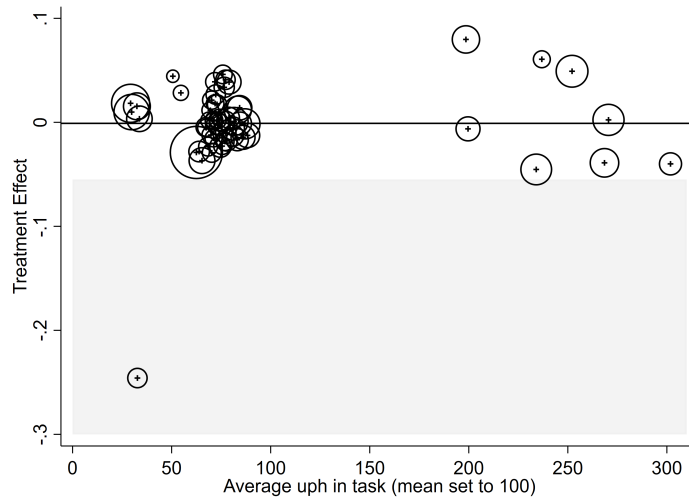
Notes: The graph plots the treatment effect in the INDIVIDUAL trial for each task separately. The size of each bubble is proportional to the total time spent in that task. Tasks are sorted by the average number of units per hour (uph) achieved by workers. The solid line marks the overall treatment effect (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)
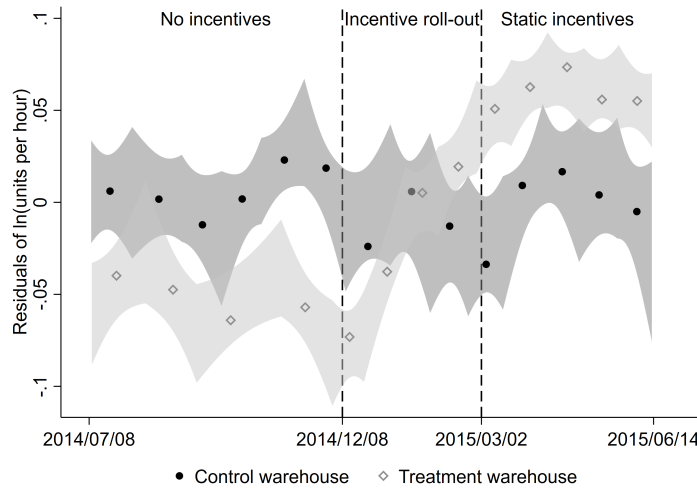
Note: Binscatter graph of the residuals of ln(units per hour) in the treated and the control warehouse, binned by week. The incentives were rolled out, task by task, between 8 December 2014 and 2 March 2015, for the treated warehouse and were always present in the control warehouse. The graph corresponds 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.

64

Figure D.3: Effect of static incentives and observed ratchet effect by task
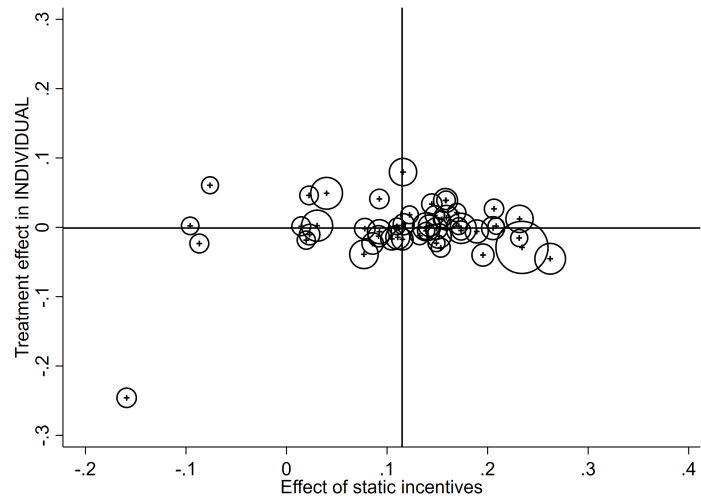
Notes: Scatter plot of the effect of static incentives (measured during the roll-out of incentives) and the observed ratchet effect (measured during the INDIVIDUAL trial). Each bubble corresponds to one task. The size of the bubble is proportional to the amount worked in this task during the INDIVIDUAL 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
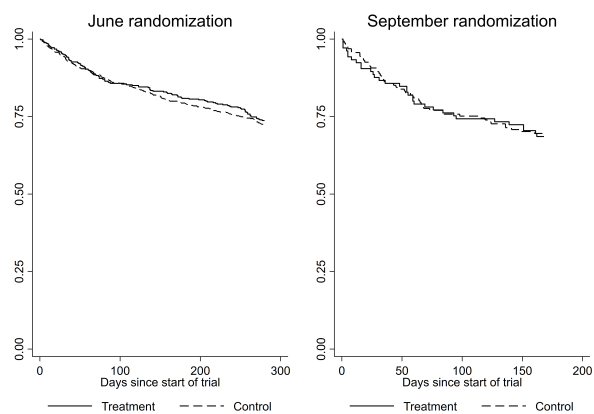
Notes: Kaplan-Meier survival estimates for the GROUP trial, shown separately for the two randomization cohorts.

