Making Reference-Dependent Preferences: Evidence from Door-to-Door Sales

Samuel Dodini*
Federal Reserve Bank of Dallas

June 2026

Abstract

This paper uses data from a door-to-door sales company and an online experiment to examine the relationship between reference-dependent daily labor supply and long-run goal achievement. In the sales data, I show that daily labor supply kinks downward at a worker's expectations and that these expectations directly correspond to bonuses paid at the end of the sales season. The bonuses induce workers to adopt long-run targets and subsequently distribute these into internalized daily goals around which they exhibit loss aversion. These dynamics explain why non-linear payment schemes increase performance: workers change their short-run behavior in response to long-run performance targets. The online experiment confirms a causal interpretation of this relationship between bonuses and short-run behavior and supports the idea that short-run reference dependence can be "made" or induced by firms by adopting non-linear compensation schemes. These dynamics increase worker output and firm profitability and can explain why non-linear compensation is so popular in the labor market.

Keywords: non-linear compensation, goals, reference dependence, loss aversion
JEL Codes: J22, J33, M52, D91

*Email: samuel.dodini@dal.frb.org. I am grateful to Michael Lovenheim, Maria Fitzpatrick, Evan Riehl, Ted O'Donaghue, Alex Rees-Jones, Seth Sanders, Sam Hirshman, Bertil Tungodden, Linh Tô, and Devin Pope for helpful comments, guidance, and support with the experimental design. Thank you to Sebastian Fest for support in implementing the online experiment. I am also thankful to seminar participants at the Cornell Behavioral Economics Workshop, Cornell Labor Economics Workshop, and Cornell Labor Work in Progress Seminar for helpful comments and discussion. This project was partially funded by the Research Council of Norway through its Centers of Excellence Scheme, FAIR project no. 262675. The experimental design was approved under the framework NHH-IRB 42/22. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of Dallas or the Federal Reserve System.

1 Introduction

People often fail to reach their personal performance targets due to motivational problems such as present bias (Dellavigna, 2009). From a firm's perspective, this is particularly problematic for the firm's objectives. One possible approach to addressing these issues is to subdivide a target into components and evaluate one's performance in a smaller window or "narrow bracket" such as a single task or a shorter time frame. In other words, people may set a short-run goal. These goals then act as reference points (Heath et al., 1999; Hsiaw, 2018).

The motivational power of narrow goal setting relies upon people behaving in ways consistent with reference-dependent preferences (Heath et al., 1999; Koch and Nafziger, 2016; 2020; Imas et al., 2017). That is, a person's utility will depend not only on her absolute performance but also on where her performance stands relative to some mental target, i.e., the goal. Under loss aversion, when the person is operating below that mental target, the negative comparison to her target induces lower utility. However, her marginal utility is higher in this state. All else equal, this leads her to exert more effort until she reaches the target. Thus, by creating reference points, sophisticated behavioral agents can use self-imposed, psychological costs in the short term to overcome self-control problems in the long term. For firms, it is useful to induce workers to such behavior if they suspect self-control problems and dynamic inconsistency might arise in their workforce.

The idea of using short-term goals as reference points in the service of larger objectives is intuitive and backed by theory (Hsiaw, 2013; 2018). However, there is limited field evidence of what this looks like in real-world labor market settings. The empirical literature on taxi drivers that has come to define much of our real-world understanding of short-run reference dependence in labor markets (e.g. Camerer et al. (1997); Crawford and Meng (2011); Thakral and Tô (2021)) provides scant guidance regarding why this behavior exists. Rather than being a cognitive bias or error, do workers leverage daily goals as a commitment device to achieve longer-run targets? While existing theories explain why forward-looking workers may choose narrow brackets to manage self-control problems, there is little evidence on whether firms can influence this process in practice. In particular, it is unclear whether workers adopt short-run reference points in response to compensation schemes that reward longer-run performance and, if so, how firms can induce such behavior without costly monitoring or period-specific incentives. Despite the fact that worker-firm interactions represent the most consistent ways in which performance targets are adopted in a person's life and where people are most likely to shirk, we know very little about how such short-run references are "made", shaped, or induced by firms, nor is there clarity about how this might relate to the firm's and the worker's longer-run objectives.

This paper investigates reference dependence and goal setting in two contexts: a context new to the literature, door-to-door sales, and an online real-effort task experiment. The main contribution of this paper is to establish two important, new empirical observations in both contexts that indicate departures from the standard labor supply model: first, workers endogenously exhibit reference dependence in their short-run labor supply and effort choices as a means of holding themselves accountable to their longer-run objectives; second, a firm's choice of a non-linear compensation scheme (i.e. discrete performance bonuses) can causally induce workers to "make" such short-run reference points by making particular longer-run targets salient and, thereby, establishing expectations. This action cost-effectively increases worker performance and firm profits. This is consistent with theories of goal-setting in which workers use gain-loss utility around reference points as a commitment device in the short run, even when contracts are not loss-framed and even in the absence of short-run monetary incentives. Bonuses, therefore, induce rational short-run reference dependence to address expected present bias—using one behavioral bias to overcome another. This result at least partially explains why non-linear compensation schemes such as bonuses are so popular while daily monetary or enforcement incentives are rarer.

Recent theories show how present bias leads to suboptimal short-term effort because present-biased agents will tend to shirk today in favor of expecting to work harder tomorrow (Hsiaw, 2013; 2018; Koch and Nafziger, 2020). In these models, goals and goal brackets can help discipline future behavior.[1] Hsiaw (2013) emphasizes goals as self-control devices, while Hsiaw (2018) models the choice between narrow and broad brackets. Koch and Nafziger (2020) show how reference dependence and loss aversion strengthen the motivational power of these goals.

In a simple theoretical framework, I propose that firms can influence workers to adopt short-run goals around which gain-loss utility is evaluated by 1) making long-run targets more salient and 2) attaching significant monetary consequences to long-run output even when a long-run target is not a loss-framed contract (e.g., Imas et al. (2017)) and even without short-run monetary incentives (e.g., Kaur et al. (2015)). These raise the incentive for workers to strategically adopt narrow brackets and short-term gain-loss utility in order to achieve their more visible and salient long-run output if they believe their future self may be tempted to shirk. I use my sales data and experiment to examine these key dynamics.

I analyze high-frequency data from a door-to-door sales company that employs fixed-term, commission-based contractors. First, I establish the baseline observation that workers do exhibit reference-dependent labor supply in this novel setting. I test for this on two margins: the extensive margin (the choice to stop working for the day) as well as the intensive or "exertion" margin (effort conditional on working). I provide clear evidence that door-to-door sales workers exhibit loss aversion around expectations in their extensive margin labor supply choices.

I use a detailed panel of observations in half-hour increments with each seller's location, cumulative service contracts generated (which I call "sales" throughout the paper), pitches presented to a prospective customer, and the probability of stopping work for the day (the extensive labor supply

[1]The model stands in contrast to the standard model, which assumes rationality in intertemporal utility affecting long-run performance (i.e., no self-control problems and, therefore, no need for short-run goals) and no utility responses to reference points in the short run. Under the behavioral model with a non-zero degree of present bias, firms find it preferable to encourage narrow brackets.

margin). My measure of intensive margin effort (or alternatively phrased "exertion") is pitches per half hour. I define the reference point as a sample proxy of expectations: each seller's own average daily number of sales for all past workdays in the season, which I show is highly correlated with revealed long-run objectives. Upon reaching their expectations-based reference point, the probability a worker stops for the day increases significantly by a factor of 2.8–4.1 relative to below the reference point, suggesting that losses loom larger than gains by a factor of approximately 3 to 4. On the "exertion" margin, the change is quantitatively small. The choice of when to stop working is the key margin at which reference-dependent daily labor supply operates.

I show three pieces of evidence to support the hypothesis that the firm's contract structure incentivizes the worker to optimize around a long-run goal at a bonus threshold and workers then distribute that goal into daily targets. First, I show that these workers are forward-looking, as evidenced by the fact that their work hours do not significantly respond to the increases in commission rates that follow their cumulative sales. Rather, they appear to optimize based on their ex ante anticipated commission rates. Second, I show that upon reaching their relevant bonus threshold, workers significantly reduce their work hours despite continuing to be paid a *higher* piece rate than they were paid before reaching the bonus. Thus, attainment of the bonus is the motivating factor behind work hours persistence and daily reference dependence and not other mechanisms like simple habituation or prior planning of work hours.

Third, I show that as a result of the two behaviors above, the distributions of performance are subject to significant "bunching" around bonus thresholds. These three robust observations together provide strong evidence that the firm, through its compensation scheme, can shape the choice of long-run expectations and induce short-run targeting behavior.

One caveat to my sales setting is that there is no state of the world in that context in which the bonus scheme is absent. To bolster a causal interpretation that the bonus scheme is the key driver, I conduct an online experiment designed to mirror in a short period the dynamics of the sales data. In the experiment, participants in a button-pushing task divided into four 2.5 minute rounds are randomly offered either a piece rate or a bonus upon reaching a certain performance level, where the bonus is only determined at the end of the last round. I then observe the distribution of performance in each round, where differences in behavior can be credited only to different responses to the compensation schemes. Significant excess "bunching" of the performance distribution each round in the bonus treatment provides clear evidence that even though round-specific performance is not payoff-relevant for those in the bonus payment condition, participants exhibit reference dependence with loss aversion around endogenously self-selected targets. In other words, they anchor their performance to specific targets each period to ensure that they satisfy the bonus condition at the end of the task as a commitment device. Many respondents make this strategy explicit in open-ended responses. Participants in the bonus treatment significantly outperformed those in the piece rate treatment in cost-effectiveness: for the same average performance across both conditions, the bonus group incurred a 31% lower per-person compensation cost.

The core contribution of this paper is to bring together real-world and experimental evidence that short-run reference dependence acts as a commitment device to achieve long-run goals and that firms can induce such behavior to improve worker performance. In doing so, the analysis contributes to three main strands of the literature on reference dependence.

In the first strain, several papers analyzing taxi and rideshare drivers' behaviors have found a negative relationship between daily wages and hours worked, downward shifts in labor supply at particular earnings levels, or negative labor supply responses to large tips, consistent with reference dependence.[2] However, the taxi cab literature has not empirically explored the purpose of having reference points at all. Even though the earliest taxi cab studies (e.g. Camerer et al. (1997))[3] hypothesized that income targeting may help drivers address self-control problems, none of these studies have empirically explored this dynamic. In a recent experiment in Kenya, Dupas et al. (2020) show that a person's stated income needs and expectations for earnings (rather than just total income) act as reference points. The authors suggest that such targeting motivates workers as a commitment device to perform their physically demanding jobs, though intertemporal dynamics are not explored as they do not consider measures of broader income needs.[4] This empirical literature has not considered the interaction between clearly defined long-run objectives and daily targets in reference dependence. This paper makes a significant contribution by examining both the short- and long-run in a unified way using real-world data and a controlled experiment.

The "lumpy" nature of income in my sales context and the lottery-like nature of success at each door that decouples immediate income from effort makes this setting quasi-experimental and ideal for the study of reference-dependent labor supply.[5] In addition, previous studies testing reference dependence are quite narrowly focused on routine and manual tasks like taxi driving or physical labor. This paper provides new evidence in both a simple manual setting (the online experiment) and a novel work context that uses adaptive cognitive and social skills in a developed country.

[2]See Camerer et al. (1997); Chou (2002); Crawford and Meng (2011); Farber (2015); Morgul and Ozbay (2015); Agarwal et al. (2015); Martin (2017); He et al. (2018); Schmidt (2018). Other papers that find evidence of reference-dependent labor supply analyze the behavior of bike messengers (Fehr and Goette, 2007; Goette et al., 2004) and fishermen in Hawaii (Nguyen and Leung, 2013). However, the literature is far from settled. A competing set of studies of drivers finds a positive relationship between daily wages and hours worked and concludes that the standard model performs better than prospect theory (Farber, 2005; 2008; 2015; Sheldon, 2016). Other analyses that find evidence supporting the standard model examine day laborers in Malawi (Goldberg, 2016), stadium vendors (Oettinger, 1999), fishermen in Florida and India (Stafford, 2015; Giné et al., 2016), and markets in India (Andersen et al., 2014).

[3]"Daily targets can also serve a second purpose: like many mental accounts, they help mitigate self-control problems." (Camerer et al., 1997), pp. 426. Setting a daily target can be a method of ensuring that longer-run monthly payment obligations can be successfully managed.

[4]Their questionnaire and definition of "income needs" is specific to each day's idiosyncratic needs, which often exceed their average income. Both idiosyncratic stated income needs and expectations affect labor supply (their Appendix E). Stated daily income needs intensify reference dependence.

[5]Encountering one extra resident willing to purchase the sellers' services leads to an increase in income of \$100–\$250. An extra sale or two by a seller is worth the same amount as an entire shift for a taxi driver (\$270) but takes roughly the same amount of time as 1-2 taxi trips (16–32 minutes) (Thakral and Tô, 2021).

Understanding this skill distinction is crucial if workers in manual occupations differ significantly in their attributes from those who select into primarily social or cognitive occupations or who have the education to enter these occupations.[6]

The second strain focuses on reference dependence as expressed in the distributions of final outcomes around a single or ending target. Experimentally, Abeler et al. (2011) show that when expectations are exogenously manipulated, subjects performing a real-effort task behave in ways consistent with reference dependence around the manipulated expectations as evidenced by bunching in the probability of stopping the task at expectations. In a firm-worker setting, Kuhn and Yu (2021) examine the effects of kinks in a commission schedule on final team performance and find these act as symbolic rewards, leading to bunching in the distribution of performance. Cai et al. (2022) examine kinks in the compensation schedule at a Chinese manufacturing firm to estimate labor supply elasticities and firm cost savings. Their analysis does not consider behavioral factors such as loss aversion. Similarly, Freeman et al. (2019) analyze a shift in a performance bonus threshold at a Chinese insurance company, which led to significant increases in worker output and firm profits. Beyond knowing that these incentive schemes increased total worker output at the end of an evaluation period, little is known about how or why these approaches were effective.[7] The underlying day-to-day behavioral dynamics have immense implications. If a new compensation scheme induced workers to set short-run goals with loss aversion, this represents, from the firm's perspective, a low-cost *psychological* incentive rather than a high-cost *monetary* incentive. This analysis documents this effect in two distinct settings.

The third strand studies goals and goal bracketing as tools for overcoming self-control problems. Hsiaw (2013) and Hsiaw (2018) theoretically describe optimal goal adoption and the choice of goal brackets, respectively, while subsequent experimental work shows that daily goals, loss-framed contracts, and self-imposed penalties can increase performance. For example, laboratory and field experiments show that those instructed to set daily goals perform better than those instructed to set weekly goals (Koch and Nafziger, 2020), that workers prefer loss-framed contracts (Imas et al., 2017), and that many workers will voluntarily reduce their daily compensation if they fail to meet daily targets when offered these dominated contracts (Kaur et al., 2015). My two settings provide distinct contributions relative to this work. First, short-run goals in my setting are not exogenously assigned but are *endogenously* adopted by workers in response to the structure

[6] Given the literature suggesting significant differences in key behavioral parameters across occupation, education, or cognitive ability (e.g. discount rates, risk preferences, etc.), it is reasonable to suspect possible differences in loss aversion and reference dependence (e.g. Cadena and Keys (2015); Bellemare and Shearer (2010); Patnaik et al. (2020); Fouarge et al. (2014); Warner and Pleeter (2001)).

[7] Explicit long-run goals act as references, as seen in the distributions of finishing times among marathon runners (Allen et al., 2017; Markle et al., 2018). However, like cumulative outcomes in a firm, marathon times are the result of dynamic processes (e.g. daily training). Despite this, previous studies have not focused on how the long-run target interacts with short-run choices or preferences, with the exception of, Soetevent (2022), whose short-run analysis is limited to time intervals in the middle of a marathon race rather than multi-day training choices.

of longer-run incentives and the expectations they set (Abeler et al., 2011). Second, it is through this endogenous adoption of narrow brackets that performance increases, thus providing a dynamic mechanism for the performance increases found in Imas et al. (2017), Kuhn and Yu (2021), and Freeman et al. (2019). I show that such a mechanism can be induced by firms at a low cost, even without a loss-framed contract and without the need for strong daily performance incentives (as in Kaur et al. (2015)), as these may be costly to monitor and implement for the firm.[8] Thus, this paper contributes significantly to our understanding of the interaction between firm-provided incentives and dynamic worker choices and helps explain why non-linear payments are so popular across many industries and occupations.

My analysis of long-run objectives and short-run expectations is the first field study of which I am aware to examine how endogenous period-specific/daily reference dependence acts as a commitment device to achieve revealed or imposed long-run goals in the absence of loss-framed contracts—to address the "why" of short-run reference points. I show that firms can leverage salient long-run targets to induce this behavior and that workers are responsive to firms' broad incentives even in their daily or period-specific activities. Self-imposed, short-run reference points appear malleable when long-run incentives change. My online experiment confirms the causal interpretation of these dynamics by linking an increase in reference-dependent behavior to non-linear compensation schemes. Firms take advantage of these internal dynamics because reference dependence makes it easier to motivate a worker if she perceives herself to be in a "loss" domain, leading to significant cost savings for the firm. The question of reference-dependent labor supply is, therefore, central to our understanding of the power of incentives to motivate effort.

2 Conceptual Framework

2.1 Reference Dependence vs Standard Model for Labor Supply

One key insight of Prospect Theory is that losses loom larger than similarly sized gains (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992). The implication for labor supply under loss aversion is that workers supply greater effort while in a loss domain (before achieving a target) relative to what they supply in a gain domain (after achieving a target). This leads to a discontinuous change in marginal utility after surpassing some reference point, with the marginal utility of income falling significantly by some factor $1/\lambda$, where λ is the parameter of loss aversion. This induces a discontinuous change in labor supply, all else equal. Importantly, no such discontinuity is predicted by the standard model (see Appendix E.1 for additional background). Because loss aversion is costly due to the agent experiencing lower levels of utility while in the loss domain, this behavior is sometimes construed as a negative cognitive bias. However, under certain conditions,

[8] My analysis suggests that if the workers in Kaur et al. (2015) were paid longer-run bonuses for cumulative performance, they might have endogenously chosen daily references and goals and imposed a utility cost upon themselves without needing to be offered the dominated daily contracts. Thus, I also contribute to the literature on "insider econometrics," the use of non-linear incentives and bonuses, and their effects (Ichniowski and Shaw, 2003).

loss aversion can be leveraged to solve other behavioral biases.

2.2 Loss Aversion, Goals, and Short-Run References

Multiple theoretical treatments propose that the strategic use of loss aversion may be rational if a worker has a problem with self-control or dynamic inconsistency as a result of present-biased preferences (see Ariely and Wertenbroch (2002); Shefrin and Thaler (1992); Camerer et al. (1997); Imas et al. (2017); Kaur et al. (2015), with Koch and Nafziger (2016; 2020) presenting the most formal recent treatment). Present bias has been documented in a variety of contexts, e.g. exercise goals (DellaVigna and Malmendier, 2006), education (Ariely and Wertenbroch, 2002), credit markets (Meier and Sprenger, 2010), and savings (Ashraf et al., 2006), and it is simple to extend the concept to labor markets. In experimental and field settings, these preferences are common, so firms may respond by establishing policies and incentive systems aimed at this sizable portion of their workforce.

2.3 A Stylized Model of Goal Bracketing and Effort Provision

In this section, I present a brief conceptual framework building on the architecture of Koch and Nafziger (2020), specifying how time-inconsistent preferences lead to effort levels in the current period that fall below what a worker would have chosen for herself ex ante and how loss aversion around reference points helps overcome these self-control problems. I then describe how firms' use of non-linear payments might influence goal-setting. For ease of exposition, I will refer to my sales context.

Consider a sales worker who performs the same sales task over T working days, indexed by $t = 1, \ldots, T$. On each day t, the worker chooses effort $e_t \geq 0$, which maps monotonically into expected sales. Effort is costly in the short run, incurring costs $c(e)$, while the monetary benefits from sales accrue at the end of the evaluation period through commissions. This type of delayed final compensation is consistent with my sales context and experiment.

Instantaneous utility is

$$u_t = -c(e_t),$$

where $c'(e) > 0$ and $c''(e) > 0$. Final-period utility depends on cumulative effort:

$$u_{T+1} = \sum_{t=1}^{T} b(e_t),$$

with $b'(e) > 0$ and $b''(e) \leq 0$. Ex ante, a period 0 self sets marginal costs and benefits equal such that $b'(e_0^*) = c'(e_0^*)$. This would be the equilibrium effort under the standard model.

But suppose that preferences are quasi-hyperbolic and each self $t \geq 1$ separately chooses effort for day t.

Preferences of self t are given by

$$U_t = u_t + \beta \sum_{\tau=t+1}^{T+1} u_\tau,$$

where $\beta \in (0, 1]$ discounting (the exponential discount factor is normalized to one).

Absent goal setting, each future self $t \geq 1$ prefers effort e_t^* satisfying $\beta b'(e_t^*) = c'(e_t^*)$, implying $e_0^* > e_t^*$ whenever $\beta < 1$. In other words, effort exerted by self t is strictly less than what self 0 would have chosen. The worker always has an incentive to substitute effort from today, expecting self $t+1, ..., T$ to make up for shortfalls in t. This constitutes a self-control problem. Combating this propensity to postpone effort is the key incentive introduced by narrow bracketing.

Following Koch and Nafziger (2020), goals act as reference points that generate comparison utility relative to a reference point z. Let

$$\mu(z) = \begin{cases} z & \text{if } z < 0, \\ 0 & \text{if } z \geq 0. \end{cases}$$

Suppose a sales worker can use a daily goal to generate this gain-loss utility.[9] Under daily goals (narrow bracketing), self 0 sets daily goals $\{g_t\}_{t=1}^{T}$. Comparison utility is evaluated each day:

$$u_{T+1}^{D} = \sum_{t=1}^{T} \left[b(e_t) + \mu(e_t - g_t) \right].$$

Because these goals are set internally by the worker, they are, by definition, not binding, but are taken as reference points by future selves. This idea is closely related to the goal-setting model of Hsiaw (2013), in which present-biased agents voluntarily adopt non-binding goals to regulate future behavior via expectations and self-evaluation, and, in that framework, can improve self-control even without loss aversion. The reference-dependent formulation I adopt here follows Koch and Nafziger (2020) in emphasizing gain-loss utility, but shares with Hsiaw (2013) the key insight that workers may rationally choose goals ex ante to discipline future selves.

An important question arises concerning where the self 0 would prefer to set the reference points g_t. Kőszegi and Rabin (2006) theorize that "recent expectations" act as important guides for reference points. Their model proposes that expectations are determined in personal equilibrium, that is, by behaviors that are optimal given the worker's expectations and beliefs about the future.

[9] Hsiaw (2018) models the choice of goal brackets and lays out the trade-offs between narrowly bracketed goals (stronger self-control properties) and broader brackets (greater flexibility and risk pooling under uncertainty). Given greater impulsivity, agents will set incremental goals whenever payoff uncertainty given effort is low. Consistent with this logic, Koch and Nafziger (2020) show theoretically and experimentally that if workers are at least partially naïve about their present bias, daily goals and narrow bracketing at least weakly dominate weekly goals and broad bracketing.

This requires a correct understanding of the choice environment the worker faces and her own reaction to that choice environment. Put another way, self 0 perceives the optimal path forward based on the distribution of outcomes in her prior experience, and when her next choice is made in real-time, the planned path becomes her reference point. What this means for rational goal-setting is that daily goals must be aligned with what the worker expects about her abilities, her effort costs, and her projections about her own preferences and what self 0 *expects* to be optimal. Under this condition, $\hat{e}_{t,0} = g_t$, or in other words, the worker's reference point (goal) each period will align with her beliefs about β, e_t^*, and μ, where her beliefs are informed by her recent performance. This is because targets set too high relative to expectations lead to persistently low utility levels. Targets set too low relative to expectations reduce marginal utility at much lower levels of performance. Under Kőszegi and Rabin (2006) and Crawford and Meng (2011), recent expectations are operationalized as the mean of the distribution of prior performance, which I adopt in my empirical models.

As long as the worker has realistic expectations about her future present bias and incentives ahead, this comparison utility solves the self-control problem by creating immediate utility costs of not meeting a minimum daily performance g_t.[10]

2.3.1 Firm Incentives and Narrow Bracketing

From a firm's perspective, where workers reaching a certain level of output matters most, it is advantageous to induce workers to engage in narrow bracketing despite the lower experienced utility in the loss domain if doing so leads to higher rates of goal attainment. How might firms do this? The first possibility is simple: craft exogenous period-specific/daily targets for each worker that make daily performance the focal outcome. This can be done via monetary incentives or by close monitoring and enforcement. However, even when real-time performance is not costly to monitor, it may be costly to enforce (e.g. in morale costs, managerial time, social perceptions). Indeed, it is extremely rare to find work arrangements in which, for example, daily performance is measured and sub-optimal performance is punished. Many firms and industries use bonuses for longer-run output instead (e.g. Kuhn and Yu (2021); Cai et al. (2022)). But how might these longer-run targets solve self-control problems in the short run?

The literature suggests two distinct approaches to addressing workers' self-control problems. First, a firm might directly shape monetary incentives through the use of loss-framed contracts, where a reward for a minimum output is removed if the worker fails to produce a certain output (Imas et al., 2017; Kaur et al., 2015). Second, workers may voluntarily adopt goals and narrow brackets in order to achieve a target, even when those goals are not externally enforced (Hsiaw,

[10] I focus my analysis and conceptual framework around daily goals, but weekly goals are also an option. In my sales data, I find no evidence of workers setting or responding to weekly goals or expectations (see Table A8). I, therefore, focus my models on narrow bracketing and abstract away from the risk pooling mechanisms in Hsiaw (2018), as the evidence suggests effort is sufficiently informative about outcomes in my sales setting.

2013; Koch and Nafziger, 2020). The first approach relies on direct incentives built into contracts, while the second operates through reference dependence and bracketing. My framework asks whether bonus schemes set up by firms target the second mechanism by inducing workers to adopt short-run goals in the service of achieving longer-run thresholds.

Imas et al. (2017) show that workers actually prefer loss-framed contracts with discrete payments for total output and that these lead to greater worker effort than gain-framed contracts. The authors suggest that workers use the loss framing to keep themselves accountable dynamically to achieve a minimum output because they are otherwise dynamically inconsistent. Unfortunately, dynamic or period-specific behaviors are not observed in the study. However, two conceptually distinct mechanisms are plausible. First, there may be a salience effect: non-linear incentives in the contract (the bonus/loss threshold) establish expectations for a worker's final performance and make a final output target *salient* and easier for workers to adopt as an internal goal or reference point (Abeler et al., 2011; Hsiaw, 2013). Second, there may be an incentive effect: the threshold ties a lump-sum reward to it, thus raising the *stakes* of failing to stay on target and increasing the value of strategically maintaining short-run discipline. This potentially affects how narrowly workers choose to bracket their goals (Hsiaw, 2018). Imas et al. (2017) suggests that both of these conditions might induce workers to narrow bracket. Despite this logical extension, we have little empirical evidence of this behavior, leaving open the question of how workers make decisions in the short run when responding to non-linear payments for long-run output, ultimately leading to greater long-run output (as in Imas et al. (2017); Kuhn and Yu (2021); Cai et al. (2022)).

