Incentive Complexity, Bounded Rationality and Effort Provision*

Johannes Abeler, David Huffman, Collin Raymond

February 15, 2023

Abstract

Preliminary version, please don't circulate.

We test whether complexity of workplace incentive schemes, and worker bounded rationality, can affect effort provision, using a combination of field and laboratory experiments. The paper has four main results. (1) Complexity of workplace incentives affects effort provision, and in the specific case class of incentive schemes we consider, actually increases effort by shrouding a perverse dynamic incentive to provide low effort. (2) Heterogeneity in a particular aspect of worker bounded rationality, captured by the Cognitive Reflection Test (CRT), matters for recognizing shrouded attributes and thus effort responses. (3) Contract features that contribute to shrouding include include largely irrelevant features of the contract acting as distractors; the explanation of incentives leaving the monetary consequences of the shrouded attribute implicit; and response to the shrouded attribute requiring relatively complicated contingent thinking. (4) Shrouding can be robust, remaining intact despite significant perturbations to the structure and communication of the incentive scheme.

JEL Codes: D8, D9, J2, J3

Keywords: Complexity; Bounded Rationality; Shrouded Attribute; Field Experiments

*Abeler: University of Oxford, IZA and CESifo (e-mail: johannes.abeler@economics.ox.ac.uk); Huffman: University of Pittsburgh, IZA and CESifo (e-mail: huffmand@pitt.edu); Raymond: Cornell University (e-mail: collinbraymond@gmail.com). JA thanks the ESRC for financial support under grant ES/R011710/1. We thank the company for making the field experiments possible. We thank Steffen Altmann, Roland Benabou, Stefano Caria, Gary Charness, Stefano DellaVigna, Benjamin Enke, Armin Falk, Henry Farber, James Fenske, Daniel Garrett, Uri Gneezy, Peter Kuhn, David Laibson, Jim Malcomson, Meg Meyer, Rani Spiegler, Dmitry Taubinsky, Roberto Weber and others for helpful discussions. Many valuable comments were also received from numerous seminar and conference participants. Samuel Altmann, Andreea Antuca, Alexander Brehm, Artur Doshchyn, Antonio Greco, Aastha Ladani, Diego Lame, Luke Milsom, Leon Musolff, Shusaku Nishiguchi, Sofia Palacios, Alvaro Salamanca, Segye Shin, Kiriaki Shinas, Lu Wang,Dilan Yang and Whenzhang Zhang provided outstanding research assistance. Ethical approval was obtained from the Department of Economics Research Ethics Committee of the University of Oxford (No. 2014-33).

1 Introduction

Traditional contract theory assumes that workers are fully rational and perfectly understand the incentives they face. Complexity seems highly relevant for labor contracts, however, as these often involve nonlinearities, complicated dynamics, and nontrivial optimization problems. Workers are also, to varying degrees, bounded rational, as reflected in different cognitive abilities and education levels. Our theoretical understanding of complexity has advanced, with recent models positing that, as contracts become complex, boundedly rational individuals deviate from the optimum because the implications of certain aspects are not recognized, i.e., are "shrouded" (e.g., Gabaix and Laibson (2006); Eliaz and Spiegler 2006; DellaVigna 2009). So far, however, there is little empirical evidence on implications for labor contracts. In this paper we test the hypotheses that complexity of workplace incentive schemes, and worker bounded rationality, can affect effort provision, and we seek to understand specific contract features and aspects of cognitive ability that matter for shrouding.[1]

Our paper provides four main takeaway sets of results. (1) We show that the complexity of workplace incentives can matter for effort provision, and in the specific case we consider, actually *increase* effort by shrouding a perverse dynamic incentive to provide low effort.[2] (2) We demonstrate that heterogeneity in a specific facet of worker bounded rationality, as captured by the Cognitive Reflection Test (CRT), can matter for effort responses because it affects recognizing shrouded attributes.[3] (3) Contract features that contribute to shrouding

[1]There are various definitions of complexity of a contract (see Oprea (2020) and Jakobsen (2020) for recent contributions) as well as different proposals as to what makes a decision-maker boundedly rational. Given the empirical focus of our paper, we take a primarily behavior-based approach to defining both these terms. Fixing an individual, and their experience with the contract, we define one contract as more complex than another, the closer is behavior to acting as though a potentially shrouded attribute does not exist, and the further is behavior from the optimum of fully rational effort provision. In some of our experiments we complement these with a non-behavior based metric: Whether workers mention the shrouded attribute when explaining what they see as the optimal response to the contract. Similarly, fixing a contract, and experience with the contract, we define one individual as more boundedly rational than another if their behavior is closer to the fully shrouded benchmark, further from fully rational effort provision, and if the worker does not mention the shrouded attribute. Part of the contribution of the paper is to relate various observables about the contract, and individual, to these metrics.

[2]Shrouding could reflect various mechanisms. For example, it could be due to inattention, such that individuals just do not think of the attribute (Ellison, 2005; Gabaix and Laibson (2006)). Alternatively, it might be a more conscious choice to not reason through the implications of an attribute, because doing so is perceived to be mentally costly (see, e.g., models of endogenous depth of reasoning such as Aloui and Penta, 2021). While this paper will shed light on specific features of incentives, and worker cognitive ability, that contribute to shrouding, it will not precisely distinguish between these types of underlying psychological mechanisms.

[3]The CRT is a simple test designed to capture an individual's tendency to think carefully about a problem,

include include largely irrelevant features of the contract acting as distractors; the explanation of incentives leaving the monetary consequences of the shrouded attribute implicit; and response to the shrouded attribute requiring relatively complicated contingent thinking. (4) We provide evidence that shrouding can be robust, remaining intact in the face of significant, realistic perturbations to the structure and communication of incentives.

We arrive at our findings in the context of studying whether a challenge that arises in many forms of performance pay schemes – perverse dynamic incentives to reduce effort in the form of the so-called "ratchet effect" (Weitzman (1980), Laffont and Tirole (1988)) – can be a shrouded attribute. This problem is ubiquitous because it arises whenever workers have private information about task difficulty. Firms have an incentive to use workers' current outputs to calibrate future marginal incentives, but this leads workers to have a motive to reduce effort initially, to have more generous incentives in the future, the ratchet effect.[4] An important theoretical result is that this type of dynamic incentives can lead to very inefficient outcomes, because high ability types pool with low ability types and reduce total surplus (Laffont and Tirole (1988)). We hypothesize, however, that such dynamic incentives might not be fully taken into account by boundedly rational workers, particularly when embedded in an already complex incentive scheme. This could mean that complexity prevents reductions in effort, and potentially increases efficiency.

We focus on a specific, but widely-used class of real-world incentive schemes, which we argue is inherently complex. The incentive scheme is an example of a "standard hour plan," meaning that a worker's task output is first converted into units of Standard Productive Hours (SPH), by dividing output by a benchmark hourly rate of speed, which we denote the "target rate." SPH are then remunerated according to a nonlinear "cap and quota" scheme, which pays nothing for SPH below a minimum quota level, then pays a linear piece rate, and then pays nothing again above a cap level of SPH. Dynamic incentives can arise in this form of incentive scheme if the firm calibrates future target rates based on current worker speeds and the worker is aware of this procedure. In the context we study, the procedure for setting target rates is explained explicitly to workers ex ante.

Section 2 of the paper introduces a simple two-period model that makes predictions for

rather than jumping to a plausible but incorrect answer (see Frederick 2005).

[4]We define the ratchet effect as arising if the optimal effort is lower in the current period than it would be if performance targets were kept fixed.

effort depending on whether there is dynamic rate adjustment in this class of incentive scheme. If target rates are fixed, the positive piece rate leads to higher effort than if there were only fixed wages, and optimal effort is constant over time all else equal. If the firm sets target rates in period 2 as a function of the worker's effort level in period 1, however , then workers have an incentive to reduce effort in period 1, the well-known ratchet effect.

We hypothesize several features of this class of incentive scheme that may cause dynamic incentives to be a shrouded attribute, and test these hypotheses in our analysis.[5] First, figuring out the implications of the dynamic incentives for current effort and the level of earnings entails relatively *complicated contingent thinking*, as workers have to also consider how future effort should respond to different possible marginal incentives. Second, the way that dynamic incentives are explained, in terms of rates and SPH rather than money, leaves the financial consequences of dynamic incentives *implicit*. Third, largely irrelevant details of how the incentives are implemented in practice may act as *distractors*. 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, all else equal. Learning could also be a mechanism that leads to unshrouding over time, so we incorporate learning opportunities for workers into our experimental designs.

Sections 3 to 5 of the paper provide our empirical results using a combination of three complementary data sets, the first of which is from field experiments within a firm. The purpose of the field experiments is to test *whether* workers respond to dynamic rate adjustment in the context of a real workplace. The firm employs relatively uneducated workers who fulfill customer orders in a large warehouse. The firm initially paid fixed wages, but then introduced the incentive scheme we study, with fixed target rates of speed.[6] Our field experiments were implemented after static incentives had been in place for six months. The firm agreed to

[5]These hypotheses are inspired partly by evidence from psychology and behavioral economics: Leaving costs or prices implicit has been found to cause them to pass unnoticed (Brown, Hossain, and Morgan (2010); Chetty, Looney, and Kroft (2009)); people have trouble with relatively complicated contingent thinking in abstract settings (for recent work see, e.g., Martínez-Marquina, Niederle, and Vespa (2019); Esponda and Vespa (2019)); and extraneous information can distract individuals who have limited attention from relevant information (see, e.g., DellaVigna and Pollet 2008; Hirshleifer, Lim, and Teoh 2009).

[6]We make no claims that the firm's approach to providing incentives is optimal. For example, they might have used a tournament incentive scheme, which could also address the fact that workers know more about task difficulties than the firm. The firm was concerned, however, about fairness between workers and about the transparency of the system. Although we do not know the firm's true profit function, one could could specify a profit function for which the system used is indeed optimal.

alter its business practices to randomly assign some workers to have dynamic rate adjustment (treatment), while others faced only static incentives (control).

Our main field experiment, denoted INDIVIDUAL, involved randomizing each cohort of new hires into treatment and control, over the course of six months (N=1,294). Treatment workers could largely determine their individual target rates for a future period, lasting several weeks, because these were computed as their individual average speed over the current period and one other (largely constant) number, the average speed of experienced workers in the warehouse. We find, however, only a very small response to treatment, a (non-significant) reduction in productivity of about -0.1 percent. Thus, worker behavior is on average consistent with dynamic incentives being fully shrouded. We also compare the observed behavior to the predictions of our calibrated structural model of rational effort prevision, which incorporates motives such as effort costs, time discounting, and intrinsic motivation or concerns about firing threat that might reduce the response of rational workers, but we find that behavior is far from the predicted optimum. The model, which is set identified, predicts a much larger response if workers were fully rational, at least 4 percent under our most conservative assumptions. The deviation we observe from the model's optimum implies non-trivial utility losses for the workers, and the firm pays about 30 percent less to achieve the observed effort level than it would if workers were fully rational.

We conduct various robustness checks on our results from the field experiment. One is a second field experiment, denoted GROUP, which used experienced workers (N=1,447), had more opportunity for learning (10 months with feedback each month), and incorporated potential social pressure motives, since rates were based on the average speed of a group of treatment workers doing a task. We again find that on average, workers act as though dynamic incentives are largely shrouded, although the response is somewhat larger than in INDIVIDUAL: A statistically significant reduction in productivity, but only of about -1 percent. Other reduced form robustness checks cast further doubt on explanations such as extreme time discounting, or worker concerns about dismissal.

Our second data set involves online experiments with the warehouse workers, with the purpose of providing more direct evidence on the question of *why* workers respond only weakly to dynamic incentives. The online experiments involved a real effort task, with treatments varying the nature of incentives across workers. This controlled setting rules out, by design, motives that might work against a response to dynamic incentives, such as firing concerns, and allows testing the impact of incentive schemes that reduce complexity and make ratchet effects stronger (for obvious reasons, the firm would not allow such treatments in the warehouse). In our main treatment, denoted COMPLEX, workers faced the same type of incentive scheme, with dynamic rate adjustment, that was implemented for treatment workers in the INDIVIDUAL field experiment in the warehouse. In another treatment, SIMPLE, target rates were fixed, but earnings in the current period were subtracted from earnings in the future period. This made monetary consequences of dynamic incentives explicit, and reduced the need for complicated contingent thinking about how to respond to dynamic incentives. SIMPLE also eliminated a largely irrelevant distractor present in COMPLEX, the averaging of worker speed with a largely constant random variable, X, to determine future target rates (this corresponded to the averaging with warehouse average speed used in the INDIVIDUAL trial). We also ran a control treatment, STATIC, which had no kind of dynamic incentives, providing a benchmark measure of behavior when dynamic incentives are truly absent rather than just shrouded. A fourth treatment, STATIC_ZERO, helps calibrate our structural model, by testing how workers respond to changing the piece rate to zero. Finally, the online environment made it possible to measure various aspects of worker cognitive ability, as well as time discounting and other traits.

The online experiments with warehouse workers provide strong evidence that dynamic incentives are shrouded in the class of incentive schemes we study, due to mechanisms of complexity and worker bounded rationality. We see very little response to dynamic incentives in COMPLEX, with behavior not statistically distinguishable from behavior when target rates are truly static, as captured by our STATIC treatment. In SIMPLE, by contrast, there are strong and significant reductions in effort precisely in the periods with dynamic incentives. Comparing to the optimums predicted by our structural model of rational effort provision, calibrated and set identified using the experimental data, behavior is far from the predicted rational optimum in COMPLEX, and almost spot-on the optimum predicted for SIMPLE. Providing further evidence that most workers do not attend to dynamic incentives in COMPLEX, only about 19 percent mention dynamic incentives in an open-ended question about optimal work strategies. By contrast, more than twice as many, 44 percent, mention dynamic incentives in SIMPLE. Finally, we find that worker bounded rationality, as captured by worse scores on the CRT test, predicts weaker response to dynamic incentives, greater distance from the predicted rational optimums, and lower probability of self-reported noticing of dynamic incentives, in both COMPLEX and SIMPLE. Other aspects of cognitive ability that we measure – education, tendency to bracket choices narrowly, and difficulty with backwards induction – have limited explanatory power for worker response to dynamic incentives. Opportunities for learning show little sign of eliminating shrouding in COMPLEX, whereas the response does become stronger in SIMPLE.

The findings from the online experiments with warehouse workers build the case that the workers do not respond to dynamic incentives because of complexity and bounded rationality, and we find additional support for this explanation when we link our first two data sets. Workers who recognized dynamic incentives online also responded significantly more in our field experiments, suggesting that the same mechanism plays a role.

Our third data set is from the same type of online experiments, but conducted with Amazon Mechanical Turk (AMT) workers. We conducted three sets of experiments *(1) Replication study*: These experiments test whether the results found with warehouse workers generalize to AMT worker, an interesting population to study because of relatively higher cognitive ability. We find similar results, in that dynamic incentives are shrouded for most AMT workers in complex, but generally unshrouded in SIMPLE, and for AMT workers with higher CRT. The shrouding of dynamic incentives is less extreme for AMT workers, however, than warehouse workers, and we show that a substantial portion of this difference can be explained by the difference in CRT scores across the two populations. Thus, the same incentive scheme has different effects depending on cognitive ability of the workforce. *(2) Contract features contributing to shrouding:* While SIMPLE eliminates several hypothesized contributors to complexity at once, in these experiments we add one complication to SIMPLE at a time. We find that each of the hypothesized factors – distractors, making financial consequences implicit, and making dynamic incentives involve relatively complicated contingent thinking – contribute to shrouding. *(3) Robustness of shrouding:* A final set of experiments tests robustness of shrouding to some perturbations that a firm might realistically consider: Making the piece rate schedule linear, eliminating the SPH construct entirely in the explanation of incentives, or combining both of these. Neither cause unshrouding on their own, but there is a modest effect when they are combined. This shows that shrouding is robust in that it does not depend on one particular formulation of the firm's incentive scheme, although combining simplifications can start to lead to unshrouding. This has important practical value, showing the implications for shrouding of various modifications to this prevalent class of incentive schemes.

The findings in this paper are relevant for the theoretical literatures on information economics and optimal incentives (for seminar papers see, e.g., Rothschild and Stiglitz (1976) and Holmström (1979)). A key insight of the former literature is that incentive schemes typically deliver second-best outcomes under asymmetric information, including in dynamic contracting settings. Our findings add nuance by showing that complexity and bounded rationality can mean that workers do not fully exploit their informational advantage, bringing incentive schemes closer to the first best, at least for some period of time.[7] While learning could seemingly undo these efficiency gains in the long run, for a given scheme and population of workers, firms may be able to periodically change the incentive scheme in ways that preserve complexity, and factors like individualized work, or high turnover, may work against social learning, thereby helping to prolong efficiency gains. Regarding optimal incentives, our findings show that complexity can be an important design factor, left out of standard models, which can affect effort provision, total surplus, and distributional outcomes. Another implication is that the nature of optimal incentives may vary across jobs and industries according to varying cognitive sophistication of worker populations.

Our findings are supportive of the growing theoretical literature positing that complexity can affect responses to contracts, and models that allow for heterogeneity in individuals' bounded rationality to matter (e.g., Eliaz and Spiegler (2006), Heidhues and Kőszegi (2017)).[8] As pointed out in this literature, a concern is that firms may use contracts to exploit consumers' bounded rationality in ways that are surplus destroying, because they lead to excessive trade.[9] This may call for regulation that prohibits exploitative contracts. Our paper provides

[7]Handel (2013) provides complementary evidence that consumer inertia in health insurance markets can improve welfare, because consumers do not exploit their informational advantage and this reduces adverse selection.

[8]See also MacLeod (1996), Gabaix, Laibson, et al. (2006), Carlin (2009), Glazer and Rubinstein (2012), Kőszegi and Szeidl (2012), Gabaix (2014), and Jakobsen (2020). A related literature, such as S. Li (2017) and Börgers and J. Li (2019), studies more general mechanism design problems under concerns about complexity.

[9]For surveys see Koszegi (2014) and Spiegler (2011). There has also been theoretical work on how contract designers who face cognitive costs may leave contracts incomplete (Tirole 2009; Bolton and Faure-Grimaud 2010), on design of optimal tax systems with behavioral agents (e.g., Farhi and Gabaix 2015; Goldin 2015), and on how firms may exploit naivete of consumers using complexity of contracts. See Garicano and Prat

evidence that exploitative workplace incentive contracts can in some cases be efficiency enhancing albeit transferring surplus from workers to firms, because they address a problem of too little trade.[10] Thus, the optimal regulation of complex incentives might not be prohibition but rather ex post transfers from firms to workers.

Our paper also complements an emerging empirical literature on the importance of complexity and bounded rationality for heterogeneous effects of consumer contracts and tax systems (e.g., Chetty, Looney, and Kroft (2009); Brown, Hossain, and Morgan (2010); Taubinsky and Rees-Jones 2017).[11] Our paper adds to this literature by showing that complexity of workplace incentives matters for effort provision, and can mitigate a fundamental challenge to incentives, the ratchet effect; by providing causal evidence about specific contract features that contribute to complexity and shrouding; and by showing the particular importance of CRT for noticing shrouded attributes. A related empirical literature has tried to empirically understand what people find complex, mainly in the context of abstract laboratory experiments (e.g., Herrnstein et al. (1993), Martínez-Marquina, Niederle, and Vespa (2019), Oprea (2020)).[12] Our paper is complementary because it explores the complexity of features that occur in less abstract, real-world incentive contracts.[13]

Finally, we add to the theoretical and empirical literatures on dynamic incentives and the ratchet effect (for recent theoretical work see, e.g., Malcomson (2016)). The empirical literature includes various examples, anecdotal and empirical, of ratchet effects existing in various settings. Notably, many of these examples are from settings that arguably involved relatively transparent incentives, such as simple piece rate schemes, or relatively skilled individuals, such as machine-shop operators, college students, or teachers.[14] Our paper does not

(2013) for a survey on theories of optimal organizational structure when individuals face cognitive costs.

[10]For surveys see Koszegi (2014) and Spiegler (2011). There has also been theoretical work on how contract designers who face cognitive costs may leave contracts incomplete (Tirole 2009; Bolton and Faure-Grimaud 2010), on design of optimal tax systems with behavioral agents (e.g., Farhi and Gabaix 2015; Goldin 2015), and on how firms may exploit naivete of consumers using complexity of contracts. See Garicano and Prat (2013) for a survey on theories of optimal organizational structure when individuals face cognitive costs.

[11]See also Finkelstein 2009, Abeler and Jäger 2015, Rees-Jones and Taubinsky 2019, Anagol and Kim 2012; Agarwal, Song, and Yao 2017. Dalton, Gowrisankaran, and Town (forthcoming) provide evidence that consumers are myopic with respect to dynamic incentives embedded in Medicare Part D (but see Aron-Dine et al. (2015)).

[12]See also S. Li (2017), Jin, Luca, and Martin (2021), Esponda and Vespa (2019), Enke and Graeber (2019).

[13]Similar in spirit to our approach but focused more specifically on salience, Englmaier, Roider, and Sunde (2016) show that workers become more responsive to features of the incentive scheme if they are made more salient through a priming intervention.

[14]This includes anecdotal evidence from piece rate jobs in the early 1900's (e.g., Mathewson 1931) and from a famous sociological account about skilled workers holding back effort in a machine shop (Roy 1952).

question the existence or importance of ratchet effects, but rather contributes novel insights into factors that can lead to responses to such dynamic incentives being stronger or weaker, namely complexity and bounded rationality, and links these to specific contract features and aspects of worker cognitive ability.

The final section of our paper, Section 6, provides a concluding discussion. This includes a discussion of a boundedly rational version of our model, which allows estimating a "misconception parameter from the data, which can be compared to other estimates from the inattention literature. We also discuss how, under additional assumptions about the firm's profit function, our model allows quantitative estimates of how much total surplus is increased by the boundedly rational behavior we observe, and also shows the possibility that shrouding can under some conditions improve the welfare of both the firm and the workers.

2 Theoretical framework

We now introduce a simple two-period model of effort provision and dynamic incentives that allows us to highlight our key theoretical predictions. Laffont and Tirole (1988) define the ratchet effect as arising when high performance today will make it more difficult to earn money in the future. [15] Intuitively, when earnings in Period $t+1$ are γ_{t+1} and effort in Period t is e_t, then the ratchet effect form of dynamic incentives is that $\frac{\partial \gamma_{t+1}}{\partial e_t} < 0$. This implies that individuals who face the ratchet effect will endogenously reduce effort, relative to individuals who face the same Period t incentives, but no ratchet effect. We denote the lower effort provision as the response to "ratchet incentives". 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 as the ratchet effect.

In building our model, we will assume workers are perfectly rational (in line with our empirical null hypothesis). In Online Appendix F, we discuss how the theoretical model may

More recently, researchers have found ratchet effects in laboratory experiments, in which one implicit theme is the need for extremely simple and abstract experimental designs in order to find the effect (Chaudhuri 1998; Cooper et al. 1999; Charness, Kuhn, and Villeval 2011; Cardella and Depew 2018). There is also a small set of papers providing field evidence on responses to dynamic incentives, using behavior of serfs in Russia (Markevich and Zhuravskaya 2018), teachers (Macartney 2016), and tree planters (Bellemare and B. Shearer 2014; B. S. Shearer (2022)). Another literature documents, for various types of workers, the adjustment of the level of performance targets based on past performance (but this does not necessarily imply ratchet effects, see Matějka, Mahlendorf, and Schäffer 2022).

[15] Weitzman (1980) notes that the term "ratchet principle" was coined by Berliner (1957).

be altered in light of our findings in order to incorporate bounded rationality.