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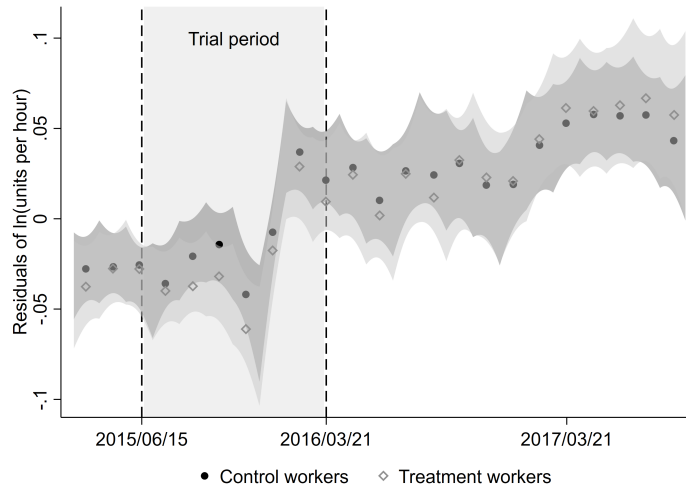
Notes: The figure shows a time series plot of worker effort for treatment and control workers. Both groups of workers face the same static incentives and treatment workers additionally face a dynamic (ratchet) incentive to reduce effort during the trial period. Effort is residualized for task fixed effects.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

71

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
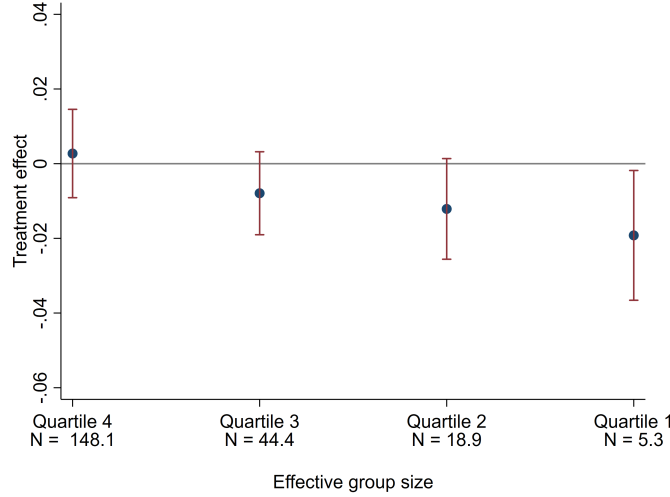
Notes: The graph plots the treatment difference on ln(units per hour), i.e., the ratchet effect, by effective group size. We calculate each treated worker's share of the time worked by all treated workers in a given task, for each rate-setting period. 1 divided by this share is the effective group size 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.	Online exp.	Online exp.
	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 × 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 (0.019)	0.0067 (0.018)	0.0067 (0.018)	-0.0327 (0.023)	-0.0375 (0.025)	-0.0375 (0.025)
Showed RE online (dummy) × treated in warehouse	-0.0651 (0.032)	-0.0876 (0.029)	-0.0876 (0.029)			
Showed RE online (PCA) × treated in warehouse				-0.0635 (0.021)	-0.0658 (0.023)	-0.0658 (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.

