Effort Momentum

Patrick DeJarnette *

JOB MARKET PAPER

First Draft: November 15, 2015
This Draft: January 9, 2016

Abstract

This paper examines how past effort can impact current effort, such as when effort is reduced following an interruption. I study incentivized real-effort experiments in which both piece rates and leisure options were manipulated and find effort displays significant stickiness, even in the absence of switching costs. I demonstrate that this intertemporal evidence is indicative of effort "momentum", rather than on-the-job learning, reciprocity, or income targeting. When employing an instrumental variables (IV) approach, approximately 50% of the effort increase persists for 5 minutes after incentives return to baseline. Thus if a worker suffers a complete interruption in productivity, it would take an average of 15 minutes to return to 90% of prior work effort. While there are serious caveats with extrapolation, these findings indicate that productivity loss due to effort momentum alone could cost the US economy as much as $200 billion annually. I further demonstrate that advanced knowledge does not significantly reduce this productivity loss. This finding of effort momentum is especially important for potential labor economics studies that intend to employ individual fixed effects.

Keywords: intertemporal labor, effort allocation, momentum, interruptions

*Wharton School at University of Pennsylvania. pdejar@wharton.upenn.edu. I am indebted to Jeremy Tobacman for invaluable support. I also thank Iwan Barankay, Daniel Gottlieb, Rob Jensen, and Judd Kessler for their guidance and feedback. This paper has also greatly benefited from the comments of Paul Bruton, Tian Cai, Amanda Chuan, Ulrich Doraszelski, Clayton Featherstone, Fernando Ferreira, Joseph Harrington, Sonia Jaffe, Todd Sinai, Charles Sprenger, Shing-Yi Wang, Kate Wilhelm, Tomoyoshi Yabu, and participants of the Wharton School Applied Economics Workshop and the Wharton School Experiments Seminar. This research was supported by a grant from the Wharton School's Mack Institute for Innovation Management. All errors are my own. The most recent version of this paper may be found at www.patdejar.net.

1 Introduction

By some estimates, interruptions disrupt 1.5 to 2.1 hours per work day for over 56 million US "knowledge workers" (Gonzalez and Mark [2004], Spira and Feintuch [2005]). Observational studies show that hospital workers are interrupted 5 times per hour (Weigl et al. [2014], Berg et al. [2013]) while software developers and managers are interrupted 25 times per day (Gonzalez and Mark [2004]).[1] In similar studies, 15-23% of interrupted work is not resumed on the same day, a particular concern within the health services literature (Westbrook et al. [2010], Mark et al. [2005]). However, this evidence is difficult to interpret when the interruptions themselves may be necessary, as within a hospital's emergency department. In other contexts, interruptions from a manager may reduce principal-agent concerns through increased monitoring or communication as formulated by Coviello et al. [2014]. Interruptions from a co-worker could increase a firm's total output, even at the expense of the interrupted worker. As a result, it might be presumptuous to target interruptions themselves as a source of productivity waste.

At the heart of the question of how interruptions affect behavior is whether there is stickiness in effort allocation. If so, interruptions could undermine productivity due to unplanned effort reduction following an interruption. This paper answers this more general question of whether productivity loss persists over time, and if so, what might be done to recover it.

This "loss of momentum" is often posited by the psychology literature,[2] media,[3] and consulting reports,[4] but has not been thoroughly examined within the economics literature. The closest literature on intertemporal labor supply tends to focus on longer time scales and hours worked rather than output (Camerer et al. [1997], Oettinger [1999], Farber [2005, 2008], Crawford and Meng [2011], Chetty et al. [2011]). Within this literature, Fehr and Goette [2007] employ a field experiment on bicycle riders and finds evidence consistent with a model in which past effort exhausts riders, making additional effort more costly. Under this model, an exogenous interruption in effort could actually boost future productivity as the worker has had a chance to "catch their breath".[5] However, the longer time scale

[1] While half of these are considered "internal" interruptions, which may be more indicative of task juggling (Coviello et al. [2014]), the remaining half of "external" interruptions are arguably the most time intensive. For example, the data in Gonzalez and Mark [2004] also shows that 1.5 hours per day are spent on unscheduled meetings such as workers stopping by or talking through cubicle walls.

[2] See an extensive psychology literature on "flow", cf. Nakamura and Csikszentmihalyi [2002], Schaffer [2013], in which clear tasks with adequate challenge and objective goals enables a continuous work state.

[3] E.g. "*First, there is the diversion itself, taking your employees off task after they have assembled the resources and thinking necessary for that particular task. Then there is the restart—reassembling the resources, thoughts, and readiness. There is the loss of momentum caused by the initial distraction from the original purpose.*" (Brown [2015])

[4] Such as "*.. even a one-minute interruption can easily cost a knowledge worker 10 to 15 minutes of lost productivity due to the time needed to reestablish mental context and reenter the flow state.*" (Nielsen [2003])

[5] While consistent with an effort exhaustion model, the authors find that individual measures of loss aversion are predictive of the effort decrease, suggesting a model of loss-aversion may be more appropriate for that setting.

and the physical nature of the task make these findings difficult to apply to interruptions among a broader class of knowledge workers.[6]

In this paper, I hypothesize and test a theory in which past effort has a direct effect on disutility from present effort. This model has theoretical similarities to a model of habit preferences, but over effort rather than consumption. I refer to this theory as effort momentum.

To test for the presence of effort momentum, I conduct a series of real-effort laboratory experiments with 577 University of Pennsylvania students at the Wharton Behavioral Lab. This controlled setting allows me to observe workers' responses to both piece rates and leisure opportunities over multiple periods. The workers complete counting or slider tasks on a computer screen but have the option to engage in leisure by viewing YouTube videos at any time.[7] I manipulate (i) the piece rate for completed problems and (ii) the leisure opportunities available (by varying subjects' access to their cell phones). Subjects are quizzed prior to every period to ensure incentives and leisure options are understood.

The laboratory setting for the experiment enabled me to accurately measure productivity. This accuracy allows me to induce variation in effort over short time scales by changing incentives quickly. In addition, while the tasks involved are somewhat artificial, exerting effort on a computer located in a cubicle closely resembles a relevant work environment for many knowledge workers (Gonzalez and Mark [2004]).[8] The laboratory also eliminates peer effect confounds that might be present in a field setting, such as fairness concerns over some workers being paid more. Moreover, the setting allowed me to replicate across multiple designs and tasks to ensure effort momentum is not limited to a single context. The laboratory environment also made it possible to accurately enforce available leisure opportunities, particularly cell phone access. This leisure variation is important for differentiating momentum from alternate theories such as reciprocity.

My results show significant evidence of effort stickiness. Workers treated with a higher (lower) piece rate exert more (less) effort in the treated period relative to control.[9] Even after financial incentives return to baseline, workers who received a higher piece rate continue to work harder than those who only received a baseline piece rate. This lingering effort differential is approximately half of the original effort increase induced by the heightened piece rate. By the same token, workers who receive a lower

[6] There is a noted lack of consensus regarding what constitutes a "knowledge worker", but in line with the literature, I characterize knowledge work "as less tangible than manual work and using the worker's brain as the means of production" (Ramírez and Nembhard [2004]). This would roughly coincide with the 56 million "management, professional, and related occupations" from the 2014 Current Population Survey.

[7] This is in line with recent work demonstrating the importance of outside leisure options for external validity of laboratory experiments. (Corgnet et al. [2014], Charness et al. [2010], Eriksson et al. [2009], Kessler and Norton [2015])

[8] This statement is not meant to downplay the notable concerns over external validity of laboratory experiments. For a detailed discussion, please see Charness and Kuhn [2011], Falk and Heckman [2009], Levitt and List [2007].

[9] I find a positive elasticity of effort with respect to that period's piece rate of approximately 5% to 10%. In comparison to previous papers, this is a small but significant elasticity (Chetty et al. [2011]). I also find a significant negative effect on effort when given access to cell phones in some specifications.

piece rate in one period continue to exert less effort in following periods relative to the control group. These findings indicate that effort allocation in one period may depend positively on recent work effort.

This evidence of effort stickiness could be a result of momentum, reciprocity, on-the-job learning, or potentially other interpretations. To identify the source of this effort stickiness, I structured the experimental design to provide additional comparisons informed by theoretical predictions. One key feature of this design is that some workers are randomly informed of future piece rate and leisure opportunities a full period in advance. Previous studies on intertemporal effort allocation feature either imperfectly anticipated shocks (Camerer et al. [1997], Oettinger [1999], Pistaferri [2003]) or fully anticipated shocks (Lozano [2011], Fehr and Goette [2007]); none (to my knowledge) intentionally manipulate the degree of anticipation of piece rate or leisure shocks.

I am able to differentiate between reciprocity and effort momentum using this variation in anticipation. First, one would expect a reciprocating worker to work harder at the time of receiving the news of an increased piece rate, not just the period in which the higher piece rate is in effect.[10] I find no such evidence. Second, for those who enjoy the additional leisure opportunity (cellphone), one would expect them to reciprocate with higher effort in surrounding periods.[11] I also uncover no evidence of this, but rather find effort is significantly reduced following phone access for those affected, as consistent with effort momentum. Thus, effort stickiness seems unlikely to be driven by reciprocity in this setting.

To address concerns about on-the-job learning, the experiments feature extensive "training" periods. Analyzing the training period data suggests that subjects reach full competency with the tasks within the first 3 minutes (see Figures 3 and 4). After this time average output is remarkably flat, as opposed to increasing output predicted by an on-the-job learning model.[12] This may not be surprising given the extreme simplicity of the tasks and is further confirmed by post-experimental surveys (see Section 3). In an experiment with multiple post-treatment periods, post-treatment effort continues to converge to baseline as predicted by momentum, rather than stay elevated as predicted by on-the-job training. Lastly, the effort stickiness implies implausibly large effects under a model of on-the-job learning, addressed in more detail in Section 6.

One might think that "switching costs" could drive this effort stickiness, but the experimental design allows me to investigate effort momentum in the absence of such switching costs. In particular,

[10] Indeed, if one expects to find evidence of reciprocity either pre- or post-treatment, Gneezy and List [2006] suggests that the pre-treatment effects should be larger than post-treatment due to the declining effects of reciprocity over time.

[11] Some recent evidence regarding the importance of non-monetary gifts suggests this might even trigger greater reciprocation than the financial rewards (Kosfeld and Neckermann [2011], Bradler et al. [2013], Kube et al. [2012]).

[12] Even if one allows for on-the-job learning to be combined with income effects in such a way to produce flat output, one would expect to see the efficiency increases result in increased leisure time. The evidence instead suggests that leisure time is also flat or even declining for the control group.

increasing the piece rate to induce greater effort would not result in switching costs as the subject would remain engaged in the task, yet there is still evidence of effort stickiness in the following period. In addition, the experimental tasks employed are able to be stopped and resumed easily as the time investment is small. Thus, to the extent that interruptions incur additional switching costs, my estimates of effort momentum could represent a lower bound of the total effort loss.[13]

Other potential explanations for effort stickiness, such as neoclassical income effects or income reference dependence, are outlined in the section on theoretical predictions and following the experimental results. In addition to not accurately matching the comparative statics found, these theories were also tested with one additional treatment involving the salience of income. I find this information salience had no effect on piece rate effects, further suggesting that effort momentum is the most parsimonious theory to explain the evidence at hand.

After addressing alternate theories, I employ an instrumental variable (IV) approach in which the previous period's piece rate and leisure options influence the previous period's effort.[14] The primary concern with this approach would be if a past period's piece rate could directly influence future effort (e.g. via a model of reciprocity). This is distinct from a model of effort momentum, where the previous period's piece rate influences this period's optimal effort only through previous period's effort.

To summarize the main findings, I find that approximately 40-50% of effort changes persist for 5 minutes even after the incentives return to baseline levels. This increase continues to decline exponentially over multiple periods. Framed another way, after an interruption of effort, it takes about 15 minutes to return to 90% of pre-interruption effort levels. Structuring the findings using this momentum parameter also provides a way to transport findings to new populations or environments (Levitt and List [2007], Falk and Heckman [2009]). For example, this estimate of 40-50% was replicated using a different "slider" task as discussed in Appendix 9.3.

To address whether this productivity loss can be prevented with knowledge, I treat some subjects with information about future piece rates and leisure opportunities. Analysis shows this advance information does not impact productivity. This suggests the average subject follows a "naive" model of momentum as opposed to a more "sophisticated" model. These models are discussed in more detail in Section 2.

To put these findings in context, research suggest US knowledge workers are interrupted somewhere between 12 and 40 times a day depending on work environment. Given the ubiquity of interruptions and the large number of US knowledge workers, it is perhaps not surprising that the resulting momentum

[13] One might expect cell phones to incur greater switching costs as there is a change in user focus. This might explain why I find larger estimates of effort stickiness for cell phones in some specifications.

[14] Using asymptotics to remove the known bias, I also employ a Panel Median Unbiased Estimator as an alternate specification and find very consistent results. See Appendix Section 10.6.

loss is quite high. If the average knowledge worker suffers 15 interruptions per work day, this will result in about 1 hour of productivity loss due to momentum alone.[15] This works out to 200 hours per year per full-time worker. If each knowledge worker earns an average of $21 per hour, then 56 million workers[16] would lose $235 billion per year from momentum loss alone.[17] One important caveat is that if the reduced productivity results in greater leisure, this figure would also not account for any welfare gains from this leisure – however, workers tend to report interruptions as a major source of stress in the workplace, making welfare gains unlikely (Mark et al. [2008]).[18] While there are serious concerns about generalizing evidence from students,[19] this back of the envelope calculation demonstrates the potential value of further research.

Additional Contributions to Literature

In addition to the broader intertemporal labor supply literature, this paper builds on an extensive literature that uses laboratory experiments to investigate labor economic theories (for a review see Charness and Kuhn [2011]). While many papers in this literature have intertemporal implications (e.g. Rabin [1993], Dickinson [1999], Gneezy and List [2006], Levitt and List [2007], Buser and Peter [2012], Kube et al. [2012], Milkman et al. [2013], Kessler and Norton [2015]), I believe this is the first laboratory study to vary incentives over short time periods specifically to investigate intertemporal spillovers.[20]

As I find evidence of effort momentum over short time periods, this paper also suggests not to estimate individual fixed effects with short time panels. This is discussed in more depth within Section 4, but follows from earlier work on the asymptotic bias from fixed effects in time recursive models, proven in Nickell [1981]. Although this has been noted when estimating effects of training programs

[15] Momentum loss might also explain why subjective reports of time wasted due to interruptions (often as high as 40% of total work time) tend to be higher than the observed time loss (roughly 20% of total work time).

[16] From 2014 Current Population Survey, number of management, professional, and related workers. Most common examples include software developers, financial managers, accountants, lawyers, school teachers, registered nurses, and chief executives. This number also corresponds with the 77 million workers that reported using a computer at work in 2003 by the Bureau of Labor Statistics.

[17] Though this is just a rough estimate for a number of reasons. One might expect that the individuals who earn higher than average wages are less prone to interruptions or momentum loss. Alternatively, perhaps wages have already been lowered to account for interruption loss, underestimating the true value of productivity loss.

[18] This may not be surprising given that interruptions are unplanned, making such 'breaks' in effort unlikely to be ex ante optimal from the interrupted worker's perspective. Thus, even if the productivity loss following interruption increases utility through leisure, it may be used as a substitute for a more relaxing (planned) break. Thus, the utility from such leisure could be a net welfare loss as it disrupts the optimal on-the-job leisure schedule.

[19] For example, students may lack the workplace experience that could help reduce momentum loss. On the other hand, one might also expect college students to be better than average at avoiding momentum loss as they have passed college admissions. In addition, while the environment studied is similar to what many knowledge workers face, the tasks employed differ, raising additional concerns about external validity.

[20] Following Corgnet et al. [2014], it is also among the first to experimentally vary leisure opportunities within the laboratory, providing additional evidence on the effect of leisure on effort (Chapela [2007], Connolly [2008], Lozano [2011], Ward [2012]).

(Card and Sullivan [1988]), this study presents new evidence that the bias may be present in more general labor settings.

This new evidence of effort momentum may also provide new interpretations of existing labor studies. While pursuing other research topics, a few[21] recent studies have uncovered intertemporal evidence consistent with momentum. In Cardella and Depew [2015], experimental subjects stuff fewer envelopes after being quantity constrained in the first period (compared to control). By itself, however, this could be evidence of on-the-job learning or reduced reciprocity due to constrained output. Bradler et al. [2015] experimentally varies payment structures in one period and also finds some persistence in effort after those incentives have been removed. For example, those who face a tournament structure exert greater effort for both creative and uncreative tasks, which significantly persists in the following period. Yet this effect was strongest among tournament winners, making it theoretically unclear whether there was a "joy of winning" effect as in Kräkel [2008] or whether tournament winners, who worked hardest, simply had the largest spillover effects. Despite these confounds, this suggests that effort momentum might fill a gap between theory and empirics that has previously gone unreported.

The burgeoning literature on multitasking within economics may also benefit from study of effort momentum. Buser and Peter [2012] employ a real-effort experiment and find that subjects forced to work sequentially were more productive than subjects forced to work on tasks simultaneously. Additionally, workers allowed to work sequentially or simultaneously were also less productive than workers forced to work sequentially – mirroring the "naive" theory of effort momentum. However as Coviello et al. [2014] outlines theoretically, this sort of task-juggling may provide incentives to work harder when effort cannot be observed (as others compete for the worker's attention).

Even though my time scale is short, understanding intertemporal labor supply has important implications for labor markets and public policy. For example, if the intertemporal substitution elasticity is large and positive, one might interpret the lower pay of "flexible" positions as resulting from compensating differentials (Goldin [2014]) or a "Rat Race" equilibrium[22] (Akerlof [1976], Landers et al. [1996]). These outcomes might invite labor market policies to increase total surplus.[23] On the other

[21] This literature review is unlikely to be comprehensive as previous studies may have suffered from bias driven by individual fixed effects or may have simply omitted reporting intertemporal spillovers.

[22] By a rat race equilibrium, I mean one in which workers work inefficient hours or effort to signal hard to observe qualities (such as ability) to employers. This was first proposed in a theoretical framework by Akerlof [1976], and there has been evidence to suggest this occurs in at least law firms (Landers et al. [1996]). As Arulampalam et al. [2007] point out, this rat race equilibrium could contribute to gender pay gaps, especially toward the top.

[23] For example, Goldin [2014] calls for "alterations in the labor market, in particular changing how jobs are structured and remunerated to enhance temporal flexibility" to reduce gender inequality in labor markets. Generally, if firms have imperfect information about worker productivity, workers may be afraid to express a desire for flexibility even though such a change would increase total surplus for the worker and firm. For example, if an hour's potential productivity is correlated with leisure opportunities, a worker's desire for flexibility could signal a desire to only work low productivity hours.

hand, if this elasticity is small or negative, then the documented wage-flexibility tradeoff may be driven by firms' production and cost functions.[24] In this case, labor restrictions on hours could reduce firm efficiency.

The experimental design I constructed provides additional evidence on this intertemporal elasticity and is the first within this literature to vary anticipation of piece rates. As Fehr and Goette [2007] stress, the anticipation of wage changes is critical to interpretation of these elasticities.[25] Yet, informing workers about wage changes has the potential to trigger reciprocity toward the employer (Rabin [1993], Fehr and Schmidt [2006]). This complicates the interpretation of previously measured intertemporal elasticities, as anticipation and reciprocation are linked in studies with anticipated wage changes.[26] Furthermore, Gneezy and List [2006] suggest reciprocity may decline over time, potentially introducing an upward bias to wage elasticities measured over a short time period (if reciprocity is a large factor).[27]

The experimental design also allows me to address whether higher piece rates induce reciprocity. While the role of reciprocity in labor markets is an area of active research (for review see Kessler [2013], Levitt and Neckermann [2014]), I believe this is the first paper to tackle this particular question. The answer is ex ante unclear because while a higher piece rate expands the budget set, the worker must still exert effort to receive the benefits. Most previous studies testing for reciprocity in labor markets employ flat hourly wage variation in a reputation free environment (Kube et al. [2012], Fehr et al. [2008], Englmaier and Leider [2010, 2012], Kessler [2013], Gneezy and List [2006], Charness [2004]). As workers have arguably no financial incentive to work harder, evidence of greater effort is taken as evidence of reciprocity. Recent work such as Kube et al. [2012], Bradler and Neckermann [2015] suggests workers may reciprocate based on their impressions of employer intentions, rather than the actual "gift".[28] While my study is more suggestive on this point, I find no evidence that additional leisure opportunities induce reciprocity.

I also address whether salience of information induces workers to engage in income targeting. In a real-effort laboratory experiment, Abeler et al. [2011] find workers exert more effort when facing a chance of a higher fixed payment. Pope and Schweitzer [2011] finds evidence of loss aversion in a high

[24]For example, this could result either from per employee fixed costs (requisite search and training, benefits, or capital) or from increasing returns to hours worked (increasing worker knowledge flows, being available for clients).

[25]In studies with imperfect knowledge of future wage variation, workers with better predictions may react differently to wage changes and this could also be correlated with ability. Pistaferri [2003] attempts to overcome this issue by incorporating elicited worker expectations into estimates and finds a larger EIS estimate. Unfortunately, output or effort levels are not directly observable in their dataset.

[26]This "total effect" may be the most policy relevant measurement, but as Kube et al. [2012] have found, the framing of the wage increase can contribute greatly to the degree of reciprocation. Thus, there may be no unified policy relevant reciprocation measurement if it varies greatly based on implementation.

[27]Though, as authors admit, whether the reciprocity effects would return on the second day of work is an important open question for interpretation.

[28]However earlier work from Charness [2004] suggests that exogenously determined wages elicit almost as much reciprocation as employer designated wages.

stakes labor market (professional sports). In these settings, the reference point is at least partially induced by the environment (i.e. the magnitude of the outside option in Abeler et al. [2011] and golf par score in Pope and Schweitzer [2011]), but there remains some uncertainty whether information about own performance can alter endogenously chosen reference points. To investigate this possibility, I vary whether the worker sees her total earnings or her past period earnings and find it does not alter effort allocation.

The remainder of the paper is organized as follows. Section 2 derives straightforward comparative statics to distinguish the theories suggested above. Section 3 outlines the experiment designs. Section 4 discusses the specifics of the estimation strategy. Section 5 presents the results. Section 6 addresses additional concerns of alternate theories and Section 7 concludes.

2 Predictions

In this section, I derive predicted changes in labor supply to inform the experimental designs. I discuss three model classes below: (i) (neoclassical) time separable utility, (ii) effort momentum, and (iii) reciprocity. Additional model discussion may be found in Section 6 and the Appendix. I find straightforward comparative statics that can then be tested by the experimental design presented in Section 3.