2.4 Goal Setting, Bonuses, and Salience

Here, I adopt the framework above and add to the model a bonus for long-run performance.

Suppose the firm offers a bonus $B > 0$ if cumulative effort exceeds the firm's target of $\bar{E}$:

$$B \cdot \mathbf{1}\left\{\sum_{t=1}^{T} e_t \geq \bar{E}\right\}.$$

This bonus serves two functions: raising the salience of a target and increasing the stakes for not reaching the target. Under the Hsiaw (2013) framework, salience is a key parameter in the effectiveness of goals in overcoming self-control problems. A salience factor of 0 means the reference point plays no role, and ill-defined goals have weaker salience. While Hsiaw (2013) takes salience as given, one key feature of my framework is that salience can respond to firm incentives for particular targets at bonus thresholds.

I define $s \geq 0$ as a weight for the salience of that threshold, and the monetary size of the bonus as θ. Salience enters utility as reference-dependent comparison utility around the bonus threshold:

$$s \cdot \mu\left(\sum_{t=1}^{T} e_t - \bar{E}\right),$$

where

$$\mu(z) = \begin{cases} \lambda z & \text{if } z < 0, \\ 0 & \text{if } z \geq 0. \end{cases}$$

, where λ is a coefficient of loss aversion. In this setup, I assume B and s are increasing in the size of the bonus: $B = B(\theta), \quad s = s(\theta), \quad$ with $B'(\theta) > 0,\ s'(\theta) > 0$

Total period $T + 1$ utility becomes:

$$u_{T+1} = \sum_{t=1}^{T} b(e_t) + B \cdot \mathbf{1}\left\{\sum e_t \geq \bar{E}\right\} + s \cdot \mu\left(\sum e_t - \bar{E}\right).$$

This expression represents the decision set for self 0 when there is a bonus in play. If salience is not a factor ($s = 0$) and there is no gain-loss utility around the bonus, then the problem is simply for the worker to maximize gains ex ante, and there is no self-control problem.

2.4.1 Present Bias and the Self-Control Problem with Bonuses

For any period $t < T$, the worker makes effort decisions based on their current accumulated effort E_t and the prospective gains from continuing toward the threshold $\overline{E}$. I defined this continuation value as:

$$V_{t+1}(E_t) = \sum_{\tau=t+1}^{T} (-c(e_\tau)) + \sum_{\tau=1}^{T} b(e_\tau) + B \cdot \mathbf{1}\left\{E_T \geq \bar{E}\right\} + s \cdot \mu(E_T - \bar{E}),$$

where E_T depends on current and future effort choices.

Period-t utility is therefore:

$$U_t = -c(e_t) + \beta V_{t+1}(E_t).$$

with $\beta \in (0, 1]$. If the bonus raises salience s and this is spread over T periods, in equilibrium, the worker will work each day until the marginal cost of effort is:

$$c'(e_t) = \beta\left[b'(e_t) + \frac{s\lambda}{T}\right] \tag{1}$$

Crucially, the psychological loss from missing the bonus threshold is realized in period $T + 1$ and therefore discounted by β from the perspective of earlier selves. Thus, even significant payments that raise the monetary reward B and salience s are discounted from the worker's perspective, meaning that despite the greater incentive arising from the bonus, the worker is still subject to self-control problems that blunt behavioral responses to the incentive.

2.4.2 Adoption of Daily Goals Under a Bonus

Following Hsiaw (2013), I assume workers may adopt goals as a means of regulating their future selves. However, my focus is not on whether goals are optimal, but on whether a firm's bonus scheme changes the attractiveness of adopting narrowly bracketed daily goals, i.e., that the bonus affects the likelihood that workers engage in goal-setting behavior.

Self 0 can adopt daily goals $\{g_t\}_{t=1}^T$, but suppose that they are costly in attention, planning, and tracking at a fixed cost $\kappa > 0$. Under daily goals, if the worker falls short of the target, they receive immediate comparison utility:

$$\sum_{t=1}^{T} \mu(e_t - g_t),$$

which is *not* discounted by present bias. Normalizing $\mu'(e_t - g_t) = \lambda$ as loss aversion, the marginal benefit of daily effort is:

$$c'(e_t) = \lambda + \beta \left[b'(e_t) + \frac{s\lambda}{T} \right]$$

Daily goals, therefore, transform the discounted salient long-run loss into short-run losses.

Because adopting daily goals is costly, the worker will adopt these daily goals only when the marginal benefits of adoption outweigh the cost, i.e. when:

$$s\lambda(1 - \beta) \geq \kappa \tag{2}$$

Adoption of these daily goals depends on the extent of self-control problems, $1 - \beta$ (or an estimate of β in the case of naïveté), the worker's sensitivity to loss, λ, and the salience response of the bonus ($s(\theta)$). Notably, if the worker has or expects no present bias ($\beta \to 1$), she will not adopt daily goals even if $s \to \infty$.

In Appendix E.2, I describe the conditions under which firms would prefer to use non-linear payments to induce short-run goal setting rather than implement a close monitoring and enforcement regime. In short, it is cheaper for firms to use bonuses to motivate workers when they have higher loss aversion parameters or the salience weight of the bonus is high. Monitoring is preferred when monitoring and enforcement costs are particularly low or salience responds weakly to bonuses.

While my empirical tests do not directly examine the question of monitoring versus the bonus scheme, my empirical tests in the door-to-door sales and experimental contexts do test for a prerequisite to this comparison to see whether bonuses do, in fact, induce this kind of short-run goal setting behavior as theorized such that self-control problems can be solved through the adoption of daily goals that leverage the power of loss aversion.[11]

[11] In the sales context, the workers are all independent contractors paid only on commission, and managers report

My framework does not seek to explain why goals or goal brackets may arise in equilibrium (e.g., Hsiaw (2013; 2018) and Koch and Nafziger (2020)). My framework instead focuses on whether firms can induce the adoption of such goals through compensation design. The empirical analysis, therefore, tests whether bonus schemes affect the formation of short-run reference points and narrow brackets. Importantly, if workers are induced to adopt short-run goals as a means of setting themselves accountable to longer-run output targets by bonuses, this constitutes empirical evidence consistent with the key features of all three of those models as well as the key insights of the Kőszegi and Rabin (2006) model. All together, these empirical features—particularly gain-loss utility and kinked labor supply —are at odds with the standard labor supply model.

I empirically explore these dynamics through the use of real-world performance data from a door-to-door sales company and a real-effort task experiment. My analysis shows that firms do, indeed, induce workers to adopt narrow brackets by establishing longer-run targets via a bonus scheme and that inducing them to do so increases their performance.

3 Door-to-Door Sales Context

The door-to-door sales industry constitutes a sizable portion of the "direct sales" industry. Workers in door-to-door sales are presented with high-powered incentives, including high commission rates that rise with performance and the use of bonuses. These are also common in a variety of sales occupations.

A large number of firms that engage in door-to-door sales are located in the Mountain West region of the United States and employ thousands of college-age workers each summer to sell their products and services.[12] These include solar panels, pest control services, knives, and home security systems. General industry practice is relatively homogeneous across these products. Recruited sellers meet with managers, listen to an explanation of the work and earnings potential, and sign independent contractor agreements that stipulate the commission structure under which they will sell and their assigned city. The work itself is unpleasantly hot in the summer and often entails distasteful interactions with local residents. To entice skilled sellers to join their teams under these conditions, most companies will advertise that sellers make an average of $40,000 during the late April to late August sales season selling six days a week. There is a high level of competition between companies seeking to land top talent, and there is an extremely wide variance in sales skills among recruits, leading to a large variance in income. The company whose data I analyze,

very few direct enforcement interventions with salespeople even when monitoring their activity is cheap via their mobile technology. In the sales data and experiment, I do not directly test for present bias but rather if workers behave and plan as if they expect present bias.

[12]One reason for locating in this region is the large supply of young college students (usually age 20-25) who have recently returned from 2-year or 18-month proselytizing missions for The Church of Jesus Christ of Latter-day Saints, which is headquartered in Salt Lake City, and whose members comprise the majority of the population in the state of Utah. These missions, in a purely practical sense, use skills very similar to a sales job: approaching strangers and striking up a conversation, connecting quickly, moving conversations toward a specific goal, and winsomely absorbing rejection. Recruiters understand this dynamic and seek to capitalize on these skills.

which I will call "PestCo," operates within these norms.

A "sale" at PestCo is recorded when a resident signs a contract for pest control services that lasts 12–18 months for services given quarterly. The contract is recorded electronically. Among pest control sellers at PestCo, the timing of sales can vary widely. On average, sellers generate one sale for every 20 pitches they present, but exactly which of those 20 pitches will result in a sale and at what time each sale will occur is highly uncertain. Within any particular time of the day, any single pitch could result in a sale, so each knock on a house door is akin to entering a type of lottery. Hitting one's expected number of sales early in the shift comes as a meaningful surprise. Because the value of each sale to the seller is large, the stakes for each sales pitch are high.

PestCo, like nearly all door-to-door sales companies, pays large commissions in the range of 18–40% on the value of the service contracts they generate. A typical sale can result in an income to the seller between $100 and $250 depending on the value of the service contract signed by the customer and the seller's commission rate. Importantly, commission rates are increasing in cumulative sales performance and jump discretely in increments of 50 sales. The final commission percentage for each sale is calculated at the *end* of the sales season. The result is a discrete bonus with a small marginal increase in per-sale commissions past the threshold. Sellers are paid an up-front portion of their commissions ($75 per sale) during two-week pay periods, similar to a regular paycheck. The balance of commission payments is calculated at the end of the season based on final performance and paid out thereafter.

Figure 1 characterizes a seller's total earnings at the end of the summer season depending on their total sales at an assumed contract value of $500. A seller who produces 149 sales receives a commission of 25% on all sales at the end of the season, while a seller who generates 151 sales receives a 27% commission. This results in a lump-sum bonus of approximately $1,500 for crossing the 150-sale threshold (plus approximately $10 more per sale above it).[13] The average first-year seller yields between 100 and 150–175 sales, while experienced sellers generate 150 to 300 on average. The highest ability sellers generate over 350 sales for incomes in the $60,000–$80,000 range. Importantly, if a worker expects to end up in a particular interval, the main incentive is a piece rate with a bonus.[14]

The prior literature has generally considered occupations in which income is a smooth function of hours worked with relatively small deviations from average income. For example, a standard deviation in wages for a taxi driver is only about 10% of the mean (Thakral and Tô, 2021). At PestCo, the standard deviation in daily sales is 100% of the mean, and the effective daily wage can double in as little as 30–60 minutes. Income is accrued in discrete units, creating more salient opportu-

[13] In addition to this de facto bonus from the commission change, the seller receives a flat "rent bonus" of about $2,000 that covers the seller's apartment rental costs for the summer upon reaching 150 sales.

[14] While the compensation is technically kinked at each threshold, the most important incentive is a flat bonus, not a kink, making this different from the recent literature. At 250 sales, sellers qualify for the company's all-expenses-paid vacation.

nities for earnings references than in the past literature.[15] The "lumpy" nature of income in this context is an advantage over existing studies because each door interaction is quasi-experimental. In contrast to the taxi literature in which drivers have unexpected success earlier in the day driven by forces like rain that might predict success later, the nature of door-to-door sales and quasi-random receptivity of households provides an even cleaner test of reference dependence. The skill requirements of the job also make this setting unique in the literature. Sellers must be able to strike up a conversation with a stranger, understand and respond to objections, communicate the value of the product, and adapt their strategy on the fly as more information about the customer is revealed. Each of these tasks is cognitively demanding and requires strong interpersonal skills.

Another unique feature of this setting is that outside considerations that might influence the formation of medium- and long-term earnings targets in other settings are absent from this setting. Most sellers are under the age of 25 and have not formed financial commitments that require set payments that might influence the formation of salient short-run "income needs" as examined in a prior study (Dupas et al., 2020) or long-run payment obligations. The cost of housing, for example, is paid for up front by PestCo and repayment is not required unless the seller fails to secure 150 sales. These needs are, therefore, baked into the performance schedule. Income needs over the short-run (e.g. the week) or over the medium-run (e.g. the month) are not fully operable because the vast majority of compensation for their work is received by workers at the *end* of the season. Similarly, because these workers live away from their normal homes and networks, there is limited scope for daily external obligations to shape their daily labor supply choices.

Through the company's internal website and mobile app, sellers can view their performance history. All workers are aware of their normal performance, including their cumulative sales and sales so far that day on the main app dashboard. The availability of this information makes references related to one's own performance highly salient. Through its website and mobile app, PestCo tracks every sale and house "knock" recorded by each seller.[16] This forms the basis of my analysis dataset. See Appendix D for more details on industry practice and contracts.

The prior empirical literature on reference dependence has been unable to examine short-run goal-setting as a response to a long-run target because the work settings analyzed to date do not provide a clear endpoint at which a worker evaluates any long-run goals she may have. The "long run" is too nebulous. On the contrary, my sales setting provides a clear end date. A second reason is that the occupations under study are measured in settings in which other factors such as income needs may form the most salient (or only) form of medium- to long-run targets, which remain

[15] E.g., it is much easier to count contracts sold than total income earned net of tips while driving, even when the tips are "large" (approximately $30) as in Schmidt (2018).

[16] One might suspect that such tracking may invoke a sense of being monitored and increase one's effort to hit certain performance targets to avoid negative appearances to the firm. This is unlikely as the sellers are independent contractors, are not paid unless they make sales, and, absent egregious personal conduct, can continue to work as sellers for multiple sales seasons if desired.

unobserved to the researcher.[17] My sales setting is a fixed-term job that is conducted far away from "home" and is paid mostly at the end of the season among young workers without major fixed-schedule financial obligations whose housing costs are baked into the bonus schedule. This limits the scope for outside income needs to dictate specific points in the earnings distribution as targets. Thus, this setting provides a unique opportunity to study these questions empirically.

4 Data

My analysis datasets come from the comprehensive sales and seller tracking databases from PestCo for 2018–2019. The company uses a common sales tracking app that documents every door at which a seller records interacting with a resident and the location and timestamp of those interactions. PestCo separately tracks the date and time each service contract is signed, the location of each customer, and the seller who generated the sale. Together, these two systems give a comprehensive view of the activities of each seller every day they are knocking on doors and selling in their work area.

Using the raw sales and knocking data, I construct two panels of individual seller performance. First, I build a daily panel of each seller's sales, work hours (defined as the time between the first knock/sale and the last knock/sale), cumulative sales over the season, and cumulative average daily sales as a measure of "recent expectations." Following the past literature (Crawford and Meng, 2011), I calculate a proxy for each seller's recent expectations by examining each seller's average past daily sales during the season. The selling week runs Monday through Saturday. Because residents are home at higher rates on Fridays and Saturdays and seller experiences differ by day of the week, I calculate each seller's average daily sales specific to each day of the week from all past days in the same sales season. These expectations can update and evolve over the course of a season, though the measure is remarkably stable after the first 2-3 weeks.[18]

In my second dataset, I construct a panel of each seller's pitches presented to a prospective customer, within-day cumulative sales, and stopping probability each half hour of their shift. This interval of observation is the same as that in the recent taxi literature (Thakral and Tô, 2021). For each seller in each half hour, I create a measure of their current distance to their daily expectations: their number of cumulative sales so far that day minus their average sales for that day of the week. For values less than zero, a seller has not yet achieved her expectations and is therefore in a loss domain, while values greater than or equal to zero indicate a seller is in a gain domain. In this dataset, I define "starting" a shift as the half hour of the day in which a seller records her first knock of the day, and I define "stopping" as the half hour of the shift when the last knock of the

[17] Bénabou and Tirole (2004) propose that self-reputation is what animates the use of daily targets. However, they do not consider the use of longer-run financial targets or needs in their models. In a variety of economic interactions, reputational considerations are intertwined with economic incentives such as promotions, bonuses, or the option value of future job prospects.

[18] Using various definitions of recent expectations such as sales in the prior month or two weeks yield similar results (see Figure A5).

day was recorded. In all, my half-hourly panel contains approximately 459,000 observations for 512 sellers across 180 days in 2018-2019 covering the late-April to mid-August season.

I supplement these panels with daily weather data from the National Oceanic and Atmospheric Administration (NOAA) National Climate Data Center (Menne et al., 2012). I include daily total precipitation, high temperature, and low temperature from the weather station nearest to each seller's working ZIP code as controls. These factors may be important because door-to-door sales is an almost exclusively outdoor job. During these summer months, heavy rain and humid heat greatly increase the marginal cost of effort, and heat can have negative effects on cognitive ability and learning (Park et al., 2020). Alternatively, these factors might keep people inside their homes if the outdoor conditions are inhospitable, making it easier to contact and sell to a population less mobile. The relationship between sales and these conditions is, therefore, ambiguous ex ante.

One theoretical concern in this context is that sellers might be differentially sorted by managers into neighborhoods that are "easier" or "harder" to sell in. I include in my analysis controls for the characteristics of each person's work area. I use ZIP code data from the American Community Survey's 5-year summary files for 2013-2017 to serve as controls. I include variables that are likely to affect demand for pest control services or the ability to pay for them.[19] However, there is essentially no evidence of sorting behaviors correlated with seller performance. Additionally, managers emphasize that making assignments to work areas based on perceived skill or other seller attributes is costly to them as managers and generates unclear returns, which significantly undermines the business case for it. In addition to ruling out sorting, the exercises in Table A3 also suggest that weather does not play a significant role in determining a seller's total output today. Given this, it is also unlikely to be an important determinant of a seller's sales expectations for tomorrow, and the sellers are unlikely to see a need to dynamically adapt their goals and expectations around weather forecasts.

Summary statistics for my two panels are in Table A1. Across all half-hour periods, the average number of sales is 0.16 based on 2.28 pitches. The average number of sales per day across all sellers is approximately two based on 6.9 hours per day, though there is substantial variation. Sellers work in relatively high-income areas. The median household income in their sales areas is \$86,000, and nearly 20% of residents in the average ZIP code have incomes between \$100,000 and \$150,000. Seller work areas are mostly single-family homes (mean of 80%), are predominantly non-Hispanic white (mean of 80%), are relatively highly educated (mean of 45% Bachelor's degree or more), and have stable populations. The number of daily pitches and the number of daily sales are slightly

[19]These variables are median household income, rates of unemployment and poverty, the share of home values in specific ranges, total housing units, the share of units that are owner-occupied or are single-family homes, the share of households with a married couple, the share of adults with a Bachelor's degree or more, and the share of the population that has not moved in the past year. These jointly explain less than 3% of the variation in sales both between and within sellers. Appendix Table A3 provides details of a regression of daily sales on weather and ZIP code characteristics. Only three coefficients are statistically significant at the 10% level. Estimates excluding these controls are nearly identical but slightly less precise. See Appendix D for additional background.

negatively correlated for two possible reasons. First, successful pitches that result in a sale take longer than unsuccessful pitches; second, if sellers exhibit reference dependence, they will work fewer hours and make fewer pitches after reaching their expected sales targets, which reduces total pitches each day. In Table A2, I show the mean and variance of daily sales stratified by deciles of daily pitches. Within deciles of daily pitches, there is a wide variance in daily sales, even within sellers when conditioning on their mean sales outcomes. This means that on days in which an individual seller changes their total effort levels between deciles, they still have a wide variance in their outcomes, consistent with the quasi-random receptivity of households.

From the half-hourly panel, Figure A1 shows the distribution of start and stop characteristics for each working day. Panel A shows that most sellers start their shift with their first knocks and sales between 1:00 PM and 2:30 PM, though there is substantial variation in start times. Some start as early as 10:00 AM, while others begin working in the late afternoon or early evening. After starting their shift, the majority of sellers stop working between the sixth and eighth hours, though a large share stop working for the day before their sixth hour of work. This context and the availability of comprehensive data provide a unique opportunity to test for real-world reference dependence connected to firm-imposed incentives.

4.1 Sales Context and Theory

What does the theoretical framework in Section 2.3 imply for my sales setting? Here, a descriptive example is helpful. When PestCo sets a bonus at 200 sales, the bonus directly affects a forward-looking worker who knows her ability on the job could reasonably yield her something close to that number (e.g. 190 sales). In the case of both the standard model and the Koch and Nafziger (2020) model, she may raise her objective for total sales at the end of the season to be at least 200 because she believes it is attainable and the $2,000 bonus is worth the extra effort. If she has time-inconsistent preferences, absent gain-loss utility, despite her optimization at the beginning of the season, she may fall short of 200 sales, missing out on the bonus that he had originally set out to achieve.

However, because the bonus at 200 has induced her to target 200 by raising the salience of that point and its stakes, she may engage in narrow-bracketing at two sales per day for 100 days, particularly if she expects any present bias. She then works with these two sales per day as her reference point, which satisfies the personal equilibrium condition. Being below her two sales generates negative comparisons and a higher marginal utility for each sale, so she will work harder or extra hours to get the remaining sales. If she does achieve her two sales, she can then quit for the day and feel satisfied with her performance as her marginal utility has declined, leading to a kink in labor supply at expectations. Achieving her two sales then keeps her on track to hit her goal of 200 by the end of the season, solving the self-control problem. In the absence of the 200-sale bonus and a single piece rate, she may have maintained her trajectory of lower total performance (190 sales) and would not have narrow bracketed absent the salience impact of the bonus.

For this behavior to be rational, the seller would need to have rational forward-looking expectations, i.e. reasonably predict her ability to achieve approximately 190-200 sales. In my sales setting, a sign that sellers are forward-looking would be that their daily labor supply does not substantially change as their cumulative performance (and therefore realized commission rate) increases because they have already optimized for their long-run expectations. She would not respond to reaching 100 or 150 sales and yielding an increase in her commission rate because she already expected to reach a total above 150.[20]

4.1.1 Theoretical Predictions

Based on the theory discussed previously, if daily reference dependence and goal setting linked to the bonus schedule occur in this setting, I expect:

(A) There will be a kink in labor supply upon surpassing daily performance expectations because of gain-loss utility around daily expectations/targets. No such shift is predicted by the standard model.

(B) If workers have established their daily targets *as a commitment device* to achieving the long-run objective (the bonus), those that have surpassed their relevant bonus will reduce their effort even though their effective piece rate for each sale has not changed. This is because of gain-loss utility around the bonus threshold.[21]

(C) As a result of both (A) and (B), the distribution of performance around expectations should be subject to bunching. If these daily references are connected to the bonus scheme, these distributions should form around the bonus schedule.

5 Empirical Strategy

5.1 Tests of Reference-Dependent Labor Supply

I first use my half-hourly panel to test for the presence of reference dependence in daily labor supply choices consistent with Prediction (A). As outcomes, I focus on stopping work for the day, a measure common to the past literature, as well as pitches presented in the next half hour, a measure of effort "exertion" conditional on continuing to work.[22]

My empirical approach approximates an experimental ideal in which sales performance relative

[20] While this assumption of a non-myopic view is reasonable, it is not certain; evidence from other contexts indicates that myopia affects the optimality of decision-making in areas like pension planning (Mitchell, 1988), health behaviors (Cawley and Ruhm, 2011), and take-up of financial aid (Bettinger et al., 2012).

[21] This rules out status quo anchoring or habituation as explanations by indicating that these workers are not simply anchoring their reference dependence to *expectations* separate from their *goals*—in other words, the two are directly connected. In the standard model, to a first approximation, rational agents have already optimized their labor supply around the present value of their predicted compensation (including the bonus), so the timing of a one-time bonus should not lead to significant changes in labor supply (Kahneman and Thaler, 1991).

[22] I also examine the probability of recording any knocks during the next half hour. This measures whether workers are more likely to take breaks as a result of their position relative to expectations. These results closely mirror the knocks-based exertion margin. See Appendix Figure A4.

to expectations is randomly assigned each half hour by netting out conditions correlated with effort costs and the number of sales a seller has generated to that point. The underlying assumption is that conditional on my various fixed effects and controls, the exact number of sales a seller has at a particular point in the day and the likelihood of getting a sale at the next door are as good as random. Given the context in Section 3 and the set of controls and fixed effects I present below, this assumption is reasonable. The sales setting presents a unique opportunity to study this behavior because whether a sale occurs or not depends strongly on who answers the door when a seller knocks—similar to a small lottery.

I first estimate a non-parametric model of labor supply with respect to each seller's distance from their sales target to trace out patterns without imposing a functional form. Following the past literature (Crawford and Meng, 2011), I define expectations and targets in all my models as the average daily sales from all past workdays in the season specific to each day of the week (i.e. a specific mean for Mondays, Tuesdays, etc). I note here that PestCo runs various competitive tournaments during the sales season of three different types. Because these significantly change the incentives faced by the sellers and may shift the workers' target for the day, I separately analyze behavior during non-tournament days and present those results in my tables and figures. For completeness, I also estimate my parametric models using a pooled sample across all tournament and non-tournament periods and interact my coefficients of interest with indicators for tournament periods. This approach modestly increases the sample size for each seller and explicitly accounts for the different incentives introduced during tournament periods, which leads to some gains in precision. These estimates are in Appendix Table A5. I report the non-tournament coefficients in Figure A8 and find even stronger evidence of reference dependence in these models, including statistically significant but quantitatively small intensive margin effort effects under this specification (see Section 6.1.1). See Appendix D for more on these tournaments.

For seller i during half hour of the shift t and half hour of the day h on day of the week d in week of the season w in year a, I estimate the following model:

$$\begin{aligned} y_{ithdwa} =& \beta_0 + \sum_{e=-k, e\neq 0}^{k} \beta_e * \boldsymbol{I}_e\{sales_{ithdwa} - \overline{Sales}_{idwa} = e\} \\ & + \alpha X_z + \sigma W_{dwa} + \mu_{it} + \eta_h + \nu_d + \omega_w + \tau_a + \varepsilon_{ithdwa} \end{aligned} \tag{3}$$

Here, y is the probability of stopping work for the day as well as the number of pitches presented to a resident in the next half hour. The expression $\{sales_{ithdwa} - \overline{Sales}_{idwa} = e\}$ represents the seller's current distance to expectations: her current cumulative sales that day ($sales_{ithdwa}$) minus the worker's average daily sales specific to the day of the week ($\overline{Sales}_{idwa}$). $\boldsymbol{I_e}$ is a dummy variable assigned to each distance value. The coefficients of interest, β_e, capture non-parametric effects of being e distance from one's expectations target. Distance values below zero are characterized as being "losses" and values above zero are "gains." Because sales are discrete values, these coefficients include values rounded to the nearest integer, with the (0,1.5) interval being included in β_1.[23] Under reference dependence, beginning with β_1 there will be an upward change in stopping probability or a downward change in exertion as the distance from expectations increases.