78

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
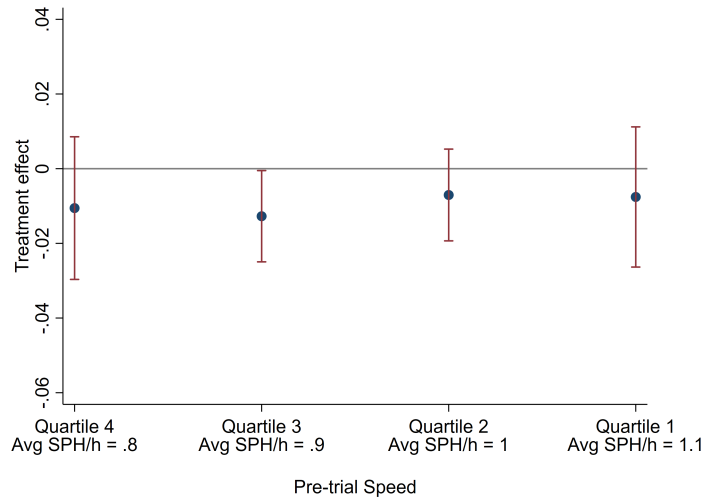
Notes: The graph plots the treatment difference on ln(units per hour), i.e., the ratchet effect, by workers' pre-trial speed. We calculate each worker's speed in the period between the roll-out of static incentives and the start of the trial and split workers into quartiles. Point estimates are from 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
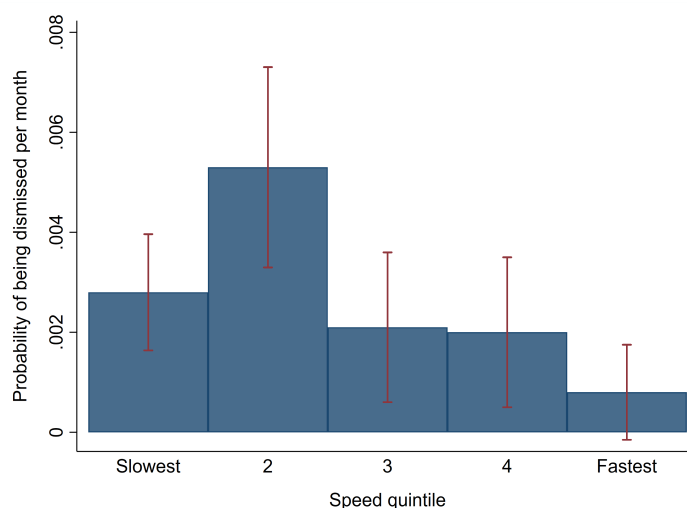
Notes: The graph shows the probability of being dismissed per month, split by worker speed. The graph only contains dismissals for unspecific, and thus potentially speed-related, reasons. All workers are divided into five quintiles based on their average speed in the last 26 weeks before being dismissed. 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
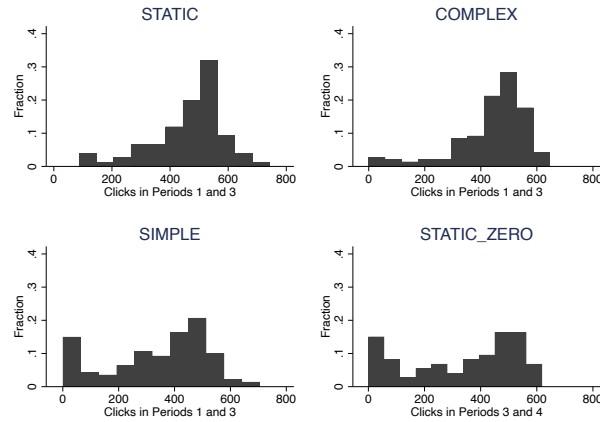
Notes: Each panel of the figure shows the distribution of clicks observed in the corresponding treatment for the specified periods. One outlier observation in STATIC is not shown in the graph.