We assume there are two time period $t = 1, 2$. Individuals discount the future by δ. We suppose that there is a single individual, denoted i, with type θ_i, drawn i.i.d. from cdf H. The individual decides every period t how much effort $e_{i,t}$ to exert and faces a convex cost $c(e_{i,t}, \theta_i)$. We suppose c is differentiable, strictly convex, $c'(0, \theta_i) = 0$ and the limit of c' is ∞. In each period the individual receives a base income o, plus a bonus γ_t. In Period 1, the bonus is simply a function of $e_{i,1}$: $\gamma_1(e_{i,1})$. In Period 2, the bonus can be a function of effort in both periods: $\gamma_2(e_{i,1}, e_{i,2})$. This, admittedly general, form can capture the form of dynamic incentive schemes that forms the basis of the ratchet effect. Weitzman (1980), for example, assumes $\gamma_2(e_{i,1}, e_{i,2}) = w(e_{i,2} - \lambda e_{i,1} - (1 - \lambda)q_1 - \delta_2)$ where w is the wage , and q_1 and δ_2 are exogenous parameters . In Laffont and Tirole (1988), both γ's are determined as best responses as part of a Nash Equilibrium in a game of incomplete information played between the firm and workers. The message, across these papers and the larger literature, is clear: ratchet incentives induce a decrease in Period 1 effort.

Our model will be tailored to our specific setting, and the details of the contract offered by the firm we work with. Thus, our assumptions reflect the actual contract in our empirical setting. In particular, the incentive scheme is a standard hour plan, so bonuses are a function of normalized effort (we imagine that effort is directly observable, rather than only output). Period t effort is normalized using a target rate of speed $\eta_{i,t}$, and then transformed into money using bonus function $\gamma_t = \gamma(\frac{e_{i,t}}{\eta_{i,t}})$. We must also specify how these rates are set. $\eta_{i,1}$ is an exogenous number set by the firm. $\eta_{i,2} = \zeta e_{i,1} + (1 - \zeta)\epsilon_1$ that is the site rate in IND, so eta would be easier to understand for some $\zeta \in [0, 1)$ (note that ζ in the experiment is actually $\frac{1}{2}$) and for some exogenous parameter ϵ_1. In other words, effort in Period 2 is normalized by the average of Period 1's effort and an exogenous parameter. Such a scheme is similar to the original scheme considered in Weitzman (1980) but uses ratios of efforts, rather than differences in effort. Our specification generates strict ratchet incentives whenever $\zeta > 0$.

In line with the actual contract used by the firm, which is an example of a commonly-used class of cap-and-quota incentive contracts, we will suppose that γ takes on a piecewise linear form.[16] No bonus is earned if effort is below the quota, $\underline{E}$. Above $\underline{E}$ the bonus is a linear

[16]Regarding the initial portion of the contract without incentive pay, 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

function of effort (proportional to w), until effort reaches the cap, $\bar{E}$, above which additional effort does not increase bonus earnings.[17]

$$\gamma(e) = \begin{cases} 0, & \text{if } e \leq \bar{E} \\ w[e - \underline{E}], & \text{if } \underline{E} \leq e \leq \bar{E} \\ w[\bar{E} - \underline{E}], & \text{if } e \geq \bar{E} \end{cases}$$

Each period individuals get flow utility that depends on their earnings, but also on non-monetary motivations and costs of effort provision. For the former, we suppose there is a differentiable, weakly concave function $a_f(e_{i,t})$, where $a_f(0) = 0$. This function can reflect intrinsic motivations on the part of the worker, altruism by the worker towards the firm, or other non-pecuniary concerns, like firing threats.

Thus, the total utility is

$$U_i = o + \gamma(\frac{e_{i,1}}{\eta_{i,1}}) - c(e_{i,1}, \theta_i) + a_f(e_{i,1}) + \delta(o + \gamma(\frac{e_{i,2}}{\eta_{i,2}}) - c(e_{i,2}, \theta_i) + a_f(e_{i,2})) \qquad (1)$$

with $\eta_{i,2} = \zeta e_{i,1} + (1 - \zeta)\epsilon_1$.

The next proposition highlights two key features that we can test in the data. First, when there are no dynamic incentives ($\zeta = 0$) workers respond to an increase in incentives by increasing effort — in other words we observe a standard labor supply response. Second, when $\zeta > 0$ we observe a ratchet effect; in other words individuals exert less effort in Period 1 compared to if $\zeta = 0$. This is because, when $\zeta = 0$, the utility function is additively time separable. In contrast, this is not true when $\zeta > 0$, and thus effort in Period 1 reduces payoffs in Period 2.

Proposition 1

1. *If $\zeta = 0$ then an increase in w increases effort in both periods.*

2. *Fixing θ_i, individuals for whom $\zeta > 0$ put in less effort in Period 1 than those individuals for whom $\zeta = 0$.*

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.

[17] Although such a scheme may be optimal for a profit-maximizing firm under some assumptions, we make no claims about the optimality for our particular firm.

The model discussed here makes many simplifying assumptions relative to the reality that the workers in our field experiment faced. First, we assume that there is a single task with a single target rate in each period. In Appendix , 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. Second, we focus on an individual ratchet effect, in line with our main experiment (the INDIVIDUAL trial). We also conducted a second field experiment, the GROUP trial, which allows us to check the robustness of our main findings. The GROUP trial features a somewhat different way of constructing $\eta_{i,t}$. In Appendix D, we explicitly model the GROUP trial and show that we also obtain a ratchet effect in that setting.

3 Results from field experiments in a firm

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

3.1 Description of the work context

3.1.1 Nature of work, worker characteristics, and personnel data

We collaborate with a firm that operates multiple warehouses in which workers fulfill the online orders of customers. The warehouse workers are involved in collecting the desired products from storage, putting these into delivery containers, and moving the containers onto vehicles for delivery to customers' homes. Our analysis focuses on one of the firm's warehouses, although we do use data from a second warehouse when we describe the response of workers to the initial introduction of static incentives.

The modal task in the warehouse involves workers moving products from shelves to containers that pass by on conveyor belts. A computer screen at the station where they are working will show a product, the worker gets the product from the shelves, scans it, and places it in the container. Workers scan each item they handle. The work is done individually and a worker's output is independent of the effort of other workers, as there are lines of containers waiting at each station, which serve as buffers between workers. Workers work on different tasks and in different locations in the warehouse throughout their shift.

The firm has provided us with access to minute-by-minute data on the activities of all workers, as captured by their scans of products. These data allow us to investigate the productivity responses of workers to the introduction of the performance pay system with static rates, and more crucially, to measure the productivity responses of workers to the treatments in our field experiments.

3.1.2 The incentive scheme

At the warehouse we study, the firm initially just paid workers an hourly wage. After about a year, the firm introduced an incentive scheme in addition to the hourly wage. (A timeline of all changes to the incentive system that we use in this paper, including the two field experiments, can be found in Figure B.1 in the appendix.) Our theoretical framework in Section 2 is modeled after this incentive scheme. The incentive scheme is an example of a "standard hour plan:" A worker's task output, i.e. the number of scans, is first converted into Standard Productive Hours (SPH). The amount of SPHs is equal to the number of scans done by the worker divided by a normalization factor, which corrects for the difficulty of the task. In line with industry jargon, we refer to this factor as the "target rate." Each week, SPH are then remunerated according to a nonlinear "cap and quota" scheme: workers receive a linear piece rate for each SPH between a minimum quota and a maximum cap level. There are no bonus payments for SPH below the minimum or above the maximum level. The average bonus is about 10 percent of weekly salary. The maximal bonus is about 38 percent of weekly salary.

The focus of this paper is on how target rates are set and how this affects worker speed. Target rates vary within the warehouse, across different tasks, but also within tasks based on area of the warehouse. Specifically, the firm has divided the warehouse into 76 different "rates areas," each with a different rate η. The total amount of a workers' SPHs is the sum of their SPHs in each rate area. One scan takes more or less effort depending on the task. Target rates that reflect the local task difficulty allow paying the same bonus for the same amount of worker effort, independent of how difficult one individual scan is.

But how should the firm set target rates? Initially, target rates in the warehouse were static. They were based on the average speed of all workers in each rates area 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. We analyze how the incentive system with such static rates affected workers' speed in Appendix C. We find that incentives strongly increased workers' speed, by about 12 percent. We will use this, and related, estimates to calibrate our structural model in Section 3.2.3.

However, task difficulties rarely stay constant over time. In our particular case, for example, the firm planned to continue adding new machinery or software, which would reduce task difficulty. Task difficulties can also change because the composition of products across different areas of the warehouse over time changes. This means that target rates become too easy or too hard over time, and do not reflect the relative difficulty of different tasks any longer.

The firm thus wanted to add dynamic rate adjustment, and in particular set next month's target rate equal to some weighted average of workers' speed in this month. Using past worker speed 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. However, such a rate-setting scheme 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 introduction of dynamic rate adjustment. To shed light on this question, we conducted field experiments within the warehouse.

3.2 Field 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 alter its business practice to conduct our field experiments, which randomly assigned some workers to face dynamic incentives within the context of the firm's scheme. The INDIVIDUAL trial is our main field experiment, described below. The GROUP trial, summarized in Section 3.3.1 and described in more detail in Appendix D, extends the analyses of the INDIVIDUAL trial and serves as a robustness check.

3.2.1 Design of the INDIVIDUAL trial

The INDIVIDUAL trial tests whether workers respond to dynamic incentives. The dynamic incentives in the trial are "individual," in two senses: Because a treatment worker can have a large and direct impact on his or her own individual future target rates, and because a treatment worker cannot influence the rates of anyone else.

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. The random allocation of workers to treatments was done by us. Due to the weekly randomization, we used a non-stratified randomization. Appendix Table B.1 contains summary statistics and randomization checks (all $p > 0.33$). In total, 1294 workers started the treatment period, which began four weeks after joining the firm. Appendix Figure B.2 and Table B.2 show that there is no differential attrition before or during the trial between treatment and control group. Workers were extensively informed about all the details outlined below, except for the fact that the trial was designed together with university researchers.

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 received training and worked in the warehouse. When they worked, they could earn a bonus, like all other workers. They faced exogenously given target rates, calibrated to the warehouse rate calculated for more experienced (more than 13 weeks of tenure) workers. During this period, the workers learned about their assignment to the treatment condition and learned how their target rates would be determined in the weeks going forward. The period of interest is weeks 4 to 6. In these weeks, workers assigned to the treatment group faced fixed rates as in baseline, but knew that their individual performances during those weeks would determine their individual target rates for weeks 7 to 9. From week 10, their rates reverted to the overall warehouse rates. Thus, the treated workers faced dynamic incentives during weeks 4 to 6, which called for lowering output relative to the benchmark of purely static incentives.

Table 1: Design of the INDIVIDUAL trial

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

The workers in the control group faced the same target rates as the treated workers in weeks 4 to 6, i.e., they faced the same static incentives. Importantly, they knew that their target rates would also be exogenously given in weeks 7 to 9, i.e., their performance in weeks 4 to 6 had no impact on their future rates. Thus, control workers did not have dynamic incentives during weeks 4 to 6. Our test for a causal effect of dynamic incentives focuses on weeks 4 to 6 of the trial for each cohort, comparing the behavior of treated to control workers. To maintain fairness and to avoid a Hawthorne effect, our design ensured that control workers also had dynamic incentives, but later on, in weeks 7 to 9.[18]

For the time periods in which workers faced dynamic incentives, the specific rule for calculating target rates was the following. An individual worker's target rate for a given activity area was given by the worker's individual average output per hour on that activity over the relevant time period, averaged with the warehouse average of experienced workers for the same activity and time period. Thus, the rate was not completely determined by an individual worker's performance, as it depended partly on the warehouse average. The worker's impact, however, was substantial, receiving a weight of 0.5. The firm was unwilling to have a larger weight to avoid extreme target rates if workers provided near-zero performance in the first period. Note that workers did not influence anyone's rate but their own, so there were no motives for social pressure and no need for collusion. The field experiment is thus a test of whether workers responded to entirely individual dynamic incentives.

[18] In weeks 7 to 9, control workers knew their individual performances 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). Thus, in weeks 7 to 9, control workers could influence future rates, and thus had dynamic incentives, whereas treated workers did not. 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.

3.2.2 Results of the INDIVIDUAL trial

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

Table 2 shows the results of OLS regressions of worker performance (measured as the log of their units (=scans) per hour) in weeks 4–6 on a treatment dummy. We control for rates-area fixed effects to account for the fact that a "unit" is harder or easier in different rates areas. We also control for shift fixed effects to account for fluctuations that are common to treatment and control group (there are two shifts per day). Finally, we control for randomization cohort and cohort $\times$ shift fixed effects to account for the fact that workers starting in different weeks, i.e., the different cohorts, could differ in many respects and will have different performance trends over time. Controlling for the cohort-related fixed effects allows us to combine the results from the 40 cohorts. In all our regressions, we use standard errors that are two-way clustered on individual workers and on shifts.

Table 2: 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)	(0.006)
Sample	Weeks 4–6	Week 6	Week 6
			Attrited after week 9
Rates area FE	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes
Shift 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 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. Significance at the 1, 5, and 10 percent level is denoted by ***, **, and *, respectively.

The treatment coefficient thus measures the "ratchet effect" in this setting. A negative coefficient would show that workers who face dynamic incentives slow down compared to workers who do not face such incentives. We find that facing dynamic incentives does indeed reduce output, but the point estimates are very small and not significantly different from zero. Table 2 column 1 is the main regression and includes the full sample of weeks 4–6. The point estimate implies a slow-down of about -0.1 percent (with a 95% confidence interval of [-1.2, 1.0]). Column 2 restricts the sample to only week 6 to allow for some learning. The (absolute) point estimate is slightly larger, implying a slow-down of -0.2 percent (CI: [-1.3, 1.0]). Column 3 further restricts the sample to only those 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 and they thus face the strongest ratchet incentives. The point estimate is again slightly larger but remains small and non-significant (-0.5 percent, CI: [-1.7, 0.8]).

3.2.3 Comparison of behavior in INDIVIDUAL trial to rational model predictions

We find a very small response to the dynamic incentive scheme in INDIVIDUAL, i.e., a very small ratchet effect. What response should we have expected if workers were fully rational? There are many factors that could affect the size of the observed ratchet effect, e.g., time discounting, the elasticity of work effort, intrinsic motivation or fear of being dismissed. In order to derive a rational benchmark we return to our model of rational effort provision from Section 2, which allows for all of these factors.

To estimate the parameters of the model, we use data on workers' effort levels and assume that workers rationally choose their effort level when facing static incentives. We thus take the observed reaction of workers to the introduction of static incentives and the effort level in Period 2 of the INDIVIDUAL trial (weeks 7–9), when workers only face static incentives, and derive what reaction they should show to the introduction of dynamic incentives in Period 1 of INDIVIDUAL (weeks 4–6).

The warehouse initially only paid hourly wages and then introduced the incentive scheme used in the INDIVIDUAL trial. However, for the first three months, target rates were fixed and this was explicitly communicated to workers. During this time, workers thus only faced static incentives. Their effort in period t did not affect their potential incentive pay in period $t+1$.

For static incentives, the worker's problem (see Equation 1) reduces to a one period optimizing of the following utility function :

$$U_i = o + \gamma(\frac{e_{i,t}}{\eta_t}) - c(e_{i,t}, \theta_i) + a_f e_{i,t}$$

For the first data moment, we are interested in the change of effort e_t from when only hourly wages were paid, i.e., when $\gamma(\cdot) = 0$, to after static incentives were introduced, when γ is a piecewise linear function with slope either 0 or w. Since a simple before-after comparison might be confounded by time trends, we estimate the impact of static incentives on effort using a difference-in-differences analysis with the firm's other main warehouse as control. All details of the analysis are provided in Appendix C. At the control warehouse nothing changed with regard to incentives during the relevant time period, and the pre-trend is very similar to the treatment warehouse (Figure C.1).

We estimate a substantial and highly significant response to static incentives: Worker effort increased by 12.4 percent due to the introduction of static incentives. Results are similar across a variety of alternative specifications (Table C.2). Thus, workers appear to find the static incentives non-trivial, and respond strongly to these. 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 use the latter estimate for the calibration of the structural model.

The second moment we will use for the calibration is the observed effort level in Period 2 of the INDIVIDUAL trial when workers again only faced static incentives.[19]

In order to estimate our model, we must make some parametric assumptions. We suppose the cost of effort is a power function: $c(e_{i,t}, \theta_i) = \theta_i \frac{e_{i,t}^{\kappa+1}}{\kappa+1}$.We also make the assumption that non-monetary motivation is a linear function of effort: $a_f(e_{i,t}) = a_f e_{i,t}$ where we abuse notation slightly so that a_f is the marginal return to effort in terms of non-monetary motivation. We assume a representative worker framework, so that there is no worker heterogeneity.[20] We assume that workers are working on a single "representative" task, and their decision is thus how much effort to provide for that task in each period. In reality, each period (as defined in the model) consists of three weeks, and SPH, bonuses, and productivity are all calculated at the weekly level. We abstract away from this and solve for the solution of a representative week and associate these with e_1 and e_2.

We have only two moments of data (the reaction to static incentives and the effort level in Period 2 of INDIVIDUAL), while we have three parameters (θ, κ, a_f) to identify. Our approach thus focuses on set identification of the parameters. We establish bounds in terms of the behavioral response and utility loss across the entire set of potential parameter combinations. In particular, for any positive value of a_f we can find corresponding values of θ and κ that rationalize the behavior under static incentives. In order to establish bounds we consider a large range of potential values of a_f, and for each value of a_f, our two moments identify a corresponding θ and κ pair.

Given a triple (θ, κ, a_f), we then simulate workers' response to the dynamic contract by

[19] One could also calibrate the model to match behavior of workers in control. Because we observe an extremely small ratchet effect, these two methods generate almost identical results.

[20] We aggregate across all individuals who work at least 20 hours a week in the warehouse, in order to avoid individuals who are part-time.

finding the optimal solution to the utility function $U = o + \gamma(\frac{e_1}{\eta_1}) - \theta\frac{e_1^{\kappa+1}}{\kappa+1} + a_f e_1 + \delta(o + \gamma(\frac{e_2}{\eta_2}) - \theta\frac{e_2^{\kappa+1}}{\kappa+1} + a_f e_2)$, with the functional forms of γ and η_2 discussed in Section 2. We also need to make an explicit choice for the value of the discount rate δ. 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 work less in Period 2 than in Period 1. We find that the average ratio of the total time workers spent working in Periods 2 vs. Period 1 is 0.97 (this includes workers who leave the firm permanently). We thus set $\delta = 0.97$. We discuss robustness to this assumption below. We then solve for the optimal effort in each period. We can compare this to observed behavior, and calculate the utility loss that observed behavior generates relative to model-optimal behavior.[21]

Finding 2 *Given their observed response to static incentives, workers should have reduced effort by at least 4 percent in Period 1 of the INDIVIDUAL trial.*

Table 3 shows the results of our exercise, separately for ten representative values of a_f covering the entire range we considered. For each set of parameter values we derive the optimal effort in Period 1 and 2, if treatment workers fully understand the incentives, and compare this to the actual efforts that the treatment group exerted. We see that treatment workers exerted higher effort than predicted by the model in Period 1, i.e., they exhibit a smaller ratchet effect than predicted by the model. Because, in the data, treatment and control workers put forth almost the same amount of effort in Period 1, we can use the difference between model-optimal effort and observed effort as the estimated treatment effect we should have seen if workers were fully rational. Although the difference varies significantly across parameterizations, treatment workers exerted at least 4 percent more effort than they should have in Period 1, i.e., we should expect to see at least a 4 percent treatment effect.[22] Recall the observed treatment effect (-0.1 percent with 95% CI [-1.2, 1.0]) was over an order of magnitude smaller.

We can also calculate the utility loss individuals in our treatment group experienced given

[21]Because γ is piecewise linear in both periods, the optimization problem is not strictly convex, and there are a variety of local minima and maxima. We thus solve for the optimum numerically.

[22]Thus, the dynamic incentives should have eliminated at least 40 percent of the effort increases from the introduction of static incentives.

their actual effort over the course of the experiment, compared to what would have happened had they behaved optimally. Again, the size of the utility loss varies by the parameter values, but never falls below 0.87 utils. Because utils are measured in the money metric (our function is quasi-linear in the bonus payments), 0.87 utils translates into approximately 4 percent of the average take-home (weekly) bonus. For a wide range of alternative values (e.g., $\delta = 0.5$) our qualitative results are robust .

Of course, the wide variation in our results raises the question of what parameter estimates are most reasonable. One way of answering this is to consider the intensive elasticity of labor supply, and use estimates of the literature to help pin down our parameters. This elasticity is subject to wide dispute. However, two recent studies using natural experiments of tax holidays in Iceland and Switzerland (Stefansson 2020; and Martinez et al. 2021) find intensive-margin elasticities of 0.07 and 0.025 respectively. Their evidence is consistent with a meta-analysis of the extensive margin by Elminejad, Havranek, and Horvath (2021). In our model, this elasticity is $\frac{1}{\kappa}$. This implies that $\kappa \geq 14$. DellaVigna and Pope (2018) find estimates of $\kappa \geq 24$. These parameter values imply utility losses and effort reductions dramatically larger than our lower bounds.

Overall, the INDIVIDUAL trial thus finds that workers, when facing individual ratchet incentives, fail to reduce effort, even though rational workers should have reduced effort strongly.

Table 3: Estimating and Simulating Optimal Behavior in INDIVIDUAL

a_f (Non-Monetary Concerns)	κ (Curvature of Effort Cost)	θ (Marginal Cost of Effort)	Optimal e_1	Optimal e_2	Observed e_1	Observed e_2	Percent Difference Between Observed and Optimal e_1	Percent Difference Between Observed and Optimal e_2	Utility Loss of Actual Relative to Optimal (in utils)
1×10^{-7}	226.99	≈ 0	0	22.4	33.59	33.27	100	33	39
1×10^{-5}	166.27	2.33×10^{-253}	3.2	31	33.59	33.27	90	7	39
.001	105.56	8.28×10^{-161}	3.3	31.1	33.59	33.27	90	7	39
.1	45.27	3.52×10^{-69}	3.7	31.4	33.59	33.27	90	6	38
1	18.28	5.7×10^{-28}	4.7	32.2	33.59	33.27	86	3	17
10	3.46	6.93×10^{-5}	32.4	33.3	33.59	33.27	4	≈ 0	.84
1000	.04	873.14	32.4	33.3	33.59	33.27	4	≈ 0	.87
1×10^{5}	4×10^{-4}	99864.23	32.4	33.3	33.59	33.27	4	≈ 0	.87
1×10^{7}	4×10^{-6}	9999864	32.4	33.3	33.59	33.27	4	≈ 0	.87

3.3 Robustness checks

In this section we first present results from a second field experiment, the GROUP trial, which tests for response to dynamic incentives when workers are experienced, have a long time horizon in which to learn, and have potential social pressure motives. We then discuss robustness checks on various alternative explanations for our findings from the two field experiments.

3.3.1 Robustness to experience, learning, and social pressure: the GROUP trial

Our second field experiment, denoted the GROUP trial, serves as a robustness check for the INDIVIDUAL trial. Participants in the INDIVIDUAL trial were new workers and thus inexperienced. They faced dynamic incentives for three weeks. At any point in time, the participants constituted only a very small fraction of the workers in the warehouse. This reduces the chance to learn, by themselves or from others, about how to best respond to dynamic incentives.

The GROUP trial, by contrast, was conducted by randomizing all workers in the warehouse into one of the treatments, and thus included many workers with substantial experience. It took place over a longer time horizon, with workers facing dynamic incentives over approximately 10 months (ten 4-week periods).[23] Target rates were determined by the average speeds of groups of workers who worked on a given task. These features gave workers a much longer time frame to learn how to respond to dynamic incentives, both by themselves and through social learning and teaching. At the same time, the GROUP trial also differed in other respects from INDIVIDUAL. Setting rates as a group weakens the financial incentives to reduce effort in order to achieve more favorable future rates, as an individual worker's impact on future rates decreases with group size. In GROUP, workers also influence the rates of other workers, which brings into play motives related to social pressure, potentially *enhancing* ratchet effects. The sociological literature on ratchet effects (e.g., Mathewson 1931, Roy 1952) very much focuses on this social pressure and documents cases in which workers are able to collude and to hold back effort, often by threatening to punish fast-working "rate busters."

In the GROUP trial, we randomized workers into two conditions, treated workers (denoted rate setters) and control workers (non-rate setters). During the 10-month trial period, only

[23] The possible extent of learning over even longer time periods is constraint by the relatively high worker turnover.

the performance of the rate setters was used to determine next month's rates, but these rates applied to all workers equally. The spot incentives were thus identical for treatment and control workers. At the same time, rate setters faced a dynamic incentive to hold back effort, while control workers could work without facing this incentive to slow down. Appendix D describesthe design of the GROUP trial, shows formally that this setup induces a ratchet incentive and contains a detailed analysis of the results, which we summarize briefly here.