2.1 Time Separable Utility

To serve as a starting point for predictions, I present a time separable model in which an agent maximizes lifetime utility

$$U_0 = \sum_{t=0}^{T} \delta^t u(c_t,\, e_t,\, \gamma_t)$$

where $\delta < 1$ represents the discount factor, $u(\cdot)$ represents the one-period utility function, c_t represents consumption, e_t is effort, and γ_t is a taste shifter that alters preferences for working in particular time periods. In my setting, γ_t can incorporate the varying leisure opportunities available, such as cell phone access. I further assume that the utility function is differentiable and $u_c > 0$, $u_e < 0$ and strictly concave in c_t and e_t. The lifetime budget constraint is given by

$$\sum_{t=0}^{T} p_t c_t (1+r)^{-t} \leq \sum_{t=0}^{T} (w_t e_t + y_t)(1+r)^{-t}$$

where p_t represents prices at time t, w_t the piece rate at time t for each unit of effort e_t, and y_t represents non-labor income. Also the interest rate r is assumed to be constant, but this does not impact the sign of the comparative statics.

As shown in Fehr and Goette [2007], along the optimal path, this model can be equivalently represented as an individual optimizing a static one period utility function that is linear in income. This can be written as:[29]

$$v(e_t, \gamma_t) \quad = \quad \lambda w_t e_t - g(e_t, \gamma_t)$$

where $g(e_t, \gamma_t)$ is strictly convex in e_t and captures the discounted disutility of effort. λ captures the marginal utility of life-time wealth. In this formulation, $\lambda w_t e_t$ represents the discounted utility from total income earned in period t.

Thus, as w_t increases, the optimal e_t^* will also increase. The effort exerted today is only influenced by past piece rates through the marginal utility of life-time wealth λ. In the literature on measuring temporary wage or piece rate shocks, this λ is assumed constant as the total impact on lifetime wealth is very small, implying small changes in λ (Fehr and Goette [2007]). Therefore, with no income effects, a single period's piece rate would have no impact on effort in surrounding periods.

If one allows for income effects, additional income would increase the attractiveness of leisure given the concavity of consumption. As a result, allowing for income effects would reduce effort in periods surrounding a piece rate increase.[30]

Lastly, if one allows for leisure technology γ_t to increase the disutility of effort (e.g. harder to work when the World Cup is on), then increasing leisure technology would decrease optimal effort e_t^* in that period. As with piece rates, in the absence of income effects there are no predicted spillovers on the surrounding periods. If one allows for income effects, then an agent would work harder in surrounding periods (say before or after the World Cup game). This follows as the reduced lifetime income (from the high leisure time) would increase the marginal utility of lifetime income, λ.

2.2 Effort Momentum

Effort momentum is a model in which past period's effort directly influences the disutility of future periods. For example, working hard may engage a flow-like state in which future effort is less costly.[31]

[29] Note that this formulation omits the price path $\{p_t\}$ and δ as these are not the objects of study. If these elements change, the corresponding g function would also change, but it could still be written in a similar format.

[30] Under the current experimental design, this would have very similar predictions to a model in which the agent has a daily income target.

[31] This is consistent with an extensive psychology literature on "flow", cf. Nakamura and Csikszentmihalyi [2002], Schaffer [2013], in which workers enter a state where the disutility of work is reduced as subjects report losing a sense of

Alternatively, if effort is interrupted for a period ($e_t = 0$), the worker may face greater disutility to start working again.[32]

To capture these ideas, I present a model in which an agent encounters lifetime utility:

$$U_M = \sum_{t=1}^{T} \delta^{t-1} u(c_t, e_t, e_{t-1}, \gamma_t)$$

where $\delta < 1$ represents the discount factor, $u(\cdot)$ represents the one-period utility function, c_t represents consumption, e_t is contemporaneous effort, and γ_t is a taste shifter that alters preferences for effort in particular time periods. In my setting, γ_t can incorporate varying leisure opportunities available and will be referred to as leisure technology.

I further assume that the utility function is twice-differentiable in its arguments with $u_1 \geq 0$, $u_2 \leq 0$, has a positive cross partial $u_{23} \geq 0$. With these assumptions, consumption is enjoyable, effort is unenjoyable, and past effort decreases the marginal disutility of effort. I also assume that leisure technology makes effort more costly in utility terms ($u_{24} \leq 0$), but also has no positive effect on consumption ($u_{14} \leq 0$).[33] Lastly, that effort does not make consumption more enjoyable ($u_{12} \leq 0$, $u_{13} \leq 0$). The lifetime budget constraint is given by

$$\sum_{t=1}^{T} p_t c_t (1+r)^{-t} \leq \sum_{t=1}^{T} (w_t e_t + y_t)(1+r)^{-t}$$

where p_t represents prices at time t, w_t the piece rate at time t for each unit of effort e_t, y_t represents non-labor income, and r is the interest rate from one period to the next.[34] I also assume there is no change in lifetime marginal utility of wealth λ is constant, as the total impact on lifetime wealth is very small. This is in line with other field and laboratory experiments in the labor economics literature (Fehr and Goette [2007], Camerer et al. [1997]).

2.2.1 Sophisticated Momentum

Sophisticated Momentum is the model as described above, in which the agent correctly realizes that today's effort will influence tomorrow's marginal disutility of effort.

self.

[32]It is worth noting that these intertemporal effects do not necessarily have to be positive – a conceptually similar model proposed by Fehr and Goette [2007] includes a cost function in which greater effort today increases the marginal disutility of effort in the next period, perhaps due to stress or physical exertion.

[33]Otherwise, increased leisure technology could boost desire for consumption to the extent that the agent works more to increase lifetime wealth.

[34]While the interest rate r is assumed to be constant for notation simplicity, this assumption does not impact the sign of the comparative statics.

Proposition 2.1 *Under the above assumptions, effort is monotonic non-decreasing in past, present, and future piece rates. Alternatively, effort is monotonic non-increasing in past, present, and future leisure technology expansions.*

Proof Application of supermodularity theorems. See Appendix Section 10.4. ■

The intuition for these comparative statics is straightforward. If a worker is aware that effort now will decrease the cost of effort in the next period, then the periods' optimal efforts will move together due to the spillover. For example, if the worker faces a higher piece rate next period, then next period's effort will become marginally more valuable. As work in the present reduces the costs of working next period, the marginal benefit of working in the present has also increased. By a similar argument, if the worker faces greater leisure opportunities next period, the benefits (due to effort momentum) of working today has also decreased.

2.2.2 Naive Momentum

Although the agent experiences the effects of momentum, it may be possible that either the agent does not realize this momentum will occur in the future, or otherwise uses an exogenous reference for future effort.[35] I call this model Naive Momentum. In this model, at period t, the agent maximizes a discounted stream of future utility:

$$U_t = u(c_t, e_t, e_{t-1}, \gamma_t) + \sum_{j=t+1}^{T} \delta^j v(c_j, e_j, \gamma_j)$$

and will formulate plans of this period and future period's effort. Note that the $v(\cdot)$ function above does not have e_{t-1} in its arguments. However, once the agent actually arrives at time $t+1$, he correctly incorporates previous period's effort into his lifetime utility:

$$U_{t+1} = u(c_{t+1}, e_{t+1}, e_t, \gamma_{t+1}) + \sum_{j=t+2}^{T} \delta^j v(c_j, e_j, \gamma_j)$$

This will cause the agent to revise his plans he made in time period t. The agent also faces the same budget constraint as before:

$$\sum_{t=0}^{T} p_t c_t (1+r)^{-t} \leq \sum_{t=0}^{T} (w_t e_t + y_t)(1+r)^{-t}$$

[35] One possible justification for this is the literature on Projection Bias, see Loewenstein et al. [2003], Conlin et al. [2007], Simonsohn [2010]. Under such projection bias, a tired individual may incorrectly project that they will always be tired – but if he started working harder he may be surprised to find he isn't as tired as expected.

Proposition 2.2 *Under the assumptions above, there is an equivalent period utility function*

$$u = \lambda w_t e_t - g(e_t, e_{t-1}, \gamma_t)$$

This form demonstrates that effort is increasing in past and present piece rates, but future piece rates have no impact. By the same token, effort is decreasing in past and present leisure technology, but future leisure technology has no impact.

Proof Proof of the $g(\cdot)$ function equivalence and its convexity is provided in Appendix 9.1, but builds on work by Browning et al. [1985] and Fehr and Goette [2007]. A brief proof for the comparative statics is provided below.

Consider the effect of an increase in w_{t+j}. In the first period, the first order condition states:

$$g_e(e_1^*, e_0, \gamma_1) \quad = \quad \lambda w_1$$

e_0 cannot be influenced by any $w_{t'}$ by construction, as time period 0 is before any information is received. γ_1 are not choice variables, they are only exogenously given. Thus when I take the derivative with respect to w_{t+j} to get:

$$g_{ee}\frac{de_1^*}{dw_{t+j}} \quad = \quad 0$$

Which, as $g_{ee} > 0$ gives us the effect in the first period of 0. In time period t, to complete the induction proof I assume $\frac{de_{t-1}^*}{dw_{t+j}} = 0$ and look to prove the same is true for $\frac{de_t^*}{dw_{t+j}}$. This follows from taking the total differential of the first order condition:

$$\begin{aligned} g_{ee}\frac{de_t^*}{dw_{t+j}} + g_{e2}\frac{de_{t-1}}{dw_{t+j}} &= 0 \\ \Rightarrow \frac{de_t^*}{dw_{t+j}} &= 0. \end{aligned}$$

Thus, by induction, optimal effort prior to a piece rate increase is unchanged when holding λ constant. This follows from the assumption of naivety that the agent does not anticipate future momentum. However, once the agent reaches the period with higher piece rates, an increase in the piece rate still elicits greater effort:

$$\frac{de_t^*}{dw_t} \quad = \quad \frac{\lambda}{g_{ee}} > 0$$

This follows from the convexity of g w.r.t. e_t^*. The same sign can be seen by looking at the total derivative with respect to past piece rates, w_{t-1}:

$$\frac{de_t^*}{dw_{t-1}} = -\frac{g_{e2}}{g_{ee}} \frac{de_{t-1}}{dw_{t-1}} > 0$$

As $g_{ee} > 0$, $\frac{de_{t-1}}{dw_{t-1}} > 0$ and $g_{e2} < 0$ (as $u_{23} > 0$). The proofs for leisure technology are the same as above with opposite signs (as leisure technology makes effort more costly, rather than less). ■

Although the above proposition gives us the required comparative statics of interest for naive momentum, considerably more can be said with an additional restriction on the period utility function. Without assuming a specific functional form, one can show that that the optimal effort will follow a linear time recursive structure.

Proposition 2.3 *Assuming further that* $u(c_t, e_t, e_{t-1}, \gamma_t) = q(c_t, e_t - \rho \cdot e_{t-1}, \gamma_t)$ *with* $|\rho| < 1$*, then optimal effort will follow a time recursive structure*

$$e_t^* = \rho \cdot e_{t-1} + z(w_t, \gamma_t)$$

with $z(\cdot)$ *increasing in* w_t *and decreasing in* γ_t.

Proof By a similar proof as above, the FOC will be

$$\begin{aligned} -q_e(c_t^*, e_t^* - \rho e_{t-1}, \gamma_t) &= \lambda w_t \\ q_c(c_t^*, e_t^* - \rho e_{t-1}, \gamma_t) &= \lambda p_t \end{aligned}$$

As q is strictly concave over the first argument, this allows for inverse of q_c:

$$c_t^* = q_c^{-1}(\lambda p_t, e_t^* - \rho e_{t-1}, \gamma_t)$$

Which can be inserted into the first FOC to give:

$$-q_e(q_c^{-1}(\lambda p_t, e_t^* - \rho e_{t-1}, \gamma_t), e_t^* - \rho e_{t-1}, \gamma_t) = \lambda w_t$$

Thus, a new utility function $\lambda w_t e_t - h(e_t^* - \rho e_{t-1}, \gamma_t)$. The convexity of $h(\cdot)$ gives us an inverse function for h_1:

$$e_t^* = \rho e_{t-1} + h_1^{-1}(\lambda w_t, \gamma_t)$$

As this is a special case of the first proposition (if $\rho > 0$), optimal effort e_t^* will still be an increasing function of w_t and decreasing in γ_t. In addition, past effort positively influences current effort and future piece rates or leisure technology does not influence current effort.■

2.3 Reciprocity

Consider instead a model in which changes in w_t and γ_t induce a desire to reciprocate. As formulated, this is similar to the time separable utility, but with an additional component of utility based on the piece rates and leisure offered across all time periods:

$$U_R = \sum_{t=0}^{T} \delta^t u(c_t, e_t, \gamma_t) + \alpha\left(\{w_t\}, \{\gamma_t\}\right) \cdot \left(\sum_{t=0}^{T} \delta^t e_t\right)$$

In which u has the same properties as outlined above ($u_1 \geq 0$, $u_{11} < 0$, $u_2 \leq 0$, $u_{22} < 0$, $u_{23} < 0$) and with $\alpha(\cdot)$ strictly increasing in its arguments. In this model, increases in future or past piece rates can increase the marginal utility of effort through the 'altruism' or 'fairness' function α. This model is similar to ones found in Rabin [1993], Fehr and Schmidt [2006]. Extending the work of Browning et al. [1985], this utility function can be reformulated as a series of period utility functions:

$$v(e_t) = [\lambda w_t + \alpha(\{w_t\}, \{\gamma_t\})] \cdot e_t - g(e_t, \gamma_t)$$

In a simple two period model for illustrative purposes, the agent receives additional marginal utility based on w_1 and w_2. For simplicity, I assume that this additional utility is linear in piece rate and effort, $\alpha(\{w_t\}, \{\gamma_t\}) = \alpha_1(w_1 + w_2) + \alpha_2(\gamma_1 + \gamma_2)$. Thus the agent is maximizing:

$$\begin{aligned}
U_R &= v(e_1) + v(e_2) \\
v(e_1) &\equiv (\lambda w_1 + \alpha_1 w_1 + \alpha_1 w_2 + \alpha_2 \gamma_1 + \alpha_2 \gamma_2) e_1 - g(e_1, \gamma_1) \\
v(e_2) &\equiv (\lambda w_2 + \alpha_1 w_1 + \alpha_1 w_2 + \alpha_2 \gamma_1 + \alpha_2 \gamma_2) e_2 - g(e_2, \gamma_2)
\end{aligned}$$

In this setting, increasing the piece rate can increase reciprocity, even in surrounding periods. Under this simple model, if $\alpha_2 > 0$ and e_1, e_2 is an interior solution, then $\frac{\partial e_2}{\partial \gamma_1} > 0$. Likewise, if $\alpha_1 > 0$ and e_1, e_2 is an interior solution, then $\frac{\partial e_2}{\partial w_1} > 0$. Similar intuitions apply for future piece rates or leisure technologies when informed in advance. For proofs, please see Appendix Section 10.4.

2.4 Summary

Owing to space limitations, several theories have been moved to a discussion following the results. To summarize the most relevant theories, I present the following table that outlines how effort at time t will respond to piece rates and leisure technologies at different times (past, present, and future):

Predictions Summary Table

		Effort at time t in response to increase in:					
		Piece Rate at time:			**Leisure Tech** at time:		
Models		$t-1$	t	$t+1$	$t-1$	t	$t+1$
Time Separable	No Income Effects	0	+	0	0	-	0
	Income Effects	-	+/-	-	+	-/+	+
Momentum	Naive	+	+	0	-	-	0
	Sophisticated	+	+	+	-	-	-
Reciprocity		+	+	+	+	-	+
On-the-job Learning		+	+	+	-	-	-
Income References	Period Target	0	-	0	0	-	0
	Total Target	-	+/-	-	+	-	+
	Previous Period	+	-	0	-	-	0
Experiment Results*		+	+	0	- or 0	- or 0	0

(*see Section 5 for details)

Please note that not all reference models give precise comparative statics in some situations, as shown in Brandon et al. [2014]. The comparative statics for reference models above are under the case that the reference or target is strong and influences the intertemporal effort allocation. For example, an income target of $1000 would not be possible to achieve in a laboratory setting, but would also not influence the intertemporal results (the agent would appear as a neoclassical time-separable agent). More adaptive models of reference dependence are discussed in Section 6.

Also please note that although On-the-job Learning and Sophisticated Momentum have the same predictions for the 6 comparative statics above, there are additional tests to distinguish these two hypotheses. For example, if the gains are primarily driven by learning, one would expect either (a) increasing quantity over time or (b) increasing leisure engagement over time. Neither of these are found to occur. There are also reasons to believe that the magnitudes involved make learning a very unlikely possibility, see Section 6 for more details. In practice, as I find evidence for Naive Momentum, these comparisons are not as crucial.

3 Experiment Design Overview

In order to test these comparative statics, I investigate how effort responds to changes in (i) past (ii) contemporaneous and (iii) future piece rates and leisure opportunities. These correspond to the 6 columns of the Prediction Summary Table above. These hypotheses were tested over two similar experiments (differences outlined below).

In both experiments, subjects complete incentivized real-effort tasks in a laboratory setting. The tasks involve counting images and are performed on a computer. This is similar to previous labor economics experiments studying effort in the laboratory, especially Abeler et al. [2011]. Subjects count particular images from a matrix of 98 images, as can be seen in Figure 1.[36] This task was selected as it requires little to no training, but is menial and requires effort.[37] In post experiment surveys, subjects often mention the task is boring (see Appendix Figure 1), in line with findings presented in Abeler et al. [2011]. Thus the primary measure of effort is the number of problems solved correctly – consistent with the experimental labor literature (Charness and Kuhn [2011], Fehr and Goette [2007]).[38]

[36] Abeler et al. [2011] has agents counting zeroes in a string of 100 numbers. This exact task was not feasible in a web browser with a "search" feature, which makes the task trivial as one can merely search for 0. As a result, I ask the worker to count either heart or drop icons (randomized at the subject level). Only one subject tried bypass the task by searching the "source code" (after being asked not to) and is dropped from analysis.

[37] Gill and Prowse [2012] employ a task with sliders that also has attractive properties (further outlined in Gill and Prowse [2011]) – this task was employed in a replication experiment with very similar results, see Appendix 9.3. However, focusing on the task similar to Abeler et al. [2011] also allows for a closer comparison to their results, including testing for possibility of reference dependence.

[38] Though output and effort may not be perfectly correlated, changes in the production function are unlikely to explain

In line with Corgnet et al. [2014], Eriksson et al. [2009], Charness et al. [2010] and to mirror many labor contexts outside of the laboratory, I introduce a baseline leisure activity. Specifically, the participants were allowed to watch YouTube.com videos at any time instead of performing counting tasks (see bottom of Figure 1). To help make YouTube videos a potentially worthwhile leisure activity, a pair of headphones was attached to every computer. However, as the video was located below the counting problem, it was difficult to engage in both simultaneously. In the appendix, I confirm that YouTube videos were indeed a time substitute for effort.

As discussed in Section 2, many models of effort allocation allow for changes in either piece rates or leisure options to impact effort. To test these models, I experimentally varied the piece rate and leisure opportunities in specific periods. Although the piece rate varied in some periods, every period contributed to final earnings. This was done to focus on intertemporal substitution as opposed to regret or risk aversion. Paying in every period also allowed me to distinguish between potential "daily" income targeting and "period" income targeting models. Final payment also included a flat $10 participation fee so long as they followed laboratory guidelines (e.g. no food, no talking).[39] However no payments were made until the end of the entire session.[40]

To vary the leisure opportunities, some subjects were randomly assigned access to their cell phones. The laboratory employed for this study, Wharton Behavioral Lab, ordinarily has a strict no phone policy to improve study compliance and concentration. This policy was put in place because participants have a tendency to want to text, browse the web, and play games on their cell phones during the lab session. Thus, phone access has the potential to represent an increase in the marginal utility of leisure (γ_t from Section 2).[41] This is conceptually similar to experiments conducted in Corgnet et al. [2014] which allowed some users to browse the internet to expand possible leisure activities participants face.[42]

Prior to being allowed to start each period, the subjects had to correctly answer questions about

evidence provided, as discussed in a Section 6.

[39]It was made clear and reiterated that they did not need to solve any problems to guarantee their $10 participation fee. In practice, every participant adequately followed the laboratory guidelines and received the $10 participation fee.

[40]Paying at the end was both a practical necessity given length of the periods and also mirrors the design of Fehr and Goette [2007].

[41]The subjects of the experiment were University of Pennsylvania undergraduates. The second experiment surveyed cellphone access – only 8 out of 422 subjects (2%) did not bring a cellphone to the laboratory. Even though phone quality may vary or some subjects may not have a phone, this will not impact estimate validity if randomization was adequately done. However, this research will only be able to answer whether access to phones already owned by the subjects influence effort rather than the effect of access to a particular phone. This was done in part because introducing a new cell phone would lead to significant learning, additional experimental cost, and may not represent the same expansion of leisure opportunities as if the individual owned the phone (e.g. no contacts, no texts, etc.)

[42]When given access, the students could also use the internet on their phones, so internet access could be seen in some way as a lower bound of the potential leisure opportunity faced by allowing phone use. Other leisure technology expansions were considered, but deemed too difficult to adequately monitor under the current lab setup. With cell phones, lab assistants were able to quickly verify whether cell phone users were allowed to use the cell phone at that time.

the upcoming period's piece rate and cell access. These procedures were implemented to ensure subjects fully understood the incentives they faced.[43] In addition, counters at the bottom kept track of current earnings (as in Abeler et al. [2011]) as well as visual indications for whether phone use was permitted.[44] In post experiment surveys, 98% of subjects report that the payments and leisure opportunities available were clear.

Although the experiments followed the general design above, I outline differences in the table below and elaborate in the following sections:

Design Summary Table

	Experiment 1	Experiment 2
Task	Image Counting	Image Counting
Location	Wharton Behavioral Lab	Wharton Behavioral Lab
Subjects	155 UPenn Undergraduates	422 UPenn Undergraduates
Number of Treatment Periods*	6	3
Duration of Treatment Periods	5 minutes	5 minutes
Duration of Pre-Treatment	5 minutes	15 minutes
Baseline Piece Rate (US $)	0.05	0.05
Piece Rate Treatments (US $)	0.15 or 0.30	0.03 or 0.08 or 0.15
Baseline Leisure Access	Youtube.com	Youtube.com
Leisure Access Treatment	Phone Access	Phone Access
Advance Information Occurs	Periods 1, 3, 5	Randomly in Period 1
Instructions followed by	30 second timer and Quiz	30 second timer and Quiz
Counters at Bottom	Period Earnings and Total Earnings	Period Earnings and either Total Earnings or Last Period Earnings
On Screen Timer	No	Yes
Image Counted	Hearts	Hearts or Drops (randomized)
Randomization	Subject Level by Computer	Subject Level by Computer

[43] In one pilot study, rather than quiz the subject on the piece rate, the website merely didn't allow them to continue until 30 seconds have passed. This allows me to investigate potential salience effects from repeating the piece rate but in the post experiment analysis did not seem to make a difference.