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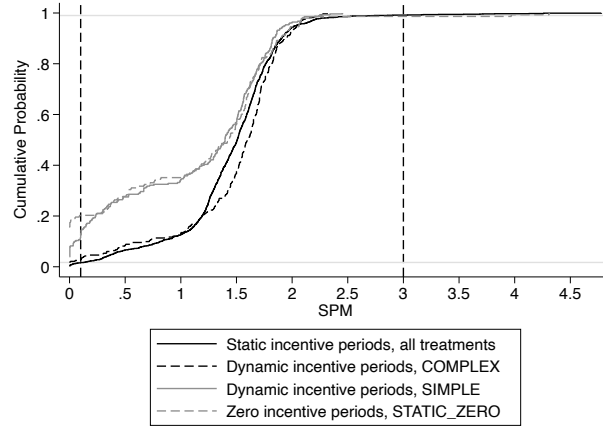
Notes: Vertical dashed lines indicate the quota and cap of 0.1 SPM and 3 SPM, respectively. Horizontal light gray lines indicate the fractions of observations below the quota, or below the cap, 0.015 and 0.99, respectively, for static incentive periods.

Figure G.3: Opacity of dynamic incentives and CRT, warehouse workers online
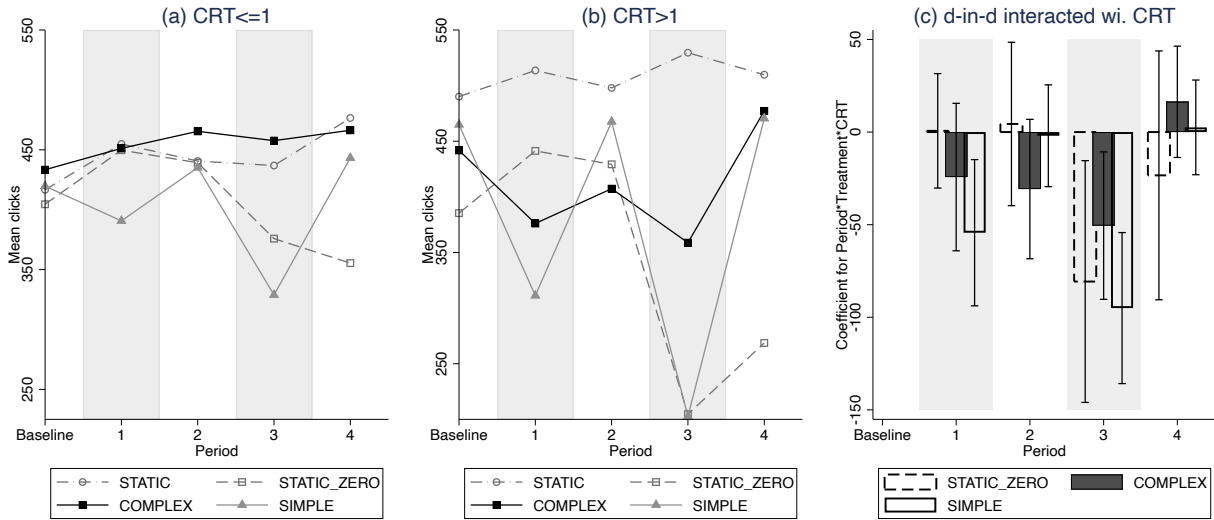
Notes: Panels (a) and (b) show average clicks in a given work period for workers with CRT ≤ 1 and CRT > 1, respectively. Shaded gray bars denote periods with dynamic incentives in COMPLEX and SIMPLE. Panel (c) plots coefficients of interaction terms, Period*Treatment*CRT, from a difference-in-differences 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
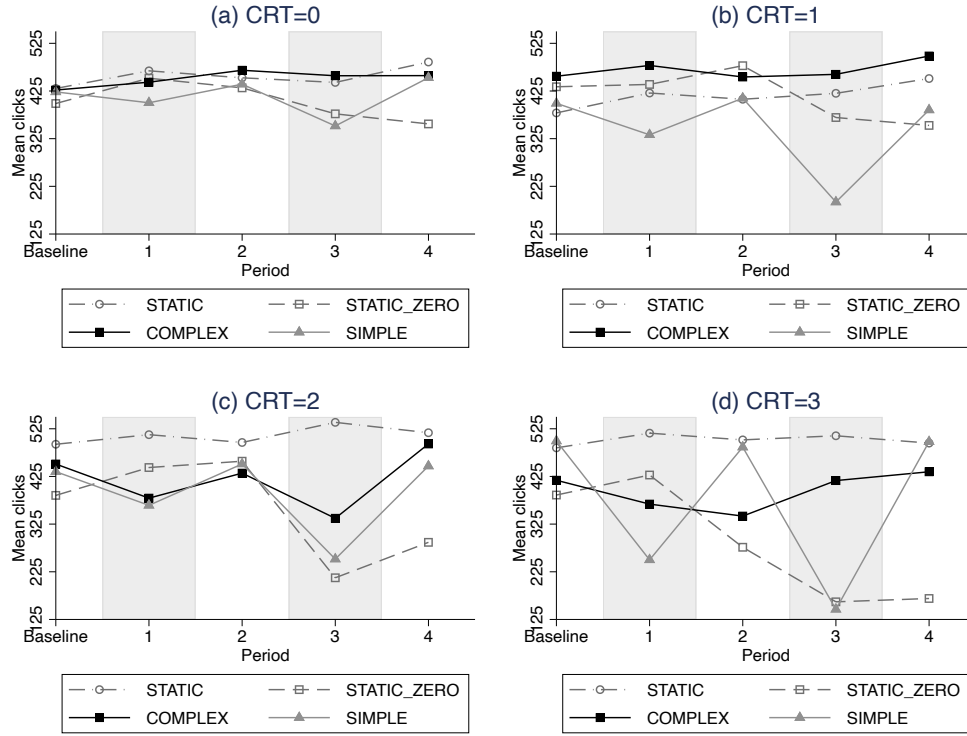