Finding 3 *The GROUP trial yields only a small ratchet effect, although somewhat larger than for INDIVIDUAL. The size of the ratchet effect stays small over time; and for large and small rate-setting groups; and for workers with longer or shorter time horizons.*

We find that the ratchet effect in the GROUP trial is larger than in INDIVIDUAL, around -1 percent, and marginally significantly different from zero. This is still much smaller, however, than the effect of the introduction of the static component of the incentive system of 12.4 percent. GROUP allows substantial opportunities for learning, and we do see the ratchet grow slowly over time, but this time trend is not significant, and even after ten 4-week rate-setting periods, the ratchet effect is still smaller than -2 percent. Any learning in our setting is thus slow and incomplete.

Could this small ratchet effect be explained by the group feature of GROUP, either because increasing group size weakens individual dynamic incentives to reduce effort, or because it potentially makes collusion more difficult? We use naturally occurring variation in group size across different areas of the warehouse and over time, and find that smaller groups indeed show a larger ratchet effect, but even in the smallest groups (of about 5 workers) the reduction in productivity is only about -2 percent. Groups of around 40 workers show a not too dissimilar ratchet effect of about -1 percent. It thus does not seem as if the size of the groups drives the weak response to dynamic incentives. Moreover, the INDIVIDUAL trial exogenously implements an effective group size of 2, and there we find an even smaller ratchet effect.

3.3.2 Additional robustness checks on time discounting and reputation concerns

Time discounting. The ratchet effect involves a trade-off between reduced earnings now and reduced effort costs in the future, and thus one explanation for a weak response 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 of effort provision in INDIVIDUAL already incorporates substantial discounting of the future over a relatively short time frame, and still the predicted reduction in effort is much larger than what we observe. We also investigate discounting in three different ways in the GROUP trial[24], and all three ways show no sign that differences in discounting affect responses to dynamic incentives, casting doubt on an explanation based on time discounting. While the lack of significant correlation between time discounting and strength of response to dynamic incentives might be surprising if all workers were rational, it is not surprising if a substantial portion of workers do not take into account dynamic incentives and thus do not react optimally to them.

Concerns about dismissal or promotion. Maybe the weak ratchet effect is due to fears of being dismissed if working too slowly? Our rational benchmark in Section 3.2.3 includes a motive to work fast (parameter a_f) due to concerns about dismissal, or ambitions to be promoted. The model nevertheless predicts a much stronger ratchet effect than we observe. The firm's personnel policies imply that workers can be substantially slower than average for an extended period of time, without attracting any special notice from management. Even then, this does not have implications for dismissal, at least in the short run, but rather leads to additional training. Additional analyses, discussed in detail in Appendix E, provide further evidence that concerns about dismissal or promotion are unlikely to be an important factor in explaining the weak response to dynamic incentives. First, dismissals are quite rare, with the vast majority of departures stemming from workers deciding to quit. Second, none of the reasons for dismissal given in the firm's records are explicitly related to speed, but rather to factors like low quality, in line with the stated personnel policies. Third, among workers for whom the reason for dismissal is vague or unspecified, and thus could potentially be about speed, we find no significant correlation of dismissal probability with speed. Taken together, workers could slow down to a non-trivial extent in the warehouse, and for a substantial time period, without implications for dismissal or promotions.

[24] We compare workers who work the entire next rate-setting period to workers who leave the firm before the end of the next period. The latter group should discount the future more. We compare permanent employees to agency/temp workers. The latter group is less likely to be called to work in the next rate-setting period and should thus discount the future more. And we measure individual time preferences for a sub-sample of workers in an online experiment. Appendix D provides the detailed analysis.

The online experiment, discussed in the next section, rules out motives related to dismissal or promotion by design.

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 suggestive of complexity and bounded rationality playing a role, we turn to online experiments with the same warehouse workers to provide sharper tests for these potential mechanisms. The controlled setting of online experiments has several advantages.

First, and most importantly, the online experiment allows us to randomly allocate workers to incentives schemes with different levels of complexity, to test whether complexity is explaining the lack of response to dynamic incentives in the firm's scheme. 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, it rules out, by design, motives such as concerns about firing threat, or social preferences towards the firm. These motives are eliminated online because workers were informed that the experiments were being conducted by outside, academic researchers, and responses would be kept confidential from the firm, and the task being done was of no intrinsic value to the firm or the researchers. Third, we can measure various aspects of cognitive ability and test directly whether lack of response to dynamic incentives is related to bounded rationality. Using the warehouse workers as subjects strengthens the external validity of the online experiments for explaining behavior in the warehouse. We also observe a sub-sample of workers who participated both online and in the warehouse field experiments, allowing a test of whether understanding and responding to dynamic incentives online is associated with responding in the warehouse. We first describe the design of our experiments, then we present results, and along the way compare these to the predictions of a model of fully-rational effort provision calibrated for the real-effort task used in the experiments.

4.1 Design of online experiments with warehouse workers

The online experiments involved workers doing a real-effort task, for an incentive scheme that is 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).[25] During the experiment, workers could work on the task for multiple periods of 90 seconds each. In a given period they could click as much or as little as they wanted. 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). Complete instructions for the online experiment, including screenshots of the task, are provided in Appendix I. There was a nonlinear payment schedule for SPM: 0 for SPM up to 0.1; $1.50 for SPM between 0.1 and 3; 0 for SPM above 3. The task, and its description, thus closely followed the incentive system in the warehouse.

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. Giving these assurances, and also the fact that the activity being performed online was not meaningful or productive, should have eliminated any motives for workers to act in certain ways to try to avoid being sanctioned by the firm, or to impress or help the firm.

After workers passed the consent stage of the online experiment, they were randomized into one of four treatments, which varied the nature of the incentive scheme. The probability was 1/3 for each of our two main treatments, and 1/6 for each of our two control treatments. In all we have 430 warehouse workers in the online experiments.

Table 4 summarizes the design of the conditions.

[25] Some studies (e.g., DellaVigna and Pope 2018) have used a task involving clicking two buttons in alternating order. We chose a simpler task of clicking only one button because we anticipated that many warehouse workers would use smartphones to participate, with the relatively small screens potentially making a two-button task too physically awkward.

Table 4: Design of online experiments, warehouse workers

Introductory phase	Consent, questions about 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			
Preference measurement	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 2 earnings subtracted	Rate is 300	Rate is 300
Cognitive ability measurement	CRT, narrow choice bracketing measure, backwards induction ability measure			
Period 3 work	Rate is 300			
Period 4 work	Rate is average of Period 1 clicks and random number X	Rate is 300, Period 2 earnings subtracted	Rate is 300	Rate is 300
Questionnaire	Open-ended question about the best strategy for Periods 3 and 4			

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 was tested in the INDIVIDUAL field experiment. 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, uniformly drawn from a narrow range of values centered around the target rate (285, 300, or 315). X was not known during Period 1 but was revealed at the beginning of Period 2, and mimicked the fact that in the INDIVIDUAL field experiment, a treatment worker's output was averaged with the average output of the rest of the warehouse to determine the worker's future rate. After doing the work for Periods 1 and 2, workers learned about periods 3 and 4, which had the same structure as Periods 1 and 2. The dynamic incentives in COMPLEX gave workers a reason to reduce clicks in Period 1 and Period 3, potentially all the way to zero, because this would achieve the easiest possible target rates in Periods 2 and 4, and facilitate making higher total earnings.

In another condition, denoted SIMPLE, workers faced an incentive scheme that also created dynamic incentives to shift effort from Period 1 to Period 2, but in a form that was intended to be more transparent. The baseline period was the same as in COMPLEX. The workers then learned about the rules for periods 1 and 2: The target rate was fixed at 300 for both periods, but any earnings in Period 1 would be subtracted from earnings in Period 2. This induces dynamic incentives, since increasing effort in Period 1 reduces earnings in

Period 2. These changes were intended to (1) describe the dynamic incentives explicitly in terms of money; (2) eliminate the need for complicated contingent thinking, because current effort affected the level of future earnings directly, rather than future marginal incentives; (3) eliminate distractors in the form of the random variable X.[26] Importantly, workers also knew that if total earnings in Period 1 exceeded earnings from Period 2 clicks, then the difference would come out of total earnings for the study; this meant that if a worker preferred to click lower in one period than the other, they had an incentive to allocate lower clicks to period 1 rather than period 2.[27] Periods 3 and 4 were identical to 1 and 2. In this treatment, clicking in periods 1 and 3 generated zero additional earnings, so the workers had less of a motive to click in these periods than in periods 2 and 4 although potentially still some motivation due to intrinsic motivation (unlike COMPLEX, clicking less in these periods did not have the benefit of lowering future target rates). We deliberately tried to design SIMPLE so that, if anything, workers would have less of a reason to reduce effort than in COMPLEX, to rule out that workers respond more in SIMPLE because dynamic incentives are somehow stronger.

We randomized the remaining workers into one of two additional treatments, STATIC, and STATIC_ZERO. In STATIC, workers had only static incentives, facing 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 shrouded. 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. Comparing to STATIC also shows how much workers respond to a reduction in the piece rate from a positive level to 0, which is informative about whether the workers noticed or cared about static incentives online, and helps calibrate our model of effort provision.

We also collected various other types of information about participating workers, at several points during the experiment (see Table 4). For example, workers were presented with

[26]We consider X to be a distractor, rather than a meaningful element of the incentive scheme, since it is drawn in a narrow range around 300 and thus has only minor implications for marginal incentives.

[27]To see this, suppose that a worker clicks 200 in Period 1, and 500 in Period 2. This will mean they are paid for $200 + (500 - 200) = 500$ clicks. By contrast, suppose a worker clicks 500 in Period 1, and 200 in Period 2. This will lead to being paid for $500 + (200 - 500) = 200$ clicks.

incentivized measures of time preferences and incentivized measures of risk aversion (see experiment instructions in Appendix I for more details on the measures and all other questions).

The experiments also measured three facets of cognitive ability and workers' educational attainment. One aspect of cognitive ability was the CRT, a test consisting of three questions, each with seemingly obvious but incorrect answers (Frederick 2005). As a measure of a tendency to think deeply, and avoid superficially plausible but incorrect answers, the CRT focuses on an aspect of cognitive ability that could be particularly important for noticing shrouded attributes. Other questions included hypothetical lottery questions 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 (a simplified version of the Hit 15 game used by Burks et al. (2009)). We implemented this measure in case difficulties in backwards induction might make it harder to reason backwards from incentives in Period 2, to the optimal choice in Period 1.

Finally, in a brief questionnaire 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 the same wording could prompt potential comments on dynamic incentives in both COMPLEX and SIMPLE. In COMPLEX this would mean commenting on the benefits of going slow in the earlier period, to have an easy target rate and make a lot of money in period 2. In SIMPLE, it would entail noting that clicking in the early period is not as useful for maximizing total earnings as clicking in the later period. As it is difficult to ask about utility, the question focused on money. 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).

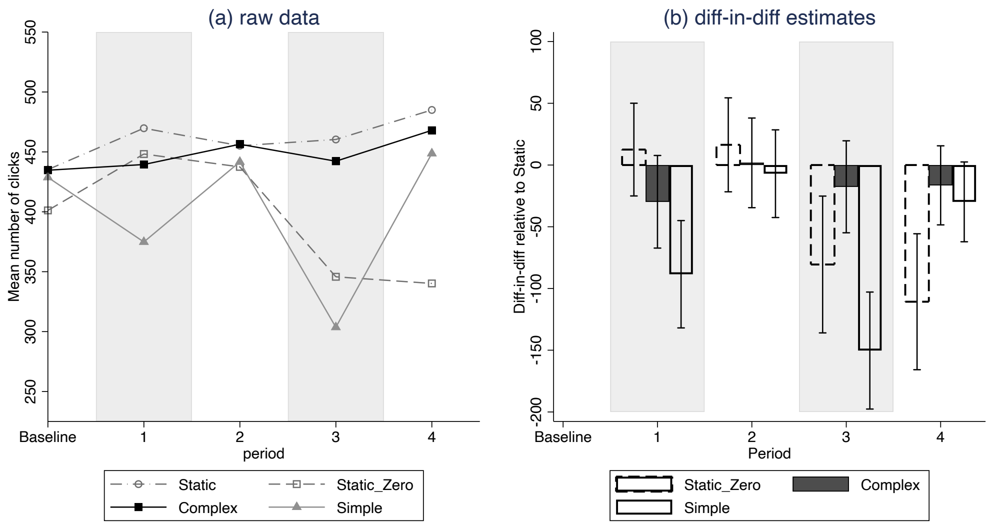
4.2 Results of online experiments with warehouse workers

Finding 4 *In COMPLEX, warehouse workers behave as though dynamic incentives are absent, and are far from the optimum predicted by a model of fully-rational effort provision. Workers in SIMPLE exhibit a significantly stronger response, and are closer to the corresponding rational optimum.*

Panel (a) of Figure 1 shows average clicks by period and treatment. In STATIC, where dynamic incentives are absent, we see that clicks are largely constant over time, although there is a slight increasing trend, potentially reflecting some learning by doing.[28] The time profile of clicks in STATIC_ZERO are very similar to STATIC initially, but then there is a sharp drop for Periods 3 and 4 in STATIC_ZERO when the piece rate is reduced to zero, with clicks significantly lower than in STATIC in each of these periods (t-tests; $p < 0.001$; $p < 0.001$). 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).

[28] The total increase is about 12 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 1: Shrouding of ratchet incentives, warehouse workers online



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 1 of Table G.1 in the appendix 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.

Turning to the COMPLEX treatment, where dynamic incentives are present in periods 1 and 3, Panel (a) shows that warehouse workers are largely unresponsive to dynamic incentives. The overall trend looks very similar to STATIC, and clicks in periods 1 and 3 are not significantly different from those in STATIC. Clicks are also statistically indistinguishable between the treatments for periods 2 and 4.[29] Factors such as reputation concerns, or social preferences towards an employer, are absent and thus cannot explain the lack of response. This further strengthens the case that, for most warehouse workers, the dynamic incentives are a shrouded attribute in this type of incentive contract.

If the response to dynamic incentives is weak due to shrouding, one would also expect that changing the scheme to make dynamic incentives more transparent might lead to a stronger response. In SIMPLE, we do see the zig-zag pattern that is consistent with workers recognizing dynamic incentives, with clicks significantly lower in SIMPLE in periods 1 and 3 compared

[29]There is also no difference in the linear time trend in COMPLEX compared to STATIC; in an OLS regression of clicks on period, a treatment dummy for COMPLEX, and the interaction of period and the treatment dummy, the interaction term is not statistically significant ($p = 0.55$).

to STATIC and COMPLEX (t-tests; all four $p < 0.001$).

Panel (b) of Figure 1 plots the coefficients from a complementary difference-in-differences regression analysis, which allows taking into account any differences in baseline productivity across treatments. The coefficients compare the change in clicks relative to the baseline period in a given treatment to the corresponding change relative to baseline in STATIC. Standard errors are clustered on worker. Panel (b) shows that there is no significant difference in how workers in COMPLEX click over time compared to workers in STATIC. By contrast, workers in SIMPLE have significantly larger drops relative to baseline in Periods 1 and 3 than workers in STATIC, but no significant difference for Periods 2 and 4 where dynamic incentives are absent. The figure also shows that clicks in STATIC_ZERO, Periods 3 and 4, are significantly lower than in STATIC. The regression underlying Panel (b) is shown in column 1 of Table G.1 in the appendix. Columns 2 and 3 show that results are robust to controls for variables that might influence clicks, such as the type of device used for the experiment (e.g., smartphone versus desktop), and measures of worker cognitive ability.

We can compare these results on worker behavior to the predictions of our model of fully-rational effort provision (details are in Appendix F). To calibrate the parameters of the model, we use the response of warehouse workers online under two levels of the piece rate (STATIC versus STATIC_ZERO). As was the case with the warehouse analysis, the model is set-identified and provides bounds. The model predicts, even under our most conservative assumptions about parameters, that workers should have clicked zero in periods 1 and 3 in COMPLEX, but around 200 in these same periods in SIMPLE. This contrasts strongly with what we observe. Conducting a fifth treatment online, e.g., with an intermediate level of the piece rate, would have allowed fully identifying the model, but we did not conduct such a treatment because we were not sure how many warehouse workers would participate in the online experiments, and were concerned about sample size and power. For our experiments with AMT workers, we do have such a treatment (see Section 5), so the model can be fully identified.

Finding 5 *Few workers mention dynamic incentives in COMPLEX in the open-ended question, while twice as many do in SIMPLE.*

Another indication that dynamic incentives are shrouded in COMPLEX, and unshrouded in SIMPLE, is that workers mention dynamic incentives in SIMPLE much more often than in COMPLEX, when asked about optimal work strategies. Three evaluators who were unaware of this paper's research question, or the hypotheses about treatment differences, independently coded responses to the open-ended question asked after Period 4, about what the worker would recommend to someone else as a strategy for working in Periods 3 and 4. The evaluators looked for any indication that the worker recognized a reason to click less in Period 3 than Period 4, 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).

The results show that only 19 percent of workers in COMPLEX indicated some awareness of the dynamic incentive, whereas in SIMPLE this share is 44 percent. The difference is highly statistically significant (Wilcoxon test; $p < 0.001$). Table G.2 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.

We hypothesized that key reasons why dynamic incentives in SIMPLE are easier to understand is because of the structure of the dynamic incentives, and the way that dynamic incentives are explained. However, an alternative explanation could be that the instruction text for SIMPLE happened to be easier for workers to understand in general, e.g., because of an easier reading grade level. The instructions for SIMPLE do have slightly fewer words than COMPLEX, but have a higher reading-grade level, 8 versus 7.1, and a lower ease of reading score, 73.1 versus 75.5, respectively (see Table G.9 in the appendix for these statistics for all treatments).[30] Thus, the greater shrouding of dynamic incentives in COMPLEX compared to SIMPLE does not seem to be explainable by differences in text difficulty of the instructions in general, but rather something specific about how SIMPLE makes dynamic incentives more transparent.[31]

[30] 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.

[31] Notably, 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 G.9, all are roughly at reading grade level 7). Thus, our experiment instructions for COMPLEX are well calibrated to the reading level of communication materials used for the corresponding warehouse incentive schemes by the firm, and SIMPLE is actually harder than what the firm uses.

In our experiments with AMT workers, discussed in Section 5, we use additional treatments to explore in more detail which contract features, and aspects of how incentives are explained, are driving the greater shrouding of dynamic incentives in COMPLEX compared to SIMPLE.

If a lack of response to dynamic incentives reflects shrouding, 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.

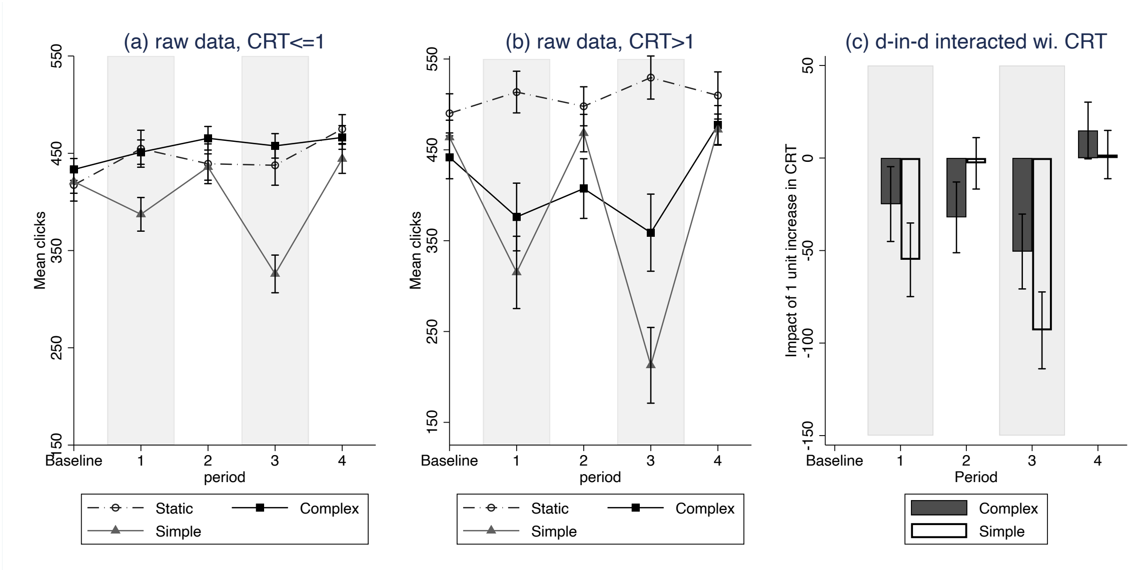
Finding 6 *Warehouse workers with higher CRT scores show a significantly stronger response to dynamic incentives. Other aspects of cognitive ability have limited explanatory power.*

Panels (a) and (b) of Figure 2 show behavior of workers in COMPLEX and SIMPLE according to CRT scores (results for all four treatments are shown in Figure G.3 and Table G.3 in the appendix). We see that workers with relatively low CRT scores, answering zero or one questions correctly, exhibit essentially no response to dynamic incentives in COMPLEX, whereas workers who answer two or three correctly do show signs of the zig-zag pattern characteristic of recognizing the dynamic incentives. [32] In SIMPLE, even low CRT workers exhibit a response to dynamic incentives, but the response is much larger for workers with higher CRT scores.[33] The weaker relationship of CRT to noticing dynamic incentives in COMPLEX compared to SIMPLE, for both high and low CRT workers, is consistent with the former incentive scheme being more complex and relatively harder to understand.

[32]For workers with low cognitive ability (CRT ≤ 1) in COMPLEX and STATIC, clicks in Period 1 and Period 3 are not significantly different. For workers with high cognitive ability (CRT > 1) clicks are significantly lower in COMPLEX than in STATIC (t-tests; $p < 0.004$; $p < 0.001$). In Figure G.2 in the appendix we also show graphs for each level of CRT separately, which supports our binarization of the CRT score for Figure 2. Specifically, we see that workers with CRT of 0, and workers with CRT of 1 do not recognize dynamic incentives in COMPLEX, whereas workers with CRT scores above 1 do respond (there are very few – only 6 – warehouse workers with CRT scores of 3).

[33]Clicks in Periods 1 and 3 are significantly lower in SIMPLE than STATIC, among workers with low cognitive ability (t-tests; $p < 0.01$; $p < 0.001$) as well as among workers with high cognitive ability (t-tests; $p < 0.001$; $p < 0.001$).

Figure 2: Shrouding of ratchet incentives and CRT, warehouse workers online



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 G.3 in the appendix for all coefficients). The vertical shaded bars in all panels denote periods with dynamic incentives to reduce effort in COMPLEX and SIMPLE.

In Panel (c) of Figure 2 we provide results from a corresponding difference-in-differences regression analysis. The figure plots coefficients of triple interactions, Period*Treatment*CRT, where we interact treatment differences with CRT scores (linearly). Panel (c) shows that in COMPLEX, a higher CRT is not associated with clicks in Period 1 and 2, but leads to a significant drop in Period 3. In SIMPLE, higher CRT matters significantly for the size of the drop in Periods 1 and 3, with no significant interactions in periods without dynamic incentives. The interaction effects of CRT with Periods 1 and 3 in SIMPLE are also larger than the corresponding interactions for COMPLEX.[34] The regression underlying Panel (c) is shown in column 1 of Table G.3 in the appendix, and subsequent columns show that results are robust to adding controls.

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

[34]The difference in interaction terms is not statistically significant for Period 1 (F-test; $p < 0.12$) but is significant for Period 3 (F-test; $p < 0.001$).

correlations; $\rho = 0.21$, $\rho = 0.32$, $p = 0.012$, $p < 0.001$). This provides another indication that the aspect of cognitive ability captured by CRT plays a role in noticing the dynamic incentive aspects of these respective incentive schemes.