[44] In addition to the subject's earnings for the current period, either (i) the total earnings or (ii) previous period earnings are displayed on the screen at all times. This treatment serves as a supplementary test for income targeting and is explained in more detail in Section 6.

*Note: Number of treatment periods is the number of all periods after the pretreatment, i.e. periods in which individuals could differ in some way. In experiment 1, every subject experienced precisely 2 of the 6 rounds had a piece rate or leisure technology that was not baseline. In experiment 2, at most 1 period had a piece rate or leisure technology that was not baseline.

3.1 Experiment 1 Design

At the beginning of the session, the subject was given a series of instructions and an example problem. This was followed by one 5 minute "Pre-treatment" period to become acquainted with the task. This Pre-treatment period had the same incentives for all subjects and serves as a proxy of worker ability, as will be discussed in Section 4. For each solved problem in that period, the subject is informed she will earn $0.05. In order to discourage random guessing, there was also a penalty for wrong answers – entering the incorrect answer three times for a single problem resulted in a deduction of $0.20, akin to Abeler et al. [2011].[45]

The participants then completed six additional periods, each 5 minutes long, with three different possible treatments:

- **Control** – Subjects receive $0.05 per completed problem for that period.
- **High Piece Rate** – Subjects receive a higher piece rate for that period, either $0.15 or $0.30.
- **High Leisure Technology** – Subjects receive the ability to access their cellphones for one period but still received $0.05 per completed problem.

The control and piece rate treatments were calibrated using a small pilot study to allow for movement in either direction, as suggested by Charness and Kuhn [2011].

Regarding the randomization, these six treatment periods are broken up into three pairs. Each pair consisted of either two periods of Control treatment; a Control treatment and a High Piece Rate treatment; or a Control treatment and High Leisure Technology treatment. Within each pair, the order of the treatments was random in order to differentiate period effects and anticipation effects. Each subject eventually receives all three treatment pairs, potentially allowing for both between and within subject analysis. This randomization was executed at the individual level by a pseudo-random number generator seeded by computer time (down to the millisecond).

To test for adequate randomization, I investigate whether pre-treatment indicators (such as gender, self-reported SAT scores, and pretreatment performance) predict the period at which the subjects faced the High Piece Rate or High Leisure treatments. As reported in Table 2A, none of these factors individually or together

[45] On average the participants entered about 0.67 problems per period incorrectly, about 10% of total problems correctly solved per period.

are predictive of the period that they receive the treatments.[46] As a result, I conclude that the treatment randomization was adequately done given the observable characteristics.

3.2 Experiment 2 Design

The second experiment simplifies the first by assigning each worker only a single primary treatment, over four periods rather than seven. The first "Pre-treatment" period lasted 15 minutes, was the same for all subjects and is used to generate proxies for worker ability (see Section 4). The following "treatment" periods were all 5 minutes. The first of these featured a baseline piece rate, but could (randomly) inform the subject about the next period piece rate and leisure opportunity. In the following period, the subject receives either the baseline, a piece rate treatment, or a high leisure treatment. In the final period, the subject is returned to baseline piece rate and no access to the cell phone. As in experiment 1, randomization was executed at the individual level by a pseudo-random number generator seeded by computer time (down to the millisecond).

This experiment also expands on the first one in a number of ways. First, a second piece rate treatment arm was included, in which the piece rate is decreased from \$0.05 to \$0.03 and another treatment arm randomizing "total" vs "previous period" earnings shown. Second, by randomizing the information available for all subjects, the design eliminates concerns about "odd-period" x treatment interaction effects present in the first experiment.[47] Third, by keeping each individual to a single treatment, there may be less concern that interactions between multiple treatments confound effects. This also allowed for a longer training period to further reduce concerns about on-the-job learning. Fourth, a timer was added in accordance with Abeler et al. [2011] to minimize concerns about time uncertainty driving results. Lastly, additional variables, including specific timing and phone usage, were collected and a timer was added to the post experiment survey to improve information quality.

4 Empirical Specifications

In the following section, the results of the experiments will be addressed, but prior to that, three important empirical notes need to be made.

1. First, the potential presence of momentum – where the previous period's effort could directly influence this period's effort – makes this a poor setting for individual fixed effects. Estimating these individual fixed effects will lead to bias in the estimate of the momentum.[48] Nickell [1981] proves this, but the intuition is

[46] For regressing "High Piece Rate" treatment period # on pre-treatment variables, the F stat corresponds to a p-value of 0.29. For regressing the "High Leisure" treatment period # on pre-treatment variables, the F stat corresponds to a p-value of 0.71. Thus, for both treatments I fail to reject the hypothesis that all coefficients are zero and that none of the observable pre-treatment variables is significantly correlated with the period in which treatments occurred.

[47] In the first design, being "surprised" can only happen on treatment periods 1, 3, and 5 and "advance knowledge" can only occur for periods 2, 4, and 6. Although period fixed effects are included in most specifications, if odd-periods were interacting with treatments in some other way besides knowledge (e.g. piece rate increases are more effective in the final period), then estimates from experiment 1 could be a combination of those odd-period interaction effects and the effect of advance knowledge.

[48] Simply omitting the momentum term will not solve this issue in general but rather can bias other coefficients.

that shocks will be partially absorbed into the fixed effect estimate rather than the coefficient estimate for the previous period's effort. This is worse when there are fewer periods as there are fewer shocks to properly distinguish the coefficient estimates.

For example, assume effort follows an AR(1) process (similar to Proposition 2.3) and there is a time-constant individual fixed component

$$e_{i,t} \quad = \quad \rho \cdot e_{i,t-1} + f_i + \beta x_{i,t} + \nu_{i,t}$$

where $e_{i,t}$ is the number of problems solved by individual i at time t, ρ captures the degree of "momentum" from the previous period, f_i is the individual ability or motivation component, $x_{i,t}$ include other shifters such as piece rate or leisure technology and $\nu_{i,t}$ is an error term. Under this model, estimating individual fixed effects will introduce an asymptotic downward bias to ρ, approximately equal to $-\frac{1+\rho}{T-1}$. In my setting with $T = 3$, even if ρ was 0.5, asymptotic estimates would become indistinguishable from 0 as $N \to \infty$. This remains an issue even though the piece rate and leisure technology are randomized.[49] To be clear, this is not an issue of error terms correlated within an individual which could bias the standard errors[50] but rather a bias in the coefficient estimates themselves.

However, this was a known issue when designing the experiment and a primary justification for the pre-treatment period. This pre-treatment period can then serve as a proxy for individual ability or motivation, taking the place of f_i. To minimize risk of overfitting the data, a non-parametric approach is employed – individuals are split into five quintiles based the number of problems solved in the pre-treatment period, then each quintile receives it's own binary indicator variable.[51] The pre-treatment period is therefore omitted from the dependent variable for all specifications. In line with Card and Sullivan [1988], I also present a random effects model with identical findings in the Appendix.

2. The second note is that, although many labor studies take logarithms of dependent and independent variables, my specifications are reported at the unit level of analysis. This is done for several reasons, first being that the theory of momentum in section 2.2 suggested a unit level of analysis of effort. Given the linearity of the task, one might think effort would be closely correlated with quantity, not log(quantity). Second, while not common, some individuals did opt to solve no problems in a given period, a common issue with log forms. Third, the unit level was my ex ante specification while designing the experiment and analysis, and interpretation of new p values would be problematic after analysis has already been completed.

On the other hand, a linear formulation with OLS may be considered problematic as effort shocks cannot be too negative if effort is bound at the lower end at 0. It's also hard to represent upward effort "caps" with a linear specification. That being said I employ a log(problems correct + 1) on log(piece rate) specification

[49] However it may be worth noting that period fixed effects or session fixed effects will not be biased by this momentum, as the error terms are not correlated across individuals.

[50] Throughout the paper all standard errors are clustered at the subject level to reduce the influence of error terms correlated within an individual.

[51] Though in practice the results are virtually identical when using a linear and quadratic term for number of problems solved in pre-treatment.

and find qualitatively very similar results. While these measurement issues are important, the hypotheses are tested by the qualitative signs.

3. Third, as explained in Section 3, experiment 1 had each subject being treated to all 3 possible treatments. This helped improve power of treatment effects given the smaller sample – but it runs the risk of multiple treatments interacting to confound estimates. For example, access to a cellphone following a high piece rate period could negate additional effort resulting from momentum or reciprocity. As a result, I also perform analysis on just the first treatment received (corresponding to the first 3 periods of treatment)[52] and find that it does influence the intertemporal results in some specifications. While the interaction of treatments may be interesting, this was not the primary goal of this research study. Given this and the above difficulties of within-individual analysis in this setting, the second experiment design was simplified so that each person received only one treatment. This also allowed for a longer "training" period to further ensure the results are not being driven by on-the-job learning.

5 Experiment Results

Within this section, experimental results are presented together, as they are overall very similar across the two experiments. I use these results, particularly the qualitative signs, to test predictions of different theories from Section 2. I begin with the contemporaneous (same period) effects and move on to intertemporal effects.

5.1 Contemporaneous Effects

The first question is whether the primary treatments impacted contemporaneous effort as predicted by most theories of intertemporal labor supply. Recall that the primary treatments (piece rate or leisure technology) were in place for only 1 period, so this is asking whether or not effort was influenced in that treated period. This is important because if there is no effect on effort in the treated period, it would be difficult to understand why they would affect earlier or later periods.[53]

Result 5.1 *In accordance with most intertemporal theories of effort allocation, an increase in the piece rate significantly raised effort in the effected period. Likewise, in some specifications, there was a significant decrease in effort when subjects are offered access to their cellphones. See Tables 3A and 3B for details.*

In the first experiment, the average worker solves 0.20 to 0.45 more problems ($p < 0.01$) when faced with a 10 cent increase in the piece rate, as seen in Table 3A. This treatment estimate corresponds roughly to an

[52] Restricting analysis to only the first treatment pair is equivalent to focusing on just periods 1 and 2; however as half of those subjects received treatment in period 2, the following period (for intertemporal analysis) would be period 3. Results change very little when limiting it to individuals who have only received baseline piece rate (0.05) and leisure (YouTube) in period 3.

[53] That is not to say that certain combinations of theories could not predict such a finding, e.g. if an agent had a period income target but also experienced reciprocity, the two effects might cancel out in the effected period but could influence outside periods. However, given the extensive literature on piece rates influencing effort in the given period, such a null finding would likely indicate the treatment or sample size is too small (Levitt and Neckermann [2014]).

effort elasticity of 2% = (0.325/7.85) / (0.10/0.05). Thus, increasing the piece rate by 50% would increase average effort in this context by approximately 1%. This elasticity is small relative to previous findings in the literature, though still significant (Card [1991], Chetty et al. [2011], Fehr and Goette [2007]). Compared to the existing literature, this low result may be best explained by an effort ceiling.[54] In other words, at a $0.15 piece rate, agents may have already been exerting close to their maximum potential effort. There is some evidence for this, as the $0.15 and $0.30 piece rates both elicited greater effort, did not significantly differ from one another. This in turn would push down the average elasticity. To examine this possibility, experiment 2 features a $0.08 piece rate (1.6x baseline) and a $0.03 piece rate (0.6x baseline) instead of the $0.30 piece rate (6x baseline). Alternatively, the low elasticity could be the result of multiple treatment interactions. As seen in Table 4A, limiting the analysis to the first treatment increases the elasticity up to about 5%. Given these issues, I believe experiment 2 represents a better estimate for the contemporaneous elasticity.

In the second experiment, I find a larger effect of a higher piece rate on effort. As can be seen in Table 3B, every 10 cents (200%) increase (decrease) in piece rate increases (decreases) the number of correct problems by 1.24 - 1.57 (17-22%). Thus, the estimated elasticity of effort with respect to contemporaneous piece rate is 9.3% for experiment 2, quite a bit higher than the 5% found in experiment 1. Although not my primary inquiry, there did not seem to be a difference in magnitude for piece rate increases compared to decreases (no significant "kink" in the slope).

As discussed in Section 2, many intertemporal labor models also predict that increasing the marginal utility of leisure detracts from effort provision. One way to test this hypothesis is by increasing the leisure options available to the subject. To the extent that these leisurely options are complements with leisure time, one would expect an increase in leisure time and a corresponding decrease in total effort. In this experiment, the agents had access to Youtube.com videos throughout the experiment, but during the "High Leisure Technology" treatment, were also given access to their cellphones. When faced with this phone access, experiment 1 subjects complete 0.43 fewer problems on average ($p < 0.05$), as seen in Table 3A. This provides support for the hypothesis that leisure opportunities can reduce effort allocation.[55] However, once the sample is restricted to the first 3 treatment periods (to eliminate potential multiple-treatment interactions), this coefficient is no longer significant (see Table 4A specifications 4 through 6). Thus it is possible then that the original effort decrease due to cellphones was driven by subjects who received access to cellphones after the increased piece rate.[56]

[54] Though it may be difficult to compare as previous literature often focuses on hour or participation elasticity rather than effort.

[55] Unfortunately the binary nature of phone access does not allow for an accurate elasticity measurement of effort with respect to leisure opportunities. However, calculations of implicit time value of the leisure opportunity might allow one derive an estimate.

[56] Alternatively this evidence might suggest that cell phones were more effective in reducing effort during the later periods, or may be due to lower power from a smaller sample sample size. Results from experiment 2 suggests the latter hypothesis, as subjects in experiment 2 are treated with cellphone access quite "late" in the session, yet the treatment does not seem to influence effort.

In the second experiment, there was no significant decrease when cell phones were permitted. This matches the finding in the first experiment once restricted to the first treatment. However, when broken down by gender, cellphones appear to reduce effort in the contemporaneous period for males, as can be seen in Tables 5A and 5B.

These contemporaneous estimates also serve to test the "period income" reference dependence model. In this model, an agent receives relatively greater disutility if she falls short of a particular income in a given period. For example, a subject might try to earn $0.50 each period and then spent the rest of the time watching YouTube videos. Under this model one would usually[57] expect to see a reduction in effort when faced with a higher piece rate (as it has become easier to earn the target income for that period). Yet, as discussed, the findings suggest the opposite direction, with an increased piece rate inducing greater effort in the period it was enacted. Therefore, the contemporaneous evidence does not support a "period income" reference point.

5.2 Intertemporal Treatment Effects

In addition to a contemporaneous treatment effect, pre-and post-treatment effects are important to differentiate the theories outlined in Section 2. For example, if workers followed a neoclassical time separable utility function, then as total impact on income is small, one would not expect to see any reduction or increase in effort in the periods surrounding the high piece rate or high leisure treatments.[58] Instead, I find significant stickiness in effort:

Result 5.2 *In the period following an increase in the piece rate, effort was also significantly higher. This is consistent with models of Effort Momentum as well as Reciprocity. See Table 4A and 4B for details. However (randomized) advance knowledge of higher piecerates did not significantly influence effort. Of the models outlined in Section 2, these results are only consistent with a model of Naive Momentum. See Table 6A and 6B for details.*

In the second experiment, the intertemporal treatment effects are quite striking, presented in Table 4B. An increase in the previous period's piece rate of $0.10 (200%) significantly increases the effort in the following period by about 0.75 problems (10%). Thus, for experiment 2 the intertemporal elasticity is about half of the contemporaneous elasticity. By itself, this intertemporal effect could be due to reciprocity or momentum, as both predict higher effort following a higher piece rate.

Experiment 1 also has similar findings once restricted to the first three treatment periods, as can be seen in Table 4A. This may be due to the fact that in the first experiment, every individual received all 3 treatment pairs. As discussed in Section 4, this suggests the presence of multiple treatment interactions that were not ex

[57] As discussed in Brandon et al. [2014], if the piece rate increase is large enough or if the target is too large, contemporaneous effort could still increase with piece rates akin to a neoclassical time separable utility model. But if the targets are not strong enough to induce behavior that differs, their predictive value is reduced.

[58] Contrary to this prediction, the literature has found some evidence of effects in surrounding periods in the pursuit of other research (Cardella and Depew [2015], Bradler et al. [2015], Connolly [2008]).

ante predicted. Therefore, to limit the analysis to the post-treatment effects of just the first treatment pair, I analyze only the first three treatment periods.[59] Upon doing so, estimates suggest that a 5 cent (100%) piece rate increase in the previous period increases effort by 0.5 - 0.66 correct problems (6 - 9%). By itself, this result could be indicative of either momentum or reciprocity, as shown in Section 2.

Also worth noting is that while cellphones do not effect effort on average, it does seem to reduce contemporaneous effort for men over both experiments (see Tables 5A and 5B). This effect also persists in experiment 2 and some specifications of experiment 1. Also worth noting is that while cellphone *access* does not significantly alter effort in future periods, that self-reported cellphone *usage* is correlated with decreased effort in future periods, even after controlling for worker ability with productivity proxies (see Appendix Table 6).[60] Therefore, the intertemporal evidence of leisure is broadly suggestive of momentum rather than reciprocity (or other theories), as reciprocity would suggest a worker work harder after use or access to an increased leisure technology, not less hard.[61]

These findings also reject the "total income" target model. In this model, an agent receives relatively greater disutility for falling short of a particular income over multiple periods (in this case, the experimental session). If subjects in this experiment exhibited a total income reference point, subjects should reduce effort following a high piece rate period, as the agent was more likely to have hit their target in the preceding period. As shown in Tables 4A and 4B, the previous piece rate is instead positively correlated with effort in this period. Thus, I conclude there was no significant evidence of a total income target in this experiment.[62]

Lastly, being informed of the upcoming piece rates one period in advance did not influence effort. In the results of experiment 1, presented in Table 6A, there seems to be some borderline significant results when focusing on early panels, but not in the full panel. In the results of experiment 2, presented in Table 6B, knowledge of future piece rates also has no effect on effort – despite having similar sized standard errors and a similar level of power to detect as the effect of past piece rates. This evidence is suggestive of naive momentum rather than sophisticated momentum or reciprocity; both of these alternatives would predict effort increases upon learning about future piece rate increases.

Indeed, for both sophisticated momentum and reciprocity, one might expect the "pre-" piece rate effect to be larger than the "post-" piece rate effect. For reciprocity, work by Gneezy and List [2006] suggests that reciprocity decreases over time, thus the "post-" period having a larger effect is unlikely. For sophisticated

[59] Restricting analysis to only the first treatment pair is equivalent to focusing on the first 3 periods; however as half of those subjects received treatment in period 3, the following period would be period 4. Results change very little when limiting it to individuals who have only received baseline piece rate (0.05) and leisure (YouTube) in period 4.

[60] However, this finding has the potential for selection effects driving omitted variable bias, suggesting the coefficients should not be taken as causal estimates.

[61] As reciprocity seems more likely to trigger with non-monetary goods (Kube et al. [2012]), one might expect cell phone access to be even more likely to generate reciprocity than increased piece rates. One possible caveat – if subjects engage in cellphone use but do not actually "enjoy" this ability to use cell phones, e.g. due to self-control problems, it may not necessarily engage in reciprocity. However, this does not seem to be the case as the subject has a number of other self-control methods for cell phones (turning phone off, pulling out battery, leaving at home) and otherwise might suffer a very similar self-control issue with YouTube.

[62] Although not detected in my setting, Abeler et al. [2011] have devised an elegant method to elicit loss aversion even in a laboratory setting by varying a random outside option. It may be that in my experiment the subject has no experience with the task prior to the experiment, making it difficult to select a total income target.

momentum, extra effort in the "pre-" piece rate period would help take full advantage of the higher piece rates in the next period, whereas "post-" piece rate effects would be a result of previously expanded effort. Thus, the small and insignificant coefficient of future knowledge is strong evidence in favor of Naive Momentum.

5.3 Instrumental Variable Approach

If naive effort momentum is occurring, Proposition 3 in Section 2 guides how one might estimate it – in particular using an AR(1) approach. As discussed in Section 4, assume the true model is of the following sort:

$$e_{i,t} = \rho \cdot e_{i,t-1} + f_i + \beta_1 w_{i,t} + \beta_2 \gamma_{i,t} + \nu_{i,t}$$

Where $e_{i,t}$ is the number of problems solved by individual i at time t, ρ captures the degree of "momentum" from the previous period, f_i is the individual ability or motivation component, $w_{i,t}$ is the piece rate at time period t and $\gamma_{i,t}$ is the leisure technology available at time t, and $\nu_{i,t}$ is an error term. This theory allows us to encapsulate the force of momentum in a single parameter, which is potentially broader in application and policy implications and allows for easier comparisons across tasks (Charness and Kuhn [2011]).

While one could design an OLS structure to estimate the above, as mentioned before, the presence of fixed effects may bias the parameter ρ. I can employ productivity proxies as in previous specifications, but there may be remaining omitted variable bias as the uncaptured component of f_i , which now resides in the error term, may be correlated with $e_{i,t-1}$.

However, natural instrumental variables for previous period's effort are present – the previous period's piece rate and leisure technology. These variables are assigned randomly, but should influence the previous period's effort directly. To achieve asymptotic consistency, the instrumental variable w_{it-1} would need to satisfy the following:

$$\begin{aligned} Cov(w_{it-1}, e_{it-1}) &\neq 0 \quad \text{("First stage")} \\ Cov(w_{it-1}, \nu_{it}) &= 0 \quad \text{("Exclusion Principle")} \end{aligned}$$

While the "first stage" is strong as piece rates do impact contemporaneous effort (see Tables 3A and 3B), one might have reasonable doubts about the exclusion principle. In particular, suppose the true data generating process was a model of reciprocity, a process in which past piece rates directly influence current effort (rather than influencing effort through past effort). For example,

$$e_{it} = \rho e_{it-1} + \alpha_1 w_{it} + \alpha_2 w_{it-1} + \alpha_3 \gamma_{it} + \nu_{it}$$

If data from this data generating process was used to estimate an AR(1) model without w_{it-1} as a regressor, then $\alpha_2 w_{it-1}$ would remain in the error term. Since w_{it-1} will be correlated with e_{it-1}, this would result in omitted variable bias. In this case, it would overestimate the magnitude of ρ, as what is actually driven by reciprocity would be misinterpreted as momentum (w_{it-1} and e_{it-1} positively correlated).

Therefore, in order to believe the asymptotic consistency of an instrumental variable (IV) approach, one must be reasonably confident that the other models where previous piece rates enters directly (such as reciprocity or "total" income targeting) are not occurring. Although momentum most closely fits the comparative statics, additional discussion of alternate theories is provided below.