Notes: Each panel shows the average number of clicks in a given work period for workers with a given CRT score. The vertical shaded bars denote periods with dynamic incentives to reduce effort in COMPLEX and SIMPLE. Note that we have very few observations for panel (d): 6 workers in STATIC, 2 in STATIC_ZERO, 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
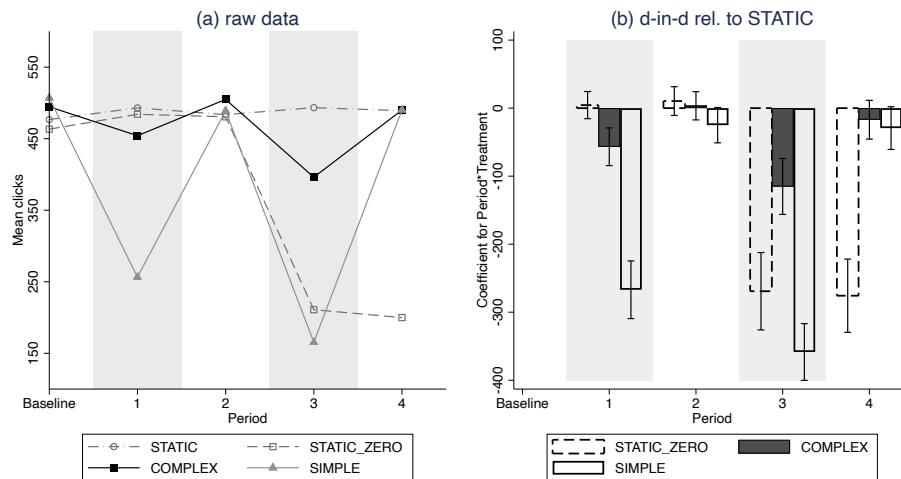
Notes: Panel (a) shows average number of clicks in a given work period. Panel (b) plots coefficients of interaction terms, Period*Treatment, from a difference-in-differences regression relative to baseline period and the treatment STATIC (see column 2 of Table G.2 for all coefficients). The vertical 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
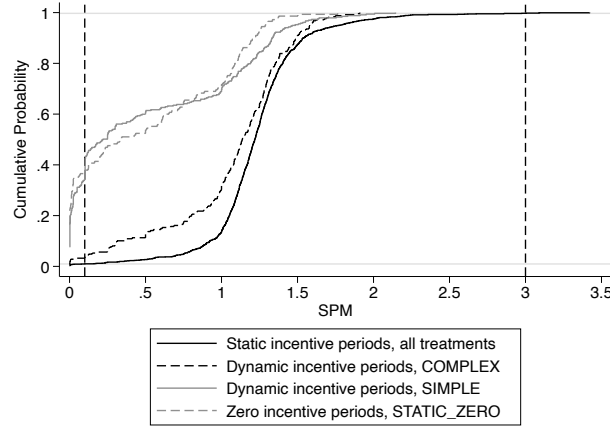
Notes: Vertical dashed lines indicate the quota and cap of 0.1 SPM and 3 SPM, respectively. Horizontal light gray lines indicate the fractions of observations below the quota, or below the cap, 0.009 and 0.997, respectively, for static incentive periods.