The various fixed effects (μ_{it}, η_h, ν_d, ω_w, τ_a) are for seller by half hour of the shift, half hour of the day, day of the week, week of the season, and year, respectively. Importantly, μ_{it} captures a seller-specific baseline hazard over the shift. That this factor is omitted by the prior literature is noted by Thakral and Tô (2021). They include a driver-specific hazard in their estimates of taxi driver behavior and conclude this is vital for unbiased estimates of labor supply responses to daily earnings. I incorporate this methodological improvement into my estimates. The X vector is the set of ZIP code characteristics from the ACS, and W is the set of weather controls from NOAA discussed in Section 4. Importantly, these controls rule out any relationship that might arise between sales and factors correlated with the length of the shift (e.g. running up against a maximum work hours limit or fatigue), the time of the day (e.g. an 8:30-9:00 PM hard stop time), weather, or cross-sectional differences between each worker in their "normal" work schedule.

In my main models of interest, I fit parametric estimates that impose a functional form to match the non-parametric estimates in Equation 3, with linear splines divided at zero:

$$\begin{aligned} y_{ithdwa} =& \beta_0 + \beta_1\{sales_{ithdwa} - \overline{Sales}_{idwa}\} \\ &+ \beta_2\{sales_{ithdwa} - \overline{Sales}_{idwa}\} * \boldsymbol{I}_{sales \geq \overline{Sales}} \\ &+ \beta_3 * \boldsymbol{I}_{sales \geq \overline{Sales}} \\ &+ \alpha X_z + \sigma W_{dwa} + \mu_{it} + \eta_h + \nu_d + \omega_w + \tau_a + \varepsilon_{ithdwa} \end{aligned} \tag{4}$$

This approach allows the slope of the relationship between labor supply and distance to one's reference point to differ in the gain and loss domains. $\boldsymbol{I}_{sales \geq \overline{Sales}}$ is a dummy for if current sales are above expectations, or in other words, for entering the gain domain. β_1 defines the slope of the relationship between one's current distance to average sales and labor supply in the loss domain. β_2 captures the change in slope upon crossing the reference and entering the gain domain. Finally, β_3 captures any discontinuous level shift in stopping probability or effort from reaching the reference. The fixed effects and controls are all the same as in Equation 3. In a standard framework, there should be no sudden change in the slope and no discrete level shift upon reaching the reference point. Under reference dependence with loss aversion, we would expect to see an upward change in the slope of stopping probability. In other words, β_2 will be significantly positive in the stopping

[23]Other studies examining reference dependence discretize earnings into ranges. The "correct" size of the earnings range has been the topic of some disagreement (Farber, 2015; Martin, 2017; Thakral and Tô, 2021). In sales, earnings are already discrete, so I do not have to impose a bin structure. Because the common support in the distance to expectations is thin outside the [-4,4] interval, I plot that interval in my figures. I report the full set of distance dummy coefficients corresponding to my figures in Appendix Table A4.

model. The coefficient β_3, while not predicted by simple loss aversion, represents a discrete penalty for "losing," or for falling short of expectations, which suggests reference dependence.[24] If β_2 and/or β_3 are significant and positive in the stopping model, this represents strong evidence of reference dependence.[25]

5.2 Tests of Goal-Setting and the Bonus Schedule

A prerequisite for setting long-run goals is that sellers must not be myopic. To test for this, I estimate how sellers adjust their labor supply as their cumulative sales increase throughout the season. Sellers only know their final earnings per sale at the *end* of the sales season after their total number of sales and total revenue are calculated. If sellers have realistic, forward-looking expectations for what they can achieve (perhaps after an initial learning period of a few weeks), perceived changes in their wages that come with entering a new 50-sale performance interval should not change their daily labor supply because they have already optimized over their chosen long-run outcome. Conversely, myopic agents would respond to an increase in their realized commission rate, which is inconsistent with long-run planning. In Appendix B, I show that daily performance across the season is remarkably stable: mean performance after the first two weeks of the season explains 75% of the variation in final season sales, while sales after the first 5 weeks explain nearly 90%. This also indicates very little switching of targets and a consistent personal equilibrium after the first few weeks. Sellers appear to select bonus thresholds early on and work consistently with those targets each day.

According to Prediction (B), sellers that have previously worked in the pursuit of their long-run targets should reduce their labor supply once they surpass (or will imminently surpass) their long-run targets. To test both of these dynamics, I use my daily panel and apply two models. In the first model, I regress hours worked per day on indicators for 10-sale intervals of current cumulative sales interacted with indicators for 50-sale bins of total sales at the end of the summer. I estimate the following equation for seller i on day of the week d in week of the season w in year a working ZIP code z:

$$\begin{aligned} y_{idwa} = &\beta_0 + \sum_{e=[0,10)}^{[320,330)} \sum_{f=[100,125)}^{[300,325)} \beta_{ef} \boldsymbol{I}_e * \boldsymbol{I}_f \\ &+ Efficiency_{idwa} + \alpha X_{zdwa} + \sigma W_{zdwa} + \mu_i + \nu_d + \omega_w + \tau_a + \varepsilon_{idwa} \end{aligned} \tag{5}$$

The outcome variable y is the number of hours worked per day. The indicators $\boldsymbol{I_e}$ and $\boldsymbol{I_f}$ are indicator variables for currently working in interval e and for having total season sales in interval f.

[24] Estimating with second-order polynomials results in small and statistically insignificant coefficients on the squared term for both outcomes.

[25] To ensure that the saturation of these fixed effects does not induce spurious correlations and that my results are not solely dependent on these, I estimate the model including only seller fixed effects in Figure A3. The finding of a discontinuous change in the slope of labor supply at expectations persists, and, in fact, the change in the slope relative to baseline values is nearly identical (2.9).

In this specification, β_{ef} captures the non-parametric effects of being in interval e for an individual whose total sales for the season were in interval f. These coefficients trace the labor supply path of those who ended with a similar total number of sales. The X and W vectors are the same as Equation 3. The $Efficiency$ variable is a time-varying measure of each seller's average sales per hour for all past workdays that season, which proxies for sales ability and may evolve as the season progresses. Changes in this measure capture learning effects over the season, which shifts the expected marginal earnings of an additional period of work. I include fixed effects for seller (μ_i), day of the week (ν_d), week of the season (ω_w), and year (τ_a). These fixed effects ensure that the β coefficients characterize within-seller choices holding constant other characteristics of the sales season, fatigue, or learning effects. Importantly, the week of the season fixed effects (ω_w) control for dynamics like being in one's final week and taking days off, holiday weeks, and any other average changes in hours and output that might vary over the weeks of the season. If the β_e coefficients are constant within different types of sellers f as they cross intermediate 50-sale intervals, then it does not appear that sellers are responsive to a change in their realized wage.

If sellers are focused on reaching a bonus threshold, the coefficients for β_e will be much smaller after crossing the worker's final bonus threshold. This would result in a significant drop in hours worked. For example, a worker who finished with 150-175 sales (just beyond the 150-sale bonus threshold) would work fewer hours in the intervals just at or after the bonus threshold. Equation 5 captures this dynamic for each bonus threshold from 100 to 300 sales.

As a final, formal test of loss aversion around the bonus threshold, I use my panel of daily sales to estimate a similar model to Equations 3 and 4, but replace distance from daily expectations with distance to the seller's nearest bonus threshold and use hours worked each day as the labor supply variable. I include the $Efficiency$ as a control variable as above and fixed effects for seller by day of the week, week of the season, and year. A discontinuous change in hours at the bonus threshold allows me to compare the relative change in slopes around the bonus to the relative change in slopes around daily expectations. Given the fact that daily expectations are based on non-binding internal targets, while the bonus has been the focus of daily targeting and comes with significant stakes, it is reasonable to expect a larger relative reduction in labor supply around the bonus.

Finally, I use my daily panel of sales to visualize how workers respond to bonus incentives with the predictions discussed above, I present kernel density estimates of cumulative sales at the end of the season and as well as throughout the season for all workers. I present densities at two-week intervals to illustrate the evolution of sales over time.

6 Sales Results

6.1 Do Workers Exhibit Reference-Dependent Labor Supply?

Figure 2 shows each of the coefficients from the non-parametric estimates from Equation 3 as well as the linear estimates from Equation 4. Panel A indicates that as sellers approach their target from the loss region, the probability that they stop working for the day is relatively flat at a slope of 0.0021. After surpassing their target number of expected sales, there is a clear upward kink in the probability of stopping work. The slope of the relationship between cumulative sales and stopping probability in the gain region for expectations-based targets is 0.0058, or 2.8 times that in the loss region. For context, the average probability of stopping right at the reference point is 0.079, so an increase in this probability of 0.0058 for each sale past the reference point represents a 7.3% increase.[26]

Panel B of Figure 2 shows the same estimates for pitches during the next half hour, which is a measure of exertion conditional on continuing to work. In contrast to the results for stopping probability, there is a relatively smooth relationship between exertion and sales each day and minimal change in this measure at the reference point. The size of the decline is small in percentage terms: each sale reduces effort conditional on continuing to work by approximately 1% from a baseline mean of 2.38 pitches. These results suggest that reference dependence in exertion is negligible, but that there is a steady decline in effort as sales increase. Reference dependence is most apparent at the extensive margin. In other words, if sellers stay on the job after reaching their expectations, their exerted effort is similar.[27]

An alternative theory behind the behavior in Figure 2 might be that sellers are actually focused on a target labor supply (e.g. pitches) in order to smooth effort across days and that their target pitches might be correlated with their target sales. As a simple test, I estimate my stopping model using distance to the expected number of pitches as the operative independent variable. I present the results in Figure A2. This shows that there is no meaningful relationship between distance to expected number of pitches (e.g. expected labor supply) and stopping behavior, consistent with the idea that sellers are targeting expected output rather than expected inputs.[28] Furthermore, if sellers are targeting hours rather than pitches, my non-parametric controls for the amount of elapsed time in the shift would result in no change in the relationship between sales and stopping behavior at sales expectations.

A related question is whether workers instead target weekly sales rather than daily sales. I use a similar estimation framework to estimate seller responses to crossing weekly expectations using

[26] Likelihood ratio tests confirm that this functional form significantly outperforms a smooth quadratic relationship across the zero cutoff, with a p-value of 0.0076, meaning that a quadratic form cannot accurately describe the relationship between sales performance and observed stopping behavior.

[27] In two of my later specifications, there is a downward kink and a discontinuity in effort. I also examine the probability of actively knocking, meaning recording any knocks during the next half hour. These results in Appendix Figure A4 closely mirror the exertion margin. At a mean active knocking share of 80%, the slope estimates are quantitatively small and not economically meaningful. Sellers do not appear more likely to take breaks during their work as a function of their position relative to expectations.

[28] Across all sellers, the correlation between expected sales and expected number of pitches in a day is slightly negative (-0.18 for averages specific to the day of the week and -0.22 for averages across all days of the week). This may be because successful pitches that result in a sale take longer than pitches that result in a rejection, allowing the seller to move on to the next house.

my daily sales panel. I estimate daily total hours and daily pitches as a function of distance to weekly total expectations and include fixed effects for seller by day of the week, week of the year, and year. Table A8 shows that there is no meaningful shift in the slope of the relationship between measures of effort and distance to weekly goals, indicating that the sellers are not exhibiting loss aversion around weekly goals and, therefore, do not appear to be targeting weekly goals.

One notable feature of the Koch and Nafziger (2020) and Kőszegi and Rabin (2006) models is that personal equilibrium is established immediately following the formation of expectations, even when the final payoff is far away. Under daily goal setting, then, there should be evidence of kinked labor supply even early in the sales season when exact performance relative to the bonus threshold is not immediately payoff relevant. To investigate this, I separately estimate my models during the May, June, and July months. In May, opportunities for effort substitution are more plentiful. Appendix Figure A7 presents these estimates. The evidence is consistent with my overall results, even in May and June. Similarly, if expectations are stable throughout the year, we might expect similar behaviors from sellers who have experience and those who are new to the job. Panel D shows that this is the case. There is little difference in behavior around expectations for those with more experience with the possible exception of experienced workers (who may be more sophisticated) being slightly *more* likely to stop working for the day right at the zero cutoff.

I use my estimates to calculate the parameter of loss aversion. Köbberling and Wakker (2005) suggest an approach to measuring loss aversion focusing on the difference in the slopes of the utility function in the gain domain and the loss domain, which is invariant to outcome (sales) scale. Because my empirical model partials out all covariates correlated with effort costs and because the timing of sales is conditionally random, the only difference between the gain and loss domains is the difference in the marginal benefit, i.e. the ratio of slopes for each outcome measures $U'(0)_{\uparrow}/U'(0)_{\downarrow}$. The ratio of slopes in the stopping model is $0.0058/0.0021$, or 2.8, and the slopes at the "exertion" margin have a ratio of 1.5. Using a bootstrap with 250 replications for inference on the ratio of these slopes in the stopping model yields 95% confidence intervals for my estimates of loss aversion of 1.2 to 4.4 (1.5 to 5.3 for the bias-corrected interval). My estimate of 2.8 in my baseline models is the most conservative of my stopping model estimates, and my non-parametric estimates imply an even larger ratio. Other specifications, which I detail in Section 6.1.1 yield estimates as high as 4.1 or 5 for the stopping model and 3.9 at the margin of effort conditional on continuing. A coefficient of loss aversion in my results of 2.8 is, therefore, very consistent with the prior literature.[29]

[29] In their review, Gächter et al. (2007) find loss aversion of approximately 1.4 to 4.8 across various measurements, with an average value of 2.6.

6.1.1 Robustness and Alternative Specifications

My first alternative specification pools together all tournament and non-tournament periods and interacts each of my key measures of distance to daily expectations with indicators for what kind of tournament or non-tournament is in operation. This allows the effect of crossing the reference to differ based on period type. The results of this specification are in Figure A8. The result for non-tournament periods is a more pronounced upward kink in stopping probability and the emergence of a downward kink in pitches per half hour. The ratio of the slopes across the target is 4.1 times in this stopping model. At the exertion margin, the slope in the gain domain is 3.9 times that in the loss domain. This specification confirms the results of my baseline model and provides even stronger evidence for reference dependence. On pitches, the results are still qualitatively small at a change of less than 2% per sale above the reference.

My estimates impose a linear structure with a cutoff at each seller's average daily sales, in line with (Kőszegi and Rabin, 2006). As a robustness check against incorrect specifications of the cutoff at zero, I estimate my models again using non-linear least squares. I first residualize the probability of stopping with my full battery of fixed effects and controls and use the residuals in my non-linear least squares estimate. Rather than imposing slope and intercept coefficients at zero, I allow the cutoff itself to be a parameter of the model. The results are in Table A6. The non-linear least squares estimates confirm that there is, indeed, a structural break at the worker's average cumulative sales and a strong upward tilt in stopping probability. The exact cutoff in the non-linear least squares estimate is 0.11, approximately one-tenth of a sale from my measure of the seller's expectations, which may be consistent with the "buffer" idea against later fatigue, which is explicitly invoked by respondents in my online experiment.[30] The ratio of slopes in the stopping model is 5.1, meaning that my baseline estimates may be quite conservative. For pitches per half hour, even though the estimates show a statistically significant kink downward and that the ratio of the slopes across the reference is 3.3, the magnitude is small in percentage terms; each sale past the reference leads to a 1.8% decline in effort conditional on continuing to work.

As an additional test, I estimate my baseline model but include exertion effort on the right-hand side: cumulative pitches that day as a measure of total exerted effort. If a worker is exerting a high level of effort on the job and becomes fatigued, the fatigue could be affecting her willingness to continue working or to exert effort in the next half hour. Table A7 presents these estimates for my parametric models. The results are nearly identical to my baseline model, implying that my baseline model adequately controls for effort differences at the intensive margin that may have generated differences in sales.

As a check that my specific measures of expectations based on day-of-the-week specific mean performance are not driving my results, I also estimate my models using two different versions of

[30] This is consistent with Kőszegi and Rabin (2007). Here, the anticipation of risk leads to a strong tendency toward planning and the purchase of insurance. This performance buffer may, therefore, be a form of small-scale insurance.

expectations in Figure A5: mean performance across all days in the prior month (Panel A) as well as the prior two weeks (Panel B). These result in nearly identical parametric estimates of behavior around expectations, in part because expectations from the prior month and two weeks are highly correlated (≈ 0.90) with expectations based on longer windows and days of the week.

As a final check, I perform a placebo exercise in which I assign a daily target based on a random choice of rational expectations. I assign these based on drawing from a normal distribution approximating the actual distribution of daily targets in the data. The results in Figure A6 show that there is no meaningful change in the relationship between stopping probability and sales progress at the placebo target. The estimate of the change in the slope parameter is not economically or behaviorally meaningful at 0.0008 and is not statistically significant. The non-parametric estimates confirm the pattern as well. This provides further confidence that the location of the target and the worker's stopping behavior relative to it are not capturing other aspects of the job correlated with effort or other unobserved factors.

6.2 Is the Bonus Schedule Inducing Reference Dependence?

I now provide evidence that sellers are forward-looking and that they significantly reduce their labor supply upon nearing or passing their relevant bonus threshold, relevant to prediction (B). Estimates from Equation 5 are summarized in Figure 3, which shows the predicted hours worked per day over 10-sales increments of cumulative sales from this model. Each line shows the labor supply trajectory of different bins of total sales at the end of the season. Sellers do not appear to be myopic. After an initial adjustment period, those whose sales totaled over 300 quickly began working approximately 8 hours per day, while those with fewer than 200 total sales worked approximately 7 hours per day consistently over their accumulated sales. Notably, within tiers of total sales, there is very little variation in the predicted hours worked over current cumulative sales, and labor supply does not significantly shift upon receiving a commission raise by crossing into a new 50-sale interval. These results show that workers do not change their work hours regardless of how much of a commission increase they have secured, suggesting a singular focus on long-run performance expectations.

These patterns shift significantly once the relevant bonus threshold has been reached. Even when conditioning on elapsed time in the season, weather patterns, and efficiency gains, sellers drastically reduce their work hours by 1.5 hours per day ($\approx 20\%$) after passing the bonus.[31]

Supporting this point, the results of estimating the daily analogs of Equations 3 and 4 around the bonus threshold are in Figure 4. These formally show the reduction in labor supply in action. The ratio of the slopes suggest loss aversion parameters far larger than those in the case of daily targets.[32] This is, perhaps, unsurprising given that the bonus represents the culmination of daily

[31] For the most productive groups around 250-325 sales, who represent the most talented and experienced sellers, there is a small reduction in work hours multiple intervals before they reach their bonus threshold, consistent with having unexpectedly surpassed their planned pace toward their target.

[32] Approximating the ratios with the individual endpoint coefficients at -100 and 20 suggests a ratio of approxi-

reference dependence as well as the achievement of a large monetary windfall. This sudden decline in effort in Figures 3 and 4 is not predicted by the standard model because a standard agent would have already optimized their hours around a bonus threshold ex ante, leading to constant effort across the threshold. Because the marginal returns to the next sale have not decreased–and have actually *increased*–the standard model would predict smooth daily labor supply around the threshold.

Altogether, the dynamics above lead to significant bunching in the distribution of performance around the bonus threshold, consistent with prediction (C). Figure 5 shows the results of kernel density estimates for total sales at the end of the season, representing the culmination of effort and performance each day as well as around the bonus.[33] Around each 50-sale bonus threshold, there is significant bunching, particularly at 150 and 250 sales when the bonuses include rent payments and the company vacation. This further emphasizes that the bonuses are salient for the sellers and that sellers are making choices with these in mind.[34]

Overall, the combination of clear gain-loss utility around daily targets in Section 6.1 and clear shifts in labor supply at the bonus in Section 6.2 suggest reference dependence is operating at both margins and that the two are connected in a way consistent with my theoretical framework and predictions. Specifically, sellers are: 1) able to predict their own performance very early in the sales season, forming rational recent expectations; 2) aware of and responsive to the bonus schedule; 3) setting long-run goals subject to gain-loss utility around bonus thresholds in the schedule; 4) distributing their long-run goals into daily expectations subject to gain-loss utility; and 5) significantly reducing their labor supply upon reaching or surpassing the bonuses. These provide clear evidence of these behavioral dynamics.

7 Experimental Design

One caveat to my analysis of the bonus schedule and reference dependence in my sales setting is that I do not observe a state of the world in which there is no bonus scheme in place. To address this in a causal framework, I conduct an online real-effort task experiment in which I experimentally manipulate non-linear compensation and then observe effort behavior across shorter time periods. This allows me to directly and causally attribute short-run reference dependence to non-linear compensation schemes.

mately 5-6.

[33]If a seller left relatively early in the season, their sales are included in the total as of the date they left and hold the same value as the weeks progress, so the relatively high density below 100 includes those who only worked a portion of the season.

[34]Figure A9 breaks down the evolution of the distribution of cumulative sales for each seller over the season. In Figure A10, I present the densities for seller who ended around the 200-sale bonus. These group-specific kernel density estimates show how the variance of performance for this group stays fairly constant throughout the season and then subsequently narrows as those above the bonus reduce their labor supply, as in Figures 3 and 4. For example, for the group centered around the 200 sales bonus, the distribution has a standard deviation of approximately 20 sales from weeks 6 through 12 and skewness of between 0.5 and 0.9. In week 14, the standard deviation falls to 15 and skewness to approximately zero.

I conduct an online real-effort task experiment on the Prolific platform. In the experiment, the participants engage in a simple button-pushing task in which they are asked to alternate pressing the “a” and “b” buttons on their computer keyboard, following closely the procedure in DellaVigna and Pope (2017). A successful sequence of “a” and “b” results in 1 point, which is tallied on the screen. Participants were asked to perform the task for a total of ten minutes in four rounds, each lasting two minutes and thirty seconds. The participants had a break of ten seconds between rounds. Before beginning the first round, participants were given an open (non-incentivized) test period of 50 points to familiarize themselves with the task, understand the on-screen point counter, and briefly assess their speed.

Button pushing requires constant engagement with the keyboard, and can quickly become tiresome (and boring). One might ask how self-control problems might be manifest in such a short period of ten minutes, but the prior literature shows that dynamic inconsistency can occur over time periods measured in minutes (McClure et al., 2004; Brown et al., 2009). For example, Brown et al. (2009) estimates β to be approximately 0.6 to 0.7 in a task of 45-90 minutes. Furthermore, the adoption of short-run goals might arise even if people could reasonably expect that present bias could occur across discrete periods. If the salience effect of the non-linear payment is large enough, the scheme can increase the incentive to adopt subdivided benchmarks so long as $\hat{\beta} < 1$.

A total of 1,464 recruits completed the task. Each participant was paid a flat \$3 to participate. I then randomly presented participants with one of three incentive conditions:

1. A bonus of \$1 for achieving 2,000 points (2,000 bonus condition)
2. A piece rate of \$0.05 per 100 points (the piece rate condition)
3. A bonus of \$1 for achieving 2,400 points (2,400 bonus condition)

Importantly, the payment rates were calibrated based on the distributions of performance in DellaVigna and Pope (2017) to have equal predicted mean performance over the full ten minutes, meaning the expected payoff for a performance of 2,000 points (the approximate mean performance in that experiment) is exactly equal in the first bonus condition and the piece rate condition. The core difference between the two is that the bonus makes the 2,000-point target salient and payoff-relevant for the end of the 10-minute task period. The bonus environment in my sales setting follows this setup closely. Upon achieving the bonus, participants in the bonus conditions were not given any additional incentives.

Importantly, a narrow bracketing agent and an agent in the standard model who is subject to the bonus condition are both expected to exhibit “bunching” behavior at the end of the fourth round at the bonus threshold, since their final output is payoff-relevant, and there is no additional incentive beyond the bonus in their experimental condition. The core difference between the two is that the standard agent has no payoff-relevant incentive to ensure that their exact performances in Rounds 1-3 happen to be exactly above 500 points given that they can smooth and substitute effort across rounds. By contrast, a narrow bracketing agent experiences loss utility if they fail to meet 500 points (the average they would need to accomplish the goal) in Rounds 1-3, leading to bunching in *every round*.

In the prior literature, in the case of the self-employed or pure piece rates (like in the taxi driver case), workers may exhibit reference dependence in pursuit of a longer-run target even if they do not have self-control problems but suspect that they might, so long as the salience effect of a long-run target is large. The imposition of goals by, for example, a firm can *intensify* the use of internalized gain-loss utility as a commitment device to increase performance. Exogenous manipulation of targets (e.g. by a firm) is not a necessary condition for reference dependence with goal-setting, but increases the incentive to do so. This experiment does not explicitly test for present bias, but the randomization rules out imbalances in underlying present bias across treatment conditions and, therefore, isolates differential responses to the compensation incentives.

8 Experiment Results

I now present the results from my experiment from each round in Figure 6. Here, I primarily focus on the comparison between the bonus payment at 2,000 points and the piece rate condition, which has shown comparable mean performance in prior studies (DellaVigna and Pope, 2017).[35] Panel A of Figure 6 shows the distribution of performance in the first round comparing the piece rate treatment and the bonus condition at 2,000 points. Panel B shows the same for the subsequent rounds. Panel D shows the same for total performance.

Several notable patterns emerge. First, there is substantial bunching in the distribution of performance. Importantly, the distributions in the bonus condition exhibit heaping to the right of the piece rate condition in every round. This kind of immediate, short-run bunching is consistent with reference dependence rather than the standard model.

Panel C presents the differences in the densities of round 1 performance between the two conditions and shows that the bonus conditions are heavily concentrated above the 500- and 600-point thresholds. Next, despite having the same payoff at 2,000 points, the distribution of performance is consistently narrower in the bonus condition than in the piece rate condition, not only in the distribution of end performance but also for the first three rounds. Because the only difference between the two groups is the compensation condition, the bonus condition itself is *causally* inducing participants to adopt (or increase their use of) round-specific reference points.[36]

To summarize the difference between the bonus condition and the piece rate condition more concretely, Figure 7 shows the density of average per-round performance in 5-point bins for the two conditions for the first three rounds (Panel A) and for all four rounds (Panel B). 500 points is the average each round that those in the bonus condition would have to perform in order to achieve

[35] Figure A11 shows a round-by-round comparison of the bonus condition at 2,400 points and the piece rate and exhibits remarkably similar dynamics.

[36] In the piece rate condition, participants may be anchoring to round numbers. That there is more heaping in the bonus conditions is important because the piece rate represents a counterfactual that takes into account any tendency to bunch at round numbers.

the bonus at the end of the fourth round. Each side of the 500-point cutoff is approximated with a simple quadratic function of the density, and the solid lines denote the excess mass accruing to the right side of the cutoff. From this summary measure, it is visually clear that the missing mass between 400 and 475 points is exceeded by the mass in the bonus condition between 500 and 550 points, after which the smoothed distributions are nearly identical up to 600 points.

Formally, I calculate excess mass in rounds 1-3 by comparing the estimated densities using these quadratic splines, i.e., measuring the ratio of the estimated densities in the 500-600 point range. This results in an excess mass of 9.4% (with a bootstrapped standard error of 0.504%) in the bonus condition. This is comparable to the 9.2% - 11.7% excess mass among loss-averse electronic tax filers around a zero final balance in Engström et al. (2015), which represents clear evidence of loss aversion in tax filing. Thus, workers in the 2,000-point condition are engaged in considerable bunching around this threshold every round as a result of the non-linear payment mechanism, even though both conditions have the same ex ante expected payout at the expected mean performance. In other words, even in the first three rounds when the exact performance in each round is not payoff-relevant, participants are exhibiting loss aversion in an effort to surpass *at least* the average required performance each round to get their bonus.