With this caveat in mind, I apply the instrumental variables (IV) approach using previous piece rate and phone access to predict previous period's effort. As presented in Table 7, the estimates find around 43-45% of the increased effort is retained in the following period, even once incentives revert to baseline.[63] I replicate this estimate of 43% using an alternate slider task in a replication experiment (see Appendix Table 1 and Section 10.3). Experiment 1 suffers from a weak instrument problem due to a smaller sample, but several specifications of experiment 1 are in line with this estimate of 45% (see Appendix Table 2).

6 Alternative Theories

While momentum seems to be the most parsimonious description of the contemporaneous and intertemporal results, careful consideration of other theories is warranted.

To reiterate, the comparative statics for the most part the design gives us have two tests to differentiate between momentum and reciprocity, both of which point in the direction of momentum.

1. As discussed in Section 5.2, informing a subject of a future piece rate increase or leisure option did not immediately increase effort as predicted by a model of reciprocity. Instead, I find that individuals only increase effort once the higher piece rate is applied. See Tables 6A and 6B.
2. After cell phone access, subjects exert less effort, not more as predicted by reciprocity (see Section 2.3 for predictions). See Tables 5A and 5B as well as Appendix Table 6 for results.

One possible alternative explanation for the experimental findings is that workers who experienced additional problems were able to increase their productivity in post-treatment periods ("On the Job Learning"). If there is significant on the job learning, then increased effort in an early period could result in additional problems solved in later periods. If true, this could account for stickiness detected.

However on-the-job learning seems an unlikely explanation as there was no indication of increased productivity over time as one would predict – see figures 2A and 2B. In addition, one might expect the number of incorrect problems to fall over time with learning, but this does not happen. Also, the task itself (counting 100 images dozens of times) has a limited scope for learning – indeed, Figure 4 shows a rapid convergence of problems solved per minute even within the Pre-treatment period.

Furthermore, even if on-the-job learning were occurring, it seems unlikely to explain the post-treatment effects. The relatively small increase in problems solved during period 2 (approximately 1 problem) is small

[63]This is substantially higher than the estimate of 75% given by OLS without employing an IV strategy.

relative to the total number of problems solved by that point (approximately 3%). If increasing the number of problems solved by 3% increases productivity by 10%, then period 3 should see an increase of approximately 60% even in the control group, which is inconsistent with the data. In addition, if this on the job learning was known to workers, then a future increase in piece rate would motivate additional effort in the preceding period so as to increase productivity. This was also not found in the data.

To recap why a period income targeting model does not fit the data, one would expect an increased piece rate to decrease contemporaneous effort (if the period income target is a significant component of utility). In addition, without adaptive references, a period income target model would predict no intertemporal spillovers. For more details, see the discussion in the contemporaneous effects section above.

To address why a "total" income targeting model does not fit the data, note that an increase in piece rate should reduce effort in surrounding periods. A higher piece rate makes it easier to hit a fixed "total" income target. Thus, to the extent that income targets induce effort,[64] the worker would exert less effort in the lower piece rate periods compared to control.[65] Instead, the data displays an increase in effort. For more details, see the discussion in the intertemporal results section above.

However, another possibility is period-level income reference dependence with adaptive references. In other words, by earning more in the previous period, the agent increases the income target for the following period.[66] This model may be hard to distinguish empirically from momentum, but there are two related tests that suggest adaptive income references are not driving the results.

First, if references are an important component of the utility function, this model would predict decreased effort when faced with a higher piece rate. This occurs as it is now easier to hit the income reference of the previous period. In the data, there are no such decreases when faced with higher piece rates, and workers work less hard when piece rates decrease.

Second, every subject had a counter to keep track of earnings from that period (see Figure 1). This was implemented to reduce subject uncertainty about earnings. In addition, experiment 2 subjects either had a "previous period earnings" or a "total earnings" counter located below the period earnings. This was randomly assigned at an individual level to potentially nudge period or daily income targeting. Specifically if an individual had been given the "previous period earnings" counter treatment and were driven by a period-level effort reference model, the post-treatment effects should have been stronger as they have more precise information about the effort and earnings exerted in the previous period. As can be seen in Table 8, this information did not significantly change the impact of an increased piece rate, either contemporaneously or

[64] If the target is too low or too high (e.g. $0.10 or $1000 for a laboratory study), then the agent will demonstrate behavior consistent with a neoclassical time separable model as the kink in utility will not be relevant to effort decisions.

[65] A neoclassical time separable model with income effects has similar predictions.

[66] Although not discussed extensively in this paper, a theory of adaptive income references presented in Brandon et al. [2014], Kőszegi and Rabin [2006, 2007, 2009] would also generally have an effect if information about piece rates is presented in advance. In addition, as shown by Brandon et al. [2014], Huffman and Goette [2006] workers who receive a higher lump sum early in the day should reduce their optimal effort afterward. This does not fit with the findings above, as workers treated to a higher piece rate worked harder even after incentives return to baseline.

in the previous period.[67] Though it seems to have influenced phone use slightly, this is hard to construe as evidence consistent with income reference dependence.

Another possibility is worker confusion regarding the piece rates. There are two reasons why this is unlikely to be driving results. First, before every period, the worker is presented a new instructions page which clearly outlines the piece rate in that period. This instructions page cannot be skipped for at least 30 seconds and workers must successfully type in the piece rate before they can continue. If the worker has information about future incentives, they are also quizzed on the future piece rate. Second, as mentioned above, in all experiments there was a counter that showed how much the subject had earned that period; thus even if they failed to understand the instructions, subjects would quickly see how much each problem was earning them. Of the 422 workers in the second experiment, only 4 individuals answered that the compensation was "somewhat unclear" or "unclear" in a post-experiment survey.

One last potential explanation is a lack of worker trust. If workers do not trust the promised piece rate increase, they may not reciprocate it initially, and instead wait until the piece rate is actually put in place. However, I believe two aspects make this explanation unlikely. First, 90% of subjects have previously completed 3 or more studies at the Wharton Behavioral Lab. As a dedicated experimental lab, Wharton Behavioral Lab has a reputation and incentives for upholding its promises to subjects. Secondly, if the worker did not trust promises of higher piece rates, it is unclear why higher piece rates would incentivize them to work harder in the treated period either, as there was no actual payment until the end of the experimental session.

7 Conclusion

I investigated the intertemporal elasticity of labor supply with a series of incentivized real effort experiments and find effort levels persist even once incentives return to baseline. After testing predictions to distinguish theories, I find strong evidence of effort momentum over short time scales and estimate a 5-minute momentum parameter of 0.45 across multiple experiments and tasks. This suggests it takes 15 minutes after an interruption to return to 90% of prior productivity levels, in line with observational evidence on interrupted work (Mark et al. [2005]). Providing information a full period in advance of does not seem to significantly influence this effort allocation – further suggesting a "naive" sort of momentum.

One weakness of this study is remaining uncertainty regarding the source of the momentum effects. For example, if effort momentum is a result of quickly decaying task-specific human capital (i.e. a "train of thought"), then switching tasks could be equally harmful as being interrupted. This would also be consistent with evidence that multitasking is less productive than sequential work, as found in Buser and Peter [2012]. Alternatively, it may be that momentum has a physiological component, perhaps due to adrenaline or other neurobiological processes. Lastly, the momentum could also be related to effort reference dependence (as opposed to income

[67] Unfortunately, while every subject did face a randomized counter, a small programming typo prevented the capture of this variable for the first day. As it is unclear which counter day 1 subjects faced, they are dropped from Table 7.

reference dependence), though many reference dependence models would predict that receiving information in advance should change the effort allocation (Brandon et al. [2014], Kőszegi and Rabin [2006, 2007, 2009]). Distinguishing these theories could help provide additional suggestions on how to minimize momentum loss after an interruption, e.g. a cellphone wallpaper reminding one to return to work after a phone call or doing 5 jumping jacks immediately after an interruption.

There's also some uncertainty to the extent to which workers are aware of these momentum effects. Although they do not seem to employ information to take advantage of momentum, there is still a chance workers are aware of it conceptually. As has occurred with some past studies (Price and Wolfers [2010], Pope et al. [2013]), increased awareness of the momentum effect may overturn or undo some of the effect. For example, if a worker knows they tend to work harder after working hard, they may slack off early and expect the work to "finish itself". Or as Mark et al. [2008] find, workers may work harder following an interruption to "catch up", though I find no evidence of this. One possibility to investigate the degree of self-awareness is to use costly commitment with a self-selected cut-off, akin to Kaur et al. [2010].

Another open question is whether these momentum effects would persist over longer time periods. One replication experiment with multiple periods following the treatment suggests that effort continues to decay exponentially, suggesting that the effects of momentum would disappear within 20 minutes or so. That being said, even if short-lived, measuring momentum may have direct applications to the economics of task juggling and interruptions. As outlined in Section 5.3, approximately 45% of effort momentum persists after 5 minutes. Using this estimate as a starting point, 45% of productivity is lost in the first 5 minutes after an interruption, an additional 20% in the second 5 minutes, 9% in the third 5 minutes, 4% in the next 5 minutes and so on. In total, if an interruption causes me to lose 5 minutes of productivity, I lose an additional 4 minutes of productivity due to effort momentum loss spread out over the next 30 minutes. Put in other terms, total productivity loss from effort momentum is 80% of the original interruption loss. Given estimates of the number of interruptions knowledge workers face, that suggests up to an hour of productivity per work day could be lost due to effort momentum alone.

References

Johannes Abeler, Armin Falk, Lorenz Goette, and David Huffman. Reference points and effort provision. *The American Economic Review*, 101(2):pp. 470–492, 2011. ISSN 00028282. URL `http://www.jstor.org/stable/29783680`.

George A Akerlof. The Economics of Caste and of the Rat Race and Other Woeful Tales. *The Quarterly Journal of Economics*, 90(4):599–617, November 1976. URL `http://ideas.repec.org/a/tpr/qjecon/v90y1976i4p599-617.html`.

Wiji Arulampalam, Alison L Booth, and Mark L Bryan. Is there a glass ceiling over europe? exploring the gender pay gap across the wage distribution. *Industrial & Labor Relations Review*, 60(2):163–186, 2007.

Lena M Berg, Ann-Sofie Källberg, Katarina E Göransson, Jan Östergren, Jan Florin, and Anna Ehrenberg. Interruptions in emergency department work: an observational and interview study. *BMJ quality & safety*, 22(8):656–663, 2013.

Christiane Bradler, Robert Dur, Susanne Neckermann, and Arjan Non. Employee recognition and performance: A field experiment. *ZEW-Centre for European Economic Research Discussion Paper*, (13-017), 2013.

Christine Bradler and Susanne Neckermann. The magic of the personal touch: field experimental evidence on money and gratitude as gifts. *Working Paper*, 2015.

Christine Bradler, Susanne Neckermann, and Arne Jones Warnke. Creativity is different: Comparing rewards across a creative and a routine task. *Working Paper*, 2015.

Alec Brandon, John List, Steffen Andersen, and Uri Gneezy. Toward and understanding of reference-dependent labor supply: Theory and evidence from a field experiment. Technical report, The Field Experiments Website, 2014.

Edward G. Brown. The hidden costs of interruptions at work. `http://www.fastcompany.com/3044667/work-smart/the-hidden-costs-of-interruptions-at-work`, 2015. Accessed: 2015-09-30.

Martin Browning, Angus Deaton, and Margaret Irish. A profitable approach to labor supply and commodity demands over the life-cycle. *Econometrica: journal of the econometric society*, pages 503–543, 1985.

Thomas Buser and Noemi Peter. Multitasking. *Experimental Economics*, 15(4):641–655, 2012.

Colin Camerer, Linda Babcock, George Loewenstein, and Richard Thaler. Labor supply of new york city cabdrivers: One day at a time. *The Quarterly Journal of Economics*, pages 407–441, 1997.

David Card. Intertemporal Labor Supply: An Assessment. NBER Working Papers 3602, National Bureau of Economic Research, Inc, January 1991. URL `http://ideas.repec.org/p/nbr/nberwo/3602.html`.

David Card and Daniel Sullivan. Measuring the effect of subsidized training programs on movements in and out of employment. *Econometrica*, 56(3):497–530, 1988.

Eric E. Cardella and Briggs Depew. Testing for the ratchet effect: Evidence from a real-effort work task. *Working Paper*, 2015.

Jorge Gonzalez Chapela. On the price of recreation goods as a determinant of male labor supply. *Journal of Labor Economics*, 25(4):pp. 795–824, 2007. ISSN 0734306X. URL `http://www.jstor.org/stable/10.1086/519538`.

Gary Charness. Attribution and reciprocity in an experimental labor market. *Journal of Labor Economics*, 22 (3):665–688, 2004.

Gary Charness and Peter Kuhn. Lab labor: What can labor economists learn from the lab? *Handbook of labor economics*, 4:229–330, 2011.

Gary Charness, David Masclet, and Marie Claire Villeval. Competitive preferences and status as an incentive: Experimental evidence. *Groupe d'Analyse et de Theorie Economique Working Paper*, (1016), 2010.

Raj Chetty, Adam Guren, Day Manoli, and Andrea Weber. Are micro and macro labor supply elasticities consistent? a review of evidence on the intensive and extensive margins. *The American Economic Review*, 101(3):471–475, 2011.

Michael Conlin, Ted O'Donoghue, and Timothy J Vogelsang. Projection bias in catalog orders. *The American Economic Review*, pages 1217–1249, 2007.

Marie Connolly. Here comes the rain again: Weather and the intertemporal substitution of leisure. *Journal of Labor Economics*, 26(1):pp. 73–100, 2008. ISSN 0734306X. URL `http://www.jstor.org/stable/10.1086/522067`.

Brice Corgnet, Roberto Hernan-Gonzalez, and Eric Schniter. Why real leisure really matters: incentive effects on real effort in the laboratory. *Experimental Economics*, pages 1–18, 2014. ISSN 1386-4157. doi: 10.1007/s10683-014-9401-4. URL `http://dx.doi.org/10.1007/s10683-014-9401-4`.

Decio Coviello, Andrea Ichino, and Nicola Persico. Time allocation and task juggling. *The American Economic Review*, 104(2):609–623, 2014.

Vincent P Crawford and Juanjuan Meng. New york city cab drivers' labor supply revisited: Reference-dependent preferences with rationalexpectations targets for hours and income. *The American Economic Review*, 101(5):1912–1932, 2011.

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

Florian Englmaier and Stephen Leider. Gift exchange in the lab-it is not (only) how much you give... 2010.

Florian Englmaier and Stephen Leider. Managerial payoff and gift exchange in the field. *Available at SSRN 2175776*, 2012.

Tor Eriksson, Anders Poulsen, and Marie Claire Villeval. Feedback and incentives: Experimental evidence. *Labour Economics*, 16(6):679–688, 2009.

Armin Falk and James J Heckman. Lab experiments are a major source of knowledge in the social sciences. *science*, 326(5952):535–538, 2009.

Henry S Farber. Is tomorrow another day? the labor supply of new york city cabdrivers. *Journal of Political Economy*, 113(1), 2005.

Henry S Farber. Reference-dependent preferences and labor supply: The case of new york city taxi drivers. *The American Economic Review*, 98(3):1069–1082, 2008.

Ernst Fehr and Lorenz Goette. Do workers work more if wages are high? evidence from a randomized field experiment. *The American Economic Review*, 97(1):pp. 298–317, 2007. ISSN 00028282. URL `http://www.jstor.org/stable/30034396`.

Ernst Fehr and Klaus M Schmidt. The economics of fairness, reciprocity and altruism–experimental evidence and new theories. *Handbook of the economics of giving, altruism and reciprocity*, 1:615–691, 2006.

Ernst Fehr, Lorenz Goette, and Christian Zehnder. A behavioral account of the labor market: The role of fairness concerns. *Institute for Empirical Research in Economics, University of Zurich, Working Paper*, (394), 2008.

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

David Gill and Victoria L Prowse. A novel computerized real effort task based on sliders. 2011.

Uri Gneezy and John A List. Putting behavioral economics to work: Testing for gift exchange in labor markets using field experiments. *Econometrica*, 74(5):1365–1384, 2006. ISSN 1468-0262. doi: 10.1111/j.1468-0262.2006.00707.x. URL `http://dx.doi.org/10.1111/j.1468-0262.2006.00707.x`.

Claudia Goldin. A grand gender convergence: Its last chapter. *The American Economic Review*, 104(4): 1091–1119, 2014.

Victor M Gonzalez and Gloria Mark. Constant, constant, multi-tasking craziness: managing multiple working spheres. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 113–120. ACM, 2004.

David Huffman and Lorenz Goette. Incentives and the allocation of effort over time: the joint role of affective and cognitive decision making. 2006.

Supreet Kaur, Michael Kremer, and Sendhil Mullainathan. Self-control and the development of work arrangements. *The American Economic Review*, pages 624–628, 2010.

Judd B Kessler. When will there be gift exchange? addressing the lab-field debate with laboratory gift exchange experiments. 2013.

Judd B Kessler and Michael I Norton. Tax aversion in labor supply. *Journal of Economic Behavior & Organization*, 2015.

Michael Kosfeld and Susanne Neckermann. Getting more work for nothing? symbolic awards and worker performance. *American Economic Journal: Microeconomics*, pages 86–99, 2011.

Botond Kőszegi and Matthew Rabin. A model of reference-dependent preferences. *The Quarterly Journal of Economics*, pages 1133–1165, 2006.

Botond Kőszegi and Matthew Rabin. Reference-dependent risk attitudes. *The American Economic Review*, pages 1047–1073, 2007.

Botond Kőszegi and Matthew Rabin. Reference-dependent consumption plans. *The American Economic Review*, 99(3):909–936, 2009.

Matthias Kräkel. Emotions in tournaments. *Journal of Economic Behavior & Organization*, 67(1):204–214, 2008.

Sebastian Kube, Michel André Maréchal, and Clemens Puppe. The currency of reciprocity: Gift exchange in the workplace. *The American Economic Review*, pages 1644–1662, 2012.

Renee M Landers, James B Rebitzer, and Lowell J Taylor. Rat race redux: Adverse selection in the determination of work hours in law firms. *The American Economic Review*, pages 329–348, 1996.

Steven D Levitt and John A List. Viewpoint: On the generalizability of lab behaviour to the field. *Canadian Journal of Economics/Revue canadienne d'économique*, 40(2):347–370, 2007.

Steven D. Levitt and Susanne Neckermann. What field experiments have and have not taught us about managing workers. *Oxford Review of Economic Policy*, 30(4):639–657, 2014. doi: 10.1093/oxrep/grv003. URL `http://oxrep.oxfordjournals.org/content/30/4/639.abstract`.

George Loewenstein, Ted O'Donoghue, and Matthew Rabin. Projection bias in predicting future utility. *The Quarterly Journal of Economics*, pages 1209–1248, 2003.

Fernando A Lozano. The Flexibility Of The Workweek In The United States: Evidence From The Fifa World Cup. *Economic Inquiry*, 49(2):512–529, 04 2011. URL `http://ideas.repec.org/a/bla/ecinqu/v49y2011i2p512-529.html`.

Gloria Mark, Victor M Gonzalez, and Justin Harris. No task left behind?: examining the nature of fragmented work. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 321–330. ACM, 2005.

Gloria Mark, Daniela Gudith, and Ulrich Klocke. The cost of interrupted work: more speed and stress. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 107–110. ACM, 2008.

Katherine L Milkman, Julia A Minson, and Kevin GM Volpp. Holding the hunger games hostage at the gym: An evaluation of temptation bundling. *Management Science*, 60(2):283–299, 2013.

Jeanne Nakamura and Mihaly Csikszentmihalyi. The concept of flow. *Handbook of positive psychology*, pages 89–105, 2002.

Stephen Nickell. Biases in dynamic models with fixed effects. *Econometrica: Journal of the Econometric Society*, pages 1417–1426, 1981.

Jakob Nielsen. Im, not ip (information pollution). *Queue*, 1(8):76–75, 2003.

Gerald S Oettinger. An empirical analysis of the daily labor supply of stadium venors. *Journal of political Economy*, 107(2):360–392, 1999.

Luigi Pistaferri. Anticipated and unanticipated wage changes, wage risk, and intertemporal labor supply. *Journal of Labor Economics*, 21(3):729–754, 2003.

Devin G Pope and Maurice E Schweitzer. Is tiger woods loss averse? persistent bias in the face of experience, competition, and high stakes. *The American Economic Review*, 101(1):129–157, 2011.

Devin G Pope, Joseph Price, and Justin Wolfers. Awareness reduces racial bias. Technical report, National Bureau of Economic Research, 2013.

Joseph Price and Justin Wolfers. Racial discrimination among nba referees. *The Quarterly Journal of Economics*, 125(4):1859–1887, 2010. doi: 10.1162/qjec.2010.125.4.1859. URL `http://qje.oxfordjournals.org/content/125/4/1859.abstract`.

Matthew Rabin. Incorporating fairness into game theory and economics. *The American economic review*, pages 1281–1302, 1993.

Yuri W Ramírez and David A Nembhard. Measuring knowledge worker productivity: A taxonomy. *Journal of intellectual capital*, 5(4):602–628, 2004.

Owen Schaffer. Crafting fun user experiences: A method to facilitate flow. *Human Factors International. Online White paper. http://www. humanfactors. com/FunExperiences. asp*, 2013.

Uri Simonsohn. Weather to go to college. *The Economic Journal*, 120(543):270–280, 2010.

Jonathan B. Spira and Joshua B. Feintuch. The cost of not paying attention: how interruptions impact knowledge worker productivity. Technical report, September 2005.