We also explore whether our measures of other aspects of cognitive ability and worker traits matter for shrouding (see Tables G.4 to G.6 in the appendix). Higher educational attainment, measured by years of schooling, is associated with significantly stronger responses to dynamic incentives in SIMPLE in Period 1 and Period 3, but there is not a significant relationship to response to dynamic incentives in COMPLEX. Our indicators for narrow choice bracketing, and for difficulties with backwards induction, are not significantly related to responses to dynamic incentives in any systematic way. As expected given the short time frame, time preference is unrelated to responses to dynamic incentives in the online experiments with warehouse workers, and risk aversion is also not systematically related to behavior (see Tables G.7 and G.8).

4.3 Combining Field Experiment and Online Experiment Data

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.[35] We can thus correlate their behavior in the field and in the online experiment.

Finding 7 *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, we conduct a principal component analysis of three variables: (i) a dummy indicating whether the worker reduced effort in Period 1 relative to the baseline period; (ii) a dummy indicating whether the worker reduced effort in Period 3 relative to Period 2; and (iii) 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. Table 5 replicates Table D.4 and adds an interaction of

[35] 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 B.1.

treatment with the variable "showed RE online". We find that workers who show a ratchet effect in the online experiment also show a stronger ratchet effect in the warehouse.

Table 5: Correlation of online and field experiment behavior

Dependent variable: ln(units per hour)	(1)	(2)	(3)
1 if treated	-0.0029	-0.0112	-0.0112
	(0.016)	(0.017)	(0.017)
1 if showed RE online × treated	-0.0603***	-0.0671***	-0.0671***
	(0.018)	(0.019)	(0.019)
Sample	COMPLEX/SIMPLE During trial	COMPLEX/SIMPLE During trial, period 3+	COMPLEX/SIMPLE During trial, period 3+ Working entire next period
Rates area FE	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes
Shift FE	Yes	Yes	Yes
all FE's × cohort	Yes	Yes	Yes
all FE's × showed RE	Yes	Yes	Yes
# Workers	154	153	153
# 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 D.4 but adds interactions of the treatment dummy with a variable for having shown a ratchet effect in the online experiment. This variable is the standardized first principal component of a PCA of reducing effort from baseline period to Period 1; reducing effort from Period 2 to 3; and mentioning arguments relating to ratchet effects or dynamic incentives in an open-ended question. The sample is restricted to workers who participated in the online experiment and were then randomly allocated to the COMPLEX and SIMPLE treatments. The sample is also restricted to the time during the trial, when the treatment workers 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 for the full sample. Specification 2 restricts 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. Specification 3 further restricts 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 dummy for showing a ratchet effect in the online experiment. Significance at the 1, 5, and 10 percent level is denoted by ***, **, and *, respectively.

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 like in Table 5.

5 Online experiments with AMT workers

The online experiments with warehouse workers were constraint by the number of available participants. We could thus not explore which contract features drive the treatment effect between COMPLEX and SIMPLE. To be able to do so, we conducted several additional treatments with workers recruited via the online labor market Amazon Mechanical Turk (AMT). An overview of all treatments and complete instructions are provided in Appendix J.

We first show that AMT workers react to our treatments very much like the warehouse workers. We conducted the same four treatments as with the warehouse 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.

Finding 8 *AMT workers respond very similarly to the treatments compared to warehouse workers. 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 among AMT workers.*

Appendix H.1 describes all of the replication results and the now point-identified structural estimation in detail and discusses the effect of higher cognitive ability among AMT workers.

The main focus of the AMT treatments is in the next section, where we study which contract features drive behavior. We also check whether the shrouding effect of the COMPLEX treatment is robust to realistic pertubations in Section 5.2.

5.1 Identifying contract features that contribute to shrouding

COMPLEX differs in three distinct features from SIMPLE: dynamic inventives affect the slope of future earnings (rather than their level), requiring complex contingent thinking; it frames the dynamic incentives in terms of SPM rather than directly in monetary terms; and it has noise in the target rate via the random parameter X. In this section, we present two treatments that disentangle which of these contract features contribute to the shrouding of dynamic incentives in COMPLEX.

The first treatment, NOISE, is the same as SIMPLE except that it adds noise to the target rate. Specifically, the target rate in Periods 2 and 4 is the average of the fixed target rate

400 and a random variable X $\in \{380, 400, 420\}$, just as in COMPLEX. Comparing SIMPLE to NOISE allows us to measure the effect of noise in the target rate.

The second treatment, NOISE_MARGINAL, builds on NOISE by adding a requirement of complex contingent reasoning when thinking about dynamic incentives. Working fast in Period 1 now affects the *target rate* in Period 2 and thus the marginal earnings (like in COMPLEX), rather than directly affecting the level of earnings (like in SIMPLE). Comparing NOISE to NOISE_MARGINAL allows us to measure the effect of complex contingent reasoning.[36] The only difference between NOISE_MARGINAL and COMPLEX is that COMPLEX frames dynamic incentives in terms of SPM rather than framing them directly in monetary terms. That treatment comparison thus allows us to estimate the effect of this implicit framing of the monetary consequences of effort provision.

We find that all three contract features matter.

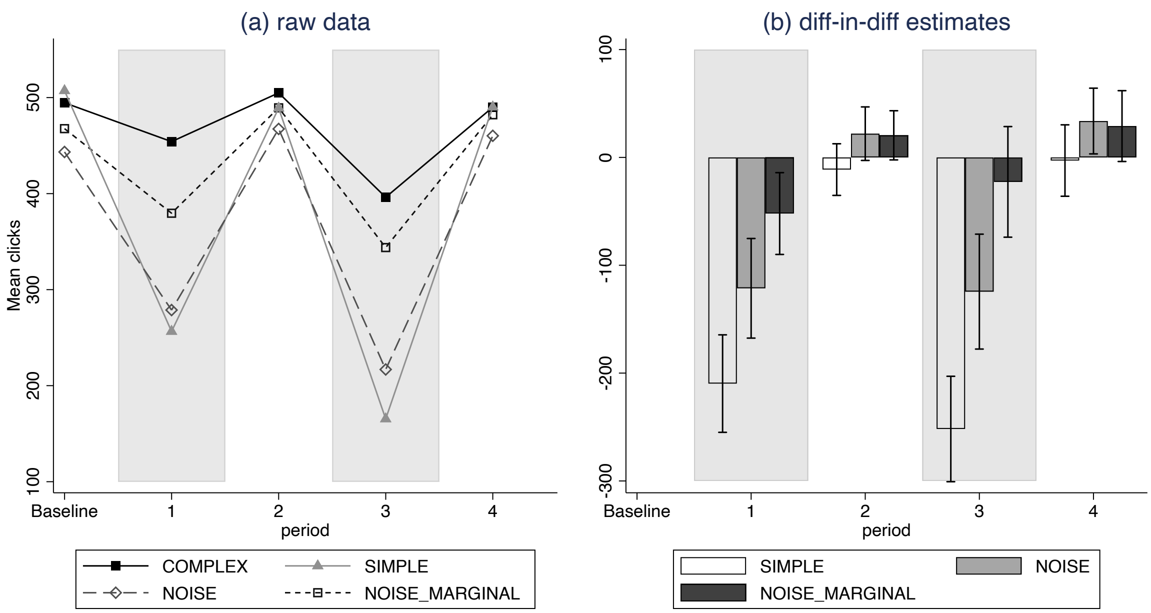
Finding 9 *Noise in the target rate, having dynamic incentives that involve relatively complex contingent thinking and making financial consequences of dynamic incentives implicit, all contribute to shrouding of dynamic incentives.*

Panel (a) of Figure 3 shows effort across treatments and periods. The reaction to dynamic incentives can be seen in Periods 1 and 3. The four treatments line up nicely, with SIMPLE inducing the strongest reaction, then NOISE, then NOISE_MARGINAL and finally COMPLEX with the weakest reaction. To test between treatments, we use difference-in-differences regressions, regressing effort in each period on treatment and period dummies and their interactions (with COMPLEX and baseline as omitted categories), clustering standard errors on worker. Such a regression guards against any randomization error between treatments by controlling for effort in the baseline period. Panel (b) shows the average coefficients for the interaction of Period 1 and 3 with each treatment. We use F-tests of these two coefficients to

[36]Depending on the level of intrinsic motivation, this treatment difference also changes the optimal level of effort provision. If intrinsic motivation is zero, then the optimal effort in all of the treatments in zero. But if intrinsic motivation is positive, which we find it is, then the optimal effort in Periods 1 and 3 of SIMPLE and NOISE is higher, while it is still zero in NOISE_MARGINAL and COMPLEX. The incentive to reduce effort is thus smaller in SIMPLE and NOISE compared to NOISE_MARGINAL and COMPLEX. Our estimate of the effect of complex contingent reasoning will thus likely be downward biased. We designed SIMPLE before we had estimates of workers' intrinsic moviation and chose a design that produced a conservative estimate for all levels of intrinsic motivation. To obtain an unbiased estimate of the effect of complex contingent reasoning, one could design a version of SIMPLE that has an optimal effort of zero for the level of intrinsic motivation that we find among warehouse and AMT workers.

compare across treatments. The full regression is in column 1 of Table H.2 in the appendix. All treatment differences are highly significant. The other columns in that table show robustness to additional control variables. Propensity to mention dynamic incentives also decreases monotonically going from SIMPLE, to NOISE, to NOISE_MARGINAL, to COMPLEX (percentages are 80, 71, 43, and 40, respectively).

Figure 3: Contract features contributing to shrouding, 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.4 in the appendix for all coefficients). The vertical shaded bars in both panels denote periods with dynamic incentives to reduce effort in all of the treatments.

5.2 Robustness of shrouding to perturbations to the incentive scheme

Another natural question is whether the shrouding result will generalize to realistic perturbations to the class of real world incentive schemes that we study, and more generally, whether shrouding is a fragile or robust phenomenon. A third set of experiments with AMT workers allows us to explore these questions.

We explore three variations on the incentive scheme. LINEAR is a condition that simplifies the scheme by paying 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 extra construct of SPM from the instructions, and explains everything in monetary terms directly, e.g., we speak of a wage per click, and how clicking more in Period 1 would reduce the wage per click in Period 2. Lastly, we implement a condition LINEAR_NOSPM, which makes the piece rate linear and eliminates SPM. In all, we had 369 AMT workers participate in these three treatments with roughly equal amounts in each.

Finding 10 *Shrouding of ratchet incentives is robust to making the scheme linear, or making monetary consequences more salient by eliminating SPH. There is a positive interaction of combining both simplifications, for increasing the response to dynamic incentives, 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, with no significant differences (see Figure H.5 and Table H.3 in the appendix).[37] Workers in LINEAR_NOSPM do have a statistically significantly stronger response to dynamic incentives than workers in COMPLEX, but the difference is modest in size, and much smaller than the response observed in SIMPLE.[38] There is also a statistically significantly stronger response in LINEAR_NOSPM compared to LINEAR or NOSPM.[39] The propensity to mention dynamic incentives mirrors the results on behavior: 40 percent, 45 percent and 46 percent mention dynamic incentives in COMPLEX, LINEAR, and NOSPM, respectively, compared to substantially higher fraction, 57 percent, in LINEAR_NOSPM.[40]

These findings have important practical implications, as they show how various plausible changes to a widely used class of incentive schemes affect shrouding of perverse dynamic incentives. Shrouding is relatively robust, in the sense that it does not depend crucially on

[37] The number of clicks is not significantly different in Period 1 or Period 3, comparing LINEAR to COMPLEX (t-tests ;$p > 0.25$; $p > 0.83$). Comparing NOSPM to COMPLEX, differences are also not statistically significant in Period 1 and Period 3 (t-tests; $p > 0.19$; $p > 0.74$).

[38] The number of clicks is significantly lower in LINEAR_NOSPM compared to COMPLEX, in Period 1, and marginally significant in Period 3 (t-tests; $p < 0.001$; $p < 0.08$). In the difference-in-differences analysis shown in Table H.3, LINEAR_NOSPM has a significantly stronger drop relative to COMPLEX, in periods 1 and 3.

[39] The interaction terms for LINEAR_NOSPM with Periods 1 and 3 are significantly different from the corresponding interaction terms for LINEAR and NOSPM, except that the comparison of LINEAR_NOSPM*P3 versus LINEAR*P3 is not significant at conventional levels (F-tests; $p < 0.001$, $p < 0.11$, $p < 0.001$, $p < 0.02$).

[40] The differences are not statistically significant, comparing LINEAR or NOSPM to COMPLEX, but the difference between LINEAR_NOSP and COMPLEX is significant (Wilcoxon test; $p < 0.01$).

one particular formulation of the scheme, although combining simplifications has a positive interaction effect that starts to lead to unshrouding.

6 Conclusion

This paper provides empirical support for the importance of contract complexity, and heterogeneity in worker bounded rationality, for understanding optimal incentives within organizations. Specifically, the paper shows that an important aspect of an organization's workplace incentive scheme – dynamic incentives in the form of the so-called ratchet effect – are a shrouded attribute that some workers neglect due to complexity. In field experiments within the firm, and in online experiments with real-effort tasks, many workers make choices consistent with being unaware of dynamic incentives. Changing the contract to make the dynamic incentives more transparent, or looking at the sub-sample of workers with high cognitive ability, a response to dynamic incentives emerges.

One implication of these findings is that organizations could find it optimal to introduce some complexity into incentive schemes. To the extent that complexity shrouds negative incentives, such as the ratchet effect, but preserves positive incentive effects, complexity could be beneficial for organizations. As an illustration, in our online experiments with the warehouse workers, COMPLEX generates higher "profits" than SIMPLE, if we calculate profits as total clicks over the five periods minus total payments from the experimenter.[41] Presumably, there may be a level of complexity that even interferes with workers recognizing the static incentives, and thus reduces profits. In line with this hypothesis, we do find that warehouse workers with lower cognitive ability have a weaker response to reducing the static incentives in the online experiments (see Table G.3).

Another implication is that optimal incentives may involve organizations tailoring the level of complexity to the cognitive sophistication of the workforce. Comparing our online experiments with warehouse workers and AMT workers, we do find signs of stronger ratchet effects among the population with higher cognitive ability, namely the AMT workers (see Table G.1). Variation in cognitive skills across occupations could be a novel mechanism for

[41] This result depends on how clicks are assumed to translate into revenues, but it is proof of concept that complexity can, for some profit functions, increase profitability.

explaining variation in incentive schemes across occupations and skill levels.

The role of bounded rationality in determining the response to incentives also has implications for thinking about the impact of incentives over time. To the extent that experience can improve understanding of complexity, incentives might have changing effects over time due to gradual unshrouding of attributes of the scheme. To the extent that learning is partially social, there are also additional interesting implications, for factors that could matter for the impact of incentives: The frequency of worker interaction, the nature of social networks and social ties among workers, the rate of worker turnover, and the proportion of the workforce with high cognitive ability, could all matter for the impact of incentives over time.

The paper also sheds some light on specific types of contract features that can contribute to complexity. These findings can provide guidance for contract design, as well as theories that try to model the nature of human bounded rationality and explain what makes contracts harder or easier to understand. Specifically, the findings point to framings that make monetary consequences less salient, distractors, and incentive structures that require complicated contingent thinking, as contributors to complexity and shrouding.

Regarding ratchet effects specifically, the message of the paper is *not* that these do not exist or are unimportant. Indeed, in our setting, we do find ratchet effects among workers with higher cognitive ability, and in incentive schemes that are more transparent. Our findings are a warning that, in worker populations that are highly sophisticated, ratchet effects could emerge quickly, and be large if contract complexity is held constant. But ratchet effects may be heterogeneous depending on worker cognitive ability. Complexity may slow down the emergence of ratchet effects even if in the long run they do emerge, and complexity could be one tool for firms to try to mitigate such effects.

References

Abeler, Johannes and Simon Jäger (2015). "Complex tax incentives". In: *American Economic Journal: Economic Policy* 7.3, pp. 1–28.

Agarwal, Sumit, Changcheng Song, and Vincent W Yao (2017). "Banking competition and shrouded attributes: evidence from the US mortgage market". In: *Georgetown McDonough School of Business Research Paper* 2900287.

Anagol, Santosh and Hugh Hoikwang Kim (2012). "The impact of shrouded fees: Evidence from a natural experiment in the Indian mutual funds market". In: *American Economic Review* 102.1, pp. 576–93.

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.

Bellemare, Charles and Bruce Shearer (2014). "Measuring Ratchet Effects within a Firm: Evidence from a Field Experiment Varying Contractual Commitment". In: *IZA Discussion Paper No. 8214*.

Bolton, Patrick and Antoine Faure-Grimaud (2010). "Satisficing contracts". In: *The Review of Economic Studies* 77.3, pp. 937–971.

Börgers, Tilman and Jiangtao Li (2019). "Strategically simple mechanisms". In: *Econometrica* 87.6, pp. 2003–2035.

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

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.

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.

Carlin, Bruce I (2009). "Strategic price complexity in retail financial markets". In: *Journal of Financial Economics* 91.3, pp. 278–287.

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: an experimental analysis”. 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.

Cooper, David J 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.

Dalton, Christina M, Gautam Gowrisankaran, and Robert J Town (forthcoming). “Salience, Myopia, and Complex Dynamic Incentives: Evidence from Medicare Part D”. In: *The Review of Economic Studies*.

DellaVigna, Stefano (2009). “Psychology and economics: Evidence from the field”. In: *Journal of Economic literature* 47.2, pp. 315–72.

DellaVigna, Stefano and Joshua Pollet (2008). “Investor inattention and Friday earnings announcements”. In: *Journal of Finance, forthcoming*.

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

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

Elminejad, M Ali, Tomas Havranek, and Roman Horvath (2021). “Publication and Identification Biases in Measuring the Intertemporal Substitution of Labor Supply”. In: *CEPR Discussion Paper No. DP16032*.

Englmaier, Florian, Andreas Roider, and Uwe Sunde (2016). “The role of communication of performance schemes: Evidence from a field experiment”. In: *Management Science* 63.12, pp. 4061–4080.

Enke, Benjamin and Thomas Graeber (2019). “Cognitive uncertainty”. In: *NBER Working Paper No. 26518*.

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

Farhi, Emmanuel and Xavier Gabaix (2015). *Optimal taxation with behavioral agents*. Tech. rep. National Bureau of Economic Research.

Finkelstein, Amy (2009). “E-ztax: Tax salience and tax rates”. In: *The Quarterly Journal of Economics* 124.3, pp. 969–1010.

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

Gabaix, Xavier (2014). “A sparsity-based model of bounded rationality”. In: *The Quarterly Journal of Economics* 129.4, pp. 1661–1710.

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.

Gabaix, Xavier, David Laibson, et al. (2006). “Costly information acquisition: Experimental analysis of a boundedly rational model”. In: *American Economic Review* 96.4, pp. 1043–1068.

Garicano, Luis and Andrea Prat (2013). “Organizational economics with cognitive costs”. In: *Advances in Economics and Econometrics* 1, p. 342.

Glazer, Jacob and Ariel Rubinstein (2012). “A model of persuasion with boundedly rational agents”. In: *Journal of Political Economy* 120.6, pp. 1057–1082.

Goldin, Jacob (2015). “Optimal tax salience”. In: *Journal of Public Economics* 131, pp. 115–123.

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

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

Heidhues, Paul and Botond Kőszegi (2017). “Naivete-based discrimination”. In: *The Quarterly Journal of Economics* 132.2, pp. 1019–1054.

Herrnstein, Richard J 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 Seongyeon Lim, and Siew Hong Teoh (2009). “Driven to distraction: Extraneous events and underreaction to earnings news”. In: *The Journal of Finance* 64.5, pp. 2289–2325.

Holmström, Bengt (1979). “Moral hazard and observability”. In: *The Bell Journal of Economics*, pp. 74–91.

Jakobsen, Alexander M (2020). “A model of complex contracts”. In: *American Economic Review* 110.5, pp. 1243–73.

Jin, Ginger Zhe, Michael Luca, and Daniel Martin (2021). “Complex disclosure”. In: *Management Science*.

Koszegi, Botond (2014). “Behavioral contract theory”. In: *Journal of Economic Literature* 52.4, pp. 1075–1118.

Kőszegi, Botond and Adam Szeidl (2012). “A model of focusing in economic choice”. In: *The Quarterly journal of economics* 128.1, pp. 53–104.

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

Li, Shengwu (2017). “Obviously strategy-proof mechanisms”. In: *American Economic Review* 107.11, pp. 3257–87.

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

MacLeod, W Bentley (1996). “Decision, contract, and emotion: Some economics for a complex and confusing world”. In: *Canadian Journal of Economics*, pp. 788–810.

Malcomson, James M (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: Evidence from the Russian Empire”. In: *American Economic Review* 108.4-5, pp. 1074–1117.

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.

Matějka, Michal, Matthias D Mahlendorf, and Utz Schäffer (2022). “The ratchet effect: Theory and empirical evidence”. In: *Management Science*.

Mathewson, Stanley Bernard (1931). “Restriction of Output among Unorganized Workers”. In.

Misra, Sanjog and Harikesh S Nair (2011). “A structural model of sales-force compensation dynamics: Estimation and field implementation”. In: *Quantitative Marketing and Economics* 9.3, 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 S and V Srinivasan (1996). "Quota-based compensation plans for multiterritory heterogeneous salesforces". In: *Management Science* 42.10, pp. 1454–1462.

Rees-Jones, Alex and Dmitry Taubinsky (2019). "Measuring Schmeduling." In: *forthcoming in Review of Economic Studies*.

Rothschild, Michael and Joseph Stiglitz (1976). "Equilibrium In Competitive Insurance Markets: An Essay On The Economics Of Imperfect Information". In: *The Quarterly Journal of Economics* 90.4, pp. 629–649.

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

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

Spiegler, Ran (2011). *Bounded rationality and industrial organization*. Oxford University Press.

Taubinsky, Dmitry and Alex Rees-Jones (2017). "Attention variation and welfare: theory and evidence from a tax salience experiment". In: *The Review of Economic Studies* 85.4, pp. 2462–2496.

Tirole, Jean (2009). "Cognition and incomplete contracts". In: *American Economic Review* 99.1, pp. 265–94.

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

Online appendix

A Theoretical appendix for field studies

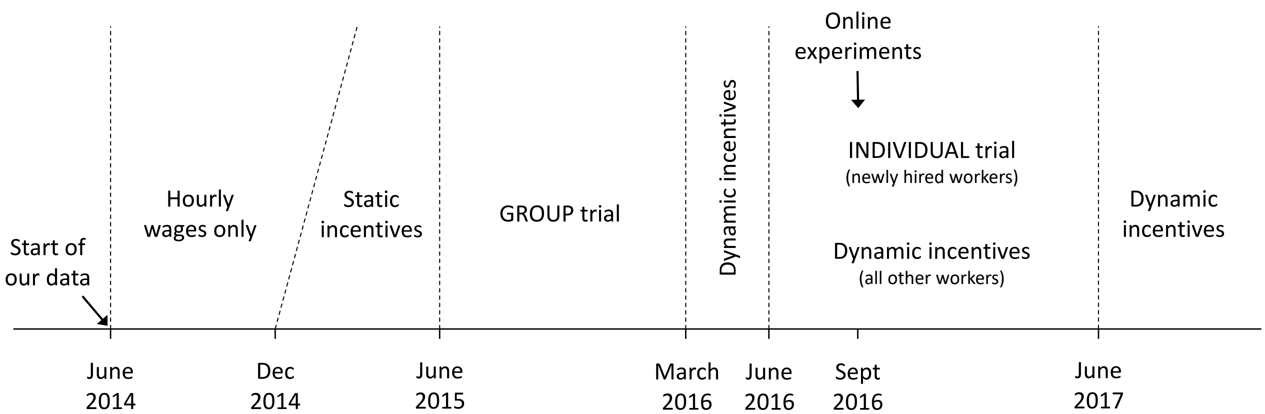
Proof of Proposition 1: We focus on a single individual and so suppress i subscripts. We first prove part (i). If individuals face only static incentives ($\zeta = 0$), then $U_i(e_1) - U_i(e_1^{'}) = a_f(e_1) - a_f(e_1^{'}) + \gamma(\frac{e_1}{\eta_1}) - \gamma(\frac{e_1^{'}}{\eta_1})$, which is (weakly) increasing in w for $e_1 > e_1^{'}$. Standard monotone comparative statics imply the optimal choice of effort is higher under $w > 0$ than $w = 0$.