Another important observation is that the left tail in the density of round-specific performance increases in mass substantially in the final round and does so only in the bonus condition (and the same is true in Figure A11 for the bonus at 2,400 points). The result is a steep increase in the variance of effort during the final round in the bonus condition, whereas the variance does not exhibit the same behavior in the piece rate scheme.

To further test the experimental analog of prediction (B) in this experimental setting, it is useful to examine final round effort based on how close the relevant bonus is at the end of the third round to examine who is filling in the left tail of the fourth round point distribution. Figure 8 shows a scatterplot of this relationship across both bonus thresholds (2,000 and 2,400). Panel A shows two interesting features. First, despite not being paid any more after surpassing the threshold, there is a noticeable number of participants who are persistently "better" at the task than others, continuing their behavior from before into round 4. Second, there is a significant reduction in output for those nearing the threshold and those who had already passed it. When I condition on round 1 output as a way of proxying for persistent ability or habit in the task, this downward shift in output near the bonus is even more apparent. Together with the significant excess mass in rounds 1-3, this behavior aligns closely with participants targeting their overall effort to the bonus threshold, distributing that bonus into early-round expectations, and exhibiting reference dependence around both sets of targets.

The panels of Figure 6 demonstrate other interesting results. There appear to be two heaping points at 500 points and 600 points and this heaping is more pronounced in the bonus conditions.[37]

[37]Payoffs above 2,000 points are also *higher* in the piece rate condition, so more bunching at 600 (above what

In Figure A11, the heaping points are higher when the bonus is set at 2,400 compared to 2,000 points (closer to 700), meaning that round-specific sub-goals are responsive to the location of the end goal. When respondents were asked an open-ended question about their strategy, many in the 2,000-point treatment responded they targeted 600 points for each round as a buffer against fatigue or surprises in later rounds in order to avoid missing the 2,000-point threshold. Similarly, many of those in the 2,400-point bonus treatment stated that 700 was their round-specific target as a hedge against risk.

Panel D of Figure 6 shows the total performance across the three experimental conditions. These densities affirm the predictions discussed above in the sales data. The variance of total performance is lower in the bonus conditions than in the piece rate condition. Performance is higher with the bonus at 2,400 than the bonus at 2,000. Interestingly, despite cumulative performance of 2,000 points not being relevant at all to payoffs in the 2,400-point bonus condition, there is still substantial distributional heaping at 2,000 total points. Thus, non-linear payments create or make salient personal targets, even if those targets are not immediately relevant for final payoffs.

Overall, this dynamic behavior is particularly interesting because the time period of the experiment is very short–ten minutes–so one would expect present bias itself to be relatively low, though still meaningful (McClure et al., 2004; Brown et al., 2009). However, these behavioral responses were triggered among participants by three simple features: a task requiring tiresome continuous effort, discrete action periods to act as simple mental accounts, and a non-linear payment.

These results have significant implications for firm costs. In the experiment, the average bonus payouts for the piece rate condition were \$1.15 per worker compared to \$0.80 for the bonus condition at 2,000 points, representing a statistically significant reduction in per-person costs of 31%. Meanwhile, the average total output for the 2,000-point bonus condition was 1.58% lower and not statistically different from the piece rate condition. In the 2,400-point bonus, total payouts were only \$0.62 per worker, while output was slightly and marginally significantly higher than the piece rate. Thus, consistent with the Koch and Nafziger (2020) model of increased performance and goal attainment through short-run reference dependence, the core reason for these differences is attributable to the bonus condition leading to gain-loss utility in each round, which increases the likelihood of attaining a minimum performance threshold in the "long run" at a lower cost. This approach to establishing expectations via non-linear payments is advantageous to the firm, meaning that, in the framework in Section E.2, these strategies would be preferred until net monitoring and enforcement costs reach nearly 30% of output costs.

Finally, after the end of the task, I asked each participant an open-ended question: "Did you have any particular strategy when performing the task across these rounds?" Most participants shared information about their hand placement or other physical movements. However, 35 respon-

would be necessary to achieve 2,000 total points if that performance continued) in early rounds in the bonus condition is notable.

dents explicitly stated unprompted that they had an internal target of 500 per round, 29 of whom were presented with the bonus conditions. An additional 14 bonus participants pinpointed 600 as their target compared to only two in the piece rate, meaning that bonus condition participants were five to seven times more likely to articulate this type of targeted goal-setting as their primary, salient focus across rounds.[38]

Taken together, these results make clear two key empirical observations from a controlled experimental setting. First, workers do exhibit reference-dependent preferences over short-run performance as a means of holding themselves accountable for achieving a certain level of performance, even during the short experimental window. Both revealed effort and explicit declarations of their internal thought processes confirm this result. Second, firms (i.e., the experimenter) can *causally induce* such behavior by using non-linear payment schemes to make long-run and short-run reference points salient and meaningful. Inducing such short-term goal-setting via these payment schemes generates substantial performance improvements relative to the monetary costs by inducing workers to impose psychological costs upon themselves in the form of gain-loss utility.

9 Discussion and Conclusion

Using novel data from a door-to-door sales company and an online experiment, this paper provides evidence of reference-dependent preferences in daily labor supply. Door-to-door sales workers exhibit loss aversion around expectations-based references in their labor supply choices. I find that the extensive margin choice (when to stop working) is the margin at which reference dependence is most operative. I show that by making particular points in the final performance distribution salient and consequential, the firm's bonus schedule for end-of-season sales facilitates this behavior: workers impose upon themselves daily goals and exhibit loss aversion around these targets in the service of attaining the bonus. Upon surpassing the relevant bonus threshold, they drastically reduce their labor supply even when they are still paid a significant piece rate.

My online real-effort task experiment confirms the causal interpretation of the sales data and shows that by simply having a bonus payment that makes a particular long-run target salient, workers respond by imposing upon themselves short-run targets that serve as reference points. They do this as a commitment device in order to achieve their (and the firm's) larger goals. That firms can induce this type of short-run adoption of gain-loss utility has not been demonstrated before in the literature.

This result provides new information about *why* workers might exhibit reference dependence in their labor supply choices—as a rational response to expected present bias. Loss aversion around daily goals acts as a commitment device, keeping these workers engaged in work when they might otherwise shirk today assuming they can catch up tomorrow. Narrow bracketing of goals, therefore,

[38] Some examples of responses include: "I ... was trying to get over 600 in the first rounds since I knew my fingers would be tired by the last round." "I made it a goal to get to 500 on each." "[I went] as fast as I possibly could ... to go over the 500 mark per round for the first three rounds."

frequently induces effort by keeping workers in a loss domain at the start of each day, leveraging one behavioral bias (loss aversion) to overcome another (dynamic inconsistency), and workers self-impose these preferences. These results provide a clear dynamic mechanism for why non-linear compensation schemes are effective in raising worker output: workers adjust their short-run labor supply choices to reflect gain-loss utility, which combats present bias. Unlike in prior settings, this daily reference dependence does not require costly real-time enforcement of daily performance or loss-framed contracts. These results also explain the daily income targeting observed in prior settings. Given the unpleasantness of door-to-door sales, self-control problems are likely widespread in the occupation. These results also reveal a key mechanism behind the effects of non-linear incentives in the workplace: the establishment of expectations. The online experiment confirms that compensation schemes used by firms can "make" or reinforce reference dependence.

These results are broadly applicable across many occupational contexts. The types of incentives in my experiment—non-linear bonuses and piece rates—are common features of a wide variety of occupations. These incentives are widely used in sales, which represents a large global market in its own right. The behaviors of door-to-door sellers, therefore, can easily be generalized to other sales and marketing occupations. Other industries and labor markets use these types of incentives. Piece rates are common in many occupations where outcomes can be objectively measured, from fruit picking (Graff-Zivin et al., 2019) to investment commissions for financial managers. The use of formal and informal bonuses at performance targets is ubiquitous, from the highest-paid CEOs to children selling coupon books to raise money for their sports or performing arts programs. That these are widely used across occupations and contexts indicates that a diverse set of actors acknowledge the motivational power of these incentives.

My results have important implications for how workers optimize their labor supply and firm efficiency. Because workers are more motivated by additional income in the loss domain and less motivated in the gain domain, the effectiveness of a wage increase depends on the worker's reference point. My results suggest that the firm, rather than trying to motivate *around* a reference point, can influence the *positioning* of the reference point itself. As the online experiment shows, via these behavioral pathways, bonuses produced significant cost savings relative to worker output. From the firm's perspective, inducing narrow goals in this way is a low-cost *psychological* alternative to high-cost *monetary* incentives or monitoring and enforcement costs. This may partially explain the use of these non-linear incentives across many workplaces.

References

Abeler, Johannes, Armin Falk, Lorenz Goette, and David Huffman. 2011. "Reference Points and Effort Provision." *American Economic Review*, 101 470–492.

Agarwal, Sumit, Mi Diao, Jessica Pan, and Tien Foo Sing. 2015. "Are Singaporean Cabdrivers Target Earners?" *Available at SSRN 2338476*.

Allen, Eric J, Patricia M Dechow, Devin G Pope, and George Wu. 2017. "Reference-Dependent Preferences: Evidence from Marathon Runners." *Management Science*, 63(6): 1657–1672.

Andersen, Steffen, Alec Brandon, Uri Gneezy, and John A List. 2014. "Toward an Understanding of Reference-Dependent Labor Supply: Theory and Evidence from a Field Experiment."Technical report, NBER.

Ariely, Dan, and Klaus Wertenbroch. 2002. "Procrastination, Deadlines, and Performance: Self-Control by Precommitment." *Psychological*

science, 13(3): 219–224.
Ashraf, Nava, Dean Karlan, and Wesley Yin. 2006. “Tying Odysseus to the Mast: Evidence from a Commitment Savings Product in the Philippines.” *The Quarterly Journal of Economics*, 121(2): 635–672.
Bellemare, Charles, and Bruce Shearer. 2010. “Sorting, Incentives and Risk Preferences: Evidence from a Field Experiment.” *Economics Letters*, 108(3): 345–348.
Bénabou, Roland, and Jean Tirole. 2004. “Willpower and Personal Rules.” *Journal of Political Economy*, 112(4): 848–886.
Bettinger, Eric P, Bridget Terry Long, Philip Oreopoulos, and Lisa Sanbonmatsu. 2012. “The Role of Application Assistance and Information in College Decisions: Results from the H&R Block FAFSA Experiment.” *The Quarterly Journal of Economics*, 127(3): 1205–1242.
Brown, Alexander L., Zhikang Eric Chua, and Colin F. Camerer. 2009. “Learning and Visceral Temptation in Dynamic Saving Experiments*.” *The Quarterly Journal of Economics*, 124(1): 197–231, URL: https://doi.org/10.1162/qjec.2009.124.1.197, DOI: http://dx.doi.org/10.1162/qjec.2009.124.1.197.
Cadena, Brian C, and Benjamin J Keys. 2015. “Human Capital and the Lifetime Costs of Impatience.” *American Economic Journal: Economic Policy*, 7(3): 126–53.
Cai, Xiqian, Wei Jiang, Hong Song, and Huihua Xie. 2022. “Pay for Performance Schemes and Manufacturing Worker Productivity: Evidence from a Kinked Design in China.” *Journal of Development Economics*, 156, p. 102840.
Camerer, Colin, Linda Babcock, George Loewenstein, and Richard Thaler. 1997. “Labor Supply of New York City Cabdrivers: One Day at a Time.” *The Quarterly Journal of Economics*, 112(2): 407–441.
Cawley, John, and Christopher J Ruhm. 2011. “The Economics of Risky Health Behaviors.” In *Handbook of Health Economics*. 2: Elsevier, 95–199.
Chou, Yuan K. 2002. “Testing Alternative Models of Labour Supply: Evidence from Taxi Drivers in Singapore.” *The Singapore Economic Review*, 47(01): 17–47.
Correia, Sergio. 2016. “Linear Models with High-Dimensional Fixed Effects: An Efficient and Feasible Estimator.”Technical report, Working Paper.
Crawford, Vincent P., and Juanjuan Meng. 2011. “New York City Cab Drivers’ Labor Supply Revisited: Reference-Dependent Preferences with Rational-Expectations Targets for Hours and Income.” *American Economic Review*, 101(5): 1912–1932, DOI: http://dx.doi.org/10.1257/aer.101.5.1912.
Dellavigna, Stefano. 2009. “Psychology and Economics: Evidence from the Field.” *Journal of Economic Literature*, 47(2): 315–372, DOI: http://dx.doi.org/10.1257/jel.47.2.315.
DellaVigna, Stefano, and Ulrike Malmendier. 2006. “Paying Not to Go to the Gym.” *American Economic Review*, 96(3): 694–719.
DellaVigna, Stefano, and Devin Pope. 2017. “What Motivates Effort? Evidence and Expert Forecasts.” *The Review of Economic Studies*, 85(2): 1029–1069.
Dupas, Pascaline, Jonathan Robinson, and Santiago Saavedra. 2020. “The Daily Grind: Cash Needs and Labor Supply.” *Journal of Economic Behavior & Organization*, 177 399–414.
Engström, Per, Katarina Nordblom, Henry Ohlsson, and Annika Persson. 2015. “Tax Compliance and Loss Aversion.” *American Economic Journal: Economic Policy*, 7(4): 132–64.
Farber, Henry S. 2005. “Is Tomorrow Another Day? The Labor Supply of New York City Cabdrivers.” *Journal of Political Economy*, 113(1): 46–82.
Farber, Henry S. 2008. “Reference-Dependent Preferences and Labor Supply: The Case of New York City Taxi Drivers.” *American Economic Review*, 98(3): 1069–1082, DOI: http://dx.doi.org/10.1257/aer.98.3.1069.
Farber, Henry S. 2015. “Why You Can’t Find a Taxi in the Rain and Other Labor Supply Lessons from Cab Drivers.” *The Quarterly Journal of Economics*, 130(4): 1975–2026, DOI: http://dx.doi.org/10.1093/qje/qjv026.
Fehr, Ernst, and Lorenz Goette. 2007. “Do Workers Work More if Wages Are High? Evidence from a Randomized Field Experiment.” *American Economic Review*, 97(1): 298–317.
Fouarge, Didier, Ben Kriechel, and Thomas Dohmen. 2014. “Occupational Sorting of School Graduates: The Role of Economic Preferences.” *Journal of Economic Behavior & Organization*, 106 335–351.
Freeman, Richard B, Wei Huang, and Teng Li. 2019. “Non-linear Incentives and Worker Productivity and Earnings: Evidence from a Quasi-experiment.” *NBER Working Paper*(25507): , URL: http://www.nber.org/papers/w25507.
Gächter, Simon, Eric J Johnson, and Andreas Herrmann. 2007. “Individual-Level Loss Aversion in Riskless and Risky Choices.”
Giné, Xavier, Monica Martinez-Bravo, and Marian Vidal-Fernández. 2016. “Are Labor Supply Decisions Consistent with Neoclassical Preferences? Evidence from Indian Boat Owners.” URL: https://elibrary.worldbank.org/doi/abs/10.1596/1813-9450-7820, DOI: http://dx.doi.org/10.1596/1813-9450-7820.
Goette, Lorenz, David Huffman, and Ernst Fehr. 2004. “Loss Aversion and Labor Supply.” *Journal of the European Economic Association*, 2(2-3): 216–228.
Goldberg, Jessica. 2016. “Kwacha Gonna Do? Experimental Evidence About Labor Supply in Rural Malawi.” *American Economic Journal: Applied Economics*, 8(1): 129–49.
Graff-Zivin, Joshua S, Lisa B Kahn, and Matthew J Neidell. 2019. “Incentivizing Learning-By-Doing: The Role of Compensation Schemes.”Technical report, NBER.
He, Shu, Liangfei Qiu, and Xusen Cheng. 2018. “Wage Elasticity of Labor Supply in Real-Time Ridesharing Markets: An Empirical Analysis.” *University of Connecticut School of Business Research Paper*(18-21): .
Heath, Chip, Richard P Larrick, and George Wu. 1999. “Goals as Reference Points.” *Cognitive Psychology*, 38(1): 79–109.
Hsiaw, Alice. 2013. “Goal-Setting and Self-Control.” *Journal of Economic Theory*, 148(2): 601–626.
Hsiaw, Alice. 2018. “Goal Bracketing and Self-Control.” *Games and Economic Behavior*, 111 100–121, DOI: http://dx.doi.org/10.1016/j.geb.2018.06.005.
Ichniowski, Casey, and Kathryn Shaw. 2003. “Beyond Incentive Pay: Insiders’ Estimates of the Value of Complementary Human Resource Management Practices.” *Journal of Economic Perspectives*, 17(1): 155–180.
Imas, Alex, Sally Sadoff, and Anya Samek. 2017. “Do People Anticipate Loss Aversion?” *Management Science*, 63(5): 1271–1284.
Kahneman, Daniel, and Richard Thaler. 1991. “Economic Analysis and the Psychology of Utility: Applications to Compensation Policy.” *The American Economic Review*, 81(2): 341–346.
Kahneman, Daniel, and Amos Tversky. 1979. “Prospect Theory: An Analysis of Decision Under Risk.” *Econometrica*, 47(2): 263–291.
Kaur, Supreet, Michael Kremer, and Sendhil Mullainathan. 2015. “Self-Control at Work.” *Journal of Political Economy*, 123(6): 1227–1277.

Köbberling, Veronika, and Peter P Wakker. 2005. "An Index of Loss Aversion." *Journal of Economic Theory*, 122(1): 119–131.

Koch, Alexander K, and Julia Nafziger. 2016. "Goals and Bracketing Under Mental Accounting." *Journal of Economic Theory*, 162 305–351.

Koch, Alexander K, and Julia Nafziger. 2020. "Motivational Goal Bracketing: An Experiment." *Journal of Economic Theory*, 185 104–149.

Kőszegi, Botond, and Matthew Rabin. 2006. "A Model of Reference-Dependent Preferences." *The Quarterly Journal of Economics*, 121(4): 1133–1165.

Kőszegi, Botond, and Matthew Rabin. 2007. "Reference-Dependent Risk Attitudes." *American Economic Review*, 97(4): 1047–1073.

Kuhn, Peter J, and Lizi Yu. 2021. "Kinks as Goals: Accelerating Commissions and the Performance of Sales Teams."

Markle, Alex, George Wu, Rebecca White, and Aaron Sackett. 2018. "Goals As Reference Points in Marathon Running: A Novel Test of Reference Dependence." *Journal of Risk and Uncertainty*, 56(1): 19–50.

Martin, Vincent. 2017. "When to Quit: Narrow Bracketing and Reference Dependence in Taxi Drivers." *Journal of Economic Behavior and Organization*, 144 166–187, URL: http://dx.doi.org/10.1016/j.jebo.2017.09.024, DOI: http://dx.doi.org/10.1016/j.jebo.2017.09.024.

McClure, Samuel M, David I Laibson, George Loewenstein, and Jonathan D Cohen. 2004. "Separate neural systems value immediate and delayed monetary rewards." *Science*, 306(5695): 503–507.

Meier, Stephan, and Charles Sprenger. 2010. "Present-Biased Preferences and Credit Card Borrowing." *American Economic Journal: Applied Economics*, 2(1): 193–210.

Menne, Matthew J, Imke Durre, Bryant Korzeniewski, Shelley McNeal, Kristy Thomas, Xungang Yin, Steven Anthony, Ron Ray, Russell S Vose, Byron E Gleason et al. 2012. "Global Historical Climatology Network-Daily (GHCN-Daily), Version 3." *NOAA National Climatic Data Center*, 10, p. V5D21VHZ, DOI: http://dx.doi.org/10.7289/V5D21VHZ.

Mitchell, Olivia S. 1988. "Worker Knowledge of Pension Provisions." *Journal of Labor Economics*, 6(1): 21–39.

Morgul, Ender Faruk, and Kaan Ozbay. 2015. "Revisiting Labor Supply of New York City Taxi Drivers: Empirical Evidence from Large-Scale Taxi Data." In *Transportation Research Board 94th Annual Meeting*. 15.

Nguyen, Quang, and Pingsun Leung. 2013. "Revenue Targeting in Fisheries: The Case of Hawaii Longline Fishery." *Environment and Development Economics*, 18(5): 559–575.

O'Donoghue, Ted, and Charles Sprenger. 2018. "Reference-Dependent Preferences." In *Handbook of Behavioral Economics: Foundations and Applications 1*. eds. by B Douglas Bernheim, Stefano DellaVigna, and David Laibson: Elsevier, , Chap. 1 1–78.

Oettinger, Gerald S. 1999. "An Empirical Analysis of the Daily Labor Supply of Stadium Vendors." *Journal of Political Economy*, 107(2): 360–392.

Park, R Jisung, Joshua Goodman, Michael Hurwitz, and Jonathan Smith. 2020. "Heat and Learning." *American Economic Journal: Economic Policy*, 12(2): 306–39.

Patnaik, Arpita, Joanna Venator, Matthew Wiswall, and Basit Zafar. 2020. "The Role of Heterogeneous Risk Preferences, Discount Rates, and Earnings Expectations in College Major Choice." *Journal of Econometrics*.

Schmidt, Marc-Antoine. 2018. "The Daily Labor Supply Response to Worker-Specific Earnings Shocks."Technical report, Working Paper.

Shefrin, Hersh M., and Richard H. Thaler. 1992. "Mental Accounting, Saving, and Self-Control." In *Choice Over Time*. eds. by George Lowenstein, and Jon Elster: New York: Russell Sage Foundation Press.

Sheldon, Michael. 2016. "Income Targeting and the Ridesharing Market." *Unpublished manuscript. Available at: https://static1. squarespace. com/static/56500157e4b0cb706005352d*, 56, p. 1457131797556.

Soetevent, Adriaan R. 2022. "Short run reference points and long run performance. (No) Evidence from running data." *Journal of Economic Psychology*, 89, p. 102471, URL: https://www.sciencedirect.com/science/article/pii/S016748702100101X, DOI: http://dx.doi.org/https://doi.org/10.1016/j.joep.2021.102471.

Stafford, Tess M. 2015. "What Do Fishermen Tell Us that Taxi Drivers Do Not? An Empirical Investigation of Labor Supply." *Journal of Labor Economics*, 33(3): 683–710.

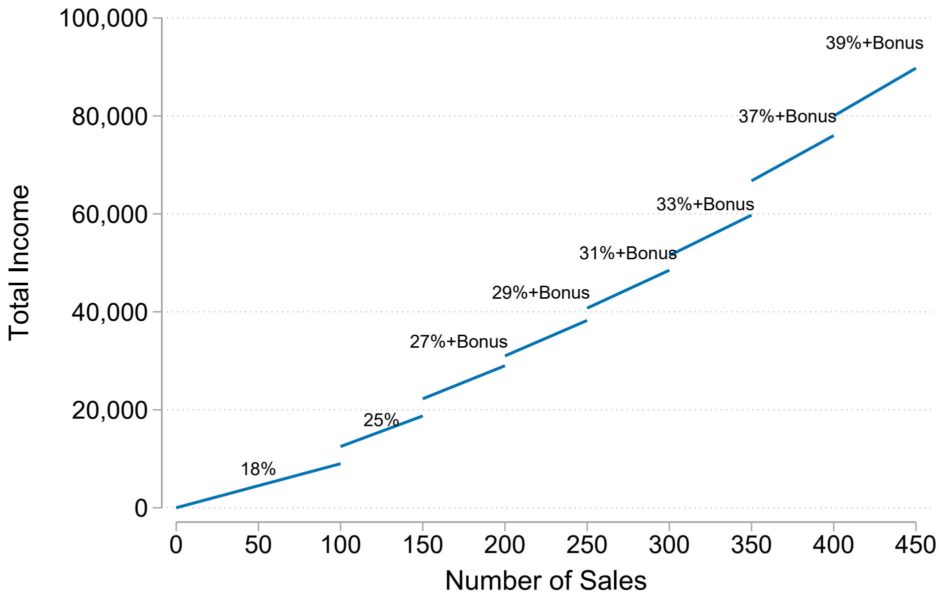
Thakral, Neil, and Linh T Tô. 2021. "Daily Labor Supply and Adaptive Reference Points." *American Economic Review*, 111(8): 2417–43.

Tversky, Amos, and Daniel Kahneman. 1992. "Advances in Prospect Theory: Cumulative Representation of Uncertainty." *Journal of Risk and Uncertainty*, 5(4): 297–323.

Warner, John T, and Saul Pleeter. 2001. "The Personal Discount Rate: Evidence from Military Downsizing Programs." *American Economic Review*, 91(1): 33–53.

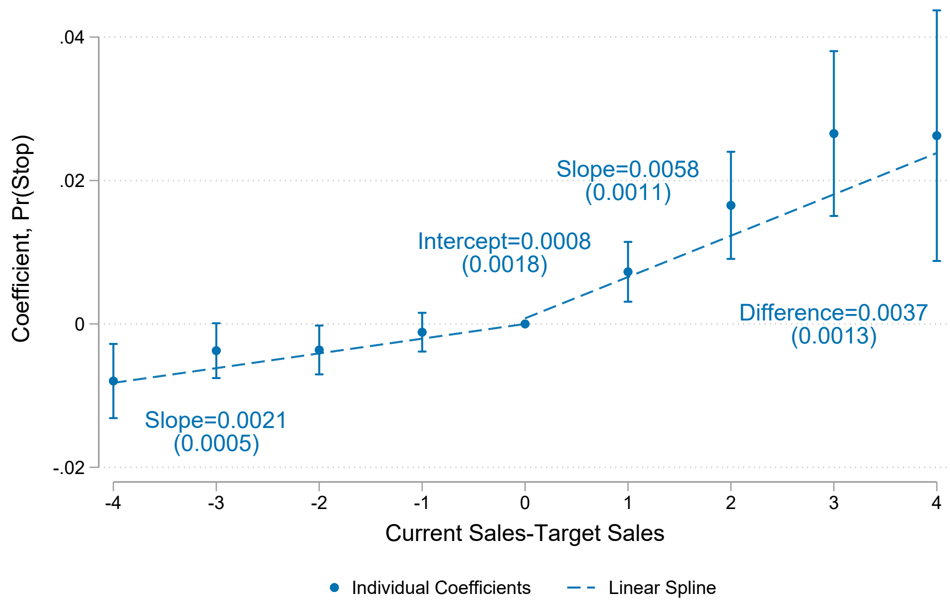
Figures

Figure 1: Contract Structure: Total Income by Sales ($500 Contract Value)

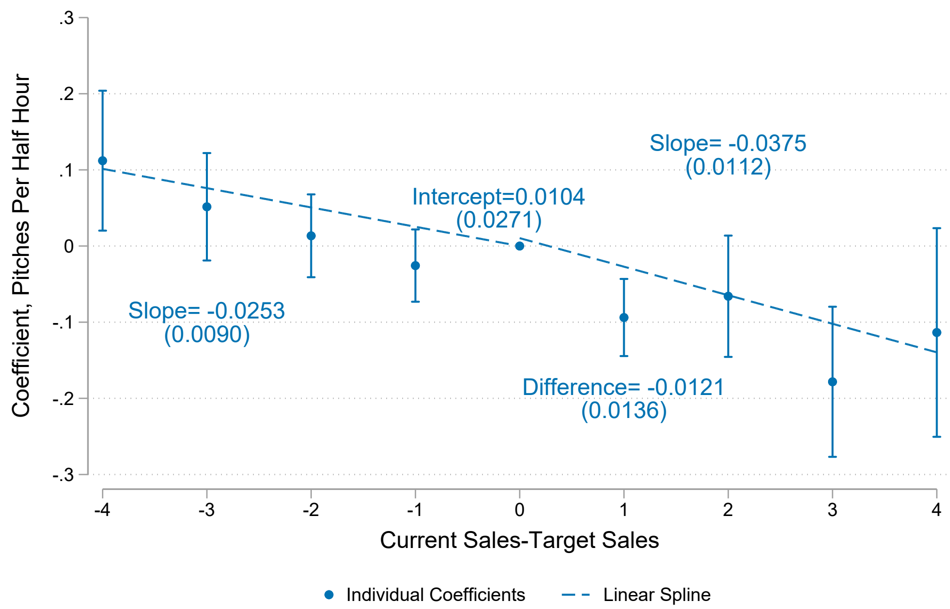


Source: Author's calculations of typical contracts from a pest control sales company.
Notes: Percentages indicate commissions as they are applied to each interval for all sales at the end of the season. At 150 sales, the "bonus" is that the company pays for the seller's rent for the summer in full (≈$2,000). At 250 sales, sellers qualify for the all-expenses-paid company vacation.