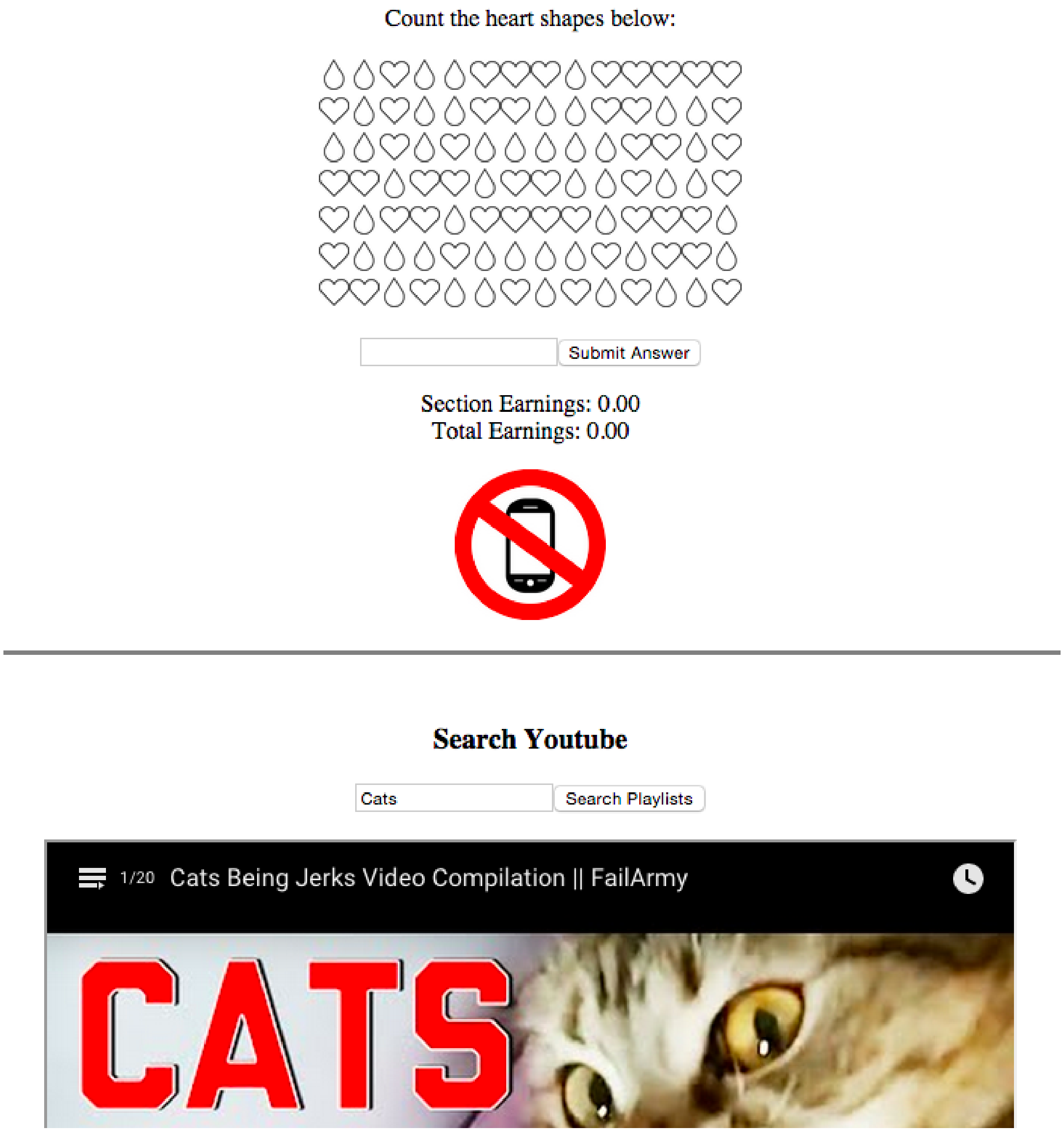
Michael R Ward. Does time spent playing video games crowd out time spent studying? 2012.

Matthias Weigl, Andreas Müller, Peter Angerer, and Florian Hoffmann. Workflow interruptions and mental workload in hospital pediatricians: an observational study. *BMC health services research*, 14(1):433, 2014.

Johanna I Westbrook, Enrico Coiera, William TM Dunsmuir, Bruce M Brown, Norm Kelk, Richard Paoloni, and Cuong Tran. The impact of interruptions on clinical task completion. *Quality and Safety in Health Care*, 19(4):284–289, 2010.

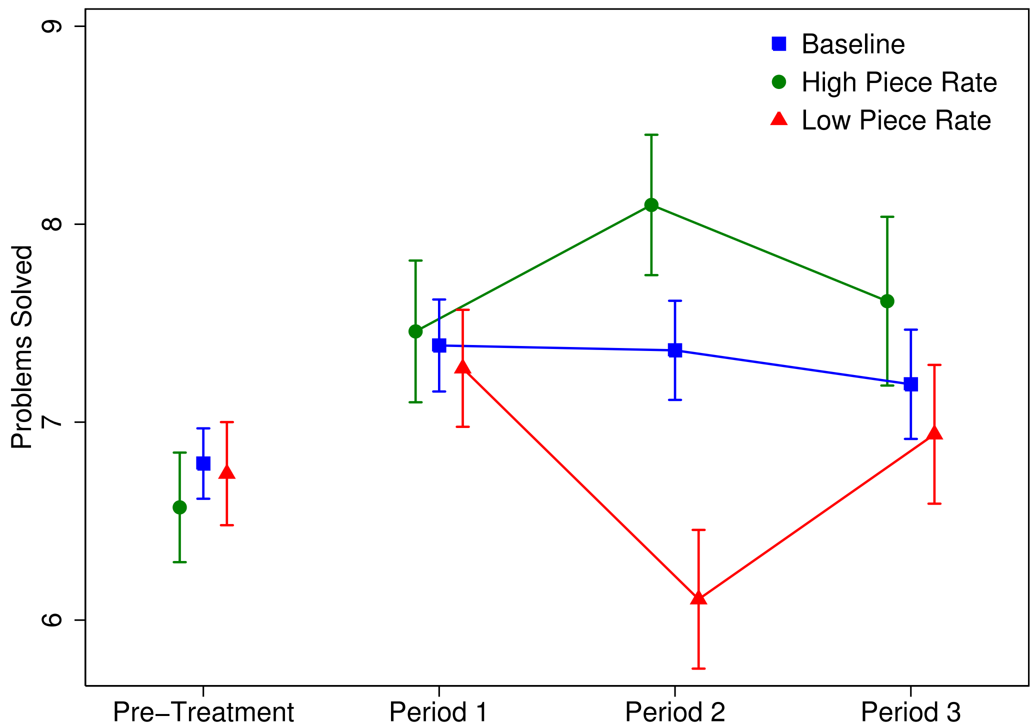
8 Figures

Figure 1: Example of Counting Problem Task



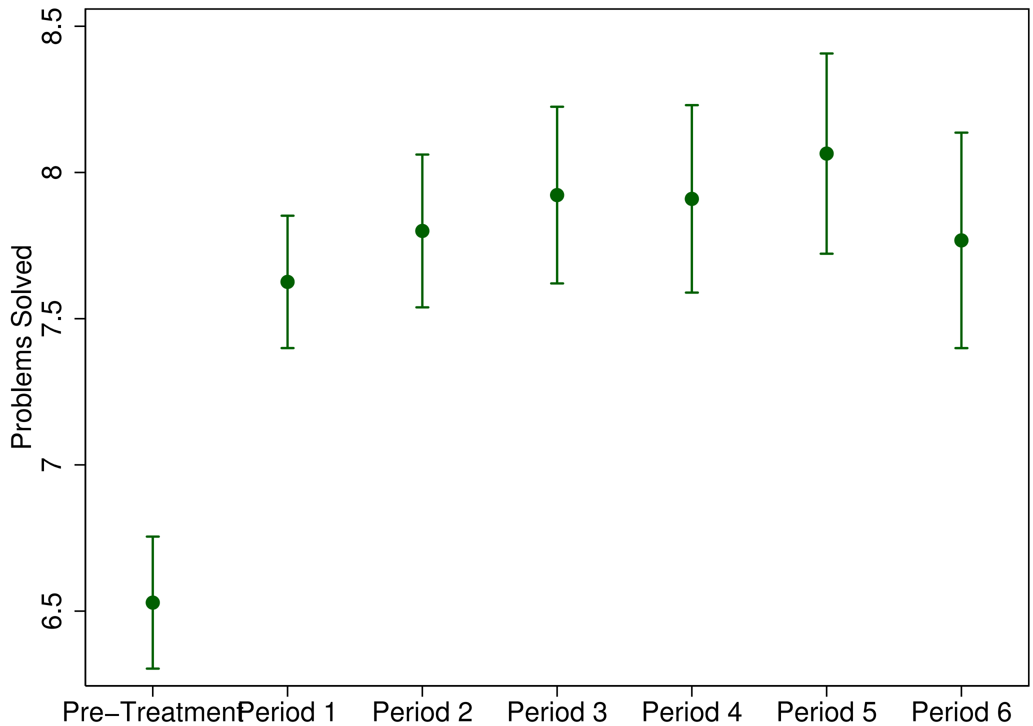
Notes: Figure demonstrates a typical counting task screen faced by subject. Whether subject was asked to count hearts or drops was randomized in experiment 2.

Figure 2: Solved Problems by Period – Experiment 2



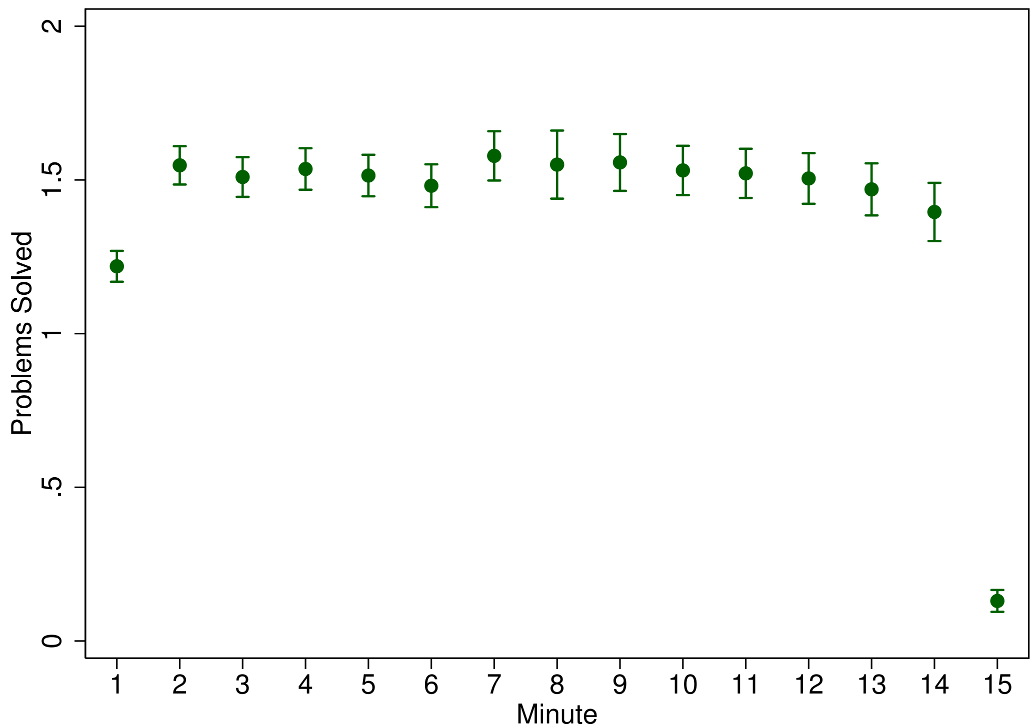
Notes: Bars represent standard errors. Vertical axis represents the number of problems solved by workers. Treatments were only in effect for period 2 (see experimental design in section 3). The Pre-Treatment period is a training period to familiarize workers with the task. Pre-Treatment lasted 3 times the duration of the other periods and thus the problems solved in Pre-Treatment is divided by 3 to provide accurate comparison.

Figure 3: Solved Problems by Period – Experiment 1



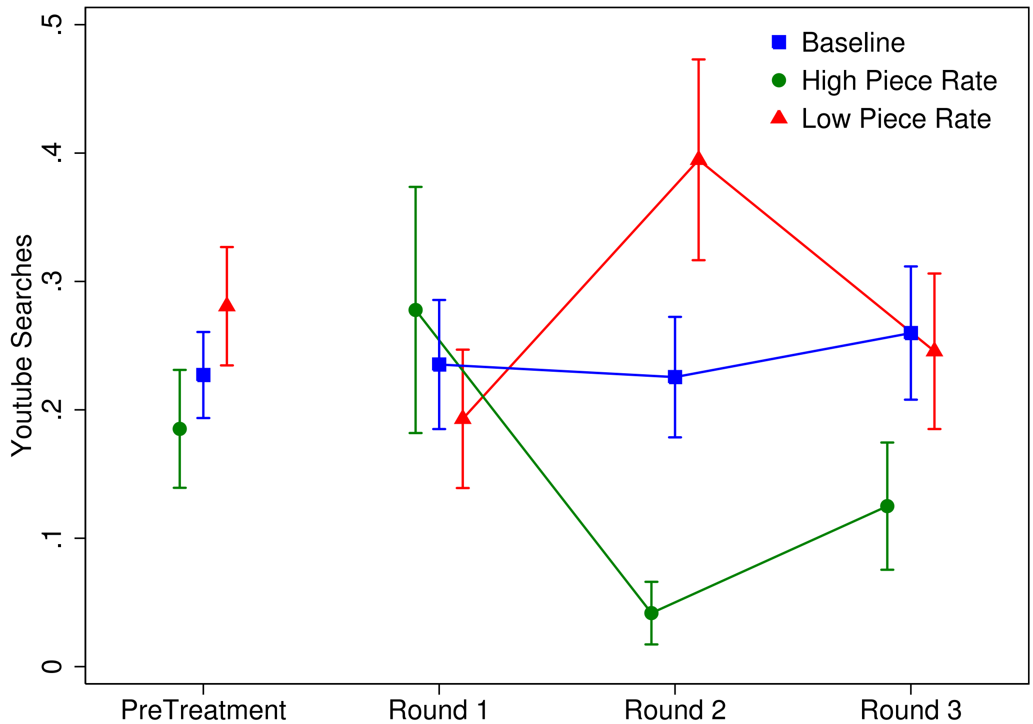
Notes: Bars represent standard errors. Three treatment pairs were applied at varying periods (see experimental design in section 3). The Pre-Treatment period is a training period to familiarize workers with the task.

Figure 4: Solved Problems within Pre-Treatment Period – Experiment 2



Notes: Bars represent standard errors. This figure demonstrates the number of problems solved by minute of the pre-treatment period. As these counting problems take about 45 seconds, the final minute was lower due a mechanical effect (of being unable to finish a problem in time) and additional uncertainty of whether one is able to finish the problem in time (perhaps due to the timer reading "0 minutes left").

Figure 5: YouTube Searches by Period – Experiment 2



Notes: Bars represent standard errors. Vertical axis represents the number of YouTube searches performed by workers (baseline leisure option). Treatments were only in effect for period 2 (see experimental design in section 3). The Pre-Treatment period is a training period to familiarize workers with the task. Pre-Treatment lasted 3 times the duration of the other periods and thus the problems solved in Pre-Treatment is divided by 3 to provide accurate comparison.

9 Tables

Table 1. Summary Statistics

	Experiment 1				Experiment 2			
	Mean	Standard dev	Min	Max	Mean	Standard dev	Min	Max
Individual Level Variables								
Female	0.72	0.45	0	1	0.71	0.45	0	1
Problems Solved in PreTreatment	6.59	2.75	0	14	20.2	7.60	0	67
Age	21.3	5.27	18	61	20.4	1.85	18	38
SAT Math Score	731	78	165	800	731	63	400	800
Total Payment	3.87	1.75	0	10.75	2.16	0.94	0	5.4
Computer Skill Test	2.01	0.08	2	3	2.01	0.10	2	3
Number of Previous Lab Studies	33.4	26.7	0	129	23.7	25.0	0	292
Period Level Variables								
Problems Solved	7.85	3.7	0	21	7.23	3.49	0	17
Problems Incorrect	0.06	0.29	0	4	0.08	0.30	0	3
Youtube Searches	0.16	0.59	0	6	0.24	0.67	0	5
Period Payment	0.60	0.65	−0.65	5.4	0.40	0.30	−0.1	2.1
High Piece Rate Indicator	0.17	0.37	0	1	0.08	0.28	0	1
Phone Access Indicator	0.17	0.37	0	1	0.09	0.29	0	1
Low Piece Rate Indicator	n.a.	n.a.	n.a.	n.a.	0.08	0.27	0	1
Number of Individuals	155				422			
Number of Treatment Periods	930				1266			

Notes: Computer Skill Test was a demographic variable collected by the Wharton Behavioral Lab prior to the experiment, however one with almost no variation. SAT Math score is missing for individuals who either took the ACT or otherwise did not wish to share that information with researchers. Indicators for treatments are presented under the period level variables – as experiment 1 had no "low piece rate" treatment, it has no such indicator.

Table 2A. Randomization Check – Experiment 1

Dependent Variable	Period # for			
	Piece Rate Treatment		Phone Treatment	
Female	−0.09	0.19	0.06	0.03
	(0.29)	(0.32)	(0.30)	(0.32)
SAT Math Score ('00s of points)		−0.002		−0.001
		(0.002)		(0.26)
PreTreatment Problems Solved		−0.047		0.011
		(0.058)		(0.048)
F-test	0.10	1.24	0.05	0.15
p value	0.75	0.30	0.83	0.93
Dependent Variable Mean	3.34	3.41	3.48	3.54
Number of Observations	930	738	930	738
Number of Individuals	155	123	155	123
Adj-R^2	0.001	0.018	0.001	0.005

Notes: Standard Errors (clustered at individual level) presented in parentheses above. As every subject in experiment 1 receives all treatments at some point, the dependent variable is the period in which they received the treatment in question. If randomization was done properly, the pre-treatment variables should not predict the period they received the treatment. Indeed, the F-stats are all large enough that I fail to reject the hypothesis that all coefficients are zero under $\alpha = 0.05$. Thus, I conclude the randomization was adequately done. SAT Math score is missing for 32 individuals who either took the ACT or otherwise did not wish to share that information with researchers.

Table 2B. Randomization Check – Experiment 2

Variable	Baseline	Piece Rate Decrease	Piece Rate Increase	Phone Access	
Female	0.68	0.71	0.74	0.71	$p < 0.76$
	(0.47)	(0.46)	(0.44)	(0.45)	(F-test = 0.39)
Age	20.26	20.13	20.73	20.39	$p < 0.11$
	(1.75)	(1.61)	(2.44)	(1.48)	(F-test = 2.05)
# Previous Studies at Lab	25.18	24.1	26.23	23.64	$p < 0.22$
	(23.9)	(20.5)	(34.5)	(18.7)	(F-test = 0.88)
Computer Skill Test	2.01	2.01	2.02	2.00	$p < 0.56$
	(0.01)	(0.09)	(0.14)	(no variation)	(F-test = 0.56)
Problems Solved in PreTreatment	20.57	20.23	19.54	20.23	$p < 0.80$
	(7.76)	(8.2)	(7.02)	(7.22)	(F-test = 0.33)
Number of Subjects Treated	103	114	104	101	

Notes: As every subject in experiment 2 receives (at most) one primary treatment, the subjects are split according to primary treatment. Means and standard deviations (in parentheses) are presented by primary treatment. If randomization was done properly, the pre-treatment variables should not differ significantly according to which treatment was received. Indeed, for all rows the F-stat corresponds to a p greater than 0.05 (fail to reject the hypothesis that all coefficients are less than zero under $\alpha = 0.05$).

Table 3A. Contemporaneous Piece Rate and Phone Access: Impact on Effort – Experiment 1

$$Problems_{i,t} \quad = \quad \alpha \cdot Piece\ Rate_{i,t} + \beta \cdot Phone\ Access_{i,t} + \gamma X_i + \epsilon_{i,t}$$

Dependent Variable Problems Solved	Specification (1)	(2)	(3)	(4)	(5)
Piece Rate	4.34***	4.62***	4.62***	2.07*	2.07*
(in cents per problem)	(1.17)	(1.15)	(1.14)	(1.09)	(1.09)
Phone Access	-0.38^{**}	-0.37^{**}	-0.39^{**}	-0.46^{**}	-0.46^{**}
	(0.19)	(0.19)	(0.19)	(0.19)	(0.19)
PreTreatment Quintiles		X	X	X	X
Period Fixed Effects			X	X	X
Session Fixed Effects				X	X
Individual Controls					X
Dependent Variable Mean	7.85	7.85	7.85	7.85	7.85
Number of Observations	930	930	930	930	930
Number of Individuals	155	155	155	155	155
Adj-R^2	0.01	0.23	0.23	0.31	0.32

Notes: The dependent variable is the number of problems solved correctly in a single period. All specifications report results from OLS regressions and also include a constant term. PreTreatment Quintiles represent five binary variables to non-parametrically control for the number of problems subject solved in the pre-treatment training period. Individual Controls include sex, age, ethnicity bins, number of sessions done, and WBL computer diagnostic scores. Standard errors are given in parentheses and clustered at the subject (individual) level. $* = p < 0.1$, $** = p < 0.05$, $*** = p < 0.01$.

Table 3B. Contemporaneous Piece Rate and Phone Access: Impact on Effort – Experiment 2

$$Problems_{i,t} \quad = \quad \alpha \cdot Piece\,Rate_{i,t} + \beta \cdot Phone\,Access_{i,t} + \gamma X_i + \epsilon_{i,t}$$

Dependent Variable:	Specification				
Problems Solved	(1)	(2)	(3)	(4)	(5)
Piece Rate	12.5***	14.9***	16.6***	16.7***	16.7***
(in cents per problem)	(3.39)	(2.67)	(3.02)	(2.98)	(2.98)
Phone Access	−0.05	−0.10	0.12	0.15	0.17
	(0.33)	(0.25)	(0.32)	(0.32)	(0.32)
PreTreatment Quintiles		X	X	X	X
Period Fixed Effects			X	X	X
Session Fixed Effects				X	X
Individual Controls					X
Dependent Variable Mean	7.23	7.23	7.23	7.23	7.23
Number of Observations	1266	1266	1266	1266	1260
Number of Individuals	422	422	422	422	420
Adj-R^2	0.01	0.41	0.41	0.43	0.45

Notes: The dependent variable is the number of problems solved correctly in a period. All specifications report results from OLS regressions and also include a constant term. PreTreatment Quintiles represent five binary variables to non-parametrically control for the number of problems subject solved in the pre-treatment training period. Individual Controls include sex, age, ethnicity bins, number of sessions done, and WBL computer diagnostic scores, but could not be matched for 2 subjects. Standard errors given in parentheses and clustered at the subject (individual) level. $* = p < 0.1$, $** = p < 0.05$, $* * * = p < 0.01$.