We now turn to part (ii). We first compare the utility from a low effort to a higher effort when $\zeta = 0$: $a_f(e_1) - a_f(e_1^{'}) + \gamma(\frac{e_1}{\eta_1}) - \gamma(\frac{e_1^{'}}{\eta_1})$. We next compare the same utility when $\zeta > 0$: $a_f(e_1) - a_f(e_1^{'}) + \gamma(\frac{e_1}{\eta_1}) - \gamma(\frac{e_1^{'}}{\eta_1}) + \gamma(\frac{e_2}{\zeta\eta_1 + (1-\zeta)\epsilon_1}) - \gamma(\frac{e_2^{'}}{\zeta\eta_1^{'} + (1-\zeta)\epsilon_1})$. We know from the static model that if $e_1^{'} > e_1$ then all else equal $e_2^{'} \leq e_2$. The difference between these two equations is: $-\gamma(\frac{e_2}{\zeta\eta_1 + (1-\zeta)\epsilon_1}) + \gamma(\frac{e_2^{'}}{\zeta\eta_1^{'} + (1-\zeta)\epsilon_1})$, which since $e_1^{'} > e_1$ and $e_2^{'} \leq e_2$ must be negative. Thus, we see that higher effort levels have a decreasing advantage compared to lower effort levels when ζ goes from 0 to positive, and so the optimum choice of effort must fall.

B Additional results for the INDIVIDUAL trial

This appendix shows a timeline of all the changes to the incentive scheme we study (Figure B.1). It then provides summary statistics for the INDIVIDUAL trial (Table B.1). Finally, it shows that there is no differential attrition before and during the INDIVIDUAL trial (Figure B.2 and Table B.2).

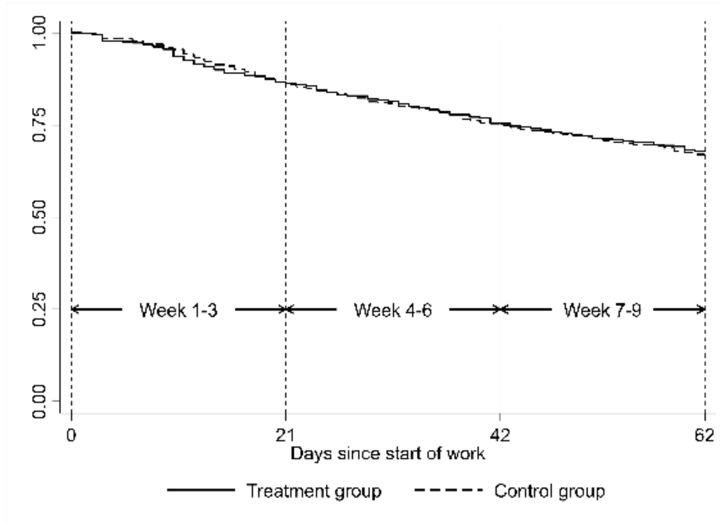
Figure B.1: Timeline of changes to the incentive scheme



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 from December 2014.

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

Figure B.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 B.2.

Table B.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. 'Treated' is 1 for workers in the treatment group. Since we have little pre-trial data for this trial, we only control for cohort fixed effects. Significance at the 1, 5, and 10 percent level is denoted by ***, **, and *, respectively.

C Analysis of the introduction of static incentives

This appendix presents the empirical analysis of the introduction of static incentives. The data on worker productivity allow us to assess whether workers were motivated to work harder by 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 calibrate our structural model of effort provision, as discussed in Section 3.2.3.

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 B.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 3.1.2. When incentives were first rolled out, target rates were based on average speeds of all workers in each rates area 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. 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. Table C.1 shows summary statistics for the data set. The control warehouse has been in service for longer than the treated warehouse. Workers in the control warehouse thus have longer tenure on average, as seen in the table. At the same time, the overall structure of the production process in the two warehouses is very similar. The table shows that, for example, the ratio of night to day shift and the allocation of workers across tasks (here: share of Zonepick) is similar. 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, incentives were rolled out gradually, rates area by rates area. From then on, all rates areas were incentivized. We thus have performance 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.

Table C.1: Summary statistics for the introduction of static incentives

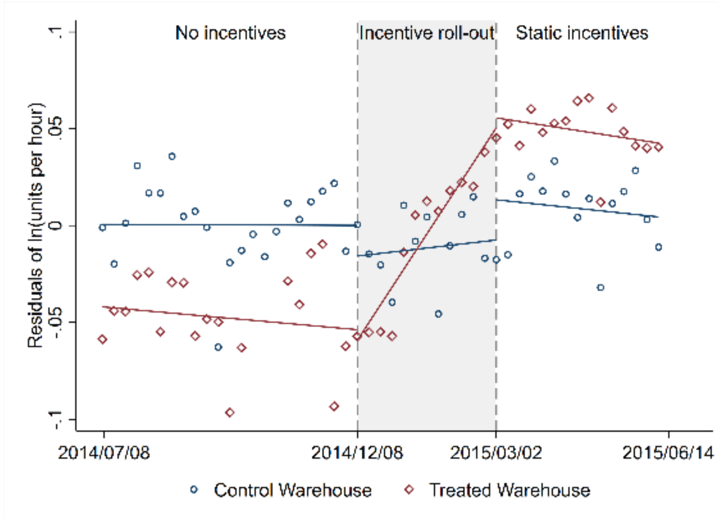
	Mean		p-value	Share of missing values	
	Control	Treatment		Control	Treatment
Tenure at end of roll-out	1528.37	232.35	0.000	0.00	0.00
Working time in 5 months before roll-out	316.91	260.68	0.000	0.09	0.15
Age at end of roll-out	31.13	33.39	0.244	0.99	0.36
1 if female	0.55	0.29	0.044	0.99	0.12
1 if non-native	0.95	0.56	0.000	0.99	0.41
Share of workers working mostly at night	0.62	0.63	0.830	0.00	0.00
Share of workers working mostly in Zonepick	0.48	0.51	0.132	0.00	0.00
# Workers	2141	2678			

Notes: p-value from OLS regressions clustering on individual workers. Tenure at start of roll-out is the number of days between the first day a worker starts working in the firm and the start of the incentives roll-out (and missing for workers starting after that date). Working time is the total time worked in hours between 8 June 2014 (start of data) and the start of the incentives roll-out (and missing for workers starting after that date).

Finding 11 *The introduction of static incentives increases worker productivity by 12.4 percent.*

Figure C.1 plots average weekly worker speed for each warehouse, measured as residuals of ln(units per hour) residualized for the control variables in column 2 of Table C.2 (see below for details). The figure shows that performance 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, performance in the control warehouse does not change much across the entire period.

Figure C.1: Visual Diff-in-Diff of introduction of static incentives (with 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, rates area by rates area, 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 2 in Table C.2. The dependent variable is thus residualized for rates-area 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. 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 speed in that warehouse. This rate setting rule was unchanged during the period shown in the graph.

The corresponding difference-in-differences regressions are shown in Table C.2. The regressions control for any time-invariant differences between warehouse 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.[42]

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

Table C.2: Diff-in-Diff analysis of introduction of static incentive on productivity

Dependent variable: ln(units per hour)	(1)	(2)	(3)	(4)	(5)	(6)
1 if static incentives	0.1276***	0.1241***	0.1359***	0.1311***	0.1039***	0.1053***
	(0.008)	(0.009)	(0.008)	(0.008)	(0.009)	(0.009)
Total time worked per warehouse and shift		-0.0371***		-0.0323**		-0.0402***
		(0.014)		(0.013)		(0.010)
Average tenure per warehouse and shift		0.0616**		0.0553**		1.5623***
		(0.027)		(0.025)		(0.315)
Sample	Full	Full	Full	Full	Restricted	Restricted
Rates Area 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	4588	4588	4588	4588	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. '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. Significance at the 1, 5, 10 percent level is denoted by ***, **, and *, respectively.

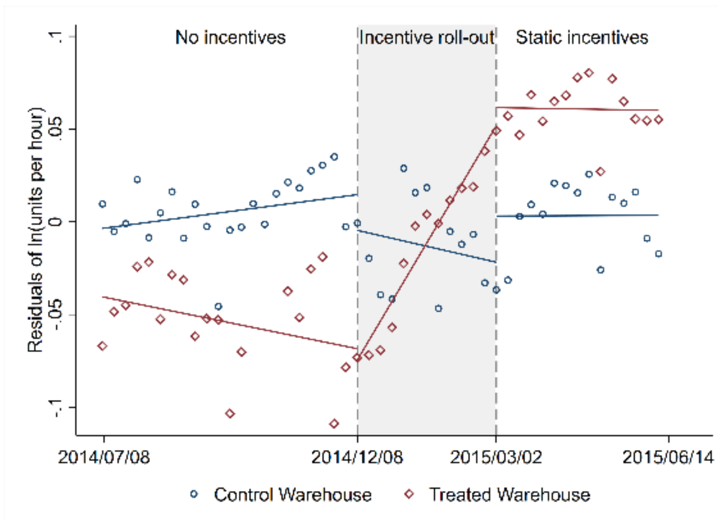
Column 2 is the specification that corresponds to Figure C.1 and is our preferred specification. It shows that the introduction of static incentives lead to a 12.4 percent increase in worker effort. The specifications in columns 1, 3, and 4 yield very similar results (Figure C.2 shows the corresponding 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 productivity in the control warehouse does not change over the time period, this should not 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 estimation in Section 3.2.3.

Figures C.3 to C.5 and Tables C.3 to C.5 analyze differential attrition between the two warehouse 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). Figure C.3 and Table C.3 consider workers who were employed on 8 July 2014. 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 speed, as this would bias the results in Table C.2. Column 2 of C.3 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 C.2, 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 incentive roll-out.

As robustness check, Figure C.4 and C.5 (and the corresponding Tables C.4 and C.5) show attrition analysis for workers who were employed on 8 December 2014 and 2 March 2015, respectively. Results are very similar to Figure C.3 and Table C.3.[43]

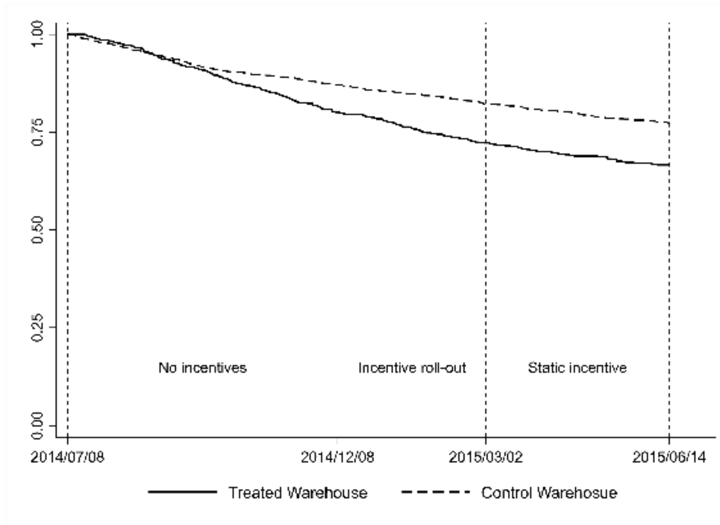
Figure C.2: Visual Diff-in-Diff of introduction of static incentives (without additional controls)



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, rates area by rates area, 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 C.2. The dependent variable is thus residualized for rates-area 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 speed in that warehouse. This rate setting rule was unchanged during the period shown in the graph.

[43] Table C.5 only contains a univariate regression, as the control variables all use pre-rollout data. The resulting regression with control variables is thus identical to column 5 in Table C.4.

Figure C.3: Attrition during the introduction of static incentives (workers employed in July 2014)



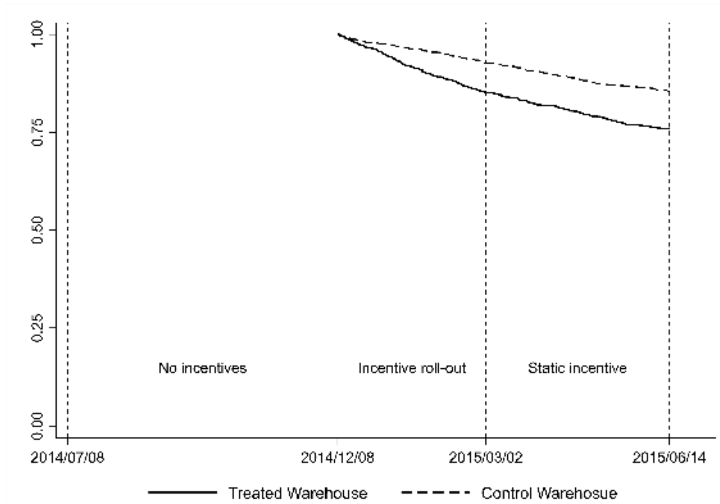
Notes: Kaplan-Meier survival estimates for the introduction of static incentives for workers who were employed on 8 July 2014. The vertical lines show the start and the end of the roll-out of static incentives in the treated warehouse. Corresponding regressions are in Table C.3.

Table C.3: 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	1.5853***	0.2220***	6.0447e+13***	1.5745***	0.9564	6.9258e+14***	1.3162	0.6893	5.4446e+15
	(0.168)	(0.073)	(4.510e+13)	(0.199)	(0.178)	(3.386e+14)	(0.247)	(0.206)	(.)
Tenure at start of baseline period		0.4362***	1.0000		0.6375***	1.0000		0.6025***	1.0000
		(0.055)	(0.608)		(0.058)	(0.272)		(0.073)	(0.545)
Tenure × treated WH		0.0816***	0.3483		1.5729*	1.1364		1.5528	1.1264
		(0.056)	(.)		(0.426)	(.)		(0.824)	(.)
Pre-roll-out speed		0.5984***	1.0000		0.8527	1.0000		0.8057	1.0000
		(0.052)	(0.090)		(0.098)	(0.072)		(0.135)	(0.103)
Pre-roll-out speed × treated WH		1.5284***	0.9757		1.1555	1.0181		1.2516	0.9969
		(0.181)	(.)		(0.155)	(.)		(0.254)	(.)
1 if female			1.0000			1.0000			1.0000
			(0.297)			(0.280)			(0.483)
1 if female × treated WH			0.7036			0.7838			0.6943
			(.)			(.)			(.)
Age at start of baseline period			1.0000			1.0000			1.0000
			(0.103)			(0.104)			(0.163)
Age × treated WH			0.9535			1.1078			0.9500
			(.)			(.)			(.)
# Workers	2270	1901	615	1916	1608	530	1781	1498	481

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. 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 rates-area fixed effects, i.e., correcting for the fact that a unit is harder or easier in different rates areas. 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. Significance at the 1, 5, and 10 percent level is denoted by ***, **, and *, respectively.

Figure C.4: Attrition during the introduction of static incentives (workers employed in December 2014)



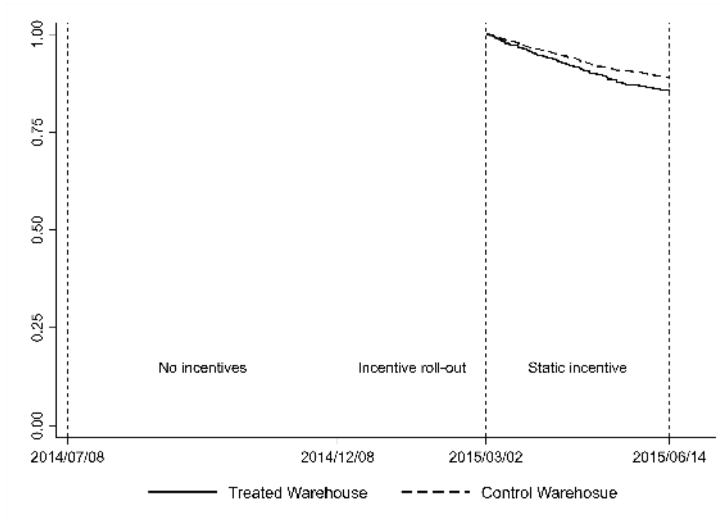
Notes: Kaplan-Meier survival estimates for the introduction of static incentives for workers who were employed on 8 December 2014. The vertical lines show the start and the end of the roll-out of static incentives in the treated warehouse. Corresponding regressions are in Table C.4.

Table C.4: Attrition during the introduction of static incentives (workers employed in December 2014)

Dependent variable: Worker left firm	Dec 2014 - Jun 2015			Mar 2015 - Jun 2015		
	(1)	(2)	(3)	(4)	(5)	(6)
1 if treated warehouse	1.7873***	0.6587*	1.0644e+14***	1.4313***	0.4957	3.0310e+14***
	(0.155)	(0.159)	(3.865e+13)	(0.184)	(0.221)	(1.009e+14)
Tenure at start of baseline period		0.5338***	1.0000		0.5184***	1.0000
		(0.046)	(0.215)		(0.061)	(0.356)
Tenure × treated WH		0.7045	1.4617		0.6598	1.5118
		(0.317)	(.)		(0.564)	(.)
Pre-roll-out speed		0.7334***	1.0000		0.7894*	1.0000
		(0.071)	(0.066)		(0.109)	(0.104)
Pre-roll-out speed × treated WH		1.2536**	1.0253		1.2229	1.0173
		(0.137)	(.)		(0.195)	(.)
1 if female			1.0000			1.0000
			(0.245)			(0.414)
1 if female × treated WH			0.6289			0.6268
			(.)			(.)
Age at start of baseline period			1.0000			1.0000
			(0.093)			(0.142)
Age × treated WH			1.0750			1.0362
			(.)			(.)
# Workers	2927	2398	849	2600	2151	778

Notes: Hazard ratios from Cox proportional hazard models for workers who were employed by the firm on 8 December 2014. Robust standard errors in parentheses. Columns 1–3 analyze the time during and after the roll-out of incentives. Columns 4–6 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 rates-area fixed effects, i.e., correcting for the fact that a unit is harder or easier in different rates areas. 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. Significance at the 1, 5, and 10 percent level is denoted by ***, **, and *, respectively.

Figure C.5: Attrition during the introduction of static incentives (workers employed in March 2015)



Notes: Kaplan-Meier survival estimates for the introduction of static incentives for workers who were employed on 2 March 2015. 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 C.5.

Table C.5: Attrition during the introduction of static incentives (workers employed in March 2015)

Dependent variable: Worker left firm	
	Mar 2015 - Jun 2015
	(1)
1 if treated warehouse	1.3180***
	(0.135)
# Workers	3019

Notes: Hazard ratios from Cox proportional hazard models for workers who were employed by the firm on 2 March 2015. Robust standard errors in parentheses. The table analyzes the time after the roll-out of incentives. Significance at the 1, 5, and 10 percent level is denoted by ***, **, and *, respectively.

D Analysis of the GROUP trial

This appendix presents more details about the GROUP trial, first the design, then the theoretical predictions, and finally the empirical analysis.

D.1 Design

Table D.1 summarizes the design of GROUP. We randomized all workers into two conditions, treated workers (denoted *rate setters*, 40 percent of workers) and control workers (denoted *non-rate setters*, 60 percent of workers), and workers kept the same roles throughout the trial.

Table D.1: Design of the GROUP trial

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

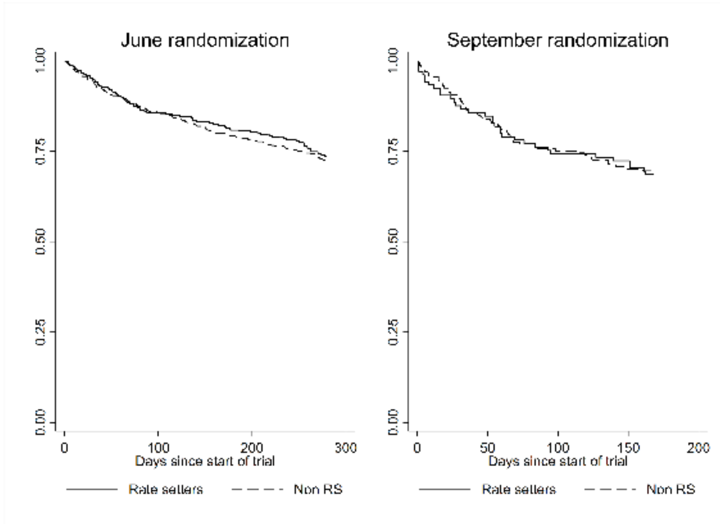
Workers were extensively informed about all the details outlined below, except for the fact that the trial was designed together with university researchers. In the baseline period, before the trial, all workers faced incentive pay with exogenous target rates. During and after the trial, rates were changed 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 rate setters group knew that their performance in that month would determine the target rates for all workers (rate setters and non-rate setters) for the second month. Specifically, in Month 2, the rate for each activity area would be the average output per hour from Month 1 in that area, with the average calculated across the group of all rate setters who worked at some point in that area. Non-rate setters knew that rates were determined by the rate setters, and that their own performances would have no impact on anyone's rates. Thus, rate setters face dynamic incentives in Month 1 whereas non-rate setters did not. In Month 2, both groups face the same rates (determined by rate-setter performance in Month 1). Rate setters again face dynamic incentives, because their performance determines rates in Month 3, while non-rate setters do not influence rates. This continues for 10 months. At that point all workers become rate setters.

In June of 2015, we randomized all workers into rate setters and non-rate setters. 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 rate setters and non-rate setters. The trial period for the second cohort was thus shorter, lasting from Month 4 to Month 11.[44] The random allocation of workers to treatments was done by us, stratifying the randomization on above median pre-trial speed, temp/agency workers, workers working mostly on the night shift, and workers working mostly in "Zonepick" (the largest part of the warehouse). Table D.2 contains summary statistics and randomization checks for the GROUP trial. Treatment and control group are not significantly different, including on characteristics we did not stratify on. Figure D.1 and Table D.3 show that there is no differential attrition between rate setters and non-rate setters.

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

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

Figure D.1: Attrition in the GROUP trial



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

Table D.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)
Batch 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' is 1 for rate setters. A worker's pre-trial speed is their average units per hour in the period before the start of the trial, controlling for rates-area fixed effects, i.e., correcting for the fact that a unit is harder or easier in different rates areas. 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. Significance at the 1, 5, and 10 percent level is denoted by ***, **, and *, respectively.

D.2 Theoretical predictions

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'll also use T to refer to the set of treatment workers). Each individual i has a type θ_i, drawn i.i.d. from H. In order to simplify exposition we suppose that types are publicly known (so that there is no learning about others' types over time). Individual i decides every period t how much effort $e_{i,t}$ to exert and faces a convex cost $c(e_{i,t}, \theta_i)$. They receive a base wage o, plus a bonus based on effort $e_{i,t}$.

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 coefficient a, which has weight ω. This captures 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: $\eta_{i,t+1} = \frac{\sum_{j \in T} e_{i,t}}{|T|}$ (in the first period the normalization rate is exogenous) .[45]

Utility is then

$$U_i = \sum_t \delta^{t-1}[o + \gamma(\frac{e_{i,t}}{\eta_t}) - c(e_{i,t}, \theta_i) + a_f e_{i,t} + \omega a(\sum_{j \neq i} \gamma(\frac{e_{j,t}}{\eta_t}))]$$

Our primary results is that we obtain the ratchet effect result in this setting:

Proposition 2 *In GROUP, fixing θ_i, Treatment puts in a lower effort in all periods than Control.*

Because both treatment and control workers face the same static incentives every period, the proposition is true regardless of whether or not the workers in treatment can coordinate. 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. If the individuals in Treatment can coordinate, then the equilibrium path will feature cycling: effort by Treatment should drop to to a very low level (potentially 0) in the first period of the cycle. In the following period, Treatment will put in the minimal amount of effort to acquire the maximal bonus, and repeating this until it is no longer optimal, at which point effort drops down to the starting point of the cycle again.