Figure 2: Labor Supply Around Expectations
Panel A: Probability of Stopping for the Day

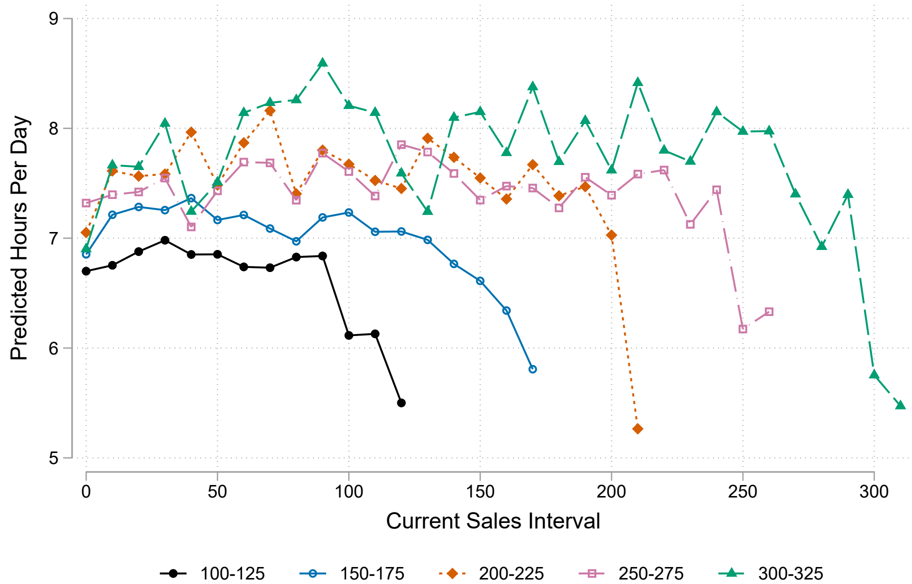


Panel B: Pitches Per Half Hour



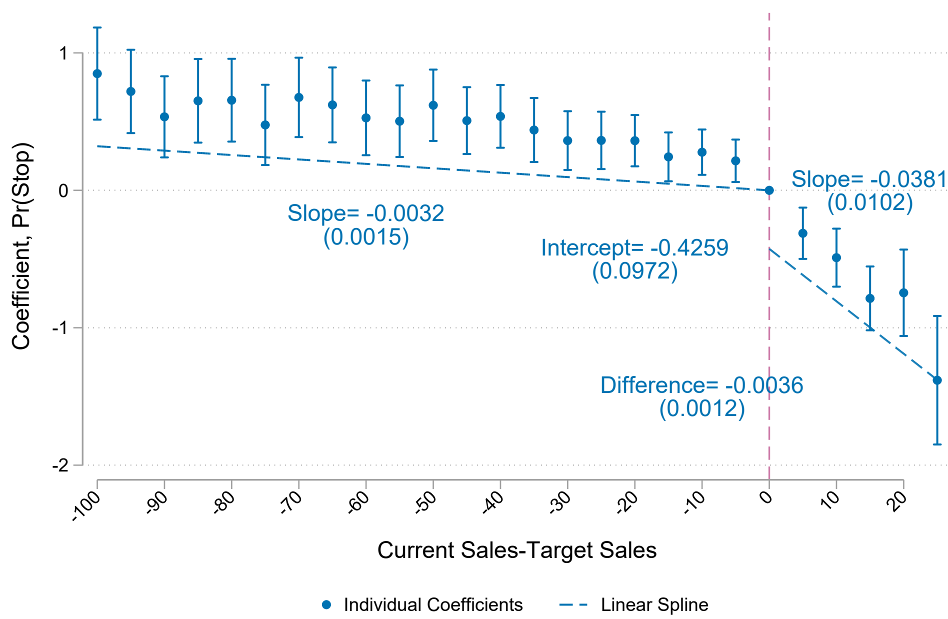
Source: Author's calculations of data from a pest control sales company.
Notes: Results are from estimates of Equations 3 and 4. Standard errors are clustered at the seller level. Individual coefficients are rounded to the nearest integer. The coefficient on +1 includes the interval (0,1.5). The omitted binary category is at zero distance to expectations, and at this point, the mean stopping probability is 0.079, and the mean number of pitches is 2.38.

Figure 3: Predicted Labor Supply Over Current Sales Interval, By Final Season Sales Interval



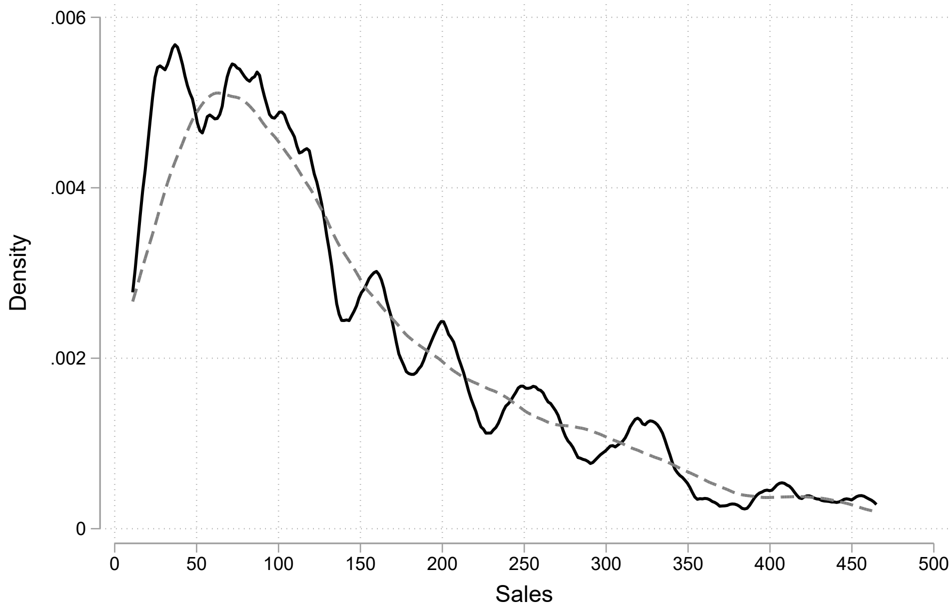
Source: Author's calculations of data from a pest control sales company.
Notes: Plot shows predicted hours from specification in Equation 5 for current sales interval (x-axis) separated by bins of total end-of-season sales.

Figure 4: Test of Reference Dependence Around Bonus Threshold



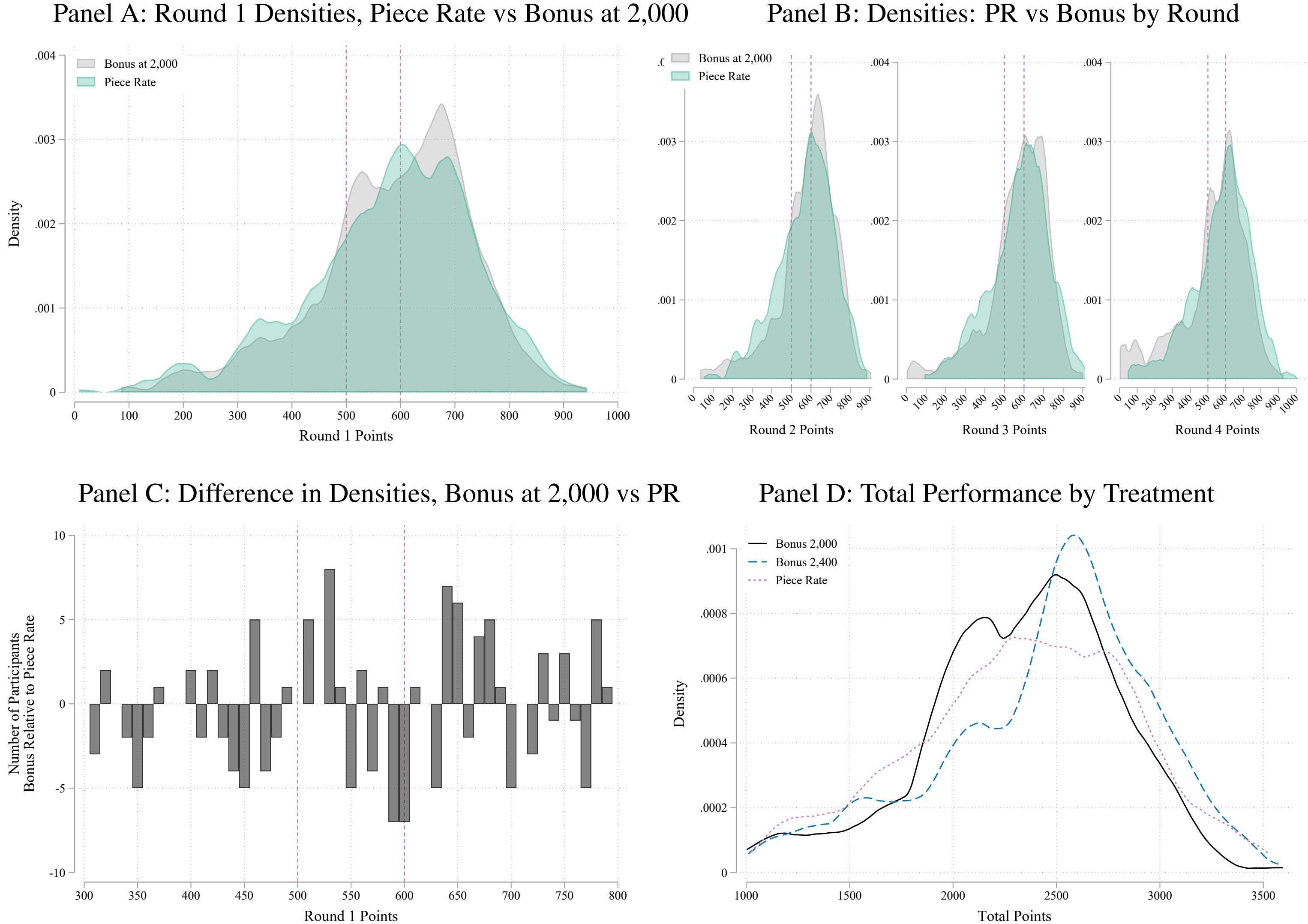
Source: Author's calculations of data from a pest control sales company.
Notes: Estimates are from a regression of hours worked each day as a function of each seller's distance from the nearest bonus threshold as of the end of their season and is a long-run analog of Equations 3 and 4 but using the daily panel. The model includes fixed effects for seller, day of the week, week of the year, and year as well as controls for daily weather, ZIP code characteristics, and $Efficiency$ from Equation 5, which is a time-varying measure of each seller's average sales per hour for all past workdays that season. The sample mean hours knocked per day is 6.9. The ratio of the slopes implies loss aversion of 11.9, while the non-parametric endpoints at -100 and 20, if linearly interpolated, suggest a ratio closer to the 5-6 range.

Figure 5: Kernel Density of Total Sales at End of Season



Source: Author's calculations of data from a pest control sales company.
Notes: Estimated using Epanechnikov kernel with bandwidth of 7 sales and 25 sales for sellers with at least ten sales and fewer than 500. The retroactive nature of the commission increases leads to a cash bonus upon hitting each 50-sale interval. At 150 sales, the company pays for the seller's rent for the summer in full (≈\$2,000). At 250 sales, sellers qualify for the all-expenses-paid company vacation.

Figure 6: Density Estimates of Experimental Performance

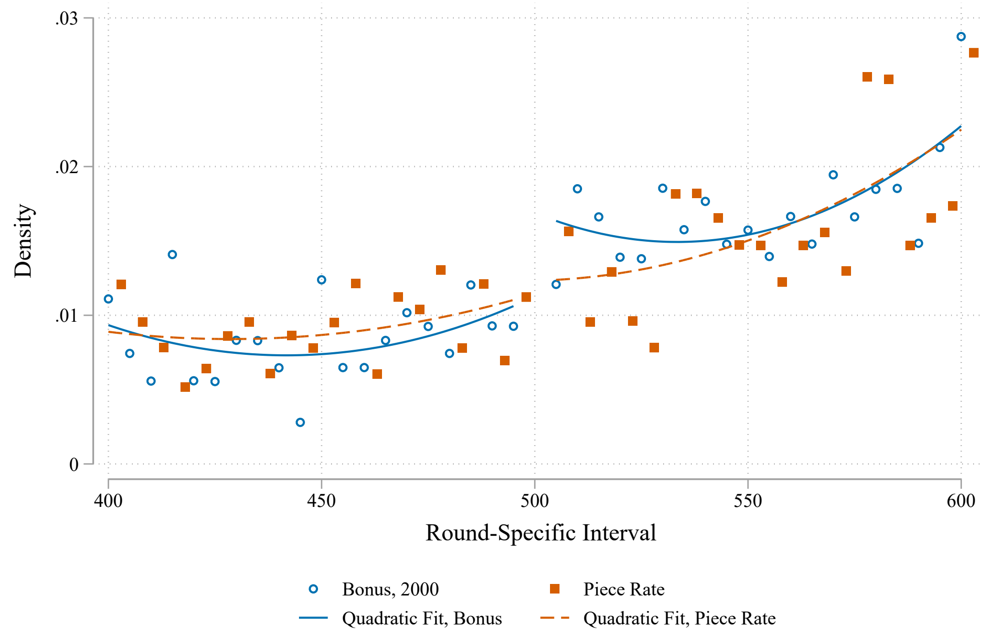


Source: Author's calculations of data from an online experiment.

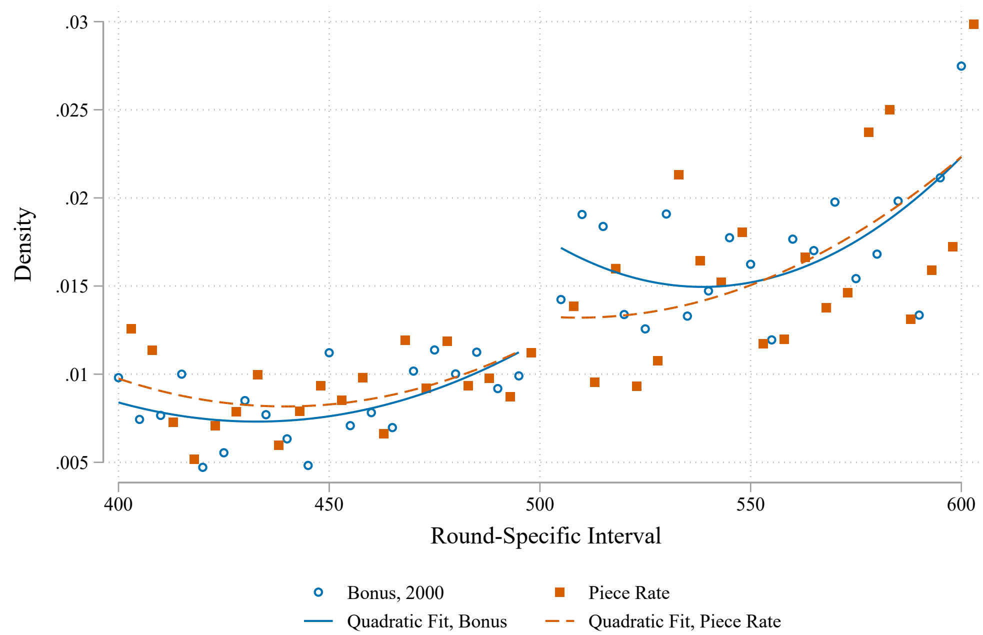
Notes: Panel C breaks frequencies down into discrete 10-point categories and presents the differences between the distributions in Panel A. In Panel C, the X-axis values reflect the minimum of each 10-point interval.

Figure 7: Density Discontinuities at 500 Points Per Round

Panel A: Rounds 1-3



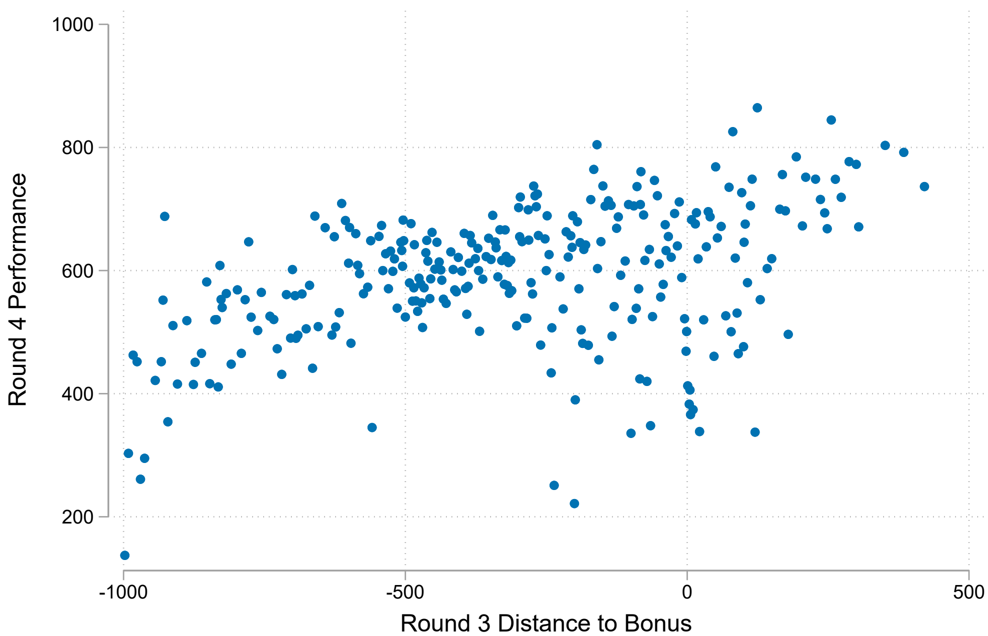
Panel B: All Rounds



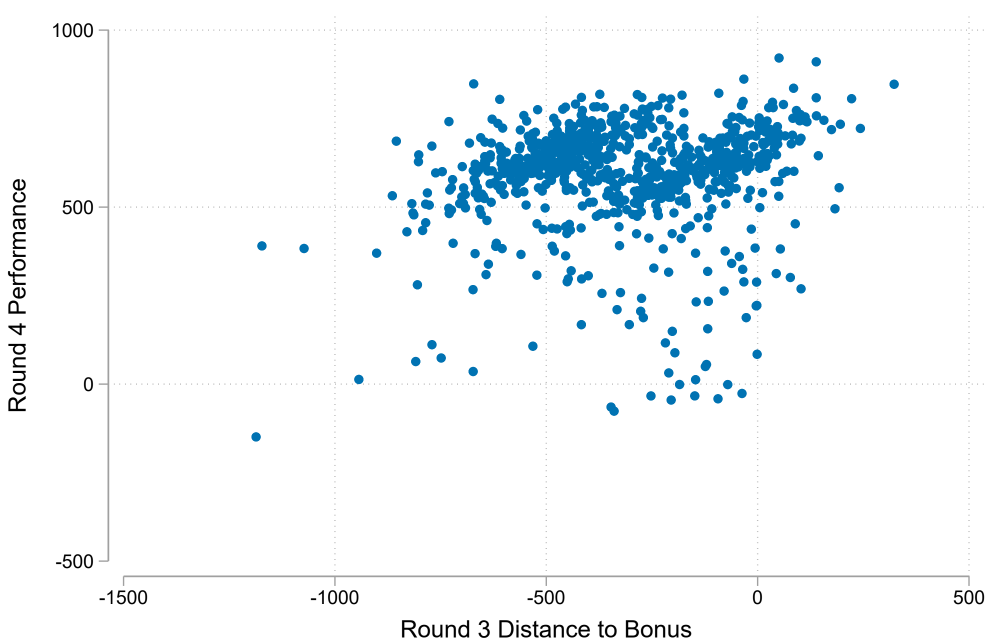
Source: Author's calculations of data from an online experiment.
Notes: The X-axis shows the average number of points achieved per round in the bonus condition and the piece rate conditions in 5-point intervals, while the Y-axis is the density. 500 points is the average each round that those in the bonus condition would have to perform in order to achieve the bonus at the end of the fourth round. Each side of the 500-point cutoff is approximated with a quadratic function of the density, and the higher solid blue lines denote the excess mass accruing to the right side of the cutoff. Using the smoothed estimates for the bonus versus the piece rate condition yields excess mass of 9.4% (bootstrapped standard error of 0.504%).

Figure 8: Effort in the Final Round by Distance to Bonus Threshold

Panel A: Not Conditioning on Round 1



Panel B: Controlling for Round 1

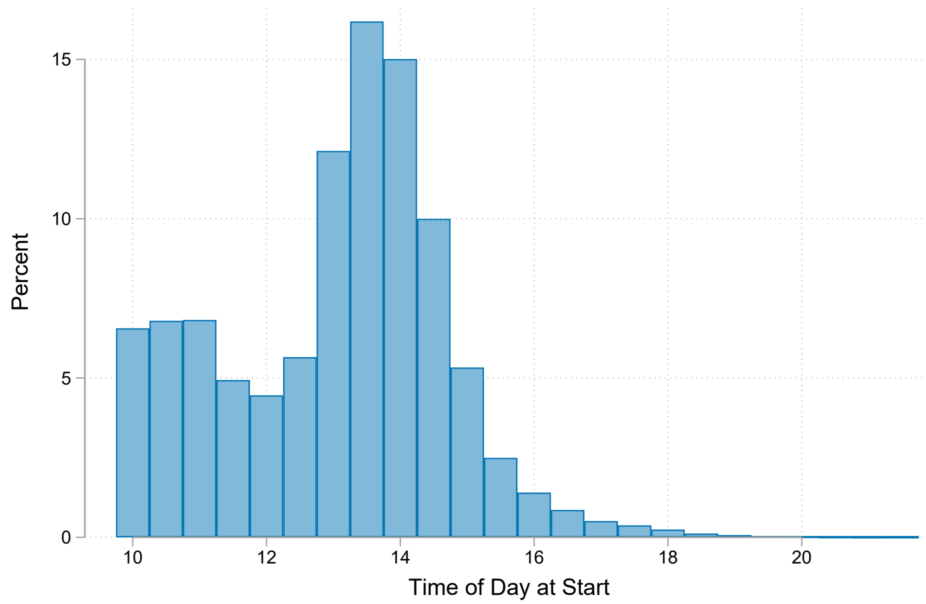


Source: Author's calculations of data from an online experiment.
Notes: The Y-axis shows the number of points achieved during Round 4, while the X-axis show the distance between the participant's Round 3 cumulative total points and their relevant bonus threshold, which is 2,000 points or 2,400 depending on the random assignment.