Table 4A. Previous Period Piece Rate and Phone Access: Impact on Effort – Experiment 1

$$Problems_{i,t} \quad = \quad \alpha_1 \cdot Piece\,Rate_{i,t} + \alpha_2 \cdot Piece\,Rate_{i,t-1} + \gamma X_i + \epsilon_{i,t}$$

Dependent Variable:	Specification					
Problems Solved	(1)	(2)	(3)	(4)	(5)	(6)
Piece Rate	5.00***	5.34***	2.19*	8.06***	7.74***	5.43***
(cents per problem)	(1.29)	(1.24)	(1.14)	(2.48)	(1.94)	(1.87)
Previous Period's Piece Rate	3.42*	3.86**	0.45	7.99**	8.04***	5.91**
(cents per problem)	(1.74)	(1.56)	(1.36)	(3.59)	(2.93)	(2.78)
Phone Access	−0.26	−0.25	−0.43**	−0.29	−0.30	−0.38
	(0.21)	(0.21)	(0.20)	(0.36)	(0.32)	(0.30)
Previous Period Phone Access	0.24	0.28	0.07	0.22	0.18	0.21
	(0.26)	(0.24)	(0.23)	(0.41)	(0.35)	(0.34)
PreTreatment Quintiles		X	X		X	X
Period Fixed Effects		X	X		X	X
Session Fixed Effects			X			X
Individual Controls			X			X
Periods 1 to 3 Only				X	X	X
Dependent Variable Mean	7.85	7.85	7.85	7.79	7.79	7.79
Number of Observations	930	930	930	465	465	465
Number of Individuals	155	155	155	155	155	155
Adj-R^2	0.01	0.24	0.32	0.04	0.34	0.39

Notes: The dependent variable is the number of problems solved correctly in a single period. All specifications report results from OLS regressions and also include a constant term. PreTreatment Quintiles represent five binary variables to non-parametrically control for the number of problems subject solved in the pre-treatment training period. "Periods 1 to 3" uses data of the first treatment period and following period to minimize treatment interactions. Individual Controls include sex, age, ethnicity bins, number of sessions done, and WBL computer diagnostic scores. Standard errors are given in parentheses and clustered at the subject (individual) level. $* = p < 0.1$, $** = p < 0.05$, $*** = p < 0.01$.

Table 4B. Previous Period Piece Rate and Phone Access: Impact on Effort – Experiment 2

$$Problems_{i,t} \quad = \quad \alpha_1 \cdot Piece\,Rate_{i,t} + \alpha_2 \cdot Piece\,Rate_{i,t-1} + \gamma X_i + \epsilon_{i,t}$$

Dependent Variable:	Specification				
Problems Solved	(1)	(2)	(3)	(4)	(5)
Piece Rate	12.66***	15.09***	16.57***	16.97***	17.09***
(cents per problem)	(3.51)	(2.74)	(3.03)	(3.05)	(3.11)
Previous Period's Piece Rate	4.09	6.51**	7.24**	7.64**	7.87**
(cents per problem)	(4.03)	(3.11)	(3.32)	(3.35)	(3.40)
Phone Access	−0.03	−0.08	0.12	0.16	0.18
	(0.36)	(0.28)	(0.32)	(0.32)	(0.33)
Previous Period Phone Access	−0.03	−0.08	0.02	0.06	0.07
	(0.41)	(0.33)	(0.36)	(0.36)	(0.36)
PreTreatment Quintiles		X	X	X	X
Period Fixed Effects			X	X	X
Session Fixed Effects				X	X
Individual Controls					X
Dependent Variable Mean	7.23	7.23	7.23	7.23	7.23
Number of Observations	1266	1266	1266	1266	1260
Number of Individuals	422	422	422	422	420
Adj-R^2	0.01	0.41	0.41	0.44	0.45

Notes: The dependent variable is the number of problems solved correctly in a period. All specifications report results from OLS regressions and also include a constant term. PreTreatment Quintiles represent five binary variables to non-parametrically control for the number of problems subject solved in the pre-treatment training period. Individual Controls include age, sex, ethnicity, computer skill test, and total # of experimental sessions done at the lab, but could not be matched for 2 subjects. Standard errors given in parentheses and clustered at the subject (individual) level. $* = p < 0.1$, $** = p < 0.05$, $*** = p < 0.01$.

Table 5A. Phone Access by Gender: Impact on Effort – Experiment 1

$$Problems_{i,t} \quad = \quad \alpha_1 \cdot Phone_{it} \cdot Female_i + \alpha_2 \cdot Phone_{it-1} \cdot Female_i + \beta_1 \cdot Phone_{it} \cdot Male_i + \beta_2 \cdot Phone_{it-1} \cdot Male_i + \gamma X_i + \epsilon_{i,t}$$

Dependent Variable:	Specification					
Problems Solved	(1)	(2)	(3)	(4)	(5)	(6)
Phone Access * Female	−0.26	−0.28	−0.29	0.08	−0.18	−0.02
	(0.22)	(0.22)	(0.22)	(0.41)	(0.35)	(0.32)
Previous Period Phone * Female	0.35	0.32	0.29	0.12	−0.39	−0.08
	(0.25)	(0.24)	(0.24)	(0.45)	(0.40)	(0.38)
Phone Access * Male	−1.10***	−1.10***	−1.11***	−2.33***	−1.78**	−2.02***
	(0.39)	(0.41)	(0.41)	(0.77)	(0.73)	(0.72)
Previous Period Phone * Male	−0.69	−0.59	−0.65	−0.42	0.42	0.18
	(0.57)	(0.55)	(0.55)	(1.05)	(0.69)	(0.71)
Male	−1.02*	−0.68	−0.10	−0.36	−0.28	0.24
	(0.58)	(0.58)	(0.56)	(0.61)	(0.50)	(0.54)
Pre-Treatment Quintiles		X	X		X	X
Period Fixed Effects		X	X		X	X
Session Fixed Effects			X			X
Individual Controls			X			X
Periods 1 to 3 Only				X	X	X
Dependent Variable Mean	7.85	7.85	7.85	7.79	7.79	7.79
Number of Observations	930	930	930	465	465	465
Number of Individuals	155	155	155	155	155	155
Adj-R^2	0.03	0.24	0.33	0.03	0.32	0.39

Notes: The dependent variable is the number of problems solved correctly in a single period. All specifications report results from OLS regressions and also include a constant term. PreTreatment Quintiles represent five binary variables to non-parametrically control for the number of problems subject solved in the pre-treatment training period. "Periods 1 to 3" uses data of the first treatment period and following period to minimize treatment interactions. Individual Controls include sex, age, ethnicity bins, number of sessions done, and WBL computer diagnostic scores. Standard errors are given in parentheses and clustered at the subject (individual) level. $* = p < 0.1$, $** = p < 0.05$, $*** = p < 0.01$.

Table 5B. Phone Access by Gender: Impact on Effort – Experiment 2

$$Problems_{i,t} \quad = \quad \alpha_1 \cdot Phone_{it} \cdot Female_i + \alpha_2 \cdot Phone_{it-1} \cdot Female_i + \beta_1 \cdot Phone_{it} \cdot Male_i + \beta_2 \cdot Phone_{it-1} \cdot Male_i + \gamma X_i + \epsilon_{i,t}$$

Dependent Variable:			Specification		
Problems Solved	(1)	(2)	(3)	(4)	(5)
Phone Access * Female	0.51	0.36	0.39	0.44	0.45
	(0.36)	(0.29)	(0.32)	(0.34)	(0.35)
Previous Period Phone * Female	0.48	0.34	0.42	0.47	0.48
	(0.44)	(0.36)	(0.39)	(0.40)	(0.40)
Phone Access * Male	−1.68**	−1.59***	−1.57***	−1.61***	−1.64***
	(0.82)	(0.51)	(0.54)	(0.51)	(0.53)
Previous Period Phone * Male	−1.61*	−1.52**	−1.43**	−1.48**	−1.51**
	(0.88)	(0.61)	(0.63)	(0.61)	(0.61)
Male	−0.45	−0.48	−0.48	−0.51*	−0.53*
	(0.36)	(0.27)	(0.27)	(0.28)	(0.28)
Pre-Treatment Quintiles		X	X	X	X
Period Fixed Effects			X	X	X
Session Fixed Effects				X	X
Individual Controls					X
Dependent Variable Mean	7.23	7.23	7.23	7.23	7.23
Number of Observations	1263	1263	1263	1263	1260
Number of Individuals	421	421	421	421	420
Adj-R^2	0.02	0.42	0.42	0.44	0.44

Notes: The dependent variable is the number of problems solved correctly in a period. All specifications report results from OLS regressions and also include a constant term. PreTreatment Quintiles represent five binary variables to non-parametrically control for the number of problems subject solved in the pre-treatment training period. Individual Controls include age, ethnicity, computer skill test, and total # of experimental sessions done at the lab. Gender could not be matched for one subject, and the controls for an additional subject. Standard errors given in parentheses and clustered at the subject (individual) level. $* = p < 0.1$, $** = p < 0.05$, $*** = p < 0.01$.

Table 6A. Next Period Piece Rate and Phone Access: Impact on Effort – Experiment 1

$$Problems_{i,t} \;=\; \alpha_1 \cdot Piece\,Rate_{i,t} + \alpha_2 \cdot Piece\,Rate_{i,t+1} \cdot Knowledge_{i,t} + \alpha_3 \cdot Piece\,Rate_{i,t-1} + \gamma X_i + \epsilon_{i,t}$$

Dependent Variable:	Specification					
Problems Solved	(1)	(2)	(3)	(4)	(5)	(6)
Piece Rate	5.33***	5.75***	2.78**	8.74***	8.62***	6.46***
(in cents)	(1.26)	(1.23)	(1.20)	(2.58)	(1.98)	(1.95)
Next Period Piece Rate	1.80	2.65	0.57	3.63*	5.01**	4.20*
(if known)	(1.73)	(2.23)	(2.35)	(2.01)	(2.14)	(2.47)
Previous Period Piece Rate	3.64**	3.90***	0.84	8.75**	8.49***	6.53**
	(1.62)	(1.45)	(1.33)	(3.73)	(2.97)	(1.62)
Pre-Treatment Quintiles		X	X		X	X
Period Fixed Effects		X	X		X	X
Shown Next Period Bin		X	X		X	X
Session Fixed Effects			X			X
Individual Controls			X			X
Periods 1 to 3 Only				X	X	X
Dependent Variable Mean	7.85	7.85	7.85	7.79	7.79	7.79
Number of Observations	930	930	930	465	465	465
Number of Individuals	155	155	155	155	155	155
Adj-R^2	0.01	0.23	0.32	0.04	0.34	0.39

Notes: The dependent variable is the number of problems solved correctly in a single period. All specifications report results from OLS regressions and also include a constant term. PreTreatment Quintiles represent five binary variables to non-parametrically control for the number of problems subject solved in the pre-treatment training period. "Only Periods 1 to 3" uses data of the first treatment period and following period to minimize treatment interactions. Individual Controls include age, sex, ethnicity, computer skill test, and total # of experimental sessions done at the lab. Standard errors are given in parentheses and clustered at the subject (individual) level. $* = p < 0.1$, $** = p < 0.05$, $*** = p < 0.01$.

Table 6B. Next Period Piece Rate and Phone Access: Impact on Effort – Experiment 2

$$Problems_{i,t} \quad = \quad \alpha_1 \cdot Piece\,Rate_{i,t} + \alpha_2 \cdot Piece\,Rate_{i,t+1} \cdot Knowledge_{i,t} + \alpha_3 \cdot Piece\,Rate_{i,t-1} + \gamma X_i + \epsilon_{i,t}$$

Dependent Variable:			Specification		
Problems Solved	(1)	(2)	(3)	(4)	(5)
Piece Rate	12.8***	15.5***	16.3***	16.65***	16.76***
(in cents)	(3.58)	(2.79)	(2.92)	(2.96)	(3.04)
Next Period Piece Rate	1.09	3.23	0.73	1.59	2.63
(if known)	(2.64)	(2.19)	(2.62)	(3.92)	(3.88)
Previous Piece Rate	4.21	6.88**	7.20**	7.52**	7.75**
	(4.12)	(3.17)	(3.27)	(3.31)	(3.38)
Pre-Treatment Quintiles		X	X	X	X
Period Fixed Effects			X	X	X
Shown Next Period Bin				X	X
Session Fixed Effects				X	X
Individual Controls					X
Dependent Variable Mean	7.23	7.23	7.23	7.23	7.23
Number of Observations	1266	1266	1266	1266	1260
Number of Individuals	422	422	422	422	420
Adj-R^2	0.01	0.41	0.41	0.44	0.44

Notes: The dependent variable is the number of problems solved correctly in a period. All specifications report results from OLS regressions and also include a constant term. PreTreatment Quintiles represent five binary variables to non-parametrically control for the number of problems subject solved in the pre-treatment training period. Individual Controls include age, sex, ethnicity, computer skill test, and total # of experimental sessions done at the lab, but could not be matched for 2 subjects. Standard errors given in parentheses and clustered at the subject (individual) level. $* = p < 0.1$, $** = p < 0.05$, $*** = p < 0.01$.

Table 7. Previous Effort Instrumental Variable: Impact on Effort – Experiment 2

$$
\begin{aligned}
Problems_{i,t-1} &= \alpha_1 \cdot Piece\,Rate_{i,t-1} + \beta_1 \cdot Phone_{i,t-1} + \gamma X_i + \epsilon_{i,t} \\
Problems_{i,t} &= \rho \cdot Problems_{i,t-1} + \alpha_2 \cdot Piece\,Rate_{i,t} + \beta_2 \cdot Phone_{i,t} + \zeta X_i + \nu_{i,t}
\end{aligned}
$$

Dependent Variable:			Specification		
Problems Solved	(1)	(2)	(3)	(4)	(5)
Problems Previous Period	0.39	0.50***	0.45***	0.44***	0.43***
	(0.26)	(0.18)	(0.17)	(0.17)	(0.17)
Piece Rate	12.7***	13.9***	15.3***	15.2***	14.9***
	(2.63)	(2.51)	(2.73)	(2.81)	(2.88)
Phone Access	0.04	0.07	0.24	0.25	0.23
	(0.24)	(0.23)	(0.27)	(0.27)	(0.27)
First Stage F Stat (IV)	6.1	14.5	15.5	16.5	16.2
PreTreatment Quintiles		X	X	X	X
Period Fixed Effects			X	X	X
Session Fixed Effects				X	X
Individual Controls					X
Dependent Variable Mean	7.22	7.22	7.22	7.22	7.22
Number of Observations	844	844	844	844	840
Number of Individuals	422	422	422	422	420
Adj-R^2	0.39	0.56	0.57	0.57	0.57

Notes: The dependent variable is the number of problems solved correctly in a single period. All specifications report results from linear Instrumental Variable regressions estimated by (iterative) GMM and also include a constant term. PreTreatment Quintiles represent five binary variables to non-parametrically control for the number of problems subject solved in the pre-treatment training period. Individual Controls include age, sex, ethnicity, computer skill test, and total # of experimental sessions done at the lab, but could not be matched for 2 subjects. Standard errors are given in parentheses and clustered at the subject (individual) level. $* = p < 0.1$, $** = p < 0.05$, $*** = p < 0.01$.

Table 8. Period or Total Earnings Salience: Impact on Earnings – Experiment 2

$$Earnings_{i,t} = \alpha_1 \cdot Piece\,Rate_{i,t} + \alpha_2 Piece\,Rate_{i,t} \cdot RS_i + \beta_1 \cdot Phone_{i,t} + \beta_2 Phone_{i,t} \cdot RS_i + \gamma Round\,Shown_i + \gamma X_i + \epsilon_{i,t}$$

Dependent Variable:	Specification				
Problems Solved	(1)	(2)	(3)	(4)	(5)
Piece Rate	9.32*	13.12***	17.05***	15.84***	15.68***
	(5.27)	(3.83)	(4.17)	(3.87)	(3.88)
Piece Rate * Period Salience	2.41	−3.82	-2.86	−1.01	−1.00
	(6.93)	(5.21)	(5.33)	(5.11)	(5.13)
Phone Access	0.92*	0.62*	0.94**	0.92**	0.88**
	(0.50)	(0.35)	(0.41)	(0.46)	(0.46)
Phone Access * Period Salience	−2.03***	−1.20**	−1.21**	−1.23**	−1.15**
	(0.73)	(0.53)	(0.54)	(0.56)	(0.54)
Period Salience	-0.32	0.34	0.40	0.22	0.25
	(0.58)	(0.40)	(0.39)	(0.39)	(0.39)
Pre-Treatment Quintiles		X	X	X	X
Period Fixed Effects			X	X	X
Session Fixed Effects				X	X
Individual Controls					X
Dependent Variable Mean	7.28	7.28	7.28	7.28	7.28
Number of Observations	894	894	894	894	891
Number of Individuals	298	298	298	298	297
Adj-R^2	0.02	0.45	0.48	0.51	0.52

Notes: The dependent variable is the total earnings from a single period. All specifications report results from OLS regressions and also include a constant term. The subject is shown either the previous period's earnings (as indicated by "Period Salience") or shown total earnings up to that period. Experiment 2 was the only one that featured this variation. Unfortunately, while every subject in Experiment 2 did face a randomized period or total counter, a small programming typo prevented the capture of this variable for the first day. As it is unclear which counter subjects faced on the first day, they are dropped from analysis above. PreTreatment Quintiles represent five binary variables to non-parametrically control for the number of problems subject solved in the pre-treatment training period. Individual Controls include age, sex, ethnicity, computer skill test, and total # of experimental sessions done at the lab, but could not be matched for one subject. Standard errors given in parentheses and clustered at the subject (individual) level. $* = p < 0.1$, $** = p < 0.05$, $*** = p < 0.01$.

10 Appendix

10.1 Appendix Tables

Appendix Table 1. Previous Effort Instrumental Variable: Impact on Effort – Replication with Sliders

$$
\begin{aligned}
Problems_{i,t-1} &= \alpha_1 \cdot Piece\,Rate_{i,t-1} + \gamma X_i + \epsilon_{i,t} \\
Problems_{i,t} &= \rho \cdot Problems_{i,t-1} + \alpha_2 \cdot Piece\,Rate_{i,t} + \zeta X_i + \nu_{i,t}
\end{aligned}
$$

Dependent Variable:			Specification		
Sliders Moved Correctly	(1)	(2)	(3)	(4)	(5)
Sliders Previous Period	0.35	0.57***	0.46**	0.42**	0.42**
(aka ρ parameter)	(0.39)	(0.19)	(0.18)	(0.21)	(0.21)
Piece Rate	41.73***	48.27***	60.26***	58.21***	57.51***
(in cents per 100 sliders)	(11.80)	(12.17)	(13.39)	(13.51)	(13.72)
First Stage F Stat (IV)	5.3	18.9	20.2	16.6	15.1
PreTreatment Quintiles		X	X	X	X
Period Fixed Effects			X	X	X
Session Fixed Effects				X	X
Individual Controls					X
Dependent Variable Mean	52.1	52.1	52.1	52.1	52.1
Number of Observations	552	552	552	552	549
Number of Individuals	184	184	184	184	183
Adj-R^2	0.45	0.66	0.63	0.64	0.65

Notes: The above table represents an additional experiment that replicated the main findings using a different (slider) task. The dependent variable is the number of sliders correctly moved to 50% in a single period. See more details in Appendix section 10.3. All specifications report results from linear Instrumental Variable regressions estimated by (iterative) GMM and also include a constant term. PreTreatment Quintiles represent five binary variables to non-parametrically control for the number of problems subject solved in the pre-treatment training period. Individual Controls include age, sex, ethnicity, computer skill test, and total # of experimental sessions done at the lab, but could not be matched for 1 subject. The First Stage F Statistic is from the instrument alone (F(1, 183)). Standard errors are given in parentheses and clustered at the subject (individual) level. $* = p < 0.1$, $** = p < 0.05$, $*** = p < 0.01$.

Appendix Table 2. Previous Effort Instrumental Variable: Impact on Effort – Experiment 1

$$
\begin{aligned}
Problems_{i,t-1} &= \alpha_1 \cdot Piece\,Rate_{i,t-1} + \beta_1 \cdot Phone_{i,t-1} + \gamma X_i + \epsilon_{i,t} \\
Problems_{i,t} &= \rho \cdot Problems_{i,t-1} + \alpha_2 \cdot Piece\,Rate_{i,t} + \beta_2 \cdot Phone_{i,t} + \zeta X_i + \nu_{i,t}
\end{aligned}
$$

Dependent Variable:	Specification					
Problems Solved	(1)	(2)	(3)	(4)	(5)	(6)
Problems Previous Period	0.49**	0.48*	−0.02	0.41	0.71*	0.47
	(0.25)	(0.26)	(0.17)	(0.34)	(0.37)	(0.40)
Piece Rate	4.50***	4.51***	2.03	7.41***	7.02***	6.13***
	(1.31)	(1.29)	(2.88)	(2.41)	(2.54)	(2.30)
Phone Access	−0.42**	−0.40**	0.23	0.14	0.09	0.17
	(0.19)	(0.18)	(0.27)	(0.34)	(0.34)	(0.33)
First Stage F Stat (IV)	9.8	9.5	4.6	2.8	3.6	2.8
PreTreatment Quintiles		X	X		X	X
Period Fixed Effects		X	X		X	X
Session Fixed Effects			X			X
Individual Controls			X			X
Period 1 and 2 Only				X	X	X
Dependent Variable Mean	7.85	7.85	7.85	7.79	7.79	7.79
Number of Observations	930	930	930	465	465	465
Number of Individuals	155	155	155	155	155	155
Adj-R^2	0.01	0.31	0.37	0.02	0.66	0.66

Notes: The dependent variable is the number of problems solved correctly in a single period. All specifications report results from linear Instrumental Variable regressions estimated by (iterative) GMM and also include a constant term. PreTreatment Quintiles represent five binary variables to non-parametrically control for the number of problems subject solved in the pre-treatment training period. Individual Controls include age, sex, ethnicity, computer skill test, and total # of experimental sessions done at the lab, but could not be matched for 2 subjects. Standard errors are given in parentheses and clustered at the subject (individual) level. $* = p < 0.1$, $** = p < 0.05$, $*** = p < 0.01$.

Appendix Table 3. Previous Period Piece Rate and Phone Access: Random Effects – Experiment 2

$$Problems_{i,t} = \alpha_1 \cdot Piece\,Rate_{i,t} + \alpha_2 \cdot Piece\,Rate_{i,t-1} + \gamma X_i + \epsilon_{i,t}$$

Dependent Variable:			Specification		
Problems Solved	(1)	(2)	(3)	(4)	(5)
Piece Rate	11.05***	12.58***	14.95***	15.07***	15.00***
	(2.40)	(2.29)	(2.67)	(2.71)	(2.78)
Previous Period's Piece Rate	2.47	4.00	5.62**	5.73**	5.78**
	(2.68)	(2.55)	(2.80)	(2.83)	(2.89)
Phone Access	−0.05	−0.08	0.25	0.27	0.28
	(0.22)	(0.22)	(0.27)	(0.27)	(0.27)
Previous Period Phone Access	−0.05	−0.08	0.15	0.17	0.17
	(0.27)	(0.27)	(0.31)	(0.31)	(0.31)
Random Effects	X	X	X	X	X
PreTreatment Quintiles		X	X	X	X
Period Fixed Effects			X	X	X
Session Fixed Effects				X	X
Individual Controls					X
Number of Observations	1266	1266	1266	1266	1260
Number of Individuals	422	422	422	422	420
Adj-R^2	0.01	0.41	0.41	0.44	0.45

Notes: The dependent variable is the number of problems solved correctly in a single period. All specifications report results from OLS regressions and also include a constant term. PreTreatment Quintiles represent five binary variables to non-parametrically control for the number of problems subject solved in the pre-treatment training period. Individual Controls include age, sex, ethnicity, computer skill test, and total # of experimental sessions done at the lab. Standard errors given in parentheses and clustered at the subject (individual) level. $* = p < 0.1$, $** = p < 0.05$, $*** = p < 0.01$.