Proof of Proposition 2: First we take the the difference in utility between effort vector e and e' for Treatment : $\sum_t \delta^{t-1}[\gamma(\frac{e_{i,t}}{\eta_t}) - \gamma(\frac{e^{'}_{i,t}}{\eta_t}) - c(e_{i,t}, \theta_i) + c(e^{'}_{i,t}, \theta_i) + a_f e_{i,t} - a_f e^{'}_{i,t} + \omega a(\sum_{j \neq i} \gamma(\frac{e_{j,t}}{\eta_t})) - \omega a(\sum_{j \neq i} \gamma(\frac{e^{'}_{j,t}}{\eta_t}))]$. We do the same for Control. Of course for Control, we also need to denote the path of effort that Treatment follows, as this determines γ, and so we fix the treatment path as e, and compare the utility at e and e'. The difference is: $\sum_t \delta^{t-1}[\gamma(\frac{e_{i,t}}{\eta_t}) -$

[45] 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 $\frac{\sum_{j \in T} e_{i,t}}{|T|} \neq 0$. If $\frac{\sum_{j \in T} e_{i,t}}{|T|} = 0$ we then suppose $\eta_{i,t+1} = \underline{\eta}$ for some small positive $\underline{\eta}$. This allows for the existence of an optimal policy.

$\gamma(\frac{e^{'}_{i,t}}{\eta_t}) - c(e_{i,t}, \theta_i) + c(e^{'}_{i,t}, \theta_i) + a_f e_{i,t} - a_f e^{'}_{i,t} + \omega a(\sum_{j \neq i} \gamma(\frac{e_{j,t}}{\eta_t})) - \omega a(\sum_{j \neq i} \gamma(\frac{e^{'}_{j,t}}{\eta_t}))]$. Taking the difference between those two differences delivers: $\sum_t \delta^{t-1}[-\gamma(\frac{e^{'}_{i,t}}{\eta_t}^{'}) + \gamma(\frac{e^{'}_{i,t}}{\eta_t}) - \omega a(\sum_{j \neq i} \gamma(\frac{e^{'}_{j,t}}{\eta_t}^{'})) + \omega a(\sum_{j \neq i} \gamma(\frac{e^{'}_{j,t}}{\eta_t}))]$. Thus, as in Proposition 1, standard monotone comparative statics delivers our results. □

It is not as straightforward to have a calibrated rational benchmark model for GROUP as for INDIVIDUAL. 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.

D.3 Results

Table D.4 mirrors Table 2 for the INDIVIDUAL trial. It shows results from OLS regressions, again using ln(units per hour), our measure of workers' performance, as dependent variable. Column 1 shows results from the contemporaneous comparison of treatment and control group, i.e., comparing the performance of rate setters to non-rate setters during the trial. The fixed effects on cohort, rates areas, shift and cohort interacted with all other fixed effects are like in Table 2. We thus flexibly control for differences between the two randomization cohorts. Treated workers are on average slower by -1.0 percent (95% 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.

Table D.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
Rates area 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 trial, when the treatment workers 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 rate-setting periods 3 and later to allow for some learning. A period lasts 4 weeks, and the trial lasted for 10 periods. Specification 3 further restricts 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. 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. Significance at the 1, 5, and 10 percent level is denoted by ***, **, and *, respectively.

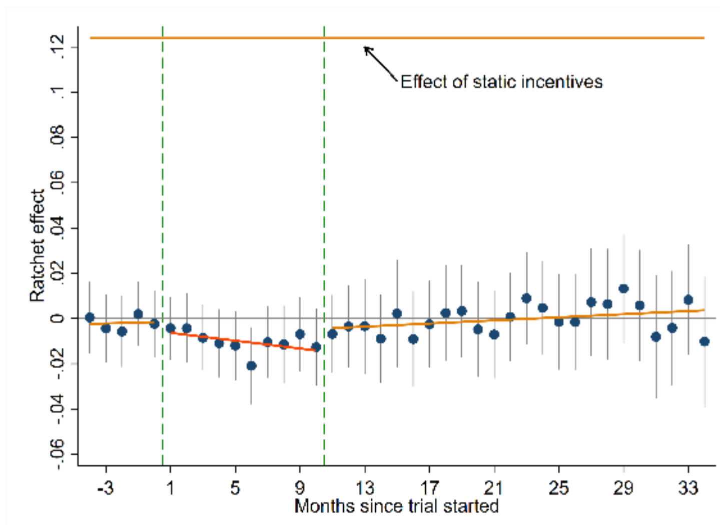
One possible explanation is that GROUP gives workers more time to learn and notice the dynamic incentives. To explore potential learning over time, Figure D.2 shows the difference between rate setters and non-rate setters for each rate-setting period separately.[46] Before the start of the trial, the performance of the two groups is extremely similar. At the start of the trial, the ratchet effect is very close to zero, like in the 3-week long INDIVIDUAL trial. Subsequently, the ratchet effect grows over time, even though this trend is not significant (p=0.475). After 10 periods of rate setting, the ratchet effect is still smaller than -2 percent and is dwarfed by the impact of static incentives, which is shown in the graph for comparison. If there were indeed learning over time, then the point estimate in column 1 of Table D.4 would underestimate the long-term ratchet effect. In column 2, we thus drop the first two

[46]For ease of exposition, this graph only contains data from the first randomization cohort. The patterns for the second, smaller, cohort look extremely similar.

months of the trial. The point estimate grows slightly to -1.2 percent (CI: [-2.5, -0.0]) but is still small.[47]

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.

Figure D.2: Ratchet effect in the GROUP trial over time



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 D.4, column 1, separately for each month. Error bars show 95% confidence intervals.

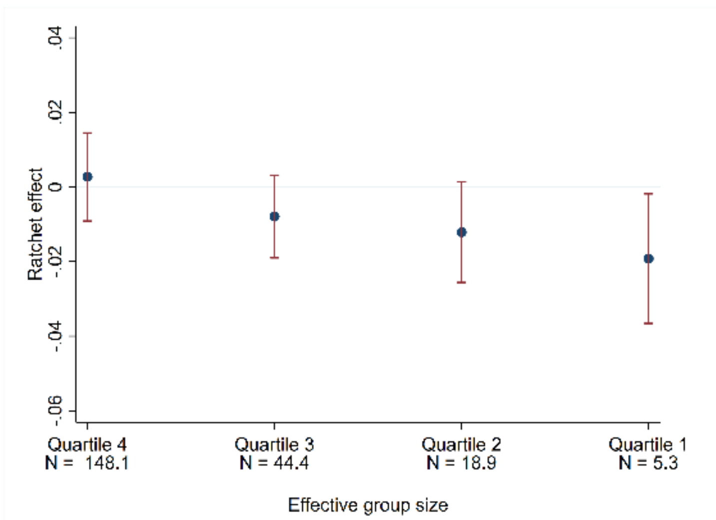
For most situations in the GROUP trial, a ratchet effect can only emerge if workers collude. The sociological literature on ratchet effects (e.g., Mathewson 1931, Roy 1952) very much focuses on this collusion and documents many cases in which workers are able to collude and to hold back effort, often by threatening to punish fast-working "rate busters". It could thus be that the ratchet effect in the GROUP trial is only small because the group sizes are

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

too large or because collusion is difficult to establish. Since we know how many workers work in a particular rates area, we know how many workers affect the corresponding target rate. We can thus study the effect of the group size on the ratchet effect. To explore the effect of group size, we calculate each rate setter's share of the time worked by all rate setters in a given rate area, for each rate-setting period. 1 divided by this share is the effective group size of rate setters. 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 D.3 plots the ratchet effect, i.e., the difference between rate setters and non-rate setters during the trial, for the four quartiles. We find that smaller groups show a larger ratchet effect. This is in line with the hypothesis that collusion is easier to achieve in smaller groups. 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.

Figure D.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 rate setter's share of the time worked by all rate setters in a given rate area, for each rate-setting period. 1 divided by this share is the effective group size of rate setters. Point estimates are from regressions as in Table D.4, column 1, separately for each group size quartile. Error bars show 95% 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 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 workers 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 D.4). Table D.5 shows this comparison. The coefficient of interest is on the interaction of not working the entire next month $\times$ 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 D.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	-0.0141	-0.0063
	(0.012)	(0.013)
Sample	During trial	During trial, periods 3+
Rates area 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 D.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. During this period, workers do not enjoy the full benefit of reducing effort in the current period and thus face weaker ratchet incentives. The direct effect of this variable drops out as it is collinear with the treatment dummy and the fixed effects. The sample is restricted to the time during the trial, when the treatment workers 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 rate-setting periods 3 and later to allow for some learning. A period lasts 4 weeks, and the trial lasted for 10 periods. 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 dummy for not working the entire next rate-setting period. Significance at the 1, 5, and 10 percent level is denoted by ***, **, and *, respectively.

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 D.6 compares the ratchet effect across these three groups. We find no significant differences between the groups.

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

Dependent variable: ln(units per hour)	(1)	(2)	(3)
1 if treated	-0.0078	-0.0108	-0.0101
	(0.009)	(0.009)	(0.009)
1 if temp/agency worker × treated	-0.0112	-0.0089	-0.0314
	(0.019)	(0.027)	(0.035)
1 if permanent worker × treated	-0.0034	-0.0031	-0.0046
	(0.012)	(0.012)	(0.013)
Sample	During trial	During trial, periods 3+	During trial, periods 3+ Working entire next period
Rates area 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 D.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. The sample is restricted to the time during the trial, when the treatment workers 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 rate-setting periods 3 and later to allow for some learning. A period lasts 4 weeks, and the trial lasted for 10 periods. Specification 3 further restricts 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. 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 dummies for being temp/agency worker and the dummy for being permanent worker. Significance at the 1, 5, and 10 percent level is denoted by ***, **, and *, respectively.

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 (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 D.7).

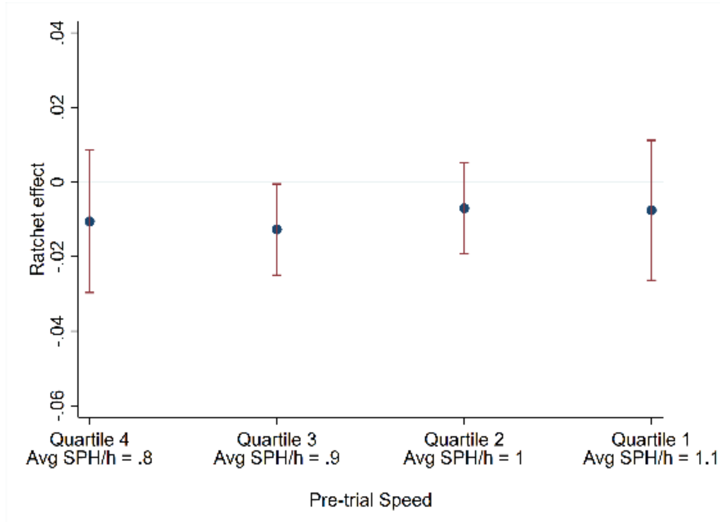
Table D.7: Ratchet effect in GROUP trial by time preferences

Dependent variable: ln(units per hour)	(1)	(2)	(3)
1 if treated	0.0070	0.0053	0.0053
	(0.019)	(0.020)	(0.020)
1 if patient × treated	-0.0039	-0.0100	-0.0100
	(0.026)	(0.027)	(0.027)
Sample	Online exp. During trial	Online exp. During trial, periods 3+	Online exp. During trial, periods 3+ Working entire next period
Rates area 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 D.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 online experiment. As in Table D.4, the sample is also restricted to the time during the trial, when the treatment workers 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 rate-setting periods 3 and later to allow for some learning. A period lasts 4 weeks, and the trial lasted for 10 periods. Specification 3 further restricts 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 dummy for showing a low discount rate. Significance at the 1, 5, and 10 percent level is denoted by ***, **, and *, respectively.

To further explore the heterogeneity of the ratchet effect between workers, Figure D.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. Finally, the ratchet effect is slightly stronger for men than for women, but not significantly so (p=0.450, in a regression akin to Table D.5, column 1).

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

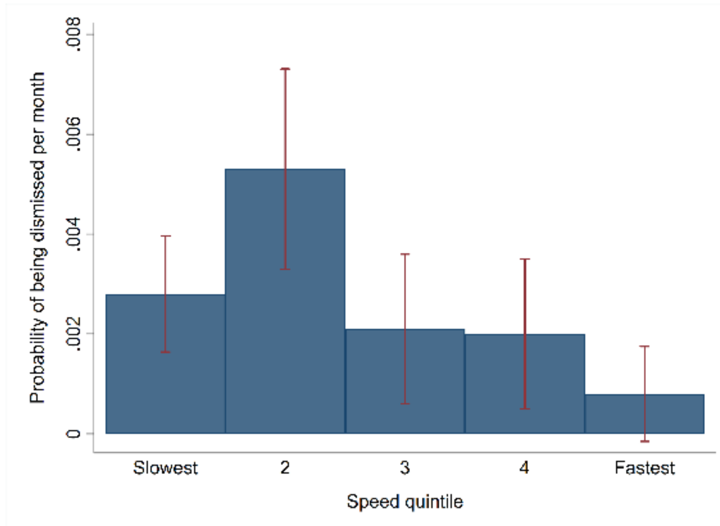


Notes: The graph plots the treatment difference on ln(units per hour), i.e., the ratchet effect, by workers' pre-trial speed. We calculate each worker's speed in the period between the roll-out of static incentives and the start of the trial and split workers into quartiles. Point estimates are from regressions as in Table D.4, column 1, separately for each pre-trial speed quartile. Error bars show 95% 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 rates areas.

E Dismissals and promotions

Maybe workers do not reduce effort when facing ratchet incentives because of the fear of being fired or because of the hope to be promoted? This turns out to not be an important concern in our setting, for several reasons. First, 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, the firm does not dismiss anybody because of speed, 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. If this happens during the probation period, the length of the probation can also be extended. 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 almost all about low quality of the produced output or about attendance. 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 speed related, despite the stated HR policies of the firm. However, we know the speed of the dismissed workers and can correlate speed and dismissal probability. Figure E.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 3.2.3 actually estimates 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_f. We show that even with levels of a_f 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.

Figure E.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 rates-area fixed effects, i.e., correcting for the fact that a unit is easier or harder in different rates areas. Error bars show 95% confidence intervals.

F Theoretical appendix for online experiments with warehouse workers

We want to consider under what conditions COMPLEX induces a (weakly) lower effort level than SIMPLE in Period 2 of the online experiments (or equivalently, Period 4). For the purposes of the model we focus on a two period model so that the the experimental Periods 2 and 4 are Period 1 in the model (and 3 and 5 correspond to Period 2 in the model).

Note that if the individual provides 0 effort in COMPLEX in Period 1 then COMPLEX must have less effect than in SIMPLE. The next proposition provides sufficient conditions for optimal effort to be zero in Period 1. Intuitively, if the individual expects to provide sufficient effort in Period 2, specifically, enough so that they have a strictly positive marginal bonus in Period 2, then so long as the piece rate (denoted w) is large enough, then optimal effort is zero in Period 1. The following proposition summarizes. The proof also also provides more detailed conditions solely in terms of primitives.

Proposition 3 *So long as the individual is earning positive marginal bonus in Period 2, if w is large enough then workers in COMPLEX provide effort of 0 in Period 1.*

Proof of Proposition 3:

Note that for SIMPLE the utility is $U_i = o - c(e_{i,1}, \theta_i) + a_f e_{i,1} + o + we_{i,2} - c(e_{i,2}, \theta) + a_f e_{i,2}$ where w is the marginal wage. Observe the FOCs condition for SIMPLE are $-c'(e_{i,1}, \theta_i) + a_f = 0$ and $g - c'(e_{i,2}, \theta) + a_f = 0$. Observe that the workers may put in effort in Period 1.

We have described complex before. For the purposes of the proof we will focus on situations where the solution is characterized by the first order conditions. Thus, the FOC condition for the COMPLEX scheme is is$\frac{1}{\eta_{i,1}}\gamma'(\frac{e_{i,1}}{\eta_{i,1}}) - c'(e_{i,1}, \theta_i) + a_f - \zeta(\frac{e_{i,2}}{(\zeta e_{i,1} + (1-\zeta)\epsilon_1)^2}\gamma'(\frac{e_{i,2}}{\zeta e_{i,1} + (1-\zeta)\epsilon_1})) = 0$.

First, we provide a proof of In order that COMPLEX subjects put forth 0 effort in Period 1, we need it to be the case that at $e_{i,1} = 0$ the marginal cost of effort exceeds the marginal benefit, or in other words $a_f - \zeta(\frac{e_{i,2}}{((1-\zeta)\epsilon_1)^2}\gamma'(\frac{e_{i,2}}{(1-\zeta)\epsilon_1})) < 0$. This is always true so long as $(\frac{e_{i,2}}{((1-\zeta)\epsilon_1)^2}\gamma'(\frac{e_{i,2}}{(1-\zeta)\epsilon_1}))$ is large enough. Two jointly sufficient conditions need to be true. First of all, $\gamma' > 0$; which means that $\underline{E} \leq \frac{e_{i,2}}{(1-\zeta)\epsilon_1} \leq \bar{E}$ and which implies that $\gamma' = w$; and second, $\frac{e_{i,2}}{(1-\zeta)\epsilon_1}w$ is large enough. Observe that so long as in Period 2 $\gamma' > 0$, then a sufficient condition to guarantee our second requirement (that $\frac{e_{i,2}}{(1-\zeta)\epsilon_1}w$) is that w is large enough. $\square$

In contrast to the previous proposition, we observe that individuals exhibit more of a ratchet effect in SIMPLE. We conjecture that this is because individuals better understand the marginal effect of today's effort on tomorrow's bonus in simple. We formalize this by supposing individuals misunderstand the incentive scheme. In particular, they underestimate the marginal effect of today's work on tomorrow's bonus. If the actual bonus tomorrow is a function of today and tomorrow's effort; i.e. $\beta(e_{i,2}, e_{i,1})$; such that the marginal effect of today's effort on tomorrow's wages is $\frac{\partial\beta(e_{i,2},e_{i,1})}{\partial e_{i,1}}$, individuals perceive the marginal effect to be $\varsigma_i \frac{\partial\beta(e_{i,2},e_{i,1})}{\partial e_{i,1}}$. We conjecture that ς_i is between 0 to 1, and is larger for individuals with higher CRT scores, and in situations where the contract scheme is less complex. In particular, observe that in both simple and complex individuals with a low ς_i will provide more effort in Period 1 than those with high ς_i. The FOC condition for the COMPLEX in Period 2 is $\frac{1}{\zeta e_{i,1}+(1-\zeta)\epsilon_1}\gamma'(\frac{e_{i,2}}{\zeta e_{i,1}+(1-\zeta)\epsilon_1}) - c'(e_{i,2},\theta) + a_f = 0$; the FOC in Period 1 is $\frac{1}{\eta_{i,1}}\gamma'(\frac{e_{i,1}}{\eta_{i,1}}) - c'(e_{i,1},\theta_i) + a_f - \zeta\varsigma_i(\frac{e_{i,2}}{(\varsigma_i\zeta e_{i,1}+(1-\zeta)\epsilon_1)^2}\gamma'(\frac{e_{i,2}}{\zeta\varsigma_i e_{i,1}+(1-\zeta)\epsilon_1})) = 0$. Note that for ς_i small enough the FOC conditions reduce to $\frac{1}{\zeta e_{i,1}+(1-\zeta)\epsilon_1}\gamma'(\frac{e_{i,2}}{\zeta e_{i,1}+(1-\zeta)\epsilon_1}) - c'(e_{i,2},\theta) + a_f = 0$; the FOC in Period 1 is $\frac{1}{\eta_{i,1}}\gamma'(\frac{e_{i,1}}{\eta_{i,1}}) - c'(e_{i,1},\theta_i) + a_f = 0$. Thus, the FOC condition in Period 1 looks exactly like that where there is no effect on Period 2 of Period 1's effort. Thus individuals provide more effort in Period 1 if ς_i is 0 rather than 1. For SIMPLE the FOC conditions is For SIMPLE the utility are $(1-\varsigma_i)g - c'(e_{i,1},\theta_i) + a_f = 0$ and $g - c'(e_{i,2},\theta) + a_f = 0$ Thus again, individuals provide more effort in Period 1 if ς_i is 0 rather than 1. Building on the previous proposition, we can then show that under relatively general conditions, individuals will provide more effort in COMPLEX with a sufficiently low ς_i than in SIMPLE with a sufficiently high ς_i.

G Additional results for online experiments with warehouse workers

Figure G.1: Screenshot of real-effort task in online experiments

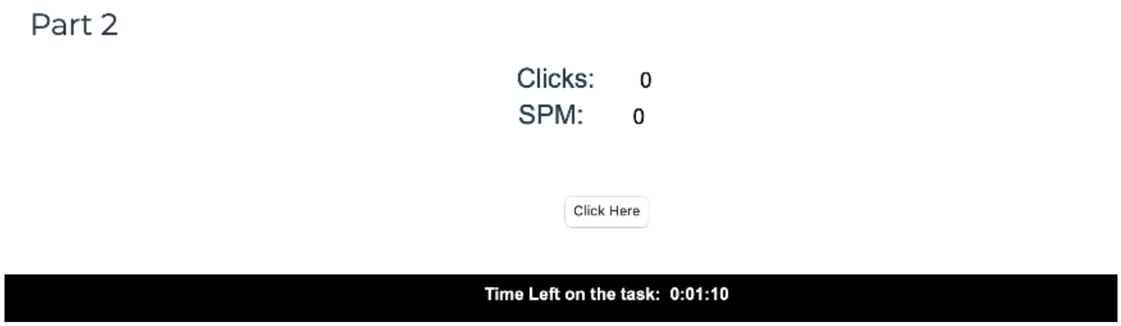


Table G.1: Diff-in-Diff of clicks relative to baseline period and STATIC

	Warehouse workers			AMT workers		
	(1)	(2)	(3)	(4)	(5)	(6)
Period1*Complex	-29.77	-29.77	-29.77	-56.76***	-56.76***	-56.76***
	(19.13)	(19.14)	(19.16)	(14.13)	(14.14)	(14.16)
Period2*Complex	1.77	1.77	1.77	3.43	3.43	3.89
	(18.52)	(18.54)	(18.55)	(10.59)	(10.60)	(10.53)
Period3*Complex	-17.64	-17.64	-17.64	-115.11***	-115.11***	-114.71***
	(19.00)	(19.02)	(19.03)	(21.00)	(21.02)	(20.99)
Period4*Complex	-16.46	-16.46	-16.46	-17.00	-17.00	-16.60
	(16.33)	(16.34)	(16.36)	(14.54)	(14.56)	(14.52)
Period1*Simple	-88.47***	-88.47***	-88.47***	-267.00***	-267.00***	-266.95***
	(22.18)	(22.19)	(22.22)	(21.63)	(21.64)	(21.68)
Period2*Simple	-7.01	-7.01	-7.01	-25.02*	-24.96*	-24.34*
	(18.11)	(18.12)	(18.14)	(13.25)	(13.26)	(13.19)
Period3*Simple	-150.27***	-150.27***	-150.27***	-358.51***	-358.52***	-358.16***
	(24.17)	(24.19)	(24.21)	(21.26)	(21.27)	(21.27)
Period4*Simple	-29.84*	-29.84*	-29.84*	-29.35*	-29.30*	-28.61*
	(16.49)	(16.50)	(16.52)	(16.02)	(16.03)	(16.00)
Period1*Static_Zero	12.51	12.51	12.51	4.46	4.46	4.46
	(19.16)	(19.17)	(19.19)	(10.19)	(10.19)	(10.21)
Period2*Static_Zero	16.38	16.38	16.38	10.39	10.40	10.85
	(19.41)	(19.43)	(19.45)	(10.79)	(10.81)	(10.73)
Period3*Static_Zero	-80.57***	-80.57***	-80.57***	-269.12***	-269.20***	-268.67***
	(28.28)	(28.30)	(28.32)	(29.00)	(29.03)	(29.02)
Period4*Static_Zero	-110.73***	-110.73***	-110.73***	-275.74***	-275.74***	-275.33***
	(28.09)	(28.11)	(28.14)	(27.45)	(27.47)	(27.48)
Constant	435.21***	454.63***	459.69***	476.74***	476.16***	484.79***
	(14.14)	(16.06)	(32.48)	(9.06)	(9.10)	(16.54)
Additional coefficients suppressed	Yes	Yes	Yes	Yes	Yes	Yes
Controls for device	No	Yes	Yes	No	Yes	Yes
Controls for cog. ability	No	No	Yes	No	No	Yes
Observations	2150	2150	2150	2221	2221	2221
Adjusted R^2	0.072	0.078	0.083	0.366	0.375	0.378

Notes: OLS regressions. Fully interacted difference-in-differences model with STATIC as the benchmark treatment. Besides the constant term, 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. Controls for device type include indicators for tablet, smartphone, and other with computer omitted. Controls for cognitive ability include CRT score, years of schooling for warehouse workers, educational attainment for AMT workers, indicator for narrow bracketer, and indicator for ability to do backwards induction. Robust standard errors in parentheses, clustering on worker. Significance at the 1, 5, and 10 percent level is denoted by ***, **, and *, respectively.