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
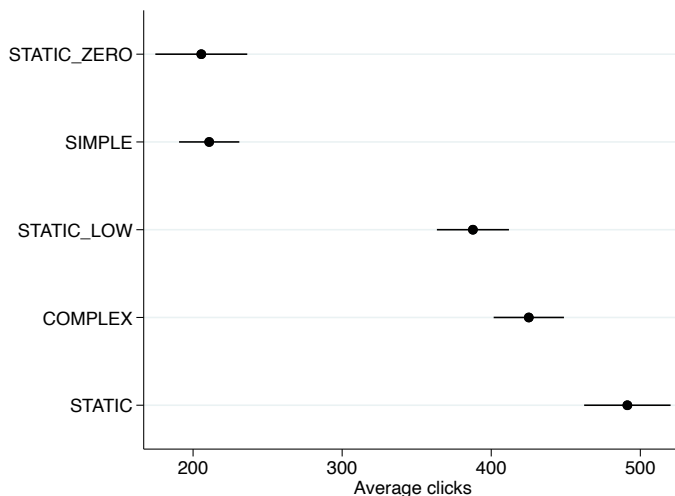
Notes: The figure shows average clicks and 95% C.I.'s. For STATIC_ZERO, STATIC_LOW, and STATIC, the averages are from periods with across-treatment variation in the piece rate (Periods 3 and 4), where piece rates are zero, $0.01, or $0.50, respectively. For comparison, the figure shows results from COMPLEX 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
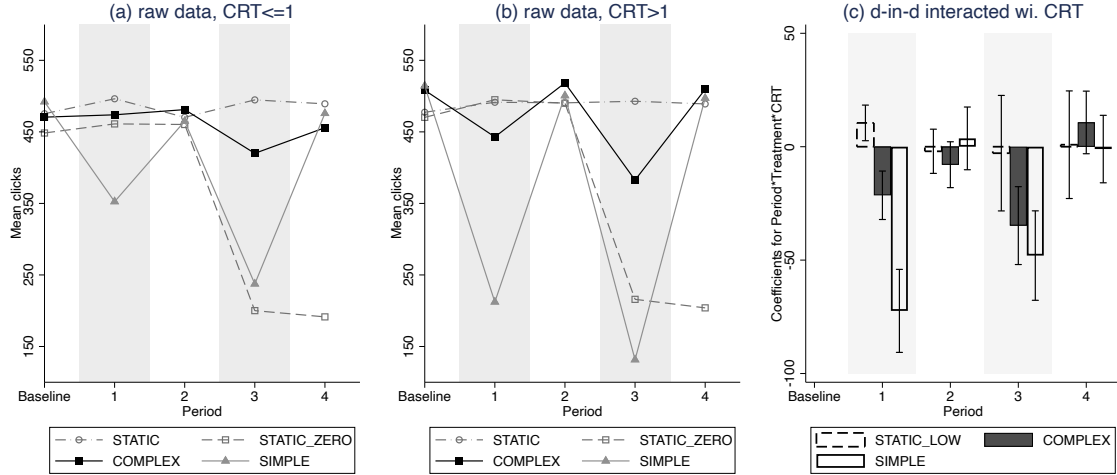
Notes: Panels (a) and (b) show the average number of clicks in a given work period for workers with CRT ≤ 1 and CRT > 1, respectively. Panel (c) plots coefficients of interaction terms, Period*Treatment*CRT, from a difference-in-differences regression relative to baseline period and the treatment STATIC (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
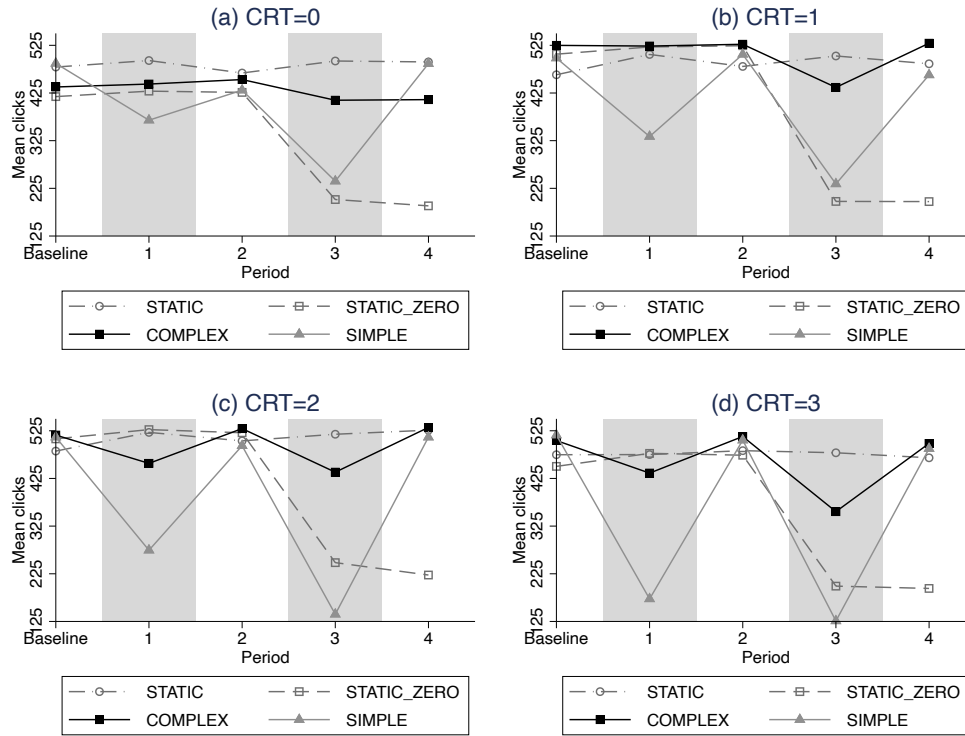
Notes: Each panel shows the average number of clicks in a given work period for workers with a given CRT score. The vertical shaded bars denote periods with dynamic incentives to reduce effort in COMPLEX and SIMPLE.