Appendix Table 4. Previous Period Piece Rate and Phone Access: Fixed Effects – Experiment 2

$$Problems_{i,t} \quad = \quad \alpha_1 \cdot Piece\,Rate_{i,t} + \alpha_2 \cdot Piece\,Rate_{i,t-1} + \gamma X_i + \epsilon_{i,t}$$

Dependent Variable:	Specification		
Problems Solved	(1)	(2)	(3)
Piece Rate	12.66***	10.46***	13.49***
	(3.51)	(2.54)	(3.00)
Previous Period's Piece Rate	4.09	1.89	4.16
	(4.03)	(2.63)	(2.94)
Phone Access	−0.03	−0.05	0.36
	(0.36)	(0.22)	(0.28)
Previous Period Phone Access	−0.03	−0.05	0.26
	(0.41)	(0.26)	(0.31)
Subject Fixed Effects		X	X
Period Fixed Effects			X
Dependent Variable Mean	7.23	7.23	7.23
Number of Observations	1266	1266	1266
Number of Individuals	422	422	422
Adj-R^2	0.01	0.72	0.72

Notes: In the presence of momentum, these estimates are **severely biased** downward. They are presented here only in the spirit of openness but are not intended to be taken as accurate estimates. The dependent variable is the number of problems solved correctly in a period. All specifications report results from OLS regressions and also include a constant term. PreTreatment Quintiles represent five binary variables to non-parametrically control for the number of problems subject solved in the pre-treatment training period. Note that session fixed effects and individual controls cannot be estimated with subject fixed effects as these variables do not vary within individual. Standard errors given in parentheses and clustered at the subject (individual) level. $* = p < 0.1$, $** = p < 0.05$, $*** = p < 0.01$.

Appendix Table 5. Previous Period Piece Rate and Phone Access: Impact on YouTube Searches – Experiment 2

$$Youtube_{i,t} \quad = \quad \alpha_1 \cdot Piece\,Rate_{i,t} + \alpha_2 \cdot Piece\,Rate_{i,t-1} + \gamma X_i + \epsilon_{i,t}$$

Dependent Variable:	Specification				
Youtube Searches	(1)	(2)	(3)	(4)	(5)
Piece Rate	−2.39***	−2.52***	−2.80***	−2.68***	−2.59***
(in cents)	(0.42)	(0.44)	(0.56)	(0.63)	(0.65)
Previous Period's Piece Rate	−1.01*	−1.15**	−0.94	−0.83	−0.72
	(0.56)	(0.55)	(0.58)	(0.66)	(0.68)
Phone Access	−0.02	−0.02	−0.06	−0.03	−0.04
	(0.07)	(0.07)	(0.08)	(0.08)	(0.08)
Previous Period Phone Access	0.09	0.09	0.12	0.14	0.14
	(0.09)	(0.08)	(0.09)	(0.09)	(0.09)
PreTreatment Quintiles		X	X	X	X
Period Fixed Effects			X	X	X
Session Fixed Effects				X	X
Individual Controls					X
Dependent Variable Mean	0.24	0.24	0.24	0.24	0.24
Number of Observations	1266	1266	1266	1266	1260
Number of Individuals	422	422	422	422	420
Adj-R^2	0.01	0.02	0.02	0.04	0.07

Notes: The dependent variable is the number of YouTube videos searched in a period. All specifications report results from OLS regressions and also include a constant term. PreTreatment Quintiles represent five binary variables to non-parametrically control for the number of problems subject solved in the pre-treatment training period. Individual Controls include age, sex, ethnicity, computer skill test, and total # of experimental sessions done at the lab, but could not be matched for 2 subjects. Standard errors given in parentheses and clustered at the subject (individual) level. $* = p < 0.1$, $** = p < 0.05$, $*** = p < 0.01$.

Appendix Table 6. Phone Usage: Relationship with Effort – Experiment 2

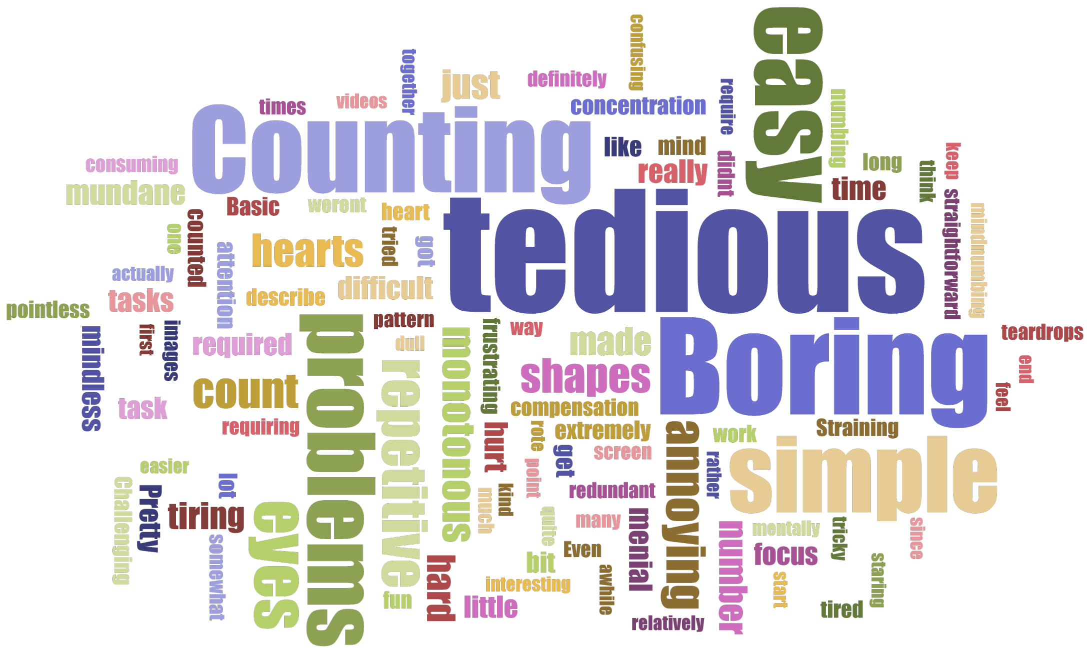
$$Problems_{i,t} \quad = \quad \alpha_1 \cdot Phone_{it} + \alpha_2 \cdot Phone_{it-1} + \beta_1 \cdot PieceRate_{it} + \beta_2 \cdot PieceRate_{it-1} + \gamma X_i + \epsilon_{i,t}$$

Dependent Variable:			Specification		
Problems Solved	(1)	(2)	(3)	(4)	(5)
Phone Usage	−3.81***	−2.19***	−2.06***	−2.11***	−1.93***
	(0.88)	(0.71)	(0.72)	(0.73)	(0.73)
Previous Period Phone Usage	−4.73***	−3.14***	−3.11***	−3.16***	−2.99***
	(0.92)	(0.65)	(0.67)	(0.69)	(0.70)
Piece Rate	12.41***	14.95***	15.99***	15.96***	16.13***
	(3.51)	(2.76)	(2.96)	(2.95)	(3.01)
Previous Period Piece Rate	3.95	6.50**	6.70**	6.67**	6.96**
	(3.93)	(3.05)	(3.23)	(3.24)	(3.29)
Pre-Treatment Quintiles		X	X	X	X
Period Fixed Effects			X	X	X
Session Fixed Effects				X	X
Individual Controls					X
Dependent Variable Mean	7.23	7.23	7.23	7.23	7.23
Number of Observations	1266	1266	1266	1266	1260
Number of Individuals	422	422	422	422	420
Adj-R^2	0.04	0.42	0.43	0.45	0.44

Notes: The dependent variable is the number of problems solved correctly in a period. Phone Usage is a self reported variable indicating use of the phone during period 3. As this is endogenously chosen, these regressions should not be taken as causal, as a subject who uses the phone may have unobservable differences. All specifications report results from OLS regressions and also include a constant term. PreTreatment Quintiles represent five binary variables to non-parametrically control for the number of problems subject solved in the pre-treatment training period. Individual Controls include age, ethnicity, computer skill test, and total # of experimental sessions done at the lab. Gender could not be matched for one subject, and the controls for an additional subject. Standard errors given in parentheses and clustered at the subject (individual) level. $* = p < 0.1$, $** = p < 0.05$, $*** = p < 0.01$.

10.2 Appendix Figures

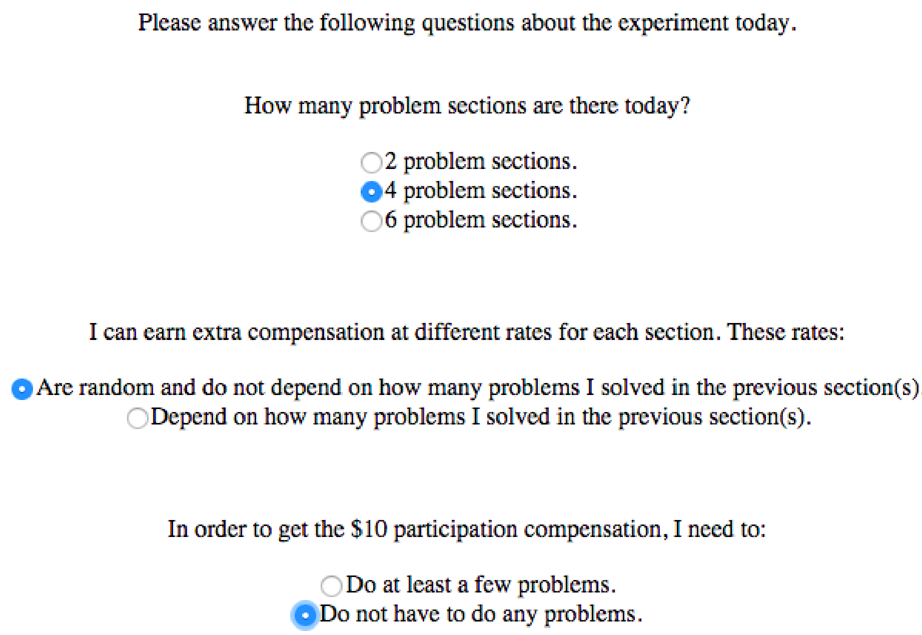
Appendix Figure 1. Word Cloud for Survey – "Opinion of Task"



Notes: Top 100 words from responses to a post experiment survey question asking "What is your opinion of the task?" Size scaled linearly with count.

Source: Jasondavies.com word cloud generator.

Appendix Figure 2. Quiz for Introduction Instructions



Please answer the following questions about the experiment today.

How many problem sections are there today?

2 problem sections.
4 problem sections.
6 problem sections.

I can earn extra compensation at different rates for each section. These rates:

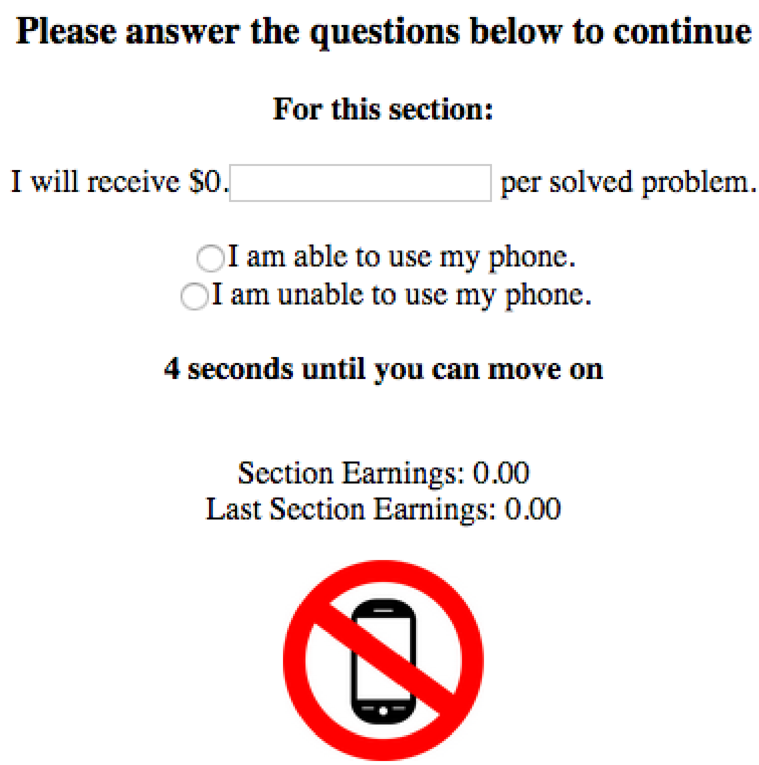
Are random and do not depend on how many problems I solved in the previous section(s).
Depend on how many problems I solved in the previous section(s).

In order to get the $10 participation compensation, I need to:

Do at least a few problems.
Do not have to do any problems.

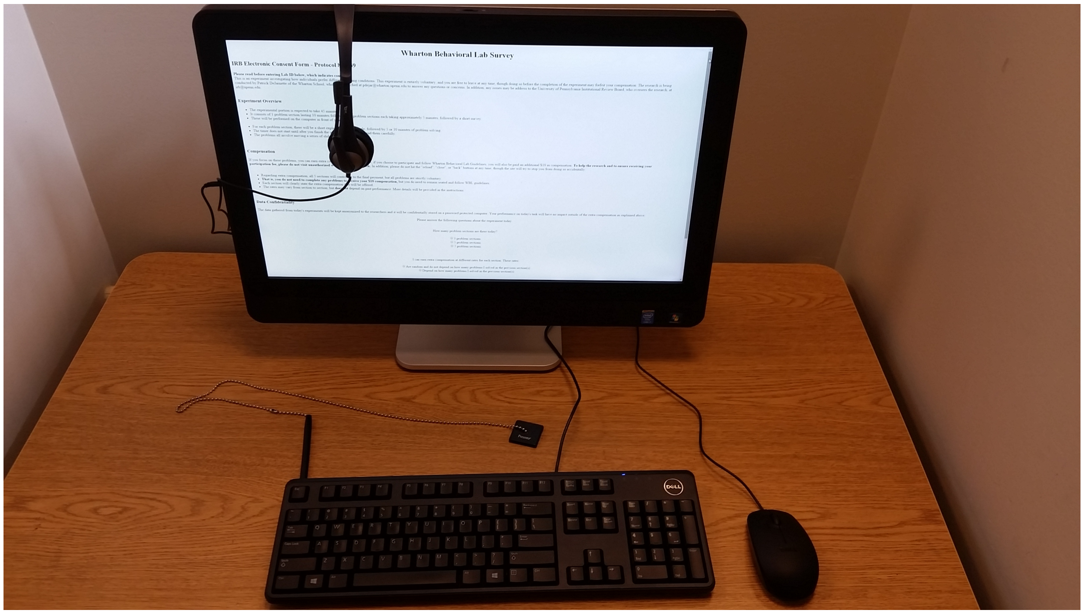
Notes: Every participant in experiment 2 had to answer the above questions after reading experiment instructions. Subjects had to answer all three questions correctly to proceed. If the subject entered the wrong answers, the browser would alert them to this and ask for them to review the instructions again.

Appendix Figure 3. Quiz for Instructions Prior to Each Period



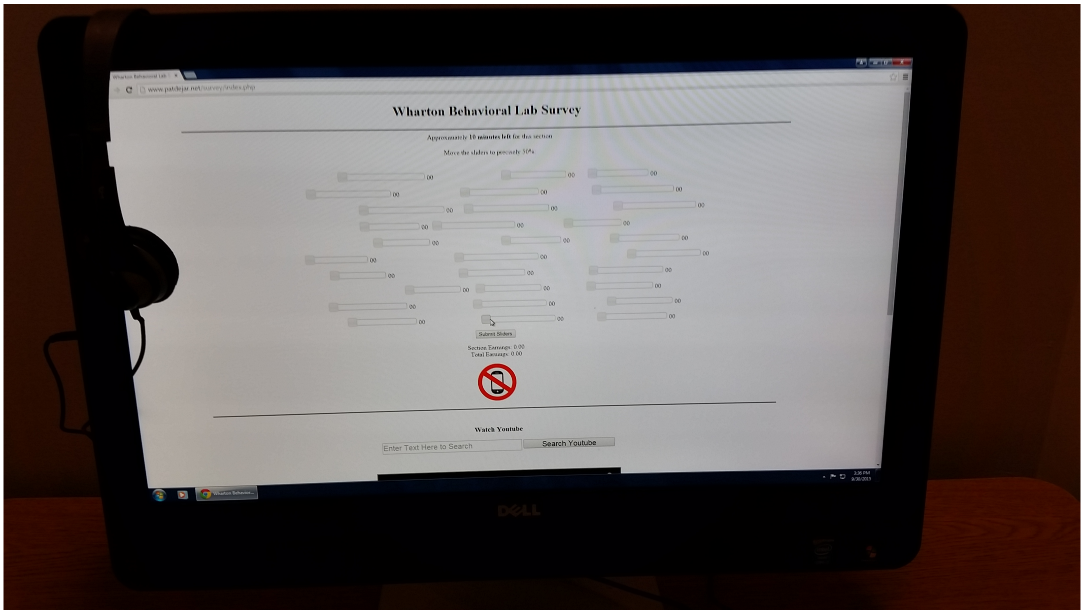
Notes: Every participant had to answer the following questions prior to every period (including Pre-Treatment). If the subject had information about future periods, they were also quizzed on the piece rate and phone access for future periods. If the subject entered the wrong answers, the browser would alert them to this and ask for them to review the instructions again.

Appendix Figure 4. Cubicle Environment



Notes: Every participant had access to an identical computer with headphones as pictured above. Screen brightness was uniformly set at 95% to ensure consistency across cubicles. It was not possible to see other subjects from within the cubicle. Google Chrome was employed as the browser during the task, with the window maximized (full screen mode). All instructions were written, but RAs were on site to answer any additional questions.

Appendix Figure 5. Slider Task Example



Notes: Slider Task employed as a replication of momentum effects. As can seen above, it would be difficult to view YouTube and move sliders at the same time on the monitor and resolution employed. Furthermore, the website code blocked any attempt to "zoom" in or out.

10.3 Slider Task – Experimental Replication

In addition to replicating findings of the first experiment with Experiment 2 (details in main text), I also ran an additional experiment with a different task to serve as an additional replication. In this setting, subjects face 30 "sliders" on a screen, as in Gill and Prowse [2011]. This can be seen in Appendix Figure 5 above. The subjects are asked to move the slider to exactly 50% of the way, with a numerical setting next to the slider indicating the current %. The exact position and length of sliders was randomized to make the task more difficult.

In this setting, as the sliders take far less time than the counting problems, the piece rate was also reduced. Subjects received a baseline of \$0.01 per 10 sliders. The "high piece rate" treatment was \$0.03 per 10 sliders (three times baseline), while the "low piece rate" treatment was \$0.0033 per 10 sliders (one third of baseline). Subjects were rounded to the nearest cent in the case they were unable to finish before time ran out. In addition, due to the small subject size available and the relatively "noisy" effect of cellphones in previous experiments, there was no phone access treatment. This affords us greater power in detecting effects through the financial incentives, but does make it harder to distinguish some theories ruled out in experiment 2.

In order to better understand the underlying model of momentum, the pre-treatment period was reduced 5 minutes and an additional 5 minute period following treatment was added. This allows one to investigate the rate of decay of effort inducement.

As predicted by a model of momentum, these gains from working harder continue to decay exponentially as time goes on. As the design and results are otherwise similar to the main findings of the paper, I have omitted most tables for brevity. See Appendix Table 1 or contact the author for additional details.

10.4 Proofs

10.4.1 Proof of Proposition 1

This proof will be using methods of supermodularity discussed in Milgrom and Shannon (1995).[68] For simplicity, I retain the assumption about the utility function being twice differentiable over c_t, e_t, e_{t-1}, however this assumption could be weakened as long as the utility function maintains increasing differences.

MS Theorem 6a: I begin by applying Theorem 6 of Milgrom and Shannon (1995), which first states that a twice differentiable function $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ has increasing differences in (x, z) if and only if $\frac{\partial f^2}{\partial x_i \partial z_j} \geq 0$ for all $i = 1, ..., n$ and $j = 1, ..., m$.

MS Theorem 6a Conditions: In this case, $x_1 \equiv c_1$, $x_2 \equiv c_2$,..., $x_T \equiv c_T$, and $x_{T+1} \equiv e_1$, $x_{T+2} \equiv e_2$,..., $x_{2T} \equiv e_T$. y are the non-choice state variables, $z_1 \equiv w_1$, $z_2 \equiv w_2$, ..., $z_T \equiv w_T$, and $z_{T+1} \equiv -\gamma_1$, $z_{T+1} \equiv -\gamma_2$, ...$z_{2T} \equiv -\gamma_T$, and f is the full Lagrangian (treating y_t, r, and p_t as fixed):

$$f(x, z) = \sum_{t=1}^{T} \delta^{t-1} u(c_t,\, e_t,\, e_{t-1},\, \gamma_t) + \lambda \left(\sum_{t=1}^{T} (w_t e_t + y_t)(1 + r)^{-t} - \sum_{t=1}^{T} p_t c_t (1 + r)^{-t} \right)$$

With variables redefined this way, we can check the conditions. For consumption and wage, it is clear that $\frac{\partial f^2}{\partial c_i \partial w_j} = 0$ as w_j does not enter the utility function and the budget constraint is linear with no term containing both c_i and w_j. For consumption and leisure technology, $\frac{\partial f^2}{\partial c_i \partial \gamma_j} = 0$ if $i \neq j$. However, without an assumption on $\frac{\partial f^2}{\partial c_t \partial \gamma_t}$, it is possible

[68] An earlier draft of the proof uses the (Dini) Multivariate Implicit Function Theorem, but supermodularity allows for fewer restrictions and also removes the need for matrix manipulation in calculating determinants.

that increased leisure technology increases the utility of consumption so much that effort rises in every period (including time period t) to satisfy the greater demand for consumption goods. However, while leisure and consumption might have complementarities (and hence effort and consumption exhibit substitutibility), leisure technology itself is constructed to not influence the utility from consumption.[69]

For effort and wage, $\frac{\partial f^2}{\partial e_i \partial w_j} = 0$ if $i \neq j$ as wage and effort only appear together in the same time period. In this case $\frac{\partial f^2}{\partial e_t \partial w_t} = \frac{\partial f^2}{\partial e_t \partial w_t} = \lambda(1+r)^{-t}$, which is ≥ 0 as consumption c_t is enjoyable and $r > -1$.