Table G.2: Categorization of open-ended responses about optimal work strategies, warehouse workers

	Complex	Simple
	(percent of responses)	
Response focused on dynamic incentives	19.15%	43.66%
Response focused on working fast or constantly	34.75%	11.97%
Response said no idea	3.55%	5.63%
Response mentioned reverse dynamic incentives	2.13%	3.52%
Response missing or nonsense	7.09%	4.23%
Response hard to interpret	33.33%	30.99%
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 based on two out of three independent evaluators categorizing 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.

Table G.3: Diff-in-Diff of clicks relative to baseline period and STATIC, interacted with CRT

	Warehouse workers		AMT workers	
	(1)	(2)	(3)	(4)
Period1*Complex*CRT	-24.28	-24.28	-21.42**	-21.42**
	(20.32)	(20.33)	(10.69)	(10.69)
Period2*Complex*CRT	-30.76	-30.76	-7.94	-7.93
	(19.20)	(19.22)	(10.11)	(10.11)
Period3*Complex*CRT	-50.52**	-50.52**	-34.77**	-34.77**
	(20.31)	(20.33)	(17.17)	(17.18)
Period4*Complex*CRT	16.35	16.35	10.71	10.71
	(15.34)	(15.35)	(13.80)	(13.81)
Period1*Simple*CRT	-54.35***	-54.35***	-72.39***	-72.41***
	(20.14)	(20.15)	(18.30)	(18.31)
Period2*Simple*CRT	-1.98	-1.98	3.66	3.69
	(14.03)	(14.04)	(13.82)	(13.83)
Period3*Simple*CRT	-95.04***	-95.04***	-48.00**	-48.00**
	(20.80)	(20.82)	(19.72)	(19.73)
Period4*Simple*CRT	2.53	2.53	-1.04	-1.02
	(13.04)	(13.05)	(14.87)	(14.88)
Period1*Static_Zero*CRT	0.66	0.66	10.49	10.49
	(15.76)	(15.77)	(7.85)	(7.86)
Period2*Static_Zero*CRT	4.39	4.39	-2.02	-2.01
	(22.50)	(22.52)	(9.75)	(9.75)
Period3*Static_Zero*CRT	-80.71**	-80.71**	-2.86	-2.92
	(33.30)	(33.32)	(25.46)	(25.48)
Period4*Static_Zero*CRT	-23.36	-23.36	0.88	0.89
	(34.28)	(34.30)	(23.72)	(23.73)
Constant	418.92***	440.31***	478.04***	478.29***
	(18.34)	(20.45)	(12.86)	(12.87)
Additional coefficients suppressed	Yes	Yes	Yes	Yes
Controls for device	No	Yes	No	Yes
Observations	2150	2150	2221	2221
Adjusted R^2	0.093	0.100	0.382	0.391

Notes: OLS regressions. Fully interacted difference-in-differences model with STATIC as the benchmark treatment. Besides the constant term, only the coefficients for the triple interaction of period with treatment and CRT score are shown. Negative coefficients mean that individuals with higher CRT scores in a given treatment and period have a larger drop in clicks relative to baseline and STATIC than individuals with lower CRT scores. Controls for device type include indicators for tablet, smartphone, and other with computer omitted. Robust standard errors in parentheses, clustering on worker. Significance at the 1, 5, and 10 percent level is denoted by ***, **, and *, respectively.

Table G.4: Diff-in-Diff of clicks relative to baseline period and STATIC, interacted with education

	Warehouse workers		AMT workers	
	(1)	(2)	(3)	(4)
Period1*Complex*Schooling	-5.56	-5.56	-4.27	-4.27
	(5.53)	(5.53)	(9.36)	(9.37)
Period2*Complex*Schooling	-0.43	-0.43	1.23	1.23
	(5.82)	(5.82)	(7.13)	(7.14)
Period3*Complex*Schooling	1.11	1.11	-16.39	-16.39
	(6.33)	(6.33)	(16.65)	(16.66)
Period4*Complex*Schooling	2.56	2.56	10.81	10.81
	(5.25)	(5.26)	(10.59)	(10.59)
Period1*Simple*Schooling	-17.86**	-17.86**	-44.42***	-44.42***
	(7.35)	(7.36)	(16.35)	(16.36)
Period2*Simple*Schooling	-2.98	-2.98	0.45	0.47
	(5.86)	(5.86)	(9.09)	(9.09)
Period3*Simple*Schooling	-14.47*	-14.47*	-25.93	-25.93
	(8.46)	(8.46)	(18.09)	(18.10)
Period4*Simple*Schooling	-1.61	-1.61	18.59	18.61
	(5.34)	(5.34)	(13.13)	(13.14)
Period1*Static_Zero*Schooling	-1.61	-1.61	4.64	4.64
	(4.84)	(4.85)	(7.72)	(7.72)
Period2*Static_Zero*Schooling	0.29	0.29	-5.87	-5.87
	(6.64)	(6.65)	(8.15)	(8.16)
Period3*Static_Zero*Schooling	4.14	4.14	-1.28	-1.33
	(10.96)	(10.97)	(24.79)	(24.80)
Period4*Static_Zero*Schooling	2.46	2.46	19.83	19.83
	(9.37)	(9.37)	(21.83)	(21.85)
Constant	383.55***	423.94***	489.79***	487.18***
	(52.51)	(55.06)	(37.89)	(38.22)
Additional coefficients suppressed	Yes	Yes	Yes	Yes
Controls for device	No	Yes	No	Yes
Observations	2150	2150	2221	2221
Adjusted R^2	0.080	0.087	0.372	0.380

Notes: OLS regressions. Fully interacted difference-in-differences model with STATIC as the benchmark treatment. Besides the constant term, only the coefficients for the triple interaction of period with treatment and education are shown. Negative coefficients mean that individuals with more education in a given treatment and period have a larger drop in clicks relative to baseline and STATIC than individuals with less education. Education is measured by years of schooling for warehouse workers, and six educational attainment categories for AMT workers (some high school; high school degree; some college; 2 year college degree; 4 year college degree; graduate or professional degree). Controls for device type include indicators for tablet, smartphone, and other with computer omitted. Robust standard errors in parentheses, clustering on worker. Significance at the 1, 5, and 10 percent level is denoted by ***, **, and *, respectively.

Table G.5: Diff-in-Diff of clicks relative to baseline period and STATIC, interacted with backwards induction ability

	Warehouse workers		AMT workers	
	(1)	(2)	(3)	(4)
Period1*Complex*Hit7	-12.59	-12.59	20.61	20.61
	(40.61)	(40.64)	(25.41)	(25.42)
Period2*Complex*Hit7	45.77	45.77	-10.38	-10.38
	(35.54)	(35.56)	(14.80)	(14.80)
Period3*Complex*Hit7	-30.29	-30.29	-17.22	-17.22
	(44.16)	(44.19)	(40.31)	(40.32)
Period4*Complex*Hit7	46.86	46.86	-10.42	-10.42
	(38.47)	(38.50)	(19.10)	(19.10)
Period1*Simple*Hit7	-15.72	-15.72	-95.99**	-95.99**
	(45.64)	(45.67)	(40.27)	(40.28)
Period2*Simple*Hit7	16.75	16.75	30.48	30.48
	(33.29)	(33.32)	(25.43)	(25.44)
Period3*Simple*Hit7	-58.88	-58.88	-17.47	-17.47
	(53.11)	(53.15)	(41.92)	(41.92)
Period4*Simple*Hit7	21.90	21.90	-8.09	-8.09
	(37.06)	(37.09)	(36.30)	(36.30)
Period1*Static_Zero*Hit7	70.30*	70.30*	-0.14	-0.14
	(37.83)	(37.86)	(56.26)	(56.27)
Period2*Static_Zero*Hit7	39.56	39.56	-62.22	-62.22
	(37.40)	(37.43)	(69.62)	(69.63)
Period3*Static_Zero*Hit7	-110.85	-110.85	18.25	18.25
	(69.28)	(69.33)	(60.78)	(60.79)
Period4*Static_Zero*Hit7	-15.13	-15.13	-54.83	-54.83
	(67.35)	(67.40)	(53.16)	(53.17)
Constant	434.43***	452.84***	479.29***	483.28***
	(16.97)	(18.69)	(7.55)	(7.70)
Additional coefficients suppressed	Yes	Yes	Yes	Yes
Controls for device	No	Yes	No	Yes
Observations	2150	2150	8915	8915
Adjusted R^2	0.073	0.079	0.118	0.123

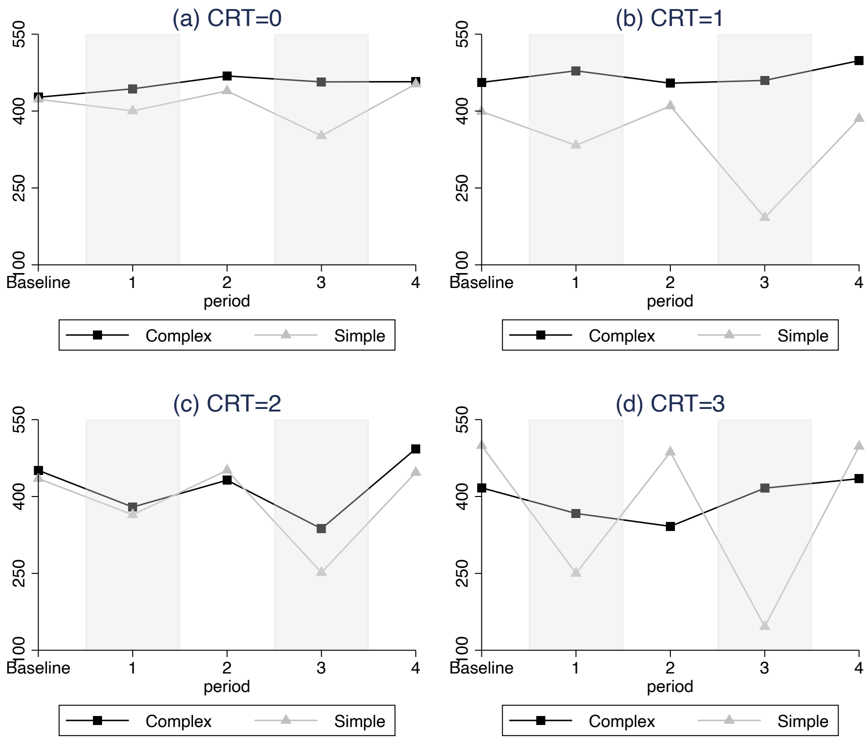
Notes: OLS regressions. Fully interacted difference-in-differences model with STATIC as the benchmark treatment. Besides the constant term, only the coefficients for the triple interaction of period with treatment and a (binary) indicator for backwards induction ability are shown. Negative coefficients mean that individuals with better ability in a given treatment and period have a larger drop in clicks relative to baseline and STATIC than individuals with lower ability. Backwards induction ability is an indicator for having won the Hit 7 game against the computer. Controls for device type include indicators for tablet, smartphone, and other with computer omitted. Robust standard errors in parentheses, clustering on worker. Significance at the 1, 5, and 10 percent level is denoted by ***, **, and *, respectively.

Table G.6: Diff-in-Diff of clicks relative to baseline period and STATIC, interacted with narrow bracketing

	Warehouse workers		AMT workers	
	(1)	(2)	(3)	(4)
Period1*Complex*Narrow	66.59*	66.59*	-39.73*	-39.73*
	(38.62)	(38.64)	(22.92)	(22.92)
Period2*Complex*Narrow	90.53**	90.53**	27.42*	27.42*
	(37.71)	(37.74)	(14.33)	(14.33)
Period3*Complex*Narrow	6.70	6.70	-35.09	-35.09
	(38.33)	(38.36)	(36.64)	(36.65)
Period4*Complex*Narrow	23.75	23.75	1.01	1.01
	(35.20)	(35.22)	(19.97)	(19.97)
Period1*Simple*Narrow	40.26	40.26	36.70	36.70
	(45.64)	(45.68)	(43.55)	(43.56)
Period2*Simple*Narrow	66.75*	66.75*	9.85	9.85
	(37.33)	(37.35)	(22.75)	(22.75)
Period3*Simple*Narrow	-20.07	-20.07	-6.13	-6.13
	(49.31)	(49.35)	(40.63)	(40.63)
Period4*Simple*Narrow	0.04	0.04	42.88	42.88
	(34.39)	(34.41)	(33.33)	(33.34)
Period1*Static_Zero*Narrow	74.67*	74.67*	-60.56	-60.56
	(39.03)	(39.06)	(75.74)	(75.75)
Period2*Static_Zero*Narrow	53.78	53.78	-62.49	-62.49
	(39.49)	(39.52)	(94.60)	(94.61)
Period3*Static_Zero*Narrow	-57.37	-57.37	57.63	57.63
	(58.58)	(58.62)	(63.53)	(63.54)
Period4*Static_Zero*Narrow	-26.79	-26.79	-9.25	-9.25
	(57.97)	(58.01)	(51.85)	(51.86)
Constant	446.14***	466.82***	470.63***	474.32***
	(16.07)	(17.46)	(5.66)	(5.51)
Additional coefficients suppressed	Yes	Yes	Yes	Yes
Controls for device	No	Yes	No	Yes
Observations	2150	2150	8915	8915
Adjusted R^2	0.070	0.077	0.113	0.118

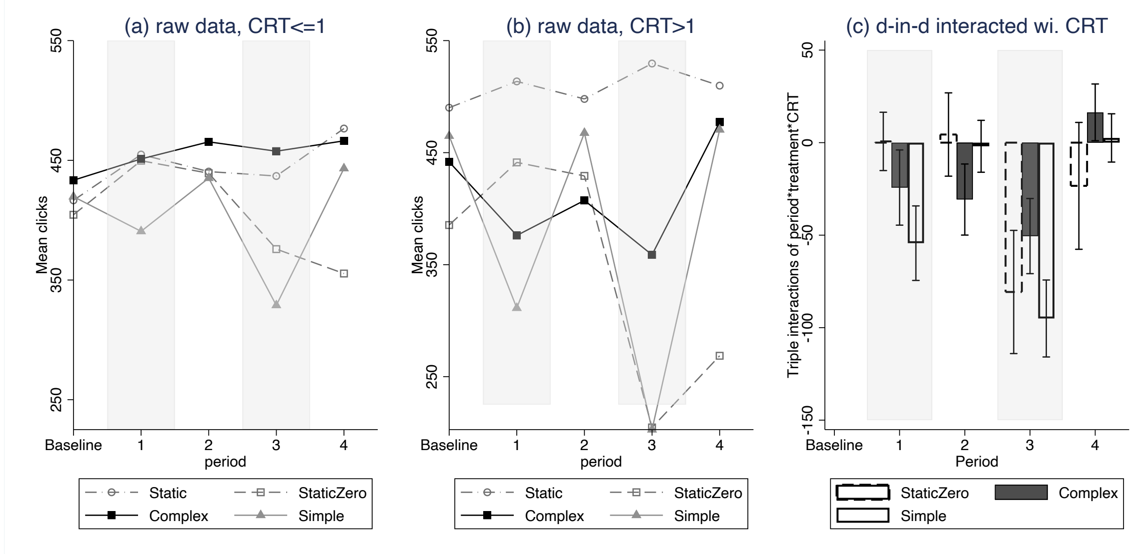
Notes: OLS regressions. Fully interacted difference-in-differences model with STATIC as the benchmark treatment. Besides the constant term, only the coefficients for the triple interaction of period with treatment and a (binary) indicator for narrow bracketing. Negative coefficients mean that individuals with narrow bracketing in a given treatment and period have a larger drop in clicks relative to baseline and STATIC than individuals with broader bracket. Narrow bracketing is an indicator for violating dominance in a set of paired lottery choices. Controls for device type include indicators for tablet, smartphone, and other with computer omitted. Robust standard errors in parentheses, clustering on worker. Significance at the 1, 5, and 10 percent level is denoted by ***, **, and *, respectively.

Figure G.2: 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.

Figure G.3: Shrouding of ratchet incentives and CRT, warehouse workers online, all treatments



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 G.3 in the appendix for all coefficients). The vertical shaded bars in all panels denote periods with dynamic incentives to reduce effort in COMPLEX and SIMPLE.

Table G.7: Diff-in-Diff of clicks relative to baseline period and STATIC, interacted with time discount rate (IRR).

	Warehouse workers		
	(1)	(2)	(3)
Period1*Complex*IRR	-2.47	-2.47	-2.47
	(21.82)	(21.84)	(21.86)
Period2*Complex*IRR	-2.33	-2.33	-2.33
	(17.68)	(17.70)	(17.71)
Period3*Complex*IRR	-13.32	-13.32	-13.32
	(19.12)	(19.14)	(19.15)
Period4*Complex*IRR	-9.33	-9.33	-9.33
	(19.22)	(19.24)	(19.26)
Period1*Simple*IRR	-15.35	-15.35	-15.35
	(20.00)	(20.01)	(20.03)
Period2*Simple*IRR	3.39	3.39	3.39
	(15.39)	(15.40)	(15.41)
Period3*Simple*IRR	4.18	4.18	4.18
	(20.68)	(20.69)	(20.71)
Period4*Simple*IRR	-5.33	-5.33	-5.33
	(15.24)	(15.25)	(15.27)
Period1*Static_Zero*IRR	-4.30	-4.30	-4.30
	(19.83)	(19.84)	(19.86)
Period2*Static_Zero*IRR	-11.12	-11.12	-11.12
	(18.27)	(18.28)	(18.30)
Period3*Static_Zero*IRR	-40.82	-40.82	-40.82
	(25.87)	(25.88)	(25.91)
Period4*Static_Zero*IRR	-26.88	-26.88	-26.88
	(29.10)	(29.12)	(29.15)
Constant	435.31***	454.96***	458.95***
	(14.20)	(16.02)	(32.48)
Additional coefficients suppressed	Yes	Yes	Yes
Controls for device	No	Yes	Yes
Controls for cog. ability	No	No	Yes
Observations	2150	2150	2150
Adjusted R^2	0.071	0.078	0.082

Notes: OLS regressions. Fully interacted difference-in-differences model with STATIC as the benchmark treatment. Besides the constant term, only the coefficients for the triple interaction of period with treatment and standardized value of an worker's time discount rate as captured by internal rate of return (IRR) in the time preference experiments. Higher IRRs correspond to greater impatience. Negative coefficients mean that individuals with a greater IRR in a given treatment and period have a larger drop in clicks relative to baseline and STATIC than individuals with lower IRR. Controls for device type include indicators for tablet, smartphone, and other with computer omitted. Robust standard errors in parentheses, clustering on worker. Significance at the 1, 5, and 10 percent level is denoted by ***, **, and *, respectively.

Table G.8: Diff-in-Diff of clicks relative to baseline period and STATIC, interacted with certainty equivalent (CE).

	Warehouse workers		
	(1)	(2)	(3)
Period1*Complex*CE	-45.46***	-45.46***	-45.46***
	(16.98)	(16.99)	(17.01)
Period2*Complex*CE	-19.94	-19.94	-19.94
	(14.90)	(14.91)	(14.92)
Period3*Complex*CE	14.02	14.02	14.02
	(20.76)	(20.77)	(20.79)
Period4*Complex*CE	-34.09**	-34.09**	-34.09**
	(15.86)	(15.87)	(15.88)
Period1*Simple*CE	-34.99**	-34.99**	-34.99**
	(17.49)	(17.50)	(17.52)
Period2*Simple*CE	0.29	0.29	0.29
	(13.10)	(13.11)	(13.12)
Period3*Simple*CE	-18.92	-18.92	-18.92
	(23.44)	(23.45)	(23.48)
Period4*Simple*CE	-4.63	-4.63	-4.63
	(15.30)	(15.31)	(15.32)
Period1*Static_Zero*CE	-13.15	-13.15	-13.15
	(13.85)	(13.86)	(13.87)
Period2*Static_Zero*CE	21.44	21.44	21.44
	(16.26)	(16.27)	(16.28)
Period3*Static_Zero*CE	58.27**	58.27**	58.27**
	(23.78)	(23.80)	(23.82)
Period4*Static_Zero*CE	24.70	24.70	24.70
	(28.72)	(28.74)	(28.77)
Constant	435.20***	454.56***	459.39***
	(14.05)	(16.04)	(32.90)
Additional coefficients suppressed	Yes	Yes	Yes
Controls for device	No	Yes	Yes
Controls for cog. ability	No	No	Yes
Observations	2150	2150	2150
Adjusted R^2	0.071	0.077	0.082

Notes: OLS regressions. Fully interacted difference-in-differences model with STATIC as the benchmark treatment. Besides the constant term, only the coefficients for the triple interaction of period with treatment and standardized value of an worker's certainty equivalent in the risk preference experiments. Higher CE's correspond to greater willingness to take risks. Negative coefficients mean that individuals with a greater CE in a given treatment and period have a larger drop in clicks relative to baseline and STATIC than individuals with lower CE. Controls for device type include indicators for tablet, smartphone, and other with computer omitted. Robust standard errors in parentheses, clustering on worker. Significance at the 1, 5, and 10 percent level is denoted by ***, **, and *, respectively.

Table G.9: 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 shrouding:			
Distractor	991	9.3	67.1
Distractor_Implicit	985	6.5	78.2
Robustness of shrouding:			
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
Monotonic	781	6.5	79.4
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 given 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. Reading grade level is measured by the Fleisch-Kincaid Grade Level test, and ease of reading is measured by the Fleisch Reading Ease test.

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, as well as the calibration and outcomes of the structural model calibrated to the AMT treatments. 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.[48]

AMT workers are an interesting population to study because they are more similar to the warehouse workers than, say, undergraduate students, in terms of age and experience. At the same time, 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.

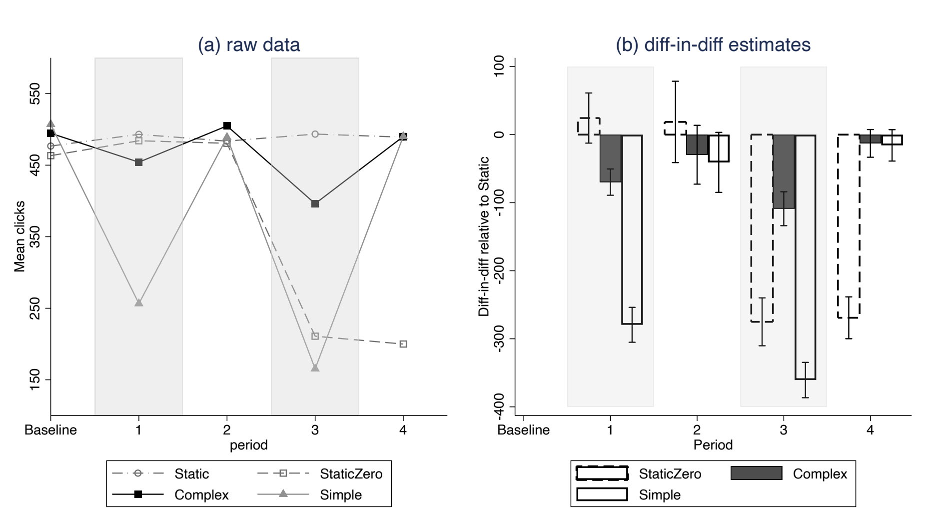
Overall, we find that AMT workers respond to treatments in a very similar way to warehouse workers. AMT workers respond only weakly to dynamic incentives in COMPLEX, and respond strongly in SIMPLE (see Figure H.1 and columns 4–6 in Table G.1).[49] AMT workers are also farther from the rational optimum in COMPLEX than in SIMPLE. We can fully identify our structural model of effort provision for the AMT workers, because STATIC_LOW

[48] 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 vale of \$1.25 used with the warehouse workers. Another difference is that the measures of risk and time preference were not incentivized.