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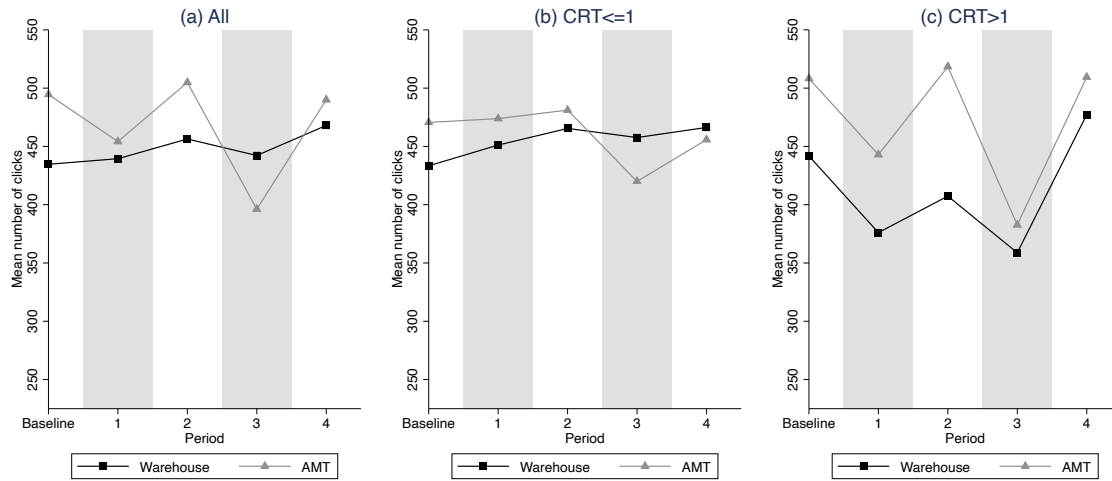
Notes: Panel (a) shows the average number of clicks in a given work period for all warehouse and AMT workers in COMPLEX. Panels (b) and (c) compare warehouse and AMT workers who have CRT $\leq$ 1, and CRT > 1, respectively.