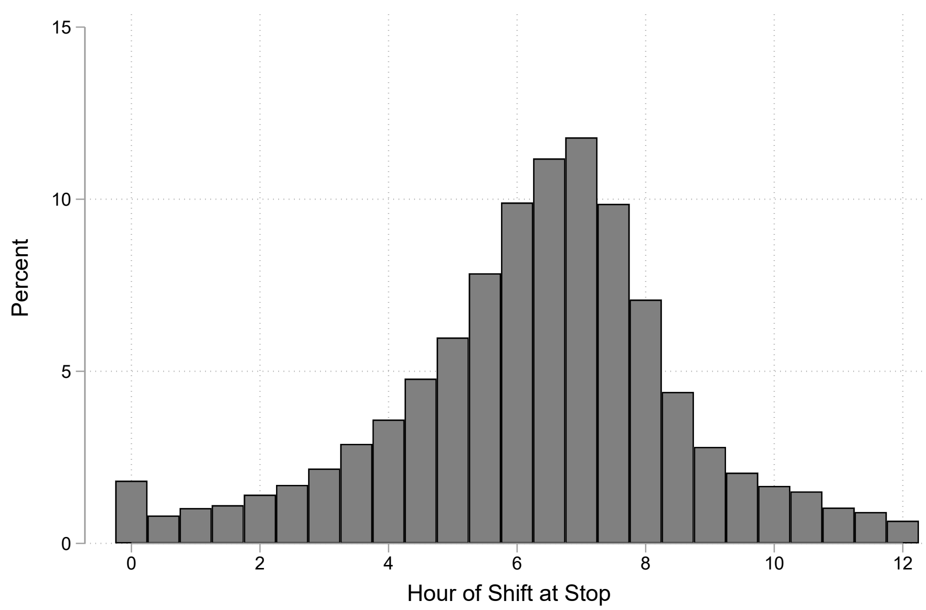
A Online Appendix: Figures and Tables

Figure A1: Distribution of Start and Stop Characteristics

Panel A: Time of Day at Start of Shift

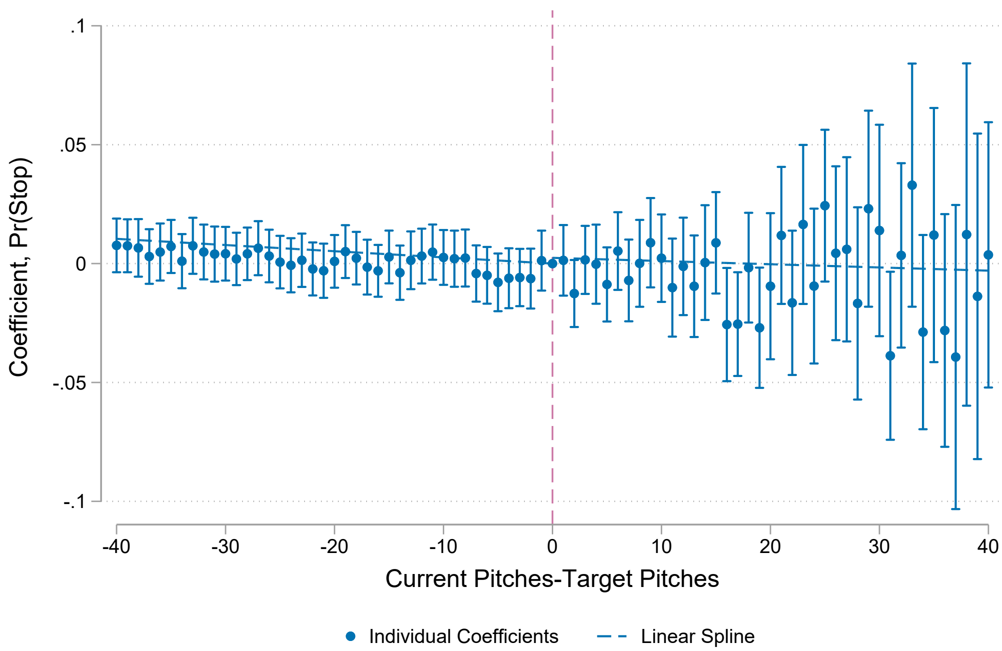


Panel B: Hour of Shift at Stop



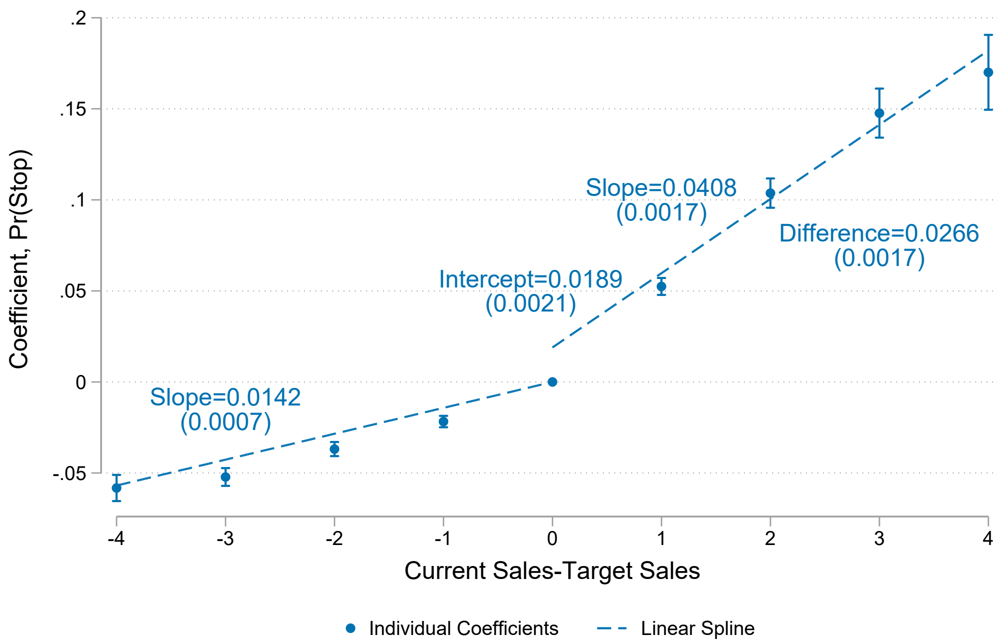
Source: Author's calculations of data from a pest control sales company.
Notes: Shifts begin during the half-hour period when a seller first registers a knock or sale on each workday. Shifts end during the half hour they record their last sale or knock for the day.

Figure A2: Do Sellers Target Numbers of Pitches?



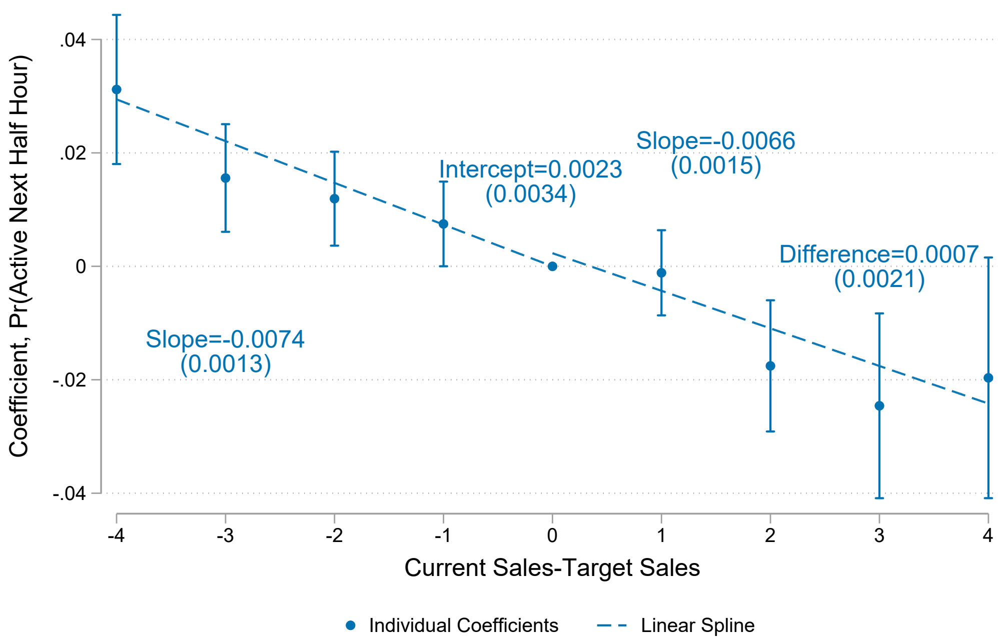
Source: Author's calculations of data from a pest control sales company.
Notes: Results are from estimates of Equations 3 and 4 for the probability of stopping for the day except replacing the target number of sales with the target number of recorded sales pitches per day (e.g. the average number of sales pitches offered each day specific to each day of the week for all prior weeks of the sales season). Standard errors are clustered at the seller level. Individual coefficients are rounded to the nearest integer. The coefficient on +1 includes the interval (0,1.5).

Figure A3: Reference-Dependent Labor Supply in Parsimonious Model



Source: Author's calculations of data from a pest control sales company.
Notes: Results are from estimates of Equations 3 and 4 for the probability of stopping for the day except for only including fixed effects for seller, i.e. cutting all other saturated fixed effects. Standard errors are clustered at the seller level. Individual coefficients are rounded to the nearest integer. The coefficient on +1 includes the interval (0,1.5). The omitted binary category is at zero distance to expectations, and at this point, the mean stopping probability is 0.079, and the mean number of pitches is 2.38.

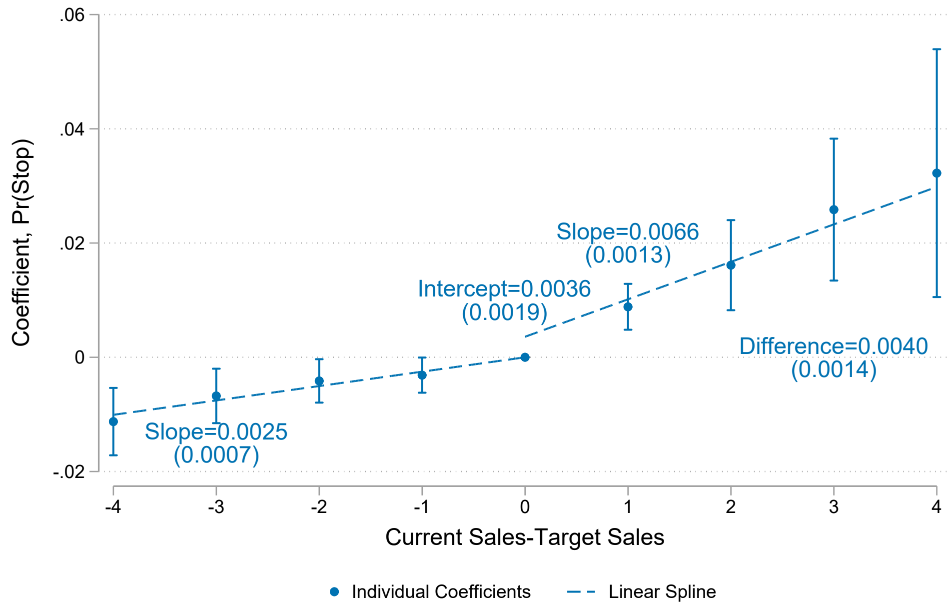
Figure A4: Reference Dependence and Probability of Active Work



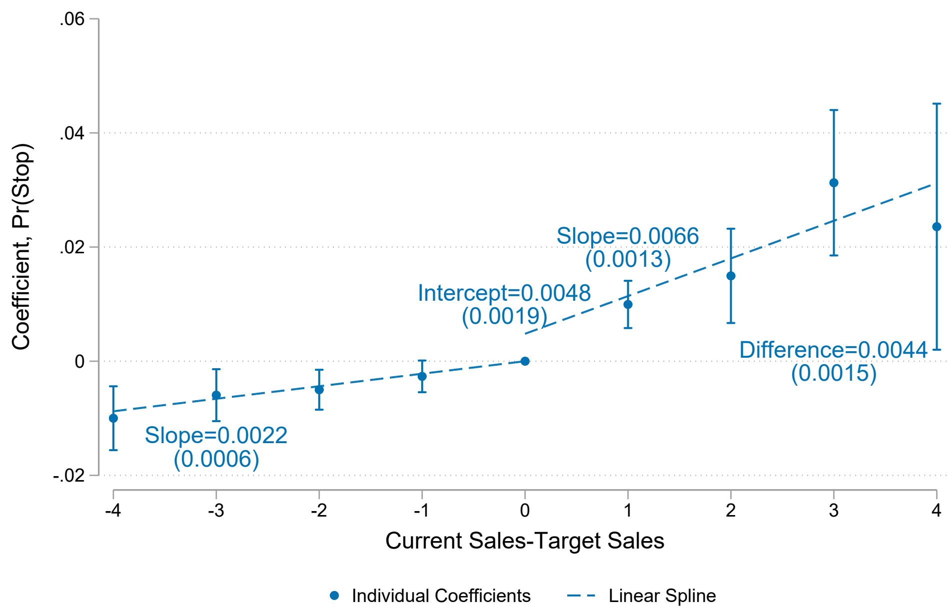
Source: Author's calculations of data from a pest control sales company.
Notes: Results are from estimates of Equations 3 and 4 for the probability of working during the next half hour of the day. Standard errors are clustered at the seller level. Individual coefficients are rounded to the nearest integer. The coefficient on +1 includes the interval (0,1.5). Model also includes an additional control for actively knocking during the current half-hour period. At a base active knocking share of 80% of all half-hour periods, an increase of 1 sale above or below expectations decreases the probability of actively knocking by approximately 0.74 percentage points, or approximately 0.093%. Sellers are not more likely to take breaks during their work as a function of their position relative to expectations.

Figure A5: Reference Dependence Tests Around Expectations from Prior Month, Two Weeks

Panel A: Expectations from Prior Month

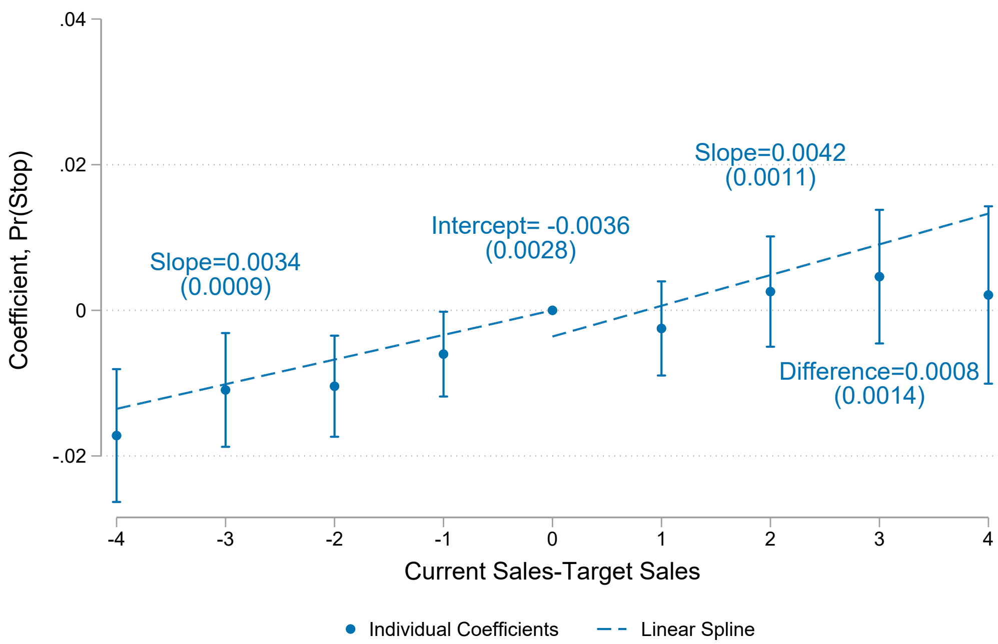


Panel B: Expectations from Prior Two Weeks



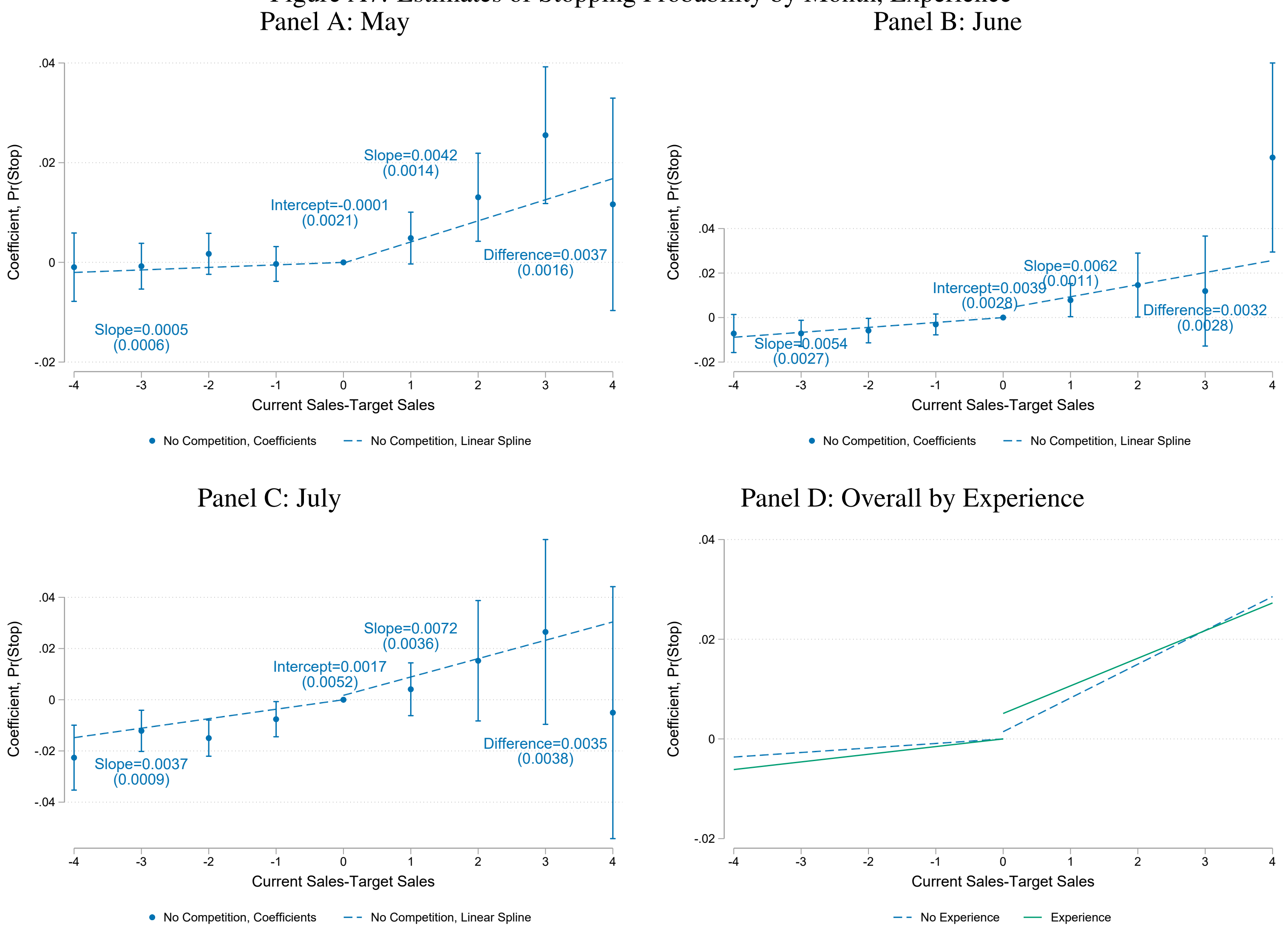
Source: Author's calculations of data from a pest control sales company.
Notes: Results are from estimates of Equations 3 and 4 for the probability of stopping for the day. Rather than using expectations based on the day-of-the-week specific mean from all prior days in the same season, these measures use all days from the prior month (Panel A) and prior two weeks (Panel B) as expectations. Standard errors are clustered at the seller level. Individual coefficients are rounded to the nearest integer. The coefficient on +1 includes the interval (0,1.5). The omitted binary category is at zero distance to placebo expectations.

Figure A6: Robustness: Placebo Targets



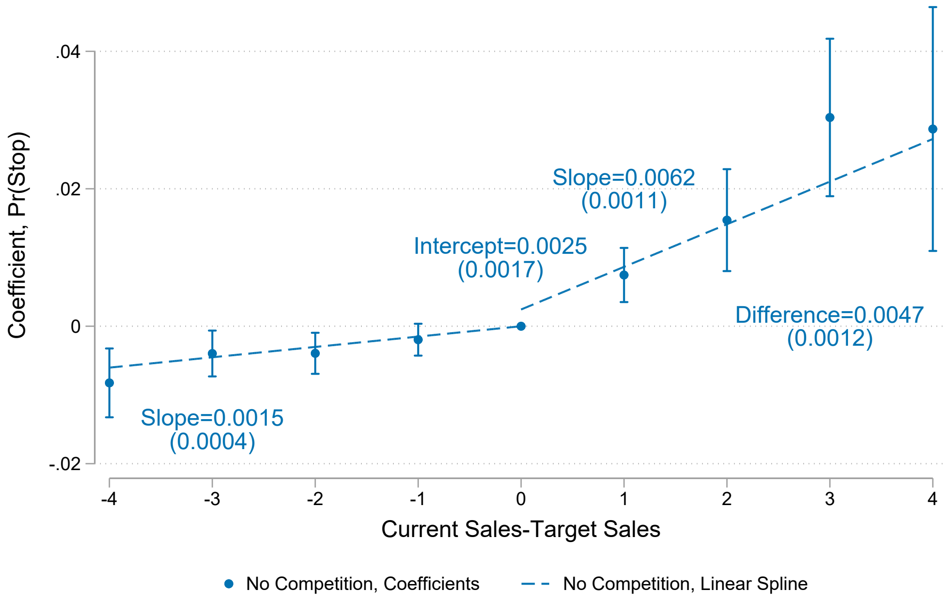
Source: Author's calculations of data from a pest control sales company.
Notes: Results are from estimates of Equations 3 and 4 for the probability of stopping for the day except replacing the target number of sales with a randomly assigned target for each seller in each year drawing from a normal distribution of mean 2 and standard deviation 2. This approximates the actual distribution of targets in the data. Standard errors are clustered at the seller level. Individual coefficients are rounded to the nearest integer. The coefficient on +1 includes the interval (0,1.5). The omitted binary category is at zero distance to placebo expectations.

Figure A7: Estimates of Stopping Probability by Month, Experience

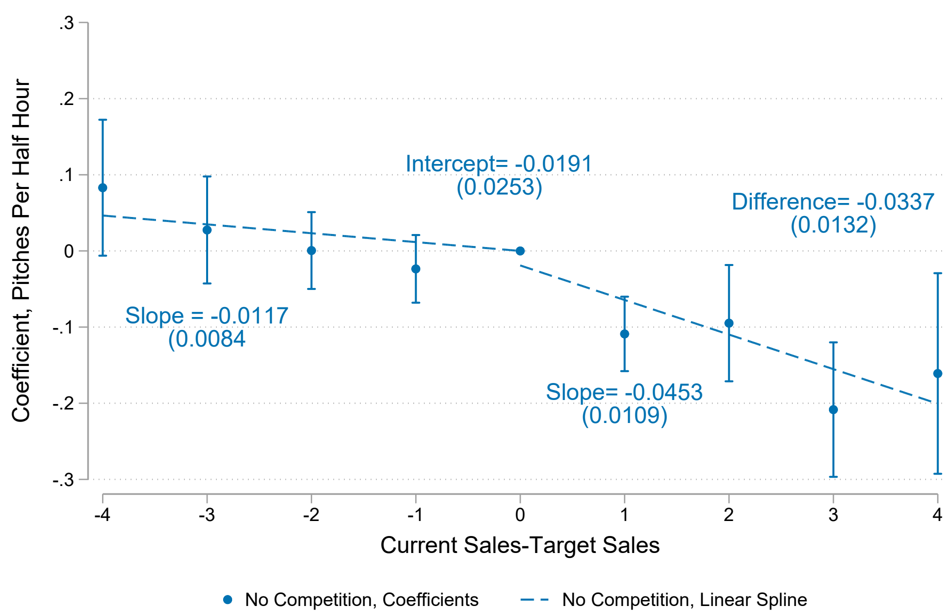


Source: Author's calculations of data from a pest control sales company.
Notes: Results are from estimates of Equations 3 and 4 separated by calendar month. Standard errors are clustered at the seller level. Individual coefficients are rounded to the nearest integer. The coefficient on +1 includes the interval (0,1.5). Panel D shows parametric estimates interacted with an indicator for whether the seller is new to the job ("no experience") or is returning for a second or third sales season in the data ("experienced"). The omitted binary category is at zero distance to expectations.

Figure A8: Robustness Test: Pooled Estimates with Tournament/Non-Tournament Interactions
Panel A: Probability of Stopping for the Day

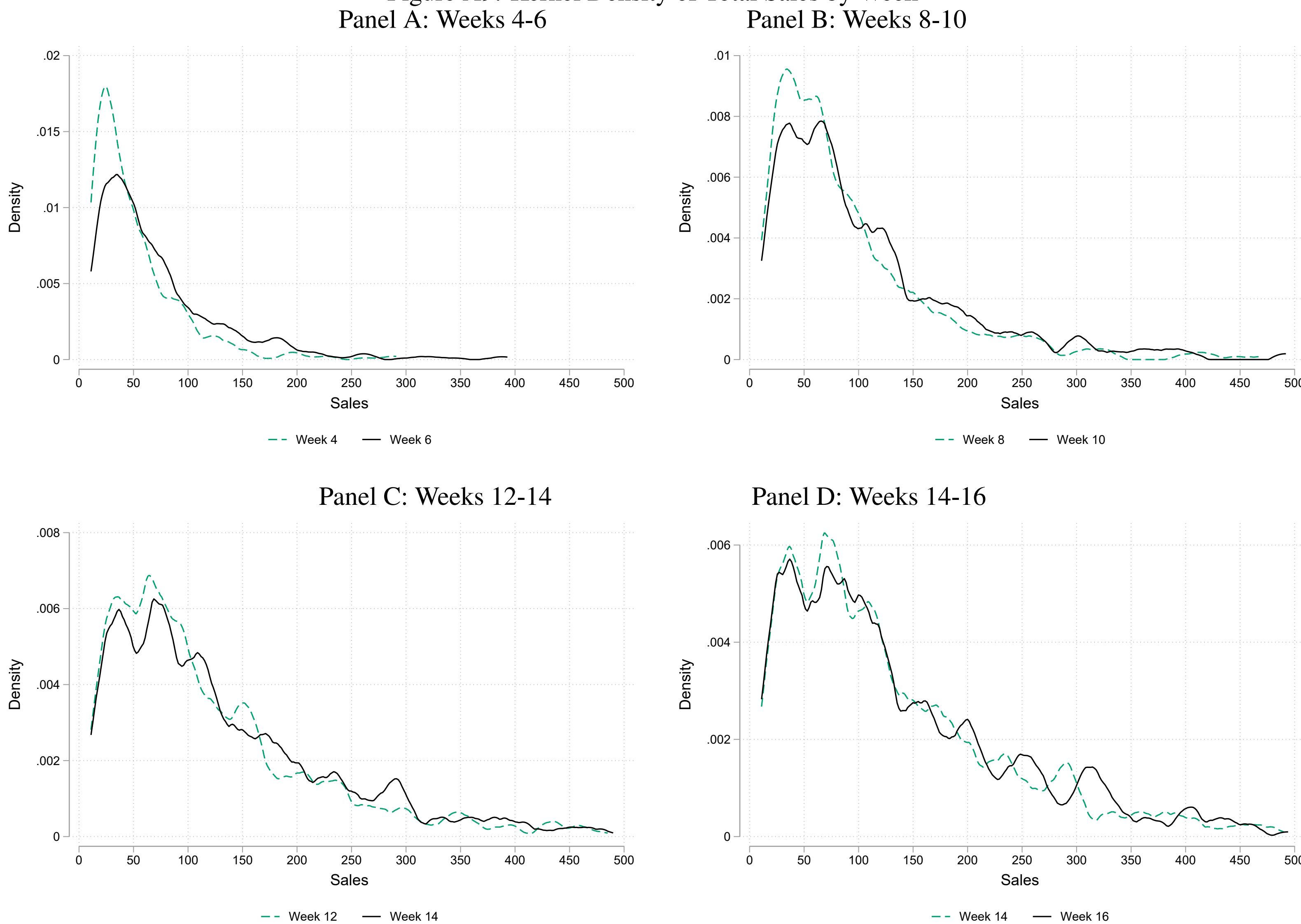


Panel B: Pitches Per Half Hour



Source: Author's calculations of data from a pest control sales company.
Notes: Results are from estimates of Equations 3 and 4 but pooling all data and including interactions between indicators for tournament type or non-tournament days and distance to the target. Reported results are for non-tournament interactions. Standard errors are clustered at the seller level. Individual coefficients are rounded to the nearest integer. The coefficient on +1 includes the interval (0,1.5). The omitted binary category is at zero distance to expectations, and at this point, the mean stopping probability is 0.079, and the mean number of pitches is 2.38.