For effort and leisure technology, $\frac{\partial f^2}{\partial e_i \partial \gamma_j} = 0$ if $i > j$ or $i + 1 < j$ as e_t only appears in two $u(\cdot)$ functions, at time t and time $t+1$. Thus, the only cases we need to check are on $\frac{\partial f^2}{\partial e_i \partial (-\gamma_i)}$ and $\frac{\partial f^2}{\partial e_i \partial (-\gamma_{i+1})}$, which are equivalent to $-\delta^{i-1}\frac{\partial u^2}{\partial e_i \partial \gamma_i}$ and $-\delta^{i}\frac{\partial u^2}{\partial e_i \partial \gamma_{i+1}}$, respectively. From the assumptions above, $\frac{\partial u^2}{\partial e_i \partial \gamma_i} \leq 0$, as leisure technology makes marginal contemporaneous effort more costly in utility terms, and $\frac{\partial u^2}{\partial e_i \partial \gamma_{i+1}} = 0$, as leisure technology does not carry over across periods. Thus, both $\frac{\partial f^2}{\partial e_i \partial (-\gamma_i)}$ and $\frac{\partial f^2}{\partial e_i \partial (-\gamma_{i+1})}$ are ≥ 0. And in conclusion, f has increasing differences in $(x,\, z)$.

MS Theorem 6b: I apply the second half of Theorem 6 of Milgrom and Shannon (1995), which states that a twice differentiable function $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is supermodular in x if and only if $\frac{\partial f^2}{\partial x_i \partial x_j} \geq 0$ for all $i \neq j$ in $1, ..., n$.

MS Theorem 6b Conditions: As before, $x_1 \equiv c_1$, $x_2 \equiv c_2$,..., $x_T \equiv c_T$, and $x_{T+1} \equiv e_1$, $x_{T+2} \equiv e_2$, ..., $x_{2T} \equiv e_T$, and f is the full Lagrangian:

$$f(x,z) \quad = \quad \sum_{t=1}^{T} \delta^{t-1} u(c_t,\, e_t,\, e_{t-1},\, \gamma_t) + \lambda \left(\sum_{t=1}^{T} (w_t e_t + y_t)(1+r)^{-t} - \sum_{t=1}^{T} p_t c_t (1+r)^{-t} \right)$$

Note that $\frac{\partial f^2}{\partial c_i \partial c_j} = 0$ for $i \neq j$ as there is no overlap in the additively separable terms. Likewise $\frac{\partial f^2}{\partial e_i \partial e_j} = 0$ if $i > j$ or $i + 1 < j$ as only last period's effort influences this period's effort.[70] Thus, for time period t we are concerned with two terms: first, $\frac{\partial f^2}{\partial e_t \partial e_{t-1}} = \delta^{t-1}\frac{\partial u^2}{\partial e_t \partial e_{t-1}} \geq 0$, as we have positive momentum (last period effort makes this period's effort marginally less costly in terms of utility). Second, $\frac{\partial f^2}{\partial e_{t+1} \partial e_t} = \delta^{t}\frac{\partial u^2}{\partial e_{t+1} \partial e_t} \geq 0$ for the same reasons.

The only remaining terms of interest are the cross-partials between consumption and effort. Similar to above, $\frac{\partial f^2}{\partial e_i \partial c_j} = 0$ if $i > j$ or $i+1 < j$ as the overlap only occurs for a c_t and e_t or c_t and e_{t-1}, as per the utility function. First, $\frac{\partial f^2}{\partial c_t \partial e_t} = \delta^{t-1}\frac{\partial u^2}{\partial c_t \partial e_t} \geq 0$, as assumed in the set up. This implies that consumption and effort are not complements. By a similar assumption $\frac{\partial f^2}{\partial c_t \partial e_{t-1}} = \delta^{t-1}\frac{\partial u^2}{\partial c_t \partial e_{t-1}} \geq 0$ as last period's effort should have no negative effect on this period's consumption. These assumptions, while not trivial, ensure the utility function is reasonably well behaved – otherwise if last period's effort greatly reduced the demand for consumption, it could theoretically reduce this period's effort as well as the demand for consumption has decreased so dramatically.[71] Thus, in conclusion, the conditions are satisfied, and f is supermodular in x.

MS Theorem 4: Using these results from Theorem 6, apply Theorem 4 of Milgrom and Shannon (1995) to achieve the main result. The theorem states that if $f : X \times Z \to \mathbb{R}$, where X is a lattice, T is a partially ordered set, and $S \subset X$, then $argmax_{x \in S} f(x,z)$ is a monotone nondecreasing function in (z, S) if and only if f is quasisupermodular in x and

[69] E.g. Popcorn is enjoyable while watching the new Star Wars movie; the new Star Wars movie does not make popcorn itself taste better on December 18th if you did not get to watch it.

[70] Though this assumption of only last period's effort influencing this period is not necessary for this proof. For example, if all previous periods' efforts entered the utitlity function as a discounted sum with non-negative weights, as long as effort is positive with respect to that sum (positive effort momentum), the result would be the same.

[71] For example, if after working a long day, the agent no longer cared for consumption. Under this example, the agent might call in sick, even though effort would have been easier due to effort momentum. In this odd model of behavior, an increase in piece rate last period could decrease effort in the next period.

satisfies the single crossing property in $(x; z)$.

MS Theorem 4 Conditions: First note that $\mathbb{R}^n$ with component-wise order forms a lattice as for $\forall x,\, y \in \mathbb{R}^n$, $x \wedge y$ and $x \vee y$ are both in $\mathbb{R}^n$. By the same token, $\mathbb{R}^m$ with component-wise ordering is a partially ordered set. Thus, using the lagrangian function above as f, with x and z defined as above, we have already established supermodularity in x, which implies quasisupermodularity in x. In addition, as f has increasing differences, it satisfies the single crossing property. Thus it follows that $\vec{c}^*,\, \vec{e}^* \in argmax_{c,e \in \mathbb{R}_+^{2T}} f(c,e,z)$ are monotone decreasing functions over $(z,\, \mathbb{R}_+^n)$ – but recall that z_{T+1} was defined as negative γ_1 , z_{T+2} as negative γ_2, and so on. Thus, consumption and effort are monotonically non-decreasing over piece rate vector $\vec{w}$ and monotonically non-increasing over leisure technology vector $\vec{\gamma}$.[72]

Alternate Proofs

It's possible to relax the assumptions of the model. The differentiability of u is unnecessary as long as the conditions of increasing differences / single crossing condition and quasi-supermodularity are satisfied. However, I felt the assumptions above are more familiar with readers compared to assumptions of increasing differences. In addition, the assumption that only last period enters the utility function is unnecessary.

It may also be worth mentioning that there are other ways to achieve a similar result. An earlier draft included a proof using the Multivariate Implicit Function Theorem and also assumed second order conditions and positive determinant Jacobian matrices to get the stronger result of effort strictly increasing in piece rates (or strictly decreasing in leisure technology). However, the matrix notation was cumbersome relative to the above proof.

10.4.2 Proposition 2: Naive Momentum g(.) function

Under the FOC for e_t, c_t:

$$\begin{aligned} -u_e(c_t^*, e_t^*, e_{t-1}, \gamma_t) &= \lambda w_t \\ u_c(c_t^*, e_t^*, e_{t-1}, \gamma_t) &= \lambda p_t \end{aligned}$$

As u is strictly concave over the first argument, this allows for inverse of u_c:

$$c_t^* = u_c^{-1}\left(\lambda p_t,\, e_t^*,\, e_{t-1},\, \gamma_t\right)$$

Which can be inserted into the first FOC to give:

$$-u_e(u_c^{-1}\left(\lambda p_t,\, e_t^*,\, e_{t-1},\, \gamma_t\right), e_t^*,\, e_{t-1}, \gamma_t) = \lambda w_t$$

The e_t^* which solves this first order condition is equivalent to the e_t^{**} which would maximize (by construction):

$$e_t^{**} = argmax_{e_t}\ \lambda w_t e_t + \int_0^{e_t} u_e(u_c^{-1}\left(\lambda p_t,\, x,\, e_{t-1},\, \gamma_t\right), x,\, e_{t-1}, \gamma_t) dx$$

This objective function can be rewritten as $U_t^{'} = \lambda w_t e_t - g\left(e_t,\, e_{t-1},\, \gamma_t\right)$. It remains to be shown that this $g(e_t,\, e_{t-1},\, \gamma_t)$ is convex in e_t, which in this case is equivalent to having a negative second derivative:

$$\frac{dg}{de_t} = -\frac{\partial u(u_c^{-1}\left(\lambda p_t,\, e_t^*,\, e_{t-1},\, \gamma_t\right), e_t^*,\, e_{t-1}, \gamma_t)}{\partial e_t}$$

[72] To be clear, as I am using component-wise ordering, increasing just one element of the piece rate vector or one element of the leisure technology vector causes the vector to be ordered as higher than the unaltered vector.

$$\frac{d^2 g}{de_t^2} = -\frac{\partial^2 u}{\partial e_t^2} - \frac{\partial^2 u}{\partial e_t \partial c_t}\frac{dc_t}{de_t}$$

Note $\frac{de_{t-1}}{de_t} = \frac{d\gamma_t}{de_t} = 0$ as e_{t-1} and γ_t are not choice variables at time t. The total differential of the first order condition for c_t yields:

$$\begin{aligned} u_{cc}dc_t + u_{ce}de_t &= 0 \\ \Rightarrow \frac{dc_t}{de_t} &= -\frac{u_{ce}}{u_{cc}} \end{aligned}$$

Thus for the second derivative:

$$\begin{aligned} \frac{d^2 g}{de_t^2} &= -u_{ee} + u_{ec}\frac{u_{ce}}{u_{cc}} \\ &= -\frac{1}{u_{cc}}\left(u_{ee}u_{cc} - u_{ec}u_{ce}\right) \\ &> 0 \end{aligned}$$

As $u_{cc} < 0$ and $u_{ee}u_{cc} - u_{ec}u_{ce} > 0$. Given this derivation, one can derive how past, present, and future piece rate and leisure technology influence effort as outlined in Section 2.

10.4.3 Proposition 3: Reciprocity Proof 1

If $\alpha_2 > 0$ and e_1, e_2 is an interior solution, then $\frac{\partial e_{t+1}}{\partial \gamma_t} > 0$

$$U = (\lambda w_1 + \alpha_1 w_1 + \alpha_1 w_2 + \alpha_2\gamma_1 + \alpha_2\gamma_2)(e_1 + e_2) - g(\gamma_1 e_1) - g(\gamma_2 e_2)$$

First order condition:

$$\begin{aligned} (\lambda w_1 + \alpha_1 w_1 + \alpha_1 w_2 + \alpha_2\gamma_1 + \alpha_2\gamma_2) &= \gamma_1 g'(\gamma_1 e_1) \\ (\lambda w_2 + \alpha_1 w_1 + \alpha_1 w_2 + \alpha_2\gamma_1 + \alpha_2\gamma_2) &= \gamma_2 g'(\gamma_2 e_2) \end{aligned}$$

Multivariate implicit function theorem (Dini) gives us:

$$\begin{aligned} \frac{\partial e_2}{\partial \gamma_1} &= -\frac{det\begin{bmatrix} \frac{\partial F_1}{\partial e_1} & \frac{\partial F_1}{\partial \gamma_1} \\ \frac{\partial F_2}{\partial e_1} & \frac{\partial F_2}{\partial \gamma_1} \end{bmatrix}}{det\begin{bmatrix} \frac{\partial F_1}{\partial e_1} & \frac{\partial F_1}{\partial e_2} \\ \frac{\partial F_2}{\partial e_1} & \frac{\partial F_2}{\partial e_2} \end{bmatrix}} \\ &= -\frac{det\begin{bmatrix} -\gamma_1^2 g^{''}(\gamma_1 e_1) & \alpha_2 - g^{'}(\gamma_1 e_1) - \gamma_1 e_1 g^{''}(\gamma_1 e_1) \\ 0 & \alpha_2 \end{bmatrix}}{det\begin{bmatrix} -\gamma_1^2 g^{''}(\gamma_1 e_1) & 0 \\ 0 & -\gamma_2^2 g^{''}(\gamma_2 e_2) \end{bmatrix}} \\ &= -\frac{-\alpha_2\gamma_1^2 g^{''}(\gamma_1 e_1)}{\gamma_1^2\gamma_2^2 g^{''}(\gamma_1 e_1) g^{''}(\gamma_2 e_2)} \\ &= \frac{\alpha_2}{\gamma_2^2 g^{''}(\gamma_2 e_2)} > 0 \end{aligned}$$

10.4.4 Proposition 3: Reciprocity Proof 2

If $\alpha_1 > 0$ and e_1, e_2 is an interior solution, then $\frac{\partial e_{t+1}}{\partial w_t} > 0$:

By Multivariate Implicit Function Theorem using the above FOC.

$$
\begin{aligned}
\frac{\partial e_2}{\partial w_1} &= -\frac{det\begin{bmatrix} \frac{\partial F_1}{\partial e_1} & \frac{\partial F_1}{\partial w_1} \\ \frac{\partial F_2}{\partial e_1} & \frac{\partial F_2}{\partial w_1} \end{bmatrix}}{det\begin{bmatrix} \frac{\partial F_1}{\partial e_1} & \frac{\partial F_1}{\partial e_2} \\ \frac{\partial F_2}{\partial e_1} & \frac{\partial F_2}{\partial e_2} \end{bmatrix}} \\
&= -\frac{det\begin{bmatrix} -\gamma_1^2 g''(\gamma_1 e_1) & \lambda + \alpha_1 \\ 0 & \alpha_1 \end{bmatrix}}{det\begin{bmatrix} -\gamma_1^2 g''(\gamma_1 e_1) & 0 \\ 0 & -\gamma_2^2 g''(\gamma_2 e_2) \end{bmatrix}} \\
&= -\frac{-\alpha_1 \gamma_1^2 g''(\gamma_1 e_1)}{\gamma_1^2 \gamma_2^2 g''(\gamma_1 e_1) g''(\gamma_2 e_2)} \\
&= \frac{\alpha_1}{\gamma_2^2 g''(\gamma_2 e_2)} > 0
\end{aligned}
$$

10.5 Secondary Outcome Treatment Effects

In addition to correct problems solved as a metric for effort, I also collected the number of YouTube videos searched as a proxy for leisure time.[73] As predicted, subjects in the first experiment search 0.2 fewer YouTube videos when the piece rate is increased (see Appendix Table 5). In some specifications, this reduction in leisure persists into the next period as well. This is consistent with the model in which YouTube videos are a leisurely activity, and when faced with a higher piece rate, the agent exerts more time and effort working (and less on leisure).

To further validate that the phone access was actually a leisure activity, subjects in the first experiment searched about 0.12 to 0.15 fewer searches ($p < 0.05$) when given access to cell phones. This is consistent with a model in which phone access and YouTube videos are both leisure activities that compete for attention. Anecdotally, both YouTube and the cellphone often rely on visual cues on different screens, making them difficult to serve as leisure complements.

10.6 Median Unbiased Estimator (MUE)

In this appendix section, I employ a Median Unbiased Estimator as an alternate specification (as opposed to instrumental variables).

Andrews (1993) outlines a method for adjusting the well known bias of using OLS to estimate an AR(1) for three cases: (i) without an intercept, (ii) with an intercept, and (iii) with an intercept and time trend. Unfortunately, the original Andrews (1993) paper does not allow for individual fixed effects or individual exogenous variables x_{it}. Estimating the model without fixed effects or individual-level covariates would likely bias the ρ parameter upwards through omitted variable bias, as individuals have time-constant heterogeneity in their effort allocation.[74] Thus the Andrews (1993)

[73] Ideally I would prefer total length of YouTube videos played, but this information could be gathered from the web browser. In addition, playing a video does not necessarily imply that the agent is actually watching a video, and length of the video may be imperfect measures if the agents skip portions. Thus, both would be noisy estimates of the total leisure time.

[74] For example, if there were two types of workers, lazy and hard working, then even if the true ρ was 0 in the model above, because the individual heterogeneity is not being addressed, you would detect a very large $\hat{\rho}$.

MUE estimator would still be biased in this setting as it would be too low, unable to account for the additional omitted variable bias.

However, there has been considerable work extending the original MUE estimator to panel data. One direction can be found in work on Panel Exactly Median-Unbiased Estimators (PEMU) by Cermeno (1999) and Phillips and Sul (2003). However, an important assumption of this early work is that the error terms are homoskedastic and i.i.d. normal.[75] In addition, these works do not allow for other exogenous regressors x_{it} aside from individual and time fixed effects. Under these assumptions, the mapping between $\hat{\rho}^{LSDV}$ from Least Squares Dummy Variable (LSDV) and the median unbiased $\hat{\rho}^{MU}$ does not depend on the individual fixed effects and can be obtained by Monte Carlo simulations.

Carree (2002) extends Andrews (1993) by allowing exogenous variables x_{it} to be included as well as individual fixed effects. This addition may be important as I have exogenous treatment variables (piece rates and leisure options) that could influence effort. As Carree (2002) proves, the Least Squares Dummy Variable estimator of ρ will still be biased downward, as in the original Nickell (1981) paper. One additional benefit of the Carree (2002) paper is that it provides closed form solutions for $T = 2$ and $T = 3$, which one of my experiments satisfies.[76] This enabled me to provide you some results below, but does not provide closed forms for the standard errors (which would be estimated by Monte Carlo simulations).

To reiterate, I will be applying the Carree (2002) results using the piece rate as the exogenous variable and number of problems solved (per 5 minutes) as the outcome variable, and differencing out the running sum to remove individual fixed effects:

$$
\begin{aligned}
e_{it} &= \rho \cdot e_{it-1} + \mu_i + \beta x_{it} + \epsilon_{it} \\
\Rightarrow e_{it} - \bar{e}_{it} &= \rho \cdot (e_{it-1} - \bar{e}_{i,t-1}) + \beta \cdot (x_{it} - \bar{x}_{it}) + (\epsilon_{it} - \bar{\epsilon}_{it}) \\
\tilde{e}_{it} &= \rho \cdot \tilde{e}_{it-1} + \beta \cdot \tilde{x}_{it-1} + \tilde{\epsilon}_{it}
\end{aligned}
$$

As in Nickell (1981) and proven in Carree (2002), the ρ estimated from the OLS of this specification is still biased downward. However, Carree provides a median unbiased estimator when $T = 3$ (as in my case). Specifically:

$$
\begin{aligned}
\hat{\rho}^{MUE} &= \frac{9\hat{\rho}^{OLS} + 2\hat{g}}{9 - \hat{g}} \qquad \text{(from equation 12b)} \\
\hat{g} &\equiv \frac{\hat{\sigma}^2_{\tilde{\epsilon}}}{(1 - \text{corr}_{\tilde{x},\tilde{y}_{t-1}}) \cdot \hat{\sigma}^2_{\tilde{y}_{t-1}}} \qquad \text{(from equation 10)}
\end{aligned}
$$

When I constructed the above OLS regression differincing out running means, I received a $\hat{\rho}^{OLS} = 0.1052$. Once I use this method for correcting the bias, I receive a $\hat{\rho}^{MUE} = 0.4610$ which is very close to the Instrumental Variable estimates I find in the paper of 0.43 to 0.45. I provide more details below on how this estimate was constructed:

[75]These assumptions are relaxed in Phillips and Sul (2003) in an estimator called Panel Feasible Generalized Least Squares Median-Unbiased Estimators (PFGLSMUE). A less important difference is that they assume the AR(1) component is actually embedded in a latent variable:

$$
\begin{aligned}
e_{it} &= \mu_i + e^*_{it} + \epsilon_{it} \\
e^*_{it} &= \rho e^*_{it-1} + \nu_{it}
\end{aligned}
$$

However, this can still be mapped to my original model above by scaling $\mu^*_i = \mu_i/(1 - \rho)$

[76]Although the paper only gets "nearly" unbiased asymptotic estimators for $T > 3$, for $T = 2$ and $T = 3$ the estimator is asymptotically exactly unbiased. However, the usual issues with sample size remains.

Term	Estimate	Origin
$\hat{\sigma}^2_{\tilde{\epsilon}}$	2.686	Estimated from residuals of OLS of differenced equation above
$\hat{\sigma}^2_{\tilde{y}_{t-1}}$	2.066	Estimated from the data of $\tilde{y}_{i,t-1}$
$\text{corr}_{\tilde{x},\tilde{y}_{t-1}}$	5.23×10^{-7}	Estimated from the data of $\tilde{x}_{i,t}$, $\tilde{y}_{i,t-1}$. Note this should be 0 (see details below).
$\hat{g}$	1.301	Transformation of above statistics using closed form solution
$\hat{\rho}^{OLS}$	0.1052	Estimated coefficient from OLS of differenced equation above (biased downward)
$\hat{\rho}^{MUE}$	0.4610	$= (9\hat{\rho}^{OLS} + 2\hat{g})/(9 - \hat{g})$ from Carree (2002), $T = 3$ case, equation 12b
$\hat{p}^{IV}$	0.43 to 0.45	From IV strategy, see Table 5

Note that $\text{corr}_{\tilde{x},\tilde{y}_{t-1}}$ is essentially 0. This is not a mistake or a sign of a weak regressor, but rather a sign that treatment was properly randomized. Because it represents the correlation between $\tilde{x}$ at time t and $\tilde{y}$ at time t-1, this is saying that last period's effort difference does not predict the piece rate treatment in the next period. This is to be expected as treatment was randomized, so last period's effort should not predict next period's piece rate treatment.

As Carree (2002) mentions, "An exogenous variable which is very highly correlated with the lagged endogenous variable and which provides little additional explanatory power will lead to worse bias." An exogenous variable x which is highly correlated with the lagged endogenous variable would result in a large $\text{corr}_{\tilde{x},\tilde{y}_{t-1}}$. As can be seen from the above equations, a large correlation would *increase* $\hat{g}$, increasing the bias. However a predictive x also helps lower $\sigma^2_{\tilde{\epsilon}}$, which reduces the bias. Thus, there is a potential trade off for including exogenous covariates. In my case, piece rate helps predict effort – the $\hat{\beta}$ from the above regression was 11.99 with a t-value of 5.49, very much in line with the Instrumental Variable results from the main specification – and does not correlate with past effort, so it helps reduce bias in both directions.

However, there are two potential issues to address with this estimation. First, as with PEMU models, this model's bias correction relies on homoskedastic and i.i.d. normal errors. Second, as with other MUE bias correction, the estimator is only median unbiased asymptotically, as $N \to \infty$. While Monte Carlo results have explored the small sample properties of these estimators in some cases, this is still a potential concern.