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
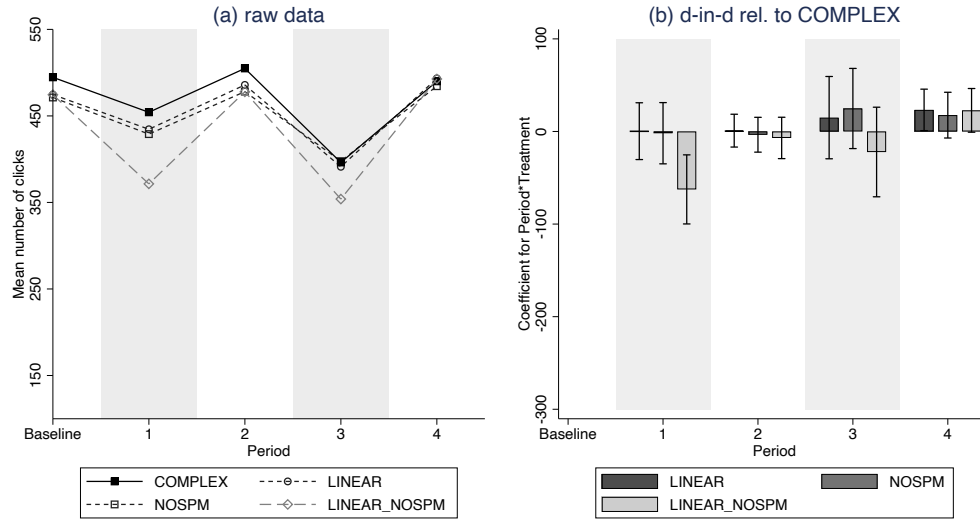
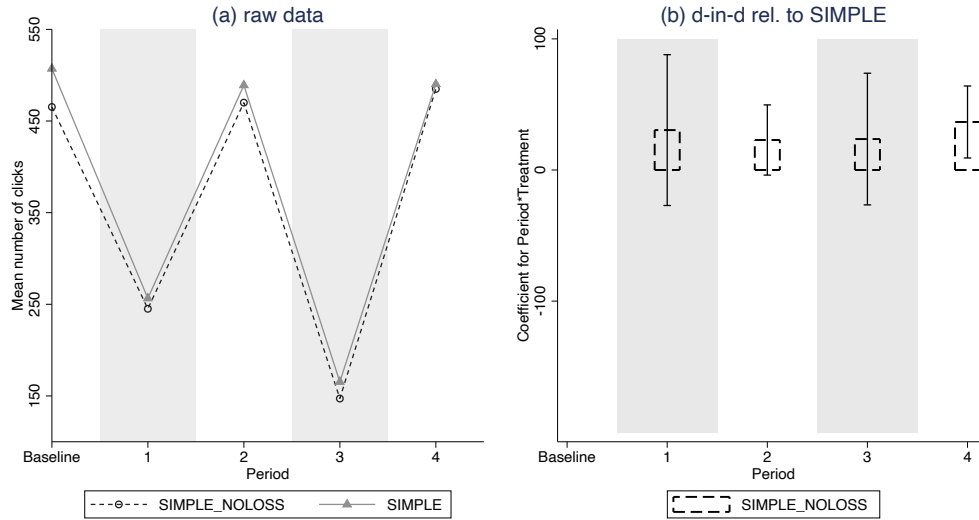Notes: Panel (a) shows average number of clicks in a given work period. Panel (b) plots coefficients of interaction terms, Period*Treatment, from a difference-in-differences regression relative to baseline period and the treatment COMPLEX (see column 1 of Table H.5 for all coefficients). The vertical 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

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?
- o Yes
- o No

I have read and understand the information above.
- o Yes
- o No

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

How does it work?
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.

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.
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.

How do you get SPM?

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

How do you get money?

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]

Iteration	Q. number	Early payment	Delayed payment	Skip logic: Go to Q x If choose Early	If choose 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]

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. **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. 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?

- o A sure gain of $3.60

- o 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?

- o A sure loss of $11.25

- o 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:**
   o We **subtract** any money you earn in round 1 from your money in round 2.
   o 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

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 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

131

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	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.

*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.

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.

*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**.**

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

*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

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

How do we set the wage per click in round 2?

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

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. 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

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. 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.

*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 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.