Figure A9: Kernel Density of Total Sales by Week

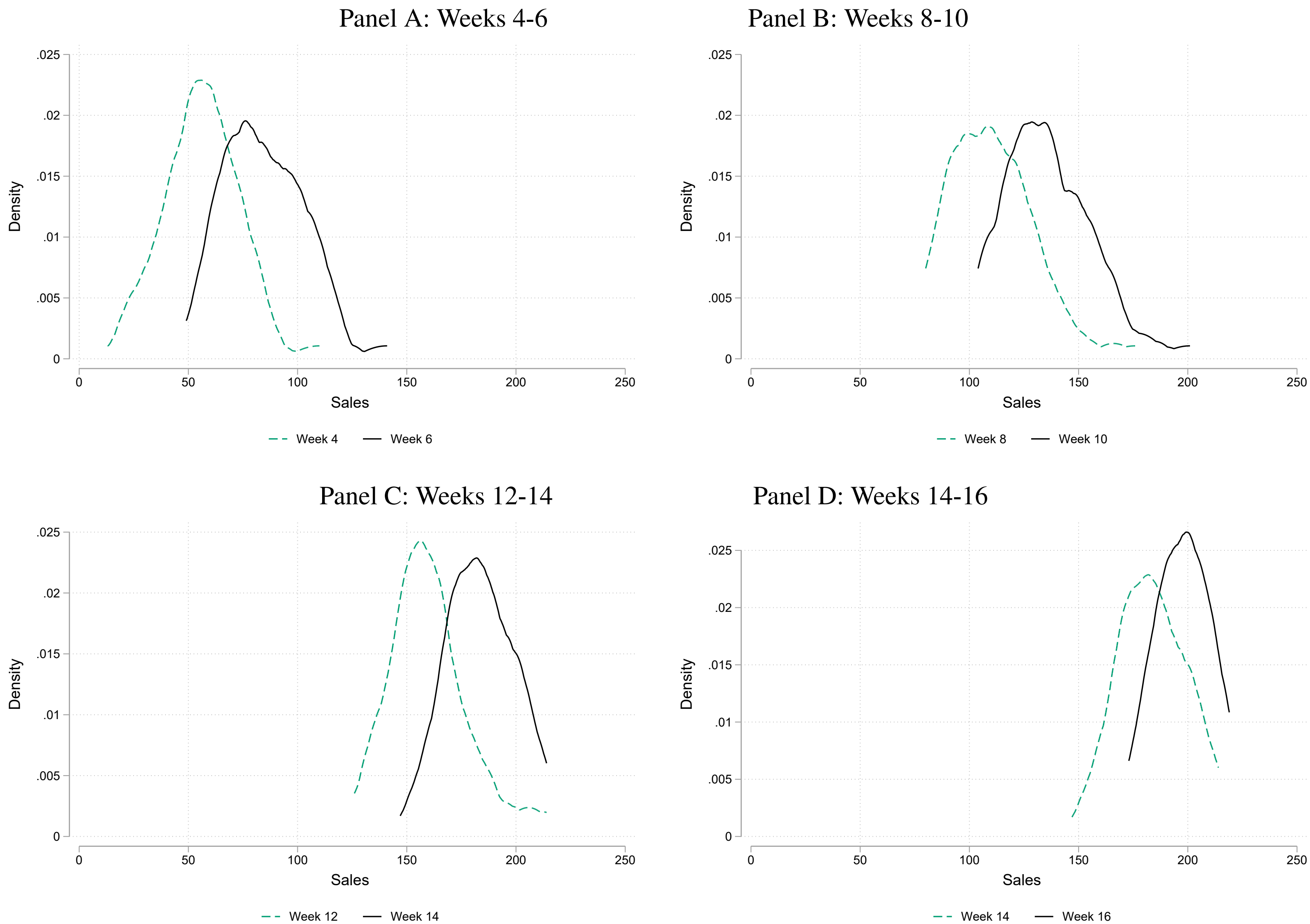


Source: Author's calculations of data from a pest control sales company.

Notes: Estimated using Epanechnikov kernel with bandwidth of 7 sales at the end of each estimated week among those with at least ten sales.

Figure A10: Kernel Density of Total Sales by Week

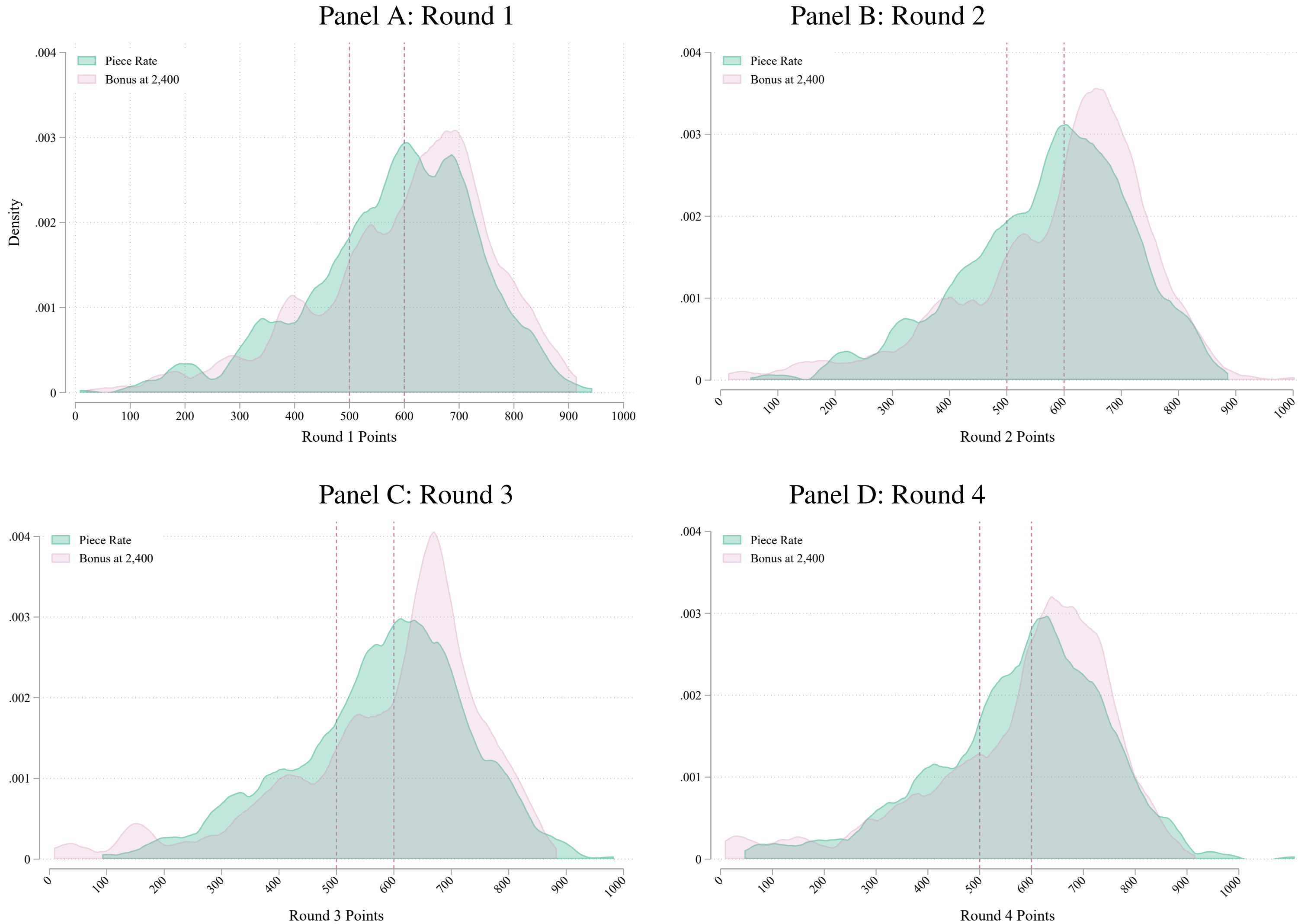
Workers with Total Sales of 175–225 at End of Season



Source: Author's calculations of data from a pest control sales company.

Notes: Estimated using Epanechnikov kernel with bandwidth of 7 sales at the end of each estimated week among those with at least ten sales. The standard deviations of the distributions were 18.96, 20.13, 20.65, 18.49 and 15.33 in weeks 6, 8, 10, 12, and 14, respectively. The skewness parameters were 0.51, 0.89, 0.79, 0.83, and 0.003, respectively.

Figure A11: Density Estimates of Experimental Performance
Piece Rate vs Bonus at 2,400



Source: Author's calculations of data from an online experiment.
Notes: These comparisons follow those in Figure 6 comparing the piece rate treatment to the bonus at 2,000 points treatment.

Table A1: Summary Statistics of Key Variables

Panel A: Half-Hourly Panel	Mean	SD
Pr(stop)	0.074	0.262
Pitches Per Half Hour	2.281	2.498
Sales Per Half Hour	0.156	0.419

Panel B: Daily Panel	Mean	SD
Sales Per Day	2.02	2.20
Labor Supply		
Pitches Per Day	31.21	19.63
Hours Per Day	6.94	2.23
Average Sales Specific to Day of the Week	1.99	1.60
Weather		
Precipitation (1/10th MM)	4.00	8.52
High Temperature (Celsius)	26.85	5.00
Low Temperature (Celsius)	15.29	4.97
Select ZIP Code Characteristics		
Median HH Income	85,945	25,385
% HH Income $100,000-$150,000	19.49	4.69
% Residents Living in Same Home From Last Year	88.19	4.41
Total Housing Units	112,203	5,766
% Housing Units Single-Family Homes	80.08	11.85
Median Home Value	258,083	107,492
% Non-Hispanic White	80.36	13.71
% Bachelors Degree or More	44.93	14.74
Corr(Sales,Pitches)	-0.080	
Total Sellers	512	
Total Days	180	
Total Half-Hourly Observations	458,558	
Total Daily Observations	37,984	

Source: Author's calculations of data from a pest control sales company, NOAA daily weather data, and ACS data on ZIP codes.

Table A2: Daily Sales Variance by Decile of Daily Pitches

	Unconditional		with Seller Fixed Effects	
Decile of Pitches Each Day	Mean	SD	Mean	SD
1	2.013	2.581	-0.271	1.825
2	2.184	2.495	0.017	1.836
3	2.211	2.322	0.145	1.766
4	2.232	2.281	0.134	1.722
5	2.193	2.192	0.141	1.687
6	2.072	2.119	0.050	1.674
7	2.040	2.044	0.028	1.638
8	1.930	2.012	-0.015	1.655
9	1.785	1.854	-0.070	1.520
10	1.537	1.720	-0.129	1.470

Source: Author's calculations of data from a pest control sales company.
Notes: Deciles of pitches per day are calculated across the whole sample of the daily panel. The values with seller fixed effects are first residualized on a fixed effect for sellers.

Table A3: Test of Location Sorting

Sales Per Day, All Significant Coefficients	(1) ACS	(2) Weather	(3) Both
% Non-Hispanic Black	0.0313*		0.0316*
	(0.0161)		(0.0162)
% Single Mothers	-0.0833**		-0.833**
	(0.0403)		(0.0403)
% House Value \$100,000-\$200,000	-0.0276**		-0.0271*
	(0.0140)		(0.0139)
Precipitation (1/10th MM)		-0.00507***	-0.00635***
		(0.00152)	(0.00147)
High Temperature (Celsius)		0.0209**	0.0188**
		(0.00774)	(0.00788)
Low Temperature (Celsius)		-0.0131	-0.0142
		(0.0107)	(0.0108)
Observations	37,508	37,943	37,467
R-squared	0.029	0.013	0.031
F-Statistic	1.59	9.724	3.782
prob>F	0.054	0	0

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Source: Author's calculations of data from a pest control sales company, the American Community Survey 2013-2017 5-year ZIP code estimates, and daily weather data from NOAA.
Notes: Results are from regression of observed ZIP code characteristics from the American Community Survey (ACS) and daily weather data on sales generated per day, including day-of-the-week, week-of-the-season, and year fixed effects. Standard errors clustered at the seller level. Non-significant coefficients on % Non-Hispanic White, % Hispanic; % of households with income \$50,000-\$75,000, \$100,000-\$150,000, and >\$200,000; median household income, poverty rate, unemployment rate, % adults with Bachelors degree or more, % households in the same home as last year; total housing units, % of housing units that are single-family homes; % homes with value \$100,000-\$200,000, \$200,000-\$300,000, \$300,000-\$500,000, and \$500,000-\$1 million and median home value.

Table A4: Non-Parametric Estimates
Expectations-Based References

Distance to Expectations		(1) Pr(Stop)	(2) Pitches Per Half Hour
	-8	-0.0128**	0.174
		(0.00535)	(0.169)
	-7	-0.00577	0.00429
		(0.00472)	(0.0765)
	-6	-0.00812*	0.0403
		(0.00421)	(0.0888)
	-5	-0.0130***	0.120*
		(0.00290)	(0.0688)
	-4	-0.00796***	0.112**
		(0.00264)	(0.0469)
	-3	-0.00373*	0.0515
		(0.00195)	(0.0360)
	-2	-0.00363**	0.0134
		(0.00174)	(0.0277)
	-1	-0.00115	-0.0257
		(0.00138)	(0.0242)
	1	0.00727***	-0.0939***
		(0.00213)	(0.0258)
	2	0.0165***	-0.0659
		(0.00381)	(0.0406)
	3	0.0266***	-0.178***
		(0.00586)	(0.0503)
	4	0.0263***	-0.114
		(0.00892)	(0.0699)
	5	0.00574	-0.224**
		(0.0132)	(0.108)
	6	0.00874	-0.244**
		(0.0177)	(0.107)
	7	0.0254	-0.420**
		(0.0365)	(0.163)
	8	0.0473	-0.209
		(0.0426)	(0.254)
	9	0.0704	0.733
		(0.0542)	(0.569)
	10	0.199***	-0.819***
		(0.0738)	(0.294)

Robust standard errors in parentheses
*** p<0.01, **p<0.05, * p<0.1

Notes: Results are from regression in Equation 3 and coincide with estimates from Figure A8. Standard errors clustered at the seller level.

Table A5: Parametric Estimates of Stopping Probability
Pooled Estimates with Interactions for Tournament/Non-Tournament

Expectations-Based References

	(1) Slope Below Reference	(2) Slope Change Above Reference	(3) Intercept Shift at Reference	(4) Ratio of Slopes [(Change Above + Below)/Below]
No Competition	0.00151***	0.00470***	0.00244	4.113
	(0.00042)	(0.00124)	(0.00170)	
Individual Competitions	0.000333	0.00003	0.00379	1.090
	(0.00058)	(0.00302)	(0.00399)	
Team Competitions	0.00242***	0.00055	0.00997***	1.227
	(0.00048)	(0.00194)	(0.00227)	
Benchmark Competitions	0.0014**	0.00011	0.00849**	1.079
	(0.00058)	(0.0030)	(0.00376)	

Robust standard errors in parentheses
*** p<0.01, **p<0.05, * p<0.1

Notes: Results are from regression in Equation 4 but include interactions between indicators for each tournament/non-tournament period and distance to the reference. Standard errors clustered at the seller level.

Table A6: Robustness Check: Non-Linear Least Squares

Model Parameters	(1) Pr(Stop)	(2) Pitches/Half Hour
Optimal Cutoff	0.11	0
Slope Before Cutoff	0.00074***	-0.0132**
	(0.00026)	(0.0056)
Slope Change After Cutoff	0.0031***	-0.0309***
	(0.0012)	(0.0105)
Intercept Shift at Cutoff	0.0054***	0.0057
	(0.0022)	(0.0270)
Constant	-0.0011***	-0.0097
	(0.0004)	(0.0117)
Ratio of Slopes	5.2	3.3

Robust standard errors in parentheses
*** p<0.01, **p<0.05, * p<0.1

Source: Author's calculations of data from a pest control sales company.
Notes: Estimates use the residualized outcome variables from a regression on all fixed effects and controls in the non-linear estimates. Standard errors clustered at the seller level.

Table A7: Robustness Check: Parametric Model Adding Exertion Margin as Control

	(1) Pr(Stop)	(2) Pitches/Half Hour
Cumulative Pitches	-0.0005***	0.0261***
	(0.000007)	(0.0013)
Slope Before Cutoff	0.0019***	-0.0175
	(0.0005)	(0.0079)
Slope Change at Cutoff	0.0035***	-0.00005
	(0.0013)	(0.012)
Intercept Shift at Cutoff	0.0007	0.0144
	(0.0018)	(0.0244)
Ratio of Slopes	2.8	0.997

Robust standard errors in parentheses
*** p<0.01, **p<0.05, * p<0.1

Source: Author's calculations of data from a pest control sales company.
Notes: Results are from estimating Equation 4 but the model includes a control for cumulative pitches that day. This adjusts for any effects of fatigue from working more intensely. Standard errors clustered at the seller level.

Table A8: Seller Response to Weekly Expectations

VARIABLES	(1) Hours Per Day	(2) Pitches Per Day
Distance to Weekly Expectations	0.0179**	-0.0914*
	(0.00752)	(0.0516)
Over Weekly Expectations	0.0978*	-0.497
	(0.0562)	(0.401)
Distance Over Weekly Expectations	-0.0166	-0.0948
	(0.0106)	(0.0663)
Observations	16,584	16,783
R-squared	0.343	0.525
Dep Var Mean	6.60	28.80
Pct Effect at Mean, Change in Slope Over Expectations	-0.25%	-0.33%

Robust standard errors in parentheses
*** p<0.01, **p<0.05, * p<0.1

Source: Author's calculations of data from a pest control sales company.
Notes: Results are from estimating a similar model to Equation 4 but at the daily level in which I omit within-day controls such as hour of the day. Model includes fixed effects for seller by day of the week, week of the year, and year. Standard errors clustered at the seller level.

B Online Appendix: Other Tests of Reference Dependence and Persistence

As an auxiliary battery of tests for reference dependence, I perform two regression exercises to examine performance persistence and the relationship between sales and hours on above- versus below-average work days. The first is a regression of each seller's total sales at the end of the season on their average sales in the first two weeks of the season as well as the first five weeks of the season. A high R-squared indicates that initial daily sales outcomes and labor supply choices have high predictive power for total cumulative sales. Though not conclusive, this indicates a high degree of persistence in performance. In addition, observed persistence in this measure shows that changing one's goal for the final outcome over time does not appear particularly common in the data.

An important question regarding any day-to-day persistence is whether there is scope for workers to adjust their performance or if there are ceiling effects in place. There are two reasons to doubt this explanation for persistence in performance: first, there is significant variation in performance from day to day within person. Conditional on individual seller fixed effects, the standard deviation of performance is 1.7 sales, meaning that a significant amount of variation is still in play, and it is difficult to attribute all of that variation to demand shocks because this would imply that all workers are exhibiting maximum effort already and success is purely defined by local demand conditional on worker skills. Second, during periods in which the sellers are subject to additional incentives (the tournament periods), they significantly increase their performance. In a regression of daily sales on seller, day of the week, week of the season, and year fixed effects with indicators for tournament status, the average seller increases their sales by 0.12 sales (6.3%) *every day* during the tournament period. This provides strong evidence that there is significant room to increase their sales on particular days if presented with different incentives. See Appendix Table B1.

The second regression is a test of the strength of the relationship between work hours and performance based on exceeding vs not exceeding expectations. I use a panel of worker-day observations to estimate a model of hours worked each day on the number of sales that day interacted with an indicator for if the day's total sales were higher or lower than expectations (average daily sales specific to each day of the week). I estimate:

$$\begin{aligned} y_{idwa} =& \beta_0 + \beta_1 Sales_{idwa} * \boldsymbol{I}^{+}_{Expectations} + \beta_2 Sales_{idwa} * \boldsymbol{I}^{-}_{Expectations} \\ &+ \alpha X_{zdwa} + \sigma W_{zdwa} + \mu_i + \nu_d + \omega_w + \tau_a + \varepsilon_{idwa} \end{aligned} \tag{6}$$

I include fixed effects for seller (μ_i), day of the week (ν_d), week of the season (ω_w), and year (τ_a). The outcome is hours worked that day, while $Sales$ is the total number of service contracts the seller sold that day. The indicators $\boldsymbol{I}^{+}_{Expectations}$ and $\boldsymbol{I}^{-}_{Expectations}$ are dummy variables for if the total sales that day were above expectations or below. In the standard model, because workers will increase their hours when the cost of effort is low, β_1 will be more strongly positive than β_2. In other words, the relationship between work hours and sales will be stronger when total sales for the day are above average (Dellavigna, 2009). The opposite is true under reference dependence.[39]

The results of my two regression exercises using my daily panel are in Table B2. In Panel A, the R-squared for the regression of total sales at the end of the season on average daily sales in weeks 1–2 is 0.752, meaning that average daily performance in the first two weeks explains over three-quarters of the variation in total cumulative sales at the end of the season. Expanding this period to the first five weeks, the R-squared is 0.872, explaining almost 90% of the variation in total sales. There is little unexplained variation in total season sales after conditioning on the first two to five weeks, and there is high congruence between sales outcomes in the first two weeks and behaviors the rest of the season. What this means in practice is that after an initial early learning period, these sellers do not appear to significantly revise their long-term or short-term targets, meaning that once expectations are formed in the first 2-4 weeks of the season, they are extremely stable.

In Panel B, the relationship between daily sales and hours worked is stronger on days that fell *below* expectations compared to days that exceeded expectations. This runs counter to the predictions of the standard model that workers will work more hours on days that have high wage returns.

One might wonder if other external commitments might drive this relationship between sales and hours in Table B2 by putting limits on how many hours a seller can be in the field and suppressing both hours and total sales. The role of these commitments is limited for multiple reasons. The first is contextual: these sales workers are young, mostly unpartnered, and live away from their normal homes, social networks, schools, and other possible employment. The second is statistical. If sellers did have external commitments that would be consistent enough to systematically place these commitments into the "bad day" category (below their mean performance), one might expect these to be correlated with the time of the day, the day of the week, or the week of the sales season. My fixed effects remove variation correlated with these factors. If external commitments are affecting work hours, one might expect these commitments to disproportionately affect workers later in the day (rather than at, for example, 2:00 PM). If external commitments do take sellers out of the field later in the day, that should *strengthen* the relationship between hours and sales on above-average days because, according to the data in Panel C of Figure C1, sales are much easier to secure during later work hours. This would work in the opposite direction from the patterns I find.

[39]That contact and sales rates increase at the end of the work day additionally "stacks the deck" against reference dependence in this model because sales late in the day are less costly to achieve.

Table B1: The Effect of Tournament Incentives on Daily Sales

VARIABLES	(1) Daily Sales
Any Tournament Incentive	0.128***
	(0.0326)
Observations	38,927
R-squared	0.434
Mean Sales	2.034
Percent Effect	6.3 %

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Source: Author's calculations of data from a pest control sales company, NOAA daily weather data, and ACS data on ZIP codes.
Notes: Regression is executed on the panel of daily sales and includes controls for weather and work area ZIP code characteristics. Estimates include fixed effects for seller, day of the week, week of the season, and year. Standard errors clustered at the seller level.

Table B2: Secondary Evidence of Persistence and Reference Dependence

Panel A: Average Daily Sales in Early Weeks

Total Sales at End of Season	Weeks 1–2	Weeks 1–5
Average Daily Sales	95.61***	91.32***
	(3.785)	(2.129)
Observations	33,728	36,857
R-squared	0.752	0.872

Panel B: Sales and Hours, Days that Exceeded Expectations or Not

Hours Worked Per Day	Did Not Exceed	Exceeded Expectations
Sales	0.441***	0.335***
	(0.0179)	(0.0101)
Observations	37,977	
R-squared	0.266	

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Source: Author's calculations of data from a pest control sales company.
Notes: Panel A is from a regression of sellers' total sales at the end of the season on average daily sales during the first two or five weeks of the season. Panel B is from Equation 6 and includes fixed effects for seller, day of the week, week of the season, and year. Standard errors are clustered at the seller level.

C Online Appendix: Data

The pest control sales company data were obtained through a data use agreement prohibiting disclosure of the company's identity or intimate details of their operations.

The data cover the entirety of all sales and knocks recorded from January 2018 to January 2020. Sales in the "off-season" are not compensated the same way as they are during the summer, and there are very few recorded knocks in their system. Most sales the company generates during the off-season are renewals of current contracts for the following year as well as follow-ups with past customers, but those contacts are typically not done in person. Most off-season knocks are those done in the service of training new sellers. The knocking data are reported using their common application, which also shows leaderboards, team performance, and the performance of all other sellers in the company. The centralized sales website also contains sales information but does not include knocking information. Competition rules, dates, and prizes were collected from raw internal company documents as well as the company website usually available only to contractors and employees.

To correctly measure the incentives and behavior of these workers at the right time, I impose a few basic restrictions to my half-hourly panel. I limit my sample to the "summer sales season" each year, which is the period from the last week of April to the third week of August. This excludes trainees who arrive early, those who stay late into the end of August or early September (who are usually managers and those not enrolled in school), and off-season sales. I exclude the last two weeks of August because participation drops precipitously as sellers return to school. Less than 50% of sellers stay past August 17th-18th, and less than 25% of sellers stay past August 25th-26th. I then exclude any sellers who stopped working altogether before late May, which effectively excludes the least able sellers who averaged less than one sale per week and decided to go home after experiencing this lack of success. This group also includes managers who record knocks for training purposes during the first month. Off-season sales during the September to April months entail a different compensation structure, and many of the sales are generated by full-time employees of the company rather than the independent contractors that work during the summer.

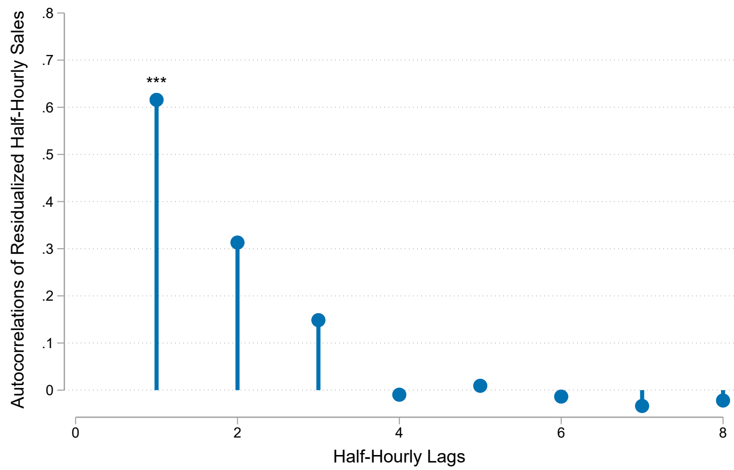
In my half-hourly panel, I exclude observations with no previous expectations, i.e. the first week a seller is active. In all, my half-hourly panel consists of approximately 459,000 observations for 512 sellers across 180 days in 2018-2019.

From an incentive standpoint, if there is a positive autocorrelation in sales each day—that is, if success now is predictive of success in the near future—then a worker having success right now faces lower marginal costs of effort in the coming hours. This will work against the downward shift in labor supply predicted by Prospect Theory. To test for this, I calculate autocorrelation on sales after conditioning on the variables in my models. I do this in two ways. First, I residualize sales each half hour by regressing sales each half hour on fixed effects for seller, day of the week, week of the season, and year as well as controls for actively knocking on doors, weather, and ZIP code characteristics. I then calculate the autocorrelation in these residuals between half-hour periods. Second, I add additional fixed effects for the number of half-hour periods elapsed since the start of the shift. The results are in Panels A and B of Figure C1. The results conditional on only the time of the day suggest that there is some autocorrelation in residualized sales approximately thirty minutes, or that success now is predictive of success at least for the next half hour. However, once the cumulative length of the shift is taken into the account, this autocorrelation falls to below 0.20 and a level that is not statistically significant or behaviorally meaningful.