[49] Clicks are significantly lower in Period 1 and Period 3 for AMT workers in COMPLEX compared to AMT workers in STATIC (t-tests; $p < 0.04$; $p < 0.001$). Clicks are significantly lower in periods 1 and 3 comparing AMT workers in SIMPLE to AMT workers in COMPLEX (t-tests; $p < 0.001$; $p < 0.001$).

provides a third data point on how hard AMT workers work under different piece rates . Our calibrated model of effort provision implies that if AMT workers were fully rational they should have reduced clicks all the way to 0 in Periods 1 and 3 in COMPLEX, in order to get easy future rates, and clicked more than 500 in each of Periods 2 and 4. The much weaker reduction in clicks exhibited by AMT workers (they did more than 400 in both Period 1 and 3) implies a utility loss of about \$0.25 relative to the optimum in the model, which is a loss of roughly 50 percent relative to average utility earned in periods without dynamic incentives. AMT workers are much closer to the optimum predicted for SIMPLE, losing only about \$0.05 in utility due to deviation. This reduced distance to the optimum is consistent with the incentives in SIMPLE being easier to understand.

Figure H.1: Replication with AMT workers

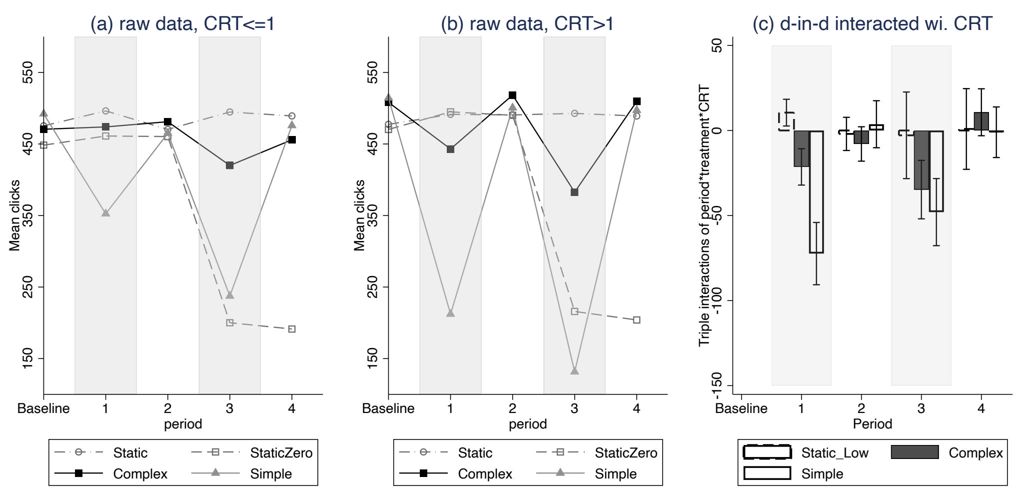


Notes: Panel (a) shows average number of clicks in a given work period. Panel (b) plots coefficients of interaction terms, Period*Treatment, from a difference-in-differences regression relative to baseline period and the treatment STATIC (see Column (1) of Table G.1 in the appendix for all coefficients).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

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 39 percent in COMPLEX versus 79 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 shrouding of dynamic incentives (see Figure H.2 and Table G.3).[50] AMT workers with higher CRT scores exhibit significantly greater responses to dynamic incentives in both Period 1 and Period 3, in both COMPLEX and in SIMPLE. Higher CRT is also associated with a smaller distance from the rational optimum in both COMPLEX and SIMPLE. Higher CRT is also significantly positively correlated with mentioning dynamic incentives, in both COMPLEX and SIMPLE (Spearman correlations; $\rho = 0.22$, $\rho = 0.16$, $p < 0.2$, $p < 0.04$). As was the case for warehouse workers, our other measures of cognitive ability have limited explanatory power for responses to dynamic incentives (see columns 3 and 4 in Tables G.4 to G.6).

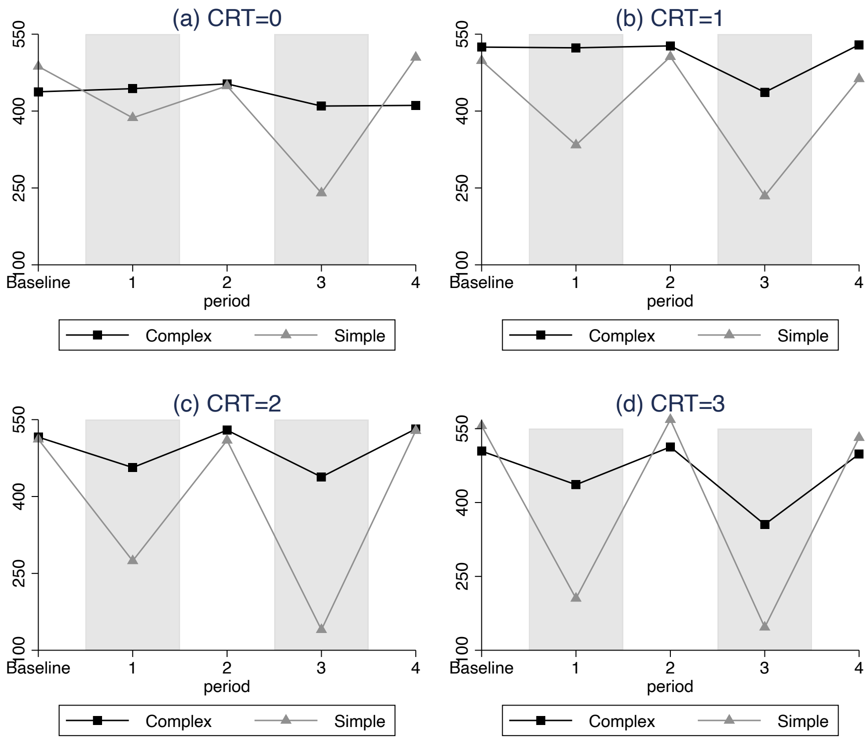
Figure H.2: Shrouding of ratchet incentives and CRT, AMT workers, all treatments



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 G.3 in the appendix for all coefficients). The vertical shaded bars in all panels denote periods with dynamic incentives to reduce effort in COMPLEX and SIMPLE.

[50]Figure H.3 shows results by each value of CRT separately, and as for warehouse workers, shows that unshrouding increases discretely when CRT surpasses 1.

Figure H.3: 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.

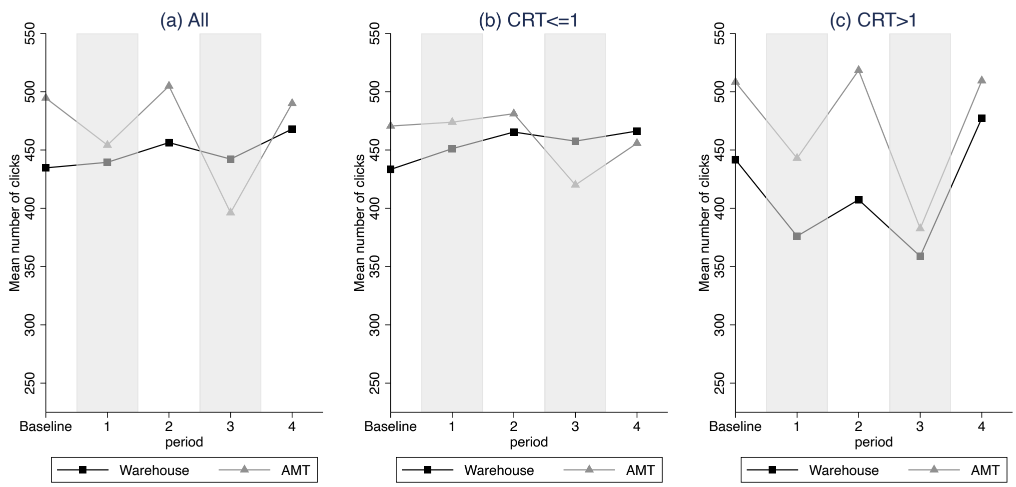
While dynamic incentives are shrouded in COMPLEX for AMT workers, AMT workers do show signs of a greater relative awareness compared to warehouse workers. AMT workers have a modest but statistically significant difference relative to STATIC in periods 1 and 3, unlike warehouse workers (see Figure H.1 and Table G.1). 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 difference is that AMT workers have higher CRT on average, an aspect of cognitive ability that we have shown matters for noticing shrouded attributes. [51] Indeed, behavior of warehouse and AMT workers is more similar if we condition on CRT.

[51]Comparing ability at backwards induction, as measured by the HIT 7 game, about 27 percent of warehouse workers win, versus 35 percent of AMT workers. Interestingly, warehouse workers are less likely to exhibit narrow bracketing than AMT workers, 39 percent versus 60 percent.

Figure H.4 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.1 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 both 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. We also see that the difference in the fractions of warehouse and AMT workers mentioning dynamic incentives is substantially smaller, if we compare within low or high CRT groups of these populations.[52] Differences in other facets of cognitive ability that we do not measure, but which might affect noticing shrouded 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 shrouded attributes.

Figure H.4: Comparing behavior in COMPLEX, warehouse versus AMT workers



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

[52]Without conditioning on CRT, the percentage of AMT workers mentioning dynamic incentives in COMPLEX is about 20 percentage points higher. Comparing within the group of warehouse and AMT workers with CRT ≤ 1, or with the group with CRT > 1, the differences are 8 percentage points and 15 percentage points, respectively.

Table H.1: Diff-in-Diff of clicks relative to baseline period and STATIC, warehouse versus AMT workers

	All workers	CRT<=1	CRT>1
	(1)	(2)	(3)
Period1*Complex*AMT	-26.99	2.90	9.86
	(23.78)	(27.76)	(45.60)
Period2*Complex*AMT	1.66	7.57	39.17
	(21.33)	(31.23)	(37.91)
Period3*Complex*AMT	-97.48***	-73.86*	-18.57
	(28.32)	(39.00)	(52.73)
Period4*Complex*AMT	-0.54	-1.57	-26.25
	(21.86)	(32.33)	(34.12)
Period1*Simple*AMT	-178.53***	-93.23**	-138.89***
	(30.97)	(43.67)	(48.30)
Period2*Simple*AMT	-17.99	-13.56	-21.66
	(22.44)	(36.89)	(22.99)
Period3*Simple*AMT	-208.24***	-162.81***	-96.48*
	(32.19)	(49.80)	(50.50)
Period4*Simple*AMT	0.51	6.32	-14.64
	(22.99)	(34.57)	(29.60)
Period1*Static_Zero*AMT	-8.14	-15.10	-21.84
	(21.80)	(27.43)	(24.23)
Period2*Static_Zero*AMT	-4.68	7.92	-29.46
	(22.26)	(31.22)	(43.63)
Period3*Static_Zero*AMT	-186.66***	-216.67***	-49.74
	(40.67)	(62.38)	(84.53)
Period4*Static_Zero*AMT	-161.61***	-157.89***	-141.56
	(39.35)	(59.37)	(86.15)
Constant	456.24***	432.43***	520.60***
	(15.63)	(18.52)	(23.68)
Additional coefficients suppressed	Yes	Yes	Yes
Controls for device	Yes	Yes	Yes
Observations	4366	2484	1882
Adjusted R^2	0.254	0.126	0.399

Notes: OLS regressions. 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 less than or equal to 1, and greater than 1, respectively. Fully interacted difference-in-differences model with COMPLEX as the benchmark treatment. Besides the constant term, 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. Controls for device type include indicators for tablet, smartphone, and other with computer omitted. Robust standard errors in parentheses, clustering on worker. Significance at the 1, 5, and 10 percent level is denoted by ***, **, and *, respectively.

H.2 Additional results on contract features contributing to shrouding

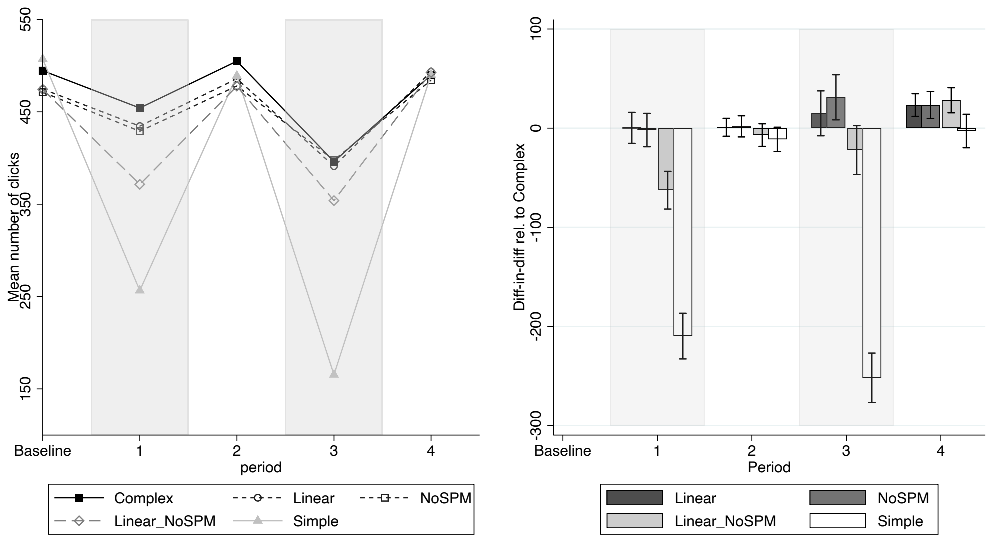
Table H.2: Diff-in-Diff of clicks relative to baseline period and COMPLEX, contract features contributing to shrouding

	AMT workers		
	(1)	(2)	(3)
Period1*Simple	-209.73***	-209.73***	-209.73***
	(23.06)	(23.07)	(23.11)
Period2*Simple	-11.27	-11.27	-11.27
	(12.21)	(12.21)	(12.23)
Period3*Simple	-251.81***	-251.81***	-251.81***
	(24.94)	(24.95)	(24.98)
Period4*Simple	-2.89	-2.89	-2.89
	(16.90)	(16.91)	(16.93)
Period1*Distractor	-121.35***	-121.35***	-121.35***
	(23.56)	(23.57)	(23.60)
Period2*Distractor	22.01*	22.01*	22.01*
	(12.67)	(12.67)	(12.69)
Period3*Distractor	-124.42***	-124.42***	-124.42***
	(27.18)	(27.19)	(27.23)
Period4*Distractor	33.74**	33.74**	33.74**
	(15.56)	(15.56)	(15.58)
Period1*Distractor_Implicit	-21.61	-21.61	-21.61
	(22.24)	(22.25)	(22.28)
Period2*Distractor_Implicit	14.52	14.52	14.52
	(13.95)	(13.95)	(13.97)
Period3*Distractor_Implicit	-50.83**	-50.83**	-50.83**
	(25.39)	(25.40)	(25.43)
Period4*Distractor_Implicit	-16.54	-16.54	-16.54
	(15.32)	(15.32)	(15.35)
Constant	494.60***	495.75***	526.80***
	(8.58)	(8.43)	(19.82)
Additional coefficients suppressed	Yes	Yes	Yes
Controls for device	No	Yes	Yes
Controls for cog. ability	No	No	Yes
Observations	2665	2665	2665
Adjusted R^2	0.208	0.209	0.211

Notes: OLS regressions. Fully interacted difference-in-differences model with COMPLEX as the benchmark treatment. Besides the constant term, 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. Controls for device type include indicators for tablet, smartphone, and other with computer omitted. Controls for cognitive ability include CRT score, educational attainment, indicator for narrow bracketer, and indicator for ability to do backwards induction. Robust standard errors in parentheses, clustering on worker. Significance at the 1, 5, and 10 percent level is denoted by ***, **, and *, respectively.

H.3 Additional results on robustness of shrouding

Figure H.5: Robustness of shrouding, AMT workers



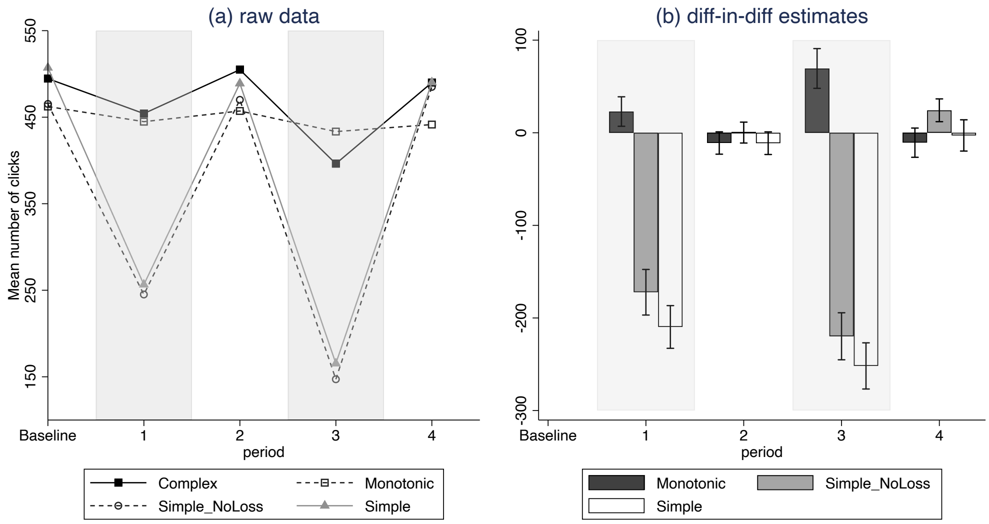
Notes: Panel (a) shows average number of clicks in a given work period. Panel (b) plots coefficients of interaction terms, Period*Treatment, from a difference-in-differences regression relative to baseline period and the treatment COMPLEX (see Column (1) of Table H.3 for all coefficients). The vertical shaded bars in both panels denote periods with dynamic incentives to reduce effort in all treatments.

Table H.3: Diff-in-Diff of clicks relative to baseline period and COMPLEX, robustness of shrouding

	AMT workers		
	(1)	(2)	(3)
Period1*Linear	0.34	0.34	0.34
	(15.64)	(15.65)	(15.67)
Period2*Linear	0.89	0.89	0.89
	(9.04)	(9.05)	(9.06)
Period3*Linear	14.93	14.93	14.93
	(22.66)	(22.67)	(22.69)
Period4*Linear	23.31**	23.31**	23.31**
	(11.43)	(11.43)	(11.44)
Period1*NoSPM	-1.89	-1.89	-1.89
	(16.86)	(16.87)	(16.89)
Period2*NoSPM	1.81	1.81	1.81
	(10.67)	(10.67)	(10.68)
Period3*NoSPM	31.11	31.11	31.11
	(22.73)	(22.74)	(22.77)
Period4*NoSPM	23.45*	23.45*	23.45*
	(13.60)	(13.61)	(13.62)
Period1*Linear_NoSPM	-62.57***	-62.57***	-62.57***
	(19.03)	(19.03)	(19.05)
Period2*Linear_NoSPM	-6.98	-6.98	-6.98
	(11.39)	(11.39)	(11.40)
Period3*Linear_NoSPM	-22.17	-22.17	-22.17
	(24.67)	(24.68)	(24.71)
Period4*Linear_NoSPM	28.14**	28.14**	28.14**
	(12.66)	(12.66)	(12.68)
Constant	494.60***	497.23***	490.38***
	(8.58)	(8.33)	(15.50)
Additional coefficients suppressed	Yes	Yes	Yes
Controls for device	No	Yes	Yes
Controls for cog. ability	No	No	Yes
Observations	3315	3315	3315
Adjusted R^2	0.202	0.211	0.216

Notes: OLS regressions. Fully interacted difference-in-differences model with COMPLEX as the benchmark treatment. Besides the constant term, only the coefficients for the interaction of period with treatment are shown (interactions of SIMPLE with period are also suppressed to save space). Negative coefficients mean that individuals in that treatment and period have a larger drop relative to baseline than individuals in COMPLEX. Controls for device type include indicators for tablet, smartphone, and other with computer omitted. Controls for cognitive ability include CRT score, educational attainment, indicator for narrow bracketer, and indicator for ability to do backwards induction. Robust standard errors in parentheses, clustering on worker. Significance at the 1, 5, and 10 percent level is denoted by ***, **, and *, respectively.

Figure H.6: Additional treatments on contract features contributing to shrouding, 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.3 for all coefficients). The vertical shaded bars in both panels denote periods with dynamic incentives to reduce effort in all treatments.

Table H.4: Diff-in-Diff of clicks relative to baseline period and COMPLEX, additional treatments on contract features contributing to shrouding

	AMT workers		
	score	score	score
Period1*Monotonic	22.92	22.92	22.92
	(15.93)	(15.94)	(15.96)
Period2*Monotonic	-11.05	-11.05	-11.05
	(12.07)	(12.08)	(12.10)
Period3*Monotonic	69.44***	69.44***	69.44***
	(21.46)	(21.48)	(21.50)
Period4*Monotonic	-10.70	-10.70	-10.70
	(15.78)	(15.78)	(15.80)
Period1*Simple_NoLoss	-172.24***	-172.24***	-172.24***
	(24.63)	(24.64)	(24.67)
Period2*Simple_NoLoss	0.20	0.20	0.20
	(11.28)	(11.29)	(11.30)
Period3*Simple_NoLoss	-219.79***	-219.79***	-219.79***
	(25.40)	(25.41)	(25.44)
Period4*Simple_NoLoss	24.26**	24.26**	24.26**
	(12.29)	(12.30)	(12.31)
Period1*Simple	-209.73***	-209.73***	-209.73***
	(23.06)	(23.07)	(23.10)
Period2*Simple	-11.27	-11.27	-11.27
	(12.21)	(12.21)	(12.23)
Period3*Simple	-251.81***	-251.81***	-251.81***
	(24.94)	(24.95)	(24.98)
Period4*Simple	-2.89	-2.89	-2.89
	(16.90)	(16.90)	(16.93)
Constant	494.60***	496.67***	496.44***
	(8.58)	(8.33)	(16.10)
Additional coefficients suppressed	Yes	Yes	Yes
Controls for device	No	Yes	Yes
Controls for cog. ability	No	No	Yes
Observations	2675	2675	2675
Adjusted R^2	0.286	0.293	0.297

Notes: OLS regressions. Fully interacted difference-in-differences model with COMPLEX as the benchmark treatment. Besides the constant term, 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. Controls for device type include indicators for tablet, smartphone, and other with computer omitted. Controls for cognitive ability include CRT score, educational attainment, indicator for narrow bracketer, and indicator for ability to do backwards induction. Robust standard errors in parentheses, clustering on worker. Significance at the 1, 5, and 10 percent level is denoted by ***, **, and *, respectively.

I Instructions for online experiments with warehouse workers

[Insert PDF here]

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

Table J.1: Descriptions of treatments with AMT workers

	Description
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 shrouding:	
Distractor	Same as Simple, except includes distractor (random variable X)
Distractor_Implicit	Same as Simple, except includes distractor (random variable X), and explanation leaves monetary consequences implicit
Robustness of shrouding:	
Linear	Dynamic incentives like Complex, but with linear piece rate schedule
NoSPM	Dynamic incentives like Complex, but whole incentive scheme explained without SPM
Linear_NoSPM	Dynamic incentives like Complex, but with linear piece rate schedule and whole scheme explained without SPM
Additional treatments	
Static_Low	Control treatment with no dynamic incentives, normal piece rate for first 3 periods, lower but non-zero for last two periods
Monotonic	Dynamic incentives like Complex, but parameters changed so that MB of reducing effort is monotonic
Simple_NoLoss	Like Simple except losses are not possible

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

	No dynamic incentives	Dynamic incentives about level of earnings	Explain Dynam. incentives with $ not SPM	No X	Entire scheme explained without SPM	Piece rate is linear	Dynam. incent. involve poss. negative earnings	MB of reducing Period 1 effort is monotonic
Main treatments (replication)								
Static	X							
Static_Zero	X							
Complex								
Simple		X	X	X		X	X	
Contract features contributing to shrouding:								
Distractor		X	X				X	X
Distractor_Implicit		X						X
Robustness of shrouding:								
Linear						X		
NoSPM					X			
Linear_NoSPM					X	X		
Additional treatments								
Static_Low	X							
Monotonic								X
Simple_NoLoss		X	X	X			X	X

[Insert PDF here]