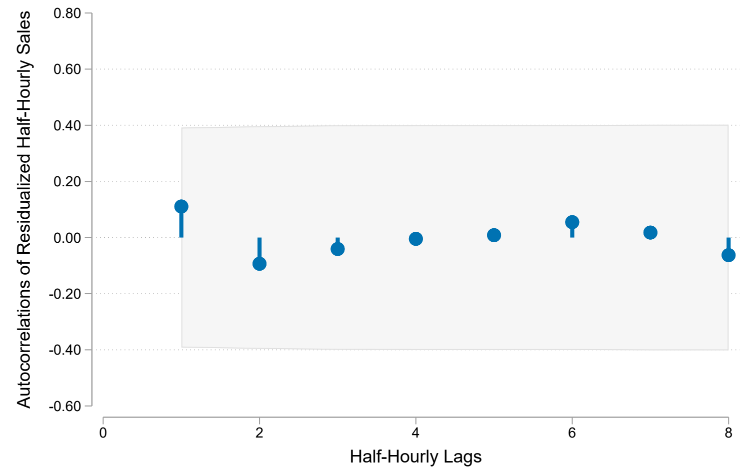
Any positive relationship in Panels A and B appears driven by the positive relationship between the time of day and sales after 5:30 PM. Panel C of Figure C1 shows that average seller performance increases as the day progresses, particularly after 5:30 PM when residents return home from work. This is not due to a change in the composition of workers, but because workers have more opportunities to make contact with residents. The marginal cost of sales falls later in the day. If sellers understand these dynamics, they have an incentive to continue to work. In short, despite these contextual features "stacking the deck" against reference dependence in terms of incentives, I still detect meaningful evidence using my formal tests, which I describe in Section 5.1.

Figure C1: Upward Pressures on Labor Supply During the Day

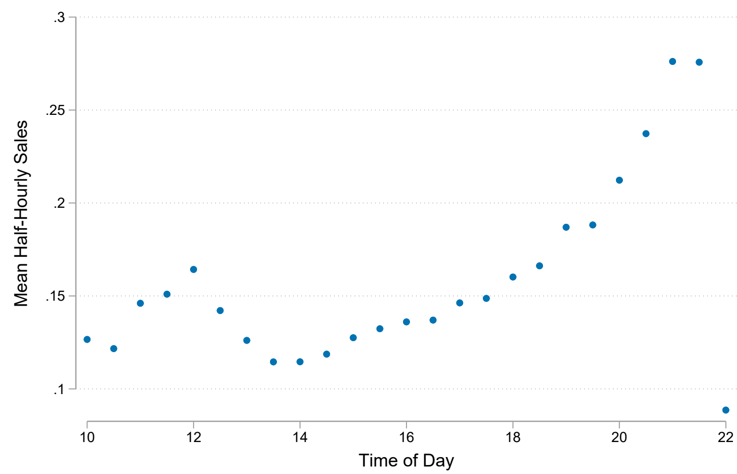
Panel A: Autocorrelation of Sales



Panel B: Autocorrelation of Sales Conditional on Shift Length



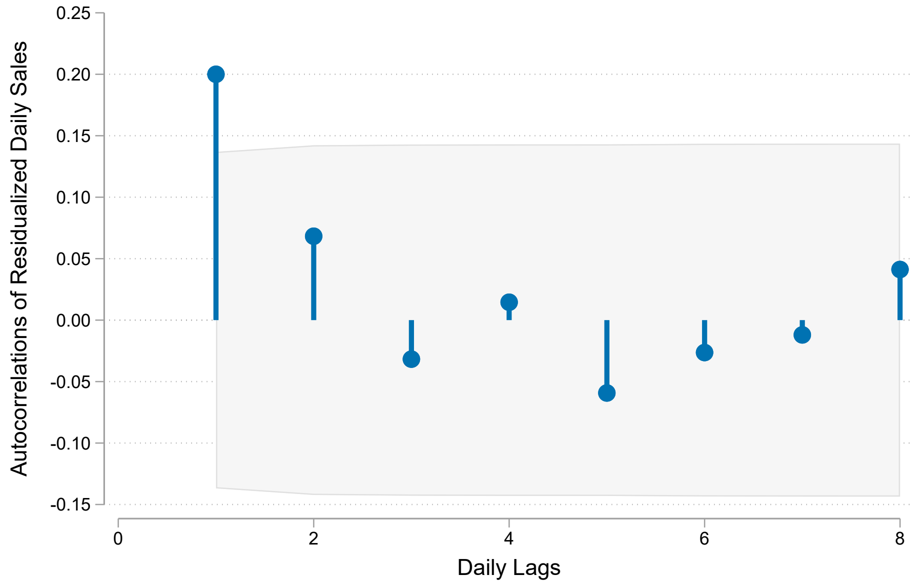
Panel C: Mean Sales by Half Hour



Source: Author's calculations of data from a pest control sales company.

Notes: In Panel A, residualized sales come from a regression of sales each half hour on seller, half-hour-of-the-day, day-of-the-week, week-of-season, and year fixed effects as well as controls for having any knocks recorded that half hour, weather, and ZIP code characteristics. I then calculate the autocorrelation for those predicted residuals for half hour lags of one through eight. Only the correlation for the one-period lag is statistically significant. Panel B adds additional non-parametric controls for the number of half-hour periods elapsed since the start of the shift. The shaded region shows Bartlett's formula for MA(q) 95% confidence bands.

Figure C2: Autocorrelation in Daily Sales



Source: Author's calculations of data from a pest control sales company.
Notes: This figure uses the seller-day panel to calculate residualized sales. I regression of sales each day on seller, day-of-the-week, week-of-the-season, and year fixed effects as well as controls for having any knocks recorded that day, weather, and ZIP code characteristics. I then calculate the autocorrelation for those predicted residuals for lags of one through eight days. The shaded region shows Bartlett's formula for MA(q) 95% confidence bands. The low autocorrelation between days indicates that performance today is not strongly predictive of performance tomorrow. In other words, individual workdays come from independent draws.

D Online Appendix: Further Background

The company whose data I use (which I call "PestCo") operates a full-service pest control service operation. In addition to removing insects, spiders, and rodents, they apply preventative treatments to prevent pests from returning or growing larvae near an individual home. There is a range of services they provide, and sellers are encouraged to "upsell" for more comprehensive services whenever they see an opportunity. Sellers are given the responsibility to generate new contracts and schedule the service with a separate wing of the company that performs the service. Most contracts last 12–18 months. Commission rates are based on the annualized value of the contracts the seller generates.

PestCo is not markedly different from the rest of the sales industry in terms of its use of incentive schemes. Their independent contractor agreements and practices are all in line with industry standards.

Sellers are paid an up-front portion ($75) of their commissions during the two-week period each sale is made, similar to a regular paycheck. The balance of commission payments is calculated at the end of the season after the status of all contracts is known. Final payouts for Spring sales are given in the Fall, and late Summer sales payouts are given at the end of the year. Most contractor agreements include penalties for leaving the selling area before the official end of the sales season or for not recording knocking activity a minimum number of days. The penalties typically stipulate that regardless of the number of sales, the commission the seller earns will return to some low base rate (usually 18–20%).

Prior to leaving for their assigned metro area, sellers at PestCo are trained in sales techniques and are given a detailed manual of behavioral tools to help them over the course of the summer. These include training on proper body language, handshaking, standards for appearance, overcoming customer objections, rephrasing customer concerns, interacting with upset neighbors, and how to look for and identify pests before approaching a door. They are provided with video examples of strong sales performance and are encouraged to review their training materials on a daily basis.

PestCo takes an active role in trying to motivate their workers. In training materials, the company encourages their sellers to be physically active and healthy, to be honest about their performance and goals, and to take accountability for their own performance and summer experience. These training materials are especially important because approximately half of the sellers who are working any given day are brand new to the company and the industry. Sellers are encouraged to learn advanced sales techniques from their more experienced teammates or roommates.

Work neighborhoods for each seller are assigned by a local team leader. Metro areas are divided into sections for each team, and within their section, team leaders assign sellers to a neighborhood. Work in each neighborhood continues until approximately 75% of doors have been marked in their tracking software, after which the seller can request a new area. Managers insist that "work area does not matter" in their training materials, and the evidence I present supports this argument. Area assignments, while not strictly random, are not correlated with sales outcomes in any meaningful way either across or within seller (see A3) and Section 4. Managers emphasize that making assignments to work areas based on perceived skill or other seller attributes is costly to them as managers and generates unclear returns, which undermines the business case for sorting. For example, assigning a better seller to a "harder" neighborhood may generate sales that would otherwise not take place. However, the marginal cost of achieving those may be high and the benefits may be smaller than the difference in sales *speed* across sellers in an "easier" neigbhorhood. In practice, managers do not typically spend large amounts of time on these assignments.

In addition to the high-powered cash incentives built into their contractor agreements, PestCo also runs frequent short-run tournaments for prizes valued from \$300 to \$3,000. These take three forms: individual rank-order, team rank-order, and what I call "benchmark" competitions. Individual rank-order tournaments pit sellers against each other for a single day, and the seller with the most head-to-head daily "wins" at the end of the two-week tournament gets a prize.

Team rank-order tournaments have a similar structure but are based on wins against another team, and "wins" are based on per-seller team revenue. During "benchmark" competitions, if a seller generates more revenue during the week-long competition period than he did during any prior week in the season, he will get a prize. Prizes include merchandise like Bluetooth headphones, apparel, and expensive grills as well as "experiences" like a cruise, resort stay, or annual ski passes, though sellers have the option to cash out the value of the prize. Importantly, these benchmark tournaments occur later in the season, and the prizes are not valuable enough that it is worthwhile for sellers to lower their effort earlier in the season in the hopes of making it easier to gain these prizes. These tournaments are not the focus this study. However, they are important for contextualizing my empirical models because they modify the incentive structure within particular sales days and thus may shift a worker's expectations on that particular day. See Table B1. However, there is no evidence of a more permanent change in expectations.

E Additional Theory Background on Reference Dependence

E.1 Reference Dependence and Recent Expectations

The basic insight of models of reference dependence and loss aversion propose that losses loom larger than gains.

O'Donoghue and Sprenger (2018) present a simple model of this idea that is instructive. A worker can choose an effort level e, which yields output $x(e)$ and has a cost of effort $c(e)$. The function $c(e)$ is increasing and convex. Utility is linear in $x(e)$. Suppose there is an output or income reference, r, which can be endogenously determined by rational expectations or exogenously imposed. Distance from the reference, $x(e) - r$, enters the utility or value function:

$$U(e) \equiv x(e) + \mu\left(x(e) - r\right) - c(e) \tag{7}$$

where

$$\mu(z) = \begin{cases} \eta z & \text{if } z \geq 0 \\ \eta \lambda z & \text{if } z \leq 0 \end{cases}$$

The μ function captures "gain-loss utility." The equilibrium labor supply for this utility function with gain/loss utility is given by:

$$\begin{array}{ll} (1+\eta)x'(e) - c'(e) = 0 & \text{if } x(e) - r > 0 \\ (1+\lambda\eta)x'(e) - c'(e) = 0 & \text{if } x(e) - r < 0 \end{array} \tag{8}$$

The shift across the reference threshold reflects the difference in the marginal value of income. At the same level e, the marginal benefit on the left side of the reference ($x(e) < r$) is scaled by a factor of $\lambda > 1$ relative to the right side of the reference ($x(e) > r$). This parameter is the coefficient of loss aversion. The parameter η is the weight of gain-loss utility in the utility function. This simple model with linear utility implies that, if current earnings, $x(e)$, are below the reference, equilibrium labor supply will be higher than if earnings are above the reference for the same value of e. For a loss-averse worker, upon reaching the reference, r, there is a downward kink in the marginal value of income, so labor supply will also kink downward, holding constant effort costs at $c(e)$. Figure E1 shows an illustration of this concept. The marginal utility when $\lambda = 1$ is the same on either side of the reference. However, when $\lambda > 1$ and income is below the reference, the marginal utility is higher and overall utility is lower because being below the reference creates a sense of loss. In the standard case in Equation 8, $\lambda = 1$ or $\eta = 0$, and there is no discontinuous change in marginal benefit across the reference.

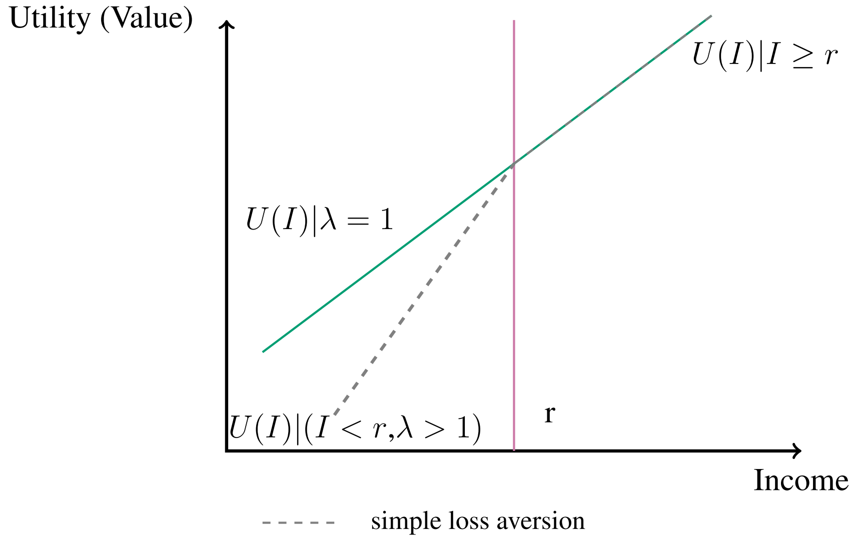
The prior literature on labor supply has almost exclusively focused on daily references. This focus simplifies the theoretical tests of reference dependence by limiting the role of income effects, which standard theory suggests may be notable in the long-run but will be negligible each day because daily income plays such a small role in long-run or lifetime earnings (O'Donoghue and Sprenger, 2018; Dellavigna, 2009). This justifies the use of linear utility in Equation 7. Reference dependence with loss aversion predicts in my context that when a seller surpasses her daily reference, the probability she stops working for the day will kink upward, holding other factors constant.

On the other hand, the standard model predicts that if the wage return, $x'(e)$, shifted upward for the same value of e, the worker would unambiguously work more hours regardless of which side of r she is on. When daily wages are high, the standard worker will increase daily labor supply, and when daily wages are low, the worker will stop working earlier in the day. These labor supply decisions will be a smooth function of $x(e)$ and $c(e)$.

As important as the parameter of loss aversion (λ) is to the model of reference dependence, equally important is the definition (or location) of the reference point itself. In an essential theoretical paper (Kőszegi and Rabin, 2006), the KR model theorizes that “recent expectations” act as important reference points. But how do people form short-term expectations? The KR model proposes that these expectations are determined in what they call “personal equilibrium,” that is, by behaviors that are optimal given the worker’s expectations about the future. Put another way, a forward-looking worker can make a plan around what she perceives to be the optimal path forward, and when the final choice is made in real-time, the planned path becomes her reference point. This “path” is her personal equilibrium. Kőszegi and Rabin (2006) posit when introducing this theory that firms can play a significant role in establishing a worker’s personal equilibrium; however, empirical evidence for this role is generally sparse.

This theoretical result has important implications. The first is that if wage increases are anticipated or predictable, a worker will respond by planning to work more hours, or by adjusting her planned path, similar to the standard model. In the context of the bonuses paid in the door-to-door sales setting and in my experiment, this means that workers make their initial daily labor supply choices based on what they determine to be optimal given what they expect to be their ability to reach a bonus threshold. If workers obtain new information about their abilities, they can quickly adjust their future goals and their short-term reference points. This creates a feedback loop between future expectations and recent experience wherein a simple measure of average past performance integrates both pieces of information. The second key implication is that workers exhibit gain-loss utility over outcomes that deviate from expectations (the path). After setting her plan for the path ahead, the worker responds each period to whether her performance is below or above what she expects for that period. Significantly, negative comparison utility and higher marginal utility while working below short-term expectations induce more effort.

Figure E1: Illustration of Basic Model of Reference Dependence with Loss Aversion



Notes: Illustration of basic loss aversion with linear utility over income. When $\lambda = 1$, the marginal utility above the reference r is the same as marginal utility below the reference.

E.2 Firm Preferences for Internal vs External Goals

An important question in this setting is whether or not the cost of the bonus that leverages this internal goal adoption is preferable from the firm's perspective to alternatives such as close external monitoring and discipline.

Let us suppose that the firm's revenue from that worker is a function of her total effort:

$$E_T = \sum_{t=1}^{T} e_t.$$

The firm earns revenue:

$$R(E_T) = p \sum_{t=1}^{T} e_t = pTe, \quad R' > 0,\ R'' \leq 0.$$

Recall from Section 2.4 that the firm offers a bonus of monetary size $\theta > 0$, which induces a transfer $B(\theta)$ to the worker upon reaching $\bar{E}$ and generates a salience weight $s(\theta)$ around the threshold, where $B'(\theta) > 0$ and $s'(\theta) > 0$. The worker's adoption condition for daily goals is:

$$s(\theta)\lambda(1-\beta) \geq \kappa.$$

The firm sets θ sufficiently large that this condition is satisfied, so that daily goals are adopted, and the worker provides equilibrium daily effort $e^B(\theta)$, implicitly defined by the daily-goals effort condition derived in Section 2.4:

$$c'(e^B) = \lambda + \beta \left[b'(e^B) + \frac{s(\theta)\lambda}{T} \right].$$

Firm profit under the bonus scheme is:

$$\Pi^B(\theta) = pTe^B(\theta) - B(\theta) - F,$$

where F is simply a fixed cost for running the bonus program.

Suppose instead that the firm enforces a per-period effort target $\tilde{g} = \bar{E}/T$ via close monitoring and enforcement. Successful enforcement ensures the worker reaches total output $\bar{E}$, with per-period effort:

$$e^M = \bar{E}/T.$$

Monitoring and enforcement entails a per-period cost $M(\bar{E}/T)$, where $M' > 0$, capturing the costs of technology, managerial review, and morale or social costs that arise from disciplining workers. Such enforcement costs can be large if, for example, managers' time is extremely valuable, or morale issues spill over to other workers. Total enforcement costs over the season are therefore $TM(\bar{E}/T)$.

Firm profit under monitoring and enforcement is:

$$\Pi^M = pTe^M - TM(\bar{E}/T) = p\bar{E} - TM(\bar{E}/T).$$

$$M(\bar{E}/T), \quad M'(E/T) > 0.$$

The firm prefers the bonus scheme over monitoring and enforcement whenever $\Pi^B(\theta) \geq \Pi^M$, i.e.:

$$pTe^B(\theta) - B(\theta) - F \geq p\bar{E} - TM(\bar{E}/T).$$

Rearranging, the bonus scheme is preferred when:

$$pT\left(e^B(\theta) - e^M\right) \geq B(\theta) + F - TM(\bar{E}/T)$$

The left-hand side is the revenue gain from the higher effort induced by the bonus and daily goal adoption relative to the monitored effort level. The right-hand side is the net cost advantage of monitoring over the bonus: the bonus transfer and fixed program costs, offset by the total monitoring and enforcement costs saved.

Recall that e^B is increasing in loss-aversion λ, salience s, and present bias $(1 - \beta)$. So bonuses will be preferred as monitoring and enforcement costs rise. Bonuses will be preferred as loss aversion rises, the salience response to the bonus $s'(\theta)$, and if workers are more present-biased. Thus, it is much cheaper for firms to use bonuses to motivate workers when they are more sensitive to losses. This is, perhaps, unsurprising, as simply increasing wages, for example, has a behavioral effect blunted by present bias in the absence of the incentives triggered by gain-loss utility. Monitoring is preferred when monitoring and discipline costs are low or salience responds weakly to bonuses.

F Online Experiment Protocol (For Online Publication)

Experimental Protocol

Start of Block: Consent

Q1
Introduction

Welcome to this research project! We very much appreciate your participation.

This is a study about decision-making. Several research institutions have provided funds for this research. It is very important for the success of our research that you answer honestly and read the questions very carefully before answering.

Procedures
You are given instructions on your screen before every decision. Please always make sure to read the instructions carefully before you continue.

Payment
Your payment will consist of the participation fee plus the amount of bonus points that you accumulate throughout the study. The exact amount of bonus points that you receive will depend on your and/or others' decision.

Your bonus will be paid to you using the bonus system within a few days after completion. Your payment for taking part in the study will be sent to you shortly after submission.

Participation
Participation in this research study is completely voluntary. You have the right to withdraw at anytime or refuse to participate entirely without jeopardy to future participation in other studies conducted by us.

Confidentiality

All data obtained from you will be kept confidential and stored on a GDPR compliant secure server. Your prolific ID will only be used for payment purposes. Any personal information that could identify you will be removed or changed before files are shared with other researchers or results are made public in open science repositories.

Verification

At the end of this survey, you will be automatically redirected to Prolific.

Questions about the Research
If you have questions regarding this study, you may contact
thechoicelab@nhh.no

Q2 If you have read and understood the instructions above and want to participate in this study, write ACCEPT in the box below.

____

End of Block: Consent

Start of Block: Attention Check

Q3 Please indicate your agreement with the following statement below:

	Strongly Disagree (1)	Disagree (0)	Agree (0)	Strongly Agree (0)
I need to cover myself in lava so that I don't freeze at night. (1)	○	○	○	○

End of Block: Attention Check

Start of Block: Treatment 1 – Bonus at 2,000 Points

Q4 As part of this study you will play a simple button-pressing task. The object of this task is to alternately press the 'a' and 'b' buttons on your keyboard as quickly as possible. You will play for a **total of 10 minutes**.

Every time you successfully press the 'a' and then the 'b' button, you will receive a point. Note that points will only be rewarded when you **alternate button pushes**: just pressing the 'a' or 'b' button without alternating between the two will not result in points. Buttons must be pressed by

hand only (key-bindings or automated button-pushing programs/scripts cannot be used) or the task will not be approved.
Feel free to score as many points as you can.

<u>You will be paid an extra 1 dollar if you score at least 2,000 points.</u> We will divide your 10-minute playing time into **4 rounds.** This means you will need to score an average of at **least 500 points per round** to receive the bonus.

You will have a 10-second break in between each round. On the next page is an example of how the task will work. Try pressing 'a' and 'b' alternately to score points. We have limited the point total below to a maximum of 50 as this is just practice, but the actual task will not have a limit. Try to reach 50 **<u>as quickly as possible.</u>**

Page Break

JS

Q6 Press 'a' then 'b'...

Points: 0

Page Break

Q7 **<u>You will be paid for your performance.</u>** Please confirm you understand this.

○ Yes (1)

○ No (2)

Page Break

Q8 Proceed to the next page when you are ready to play the task. Your 10-minute task will begin immediately when the page loads and will be divided into 4 rounds.

Page Break

Button-Pushing Task

End of Block: Treatment 3 – Bonus at 2,400 Points

Start of Block: Treatment 2 – Bonus at 2,400 Points

Q23 As part of this study you will play a simple button-pressing task. The object of this task is to alternately press the 'a' and 'b' buttons on your keyboard as quickly as possible. You will play for a **total of 10 minutes**.

Every time you successfully press the 'a' and then the 'b' button, you will receive a point. Note that points will only be rewarded when you **alternate button pushes**: just pressing the 'a' or 'b' button without alternating between the two will not result in points. Buttons must be pressed by hand only (key-bindings or automated button-pushing programs/scripts cannot be used) or the task will not be approved.
Feel free to score as many points as you can.

As a bonus, you will be paid an extra 1 dollar if you score at least 2,400 points.   We will divide your 10-minute playing time into **4 rounds.** This means you will need to score an average of at **least 600 points per round** to receive the bonus.
You will have a 10-second break in between each round.   On the next page is an example of how the task will work. Try pressing 'a' and 'b' alternately to score points. We have limited the point total below to a maximum of 50 as this is just practice, but the actual task will not have a limit. Try to reach 50 **as quickly as possible.**

Page Break

JS

Q25 Press 'a' then 'b'...

Points: 0

Page Break

Q26 **<u>You will be paid for your performance.</u>** Please confirm you understand this.

- ◯ Yes (1)
- ◯ No (2)

Page Break

Q27 Proceed to the next page when you are ready to play the task. Your 10-minute task will begin immediately when the page loads and will be divided into 4 rounds.

Page Break

<u>Button-pushing task</u>

End of Block: Treatment 3 – Bonus at 2,400 Points

Start of Block: Treatment 9 – Piece Rate

Q42 Shortly, you will play a simple button-pressing task. The object of this task is to alternately press the 'a' and 'b' buttons on your keyboard as quickly as possible. You will play for a **<u>total of 10 minutes</u>**.

Every time you successfully press the 'a' and then the 'b' button, you will receive a point. Note that points will only be rewarded when you **<u>alternate button pushes</u>**: just pressing the 'a' or 'b' button without alternating between the two will not result in points. Buttons must be pressed by hand only (key-bindings or automated button-pushing programs/scripts cannot be used) or the task will not be approved.
Feel free to score as many points as you can.

As a bonus, you will be paid an extra 5 cents for every 100 points.

We will divide your 10-minute playing time into 4 rounds. This means, for example, that if you score 2,000 points, you will receive an extra 1 dollar.

You will have a 10-second break in between each round. On the next page is an example of how the task will work. Try pressing 'a' and 'b' alternately to score points. We have limited the point total below to a maximum of 50 as this is just practice, but the actual task will not have a limit. Try to reach 50 **as quickly as possible.**

Page Break

JS

Q44 Press 'a' then 'b'...

Points: 0

Page Break

Q45 **You will be paid for your performance.** Please confirm you understand this.

- Yes (1)
- No (2)

Page Break

Q46 Proceed to the next page when you are ready to play the task. Your 10-minute task will begin immediately when the page loads and will be divided into 4 rounds.

Page Break

Button-pushing task

End of Block: Treatment 9 – Piece Rate

Start of Block: Demographics

Q61
Demographics

You are nearly finished. Please answer the remaining demographics questions below.

Page Break

Q62 Did you have any particular strategy when performing the task across these rounds?

- No (0)
- Yes (1)

Display This Question:

If Did you have any particular strategy when performing the task across these rounds? = Yes

Q63 Please briefly describe your strategy.

____

Page Break

Q64 Did you have a points goal or target for each round?

- No (1)
- Yes (2)

Display This Question:

If Did you have a points goal or target for each round? = Yes

Q65 What was it?

____

Page Break

Q66 How would you agree with the following statements?

I enjoyed this task.

- Strongly Disagree (1)
- Disagree (2)
- Neutral (3)
- Agree (4)
- Strongly Agree (5)

Page Break

Q67 I felt stress while performing this task.

- ◯ Strongly Disagree (1)
- ◯ Disagree (2)
- ◯ Neutral (3)
- ◯ Agree (4)
- ◯ Strongly Agree (5)

Page Break

Q68 What is your total household income, including all earners in your household?

▼ Less than $10,000 (1) ... More than $150,00 (12)

Q69 What is the highest level of education you have completed?

▼ Some high school (1) ... Earned graduate or professional degree (7)

End of Block: Demographics