I thank my dissertation committee members Rob Bloomfield (chair), Kevin Hallock, Mark Nelson, and Ted O'Donoghue for valuable guidance and advice. I also thank Scott Asay, Ramji Balakrishnan, Ola Bengtsson, Jeremy Bentley, Tim Brown, David De Angelis, Scott Emett, Eva Labro, Eldar Maksymov, Joe Pacelli, Harry Segal, Abdullah Shahid, Sean Wang, and Ken Whelan for helpful comments and suggestions. I gratefully acknowledge financial support from Cornell University’s Johnson Graduate School of Management.
Performance standard horizons, uncertainty, and controllability

The horizon of a performance standard specifies the time given to achieve the standard. Performance standard horizons vary considerably in practice, and the psychology literature provides mixed results on whether short or long horizons are more effective in motivating workers. In this paper I use an experiment with a real-effort task to show that the motivating effect of performance standards is determined by an interaction between the horizon of the standard and the amount of uncertainty in the relationship between effort and performance. When uncertainty was low, participants exerted significantly more effort when assigned performance standards with short horizons compared to long horizons. However, when uncertainty was high, the motivational effect of short horizons decreased because participants perceived themselves as having less ability to influence achievement of their standards through their efforts. The results demonstrate that the horizon of performance standards can affect worker motivation, and that horizons should be set as short as permitted by uncertainty in the situation. These findings provide an explanation for variation in performance standard horizons observed in practice, and suggest that investments in accounting systems that reduce noise in performance measures can allow management to set shorter horizon standards. The results also extend to the managerial accounting literature by demonstrating that the controllability of a performance standard is affected by the matching of the standard’s time horizon with the amount of noise in the worker’s ability to achieve the standard. Finally, the results clarify mixed findings about horizon effects in the goal-setting literature.
1. Introduction

A performance standard is a threshold level of performance that must be achieved within a given period of time, known as the horizon. Considerable variation in the horizon of performance standards is observed in practice (De Angelis and Grinstein 2011) and the psychology literature provides mixed results on whether short or long horizons are more effective in motivating workers. This study proposes and tests the idea that the motivating effect of horizon varies with the level of uncertainty in workers’ effort-performance relationship. Thus, while many previous studies have documented the effects that the level of a performance standard has on worker behavior (e.g. Anderson et al. 2010; Bol et al. 2010), this study demonstrates that the horizon of a performance standard can also have a measurable impact.

In this paper, I use an experiment with a real-effort task to show that the motivating effect of performance standards is determined by an interaction between the horizon of the standard and the amount of uncertainty in the relationship between effort and performance. When uncertainty was low, participants facing short horizons exerted significantly more effort than those facing long horizons, even though the rate of work demanded by the standards did not differ. However, when uncertainty was high, participants’ effort did not vary with the horizon of their assigned standards. Debriefing responses indicate that participants felt they had little ability to control standard achievement through their efforts when facing short horizons and high uncertainty. However, uncertainty did not alter their perception of controllability when they faced long-horizon standards.

The results clarify the mixed findings in the goals literature on the effect of horizon. Some studies have found that goals with short horizons are more effective at motivating workers because frequent goal achievement provides satisfaction and increases task enjoyment, and because workers increase effort as they get closer to achieving a goal (see, e.g., Bandura and Schunk 1981; Heath et al. 1999). However, other studies have found that long-horizon goals lead to higher motivation and performance (e.g. Kirschenbaum et al. 1981) and conjecture that short-horizon
goals reduce workers’ flexibility to allocate effort over time. These authors argue that the loss of flexibility can lead workers to miss many of their short-horizon goals, which then decreases motivation. Thus, the literature suggests an interaction between goal horizon and uncertainty but has never tested it. This paper provides the first direct test of this idea and thereby contributes to the goal setting literature.

The results also contribute to a central topic in managerial accounting – controllability. The controllability principle states that a worker should be compensated based on that which he can control, and that failure to adhere to this principle can lead to dysfunctional behavior, such as reduced motivation (Merchant 1985). In this paper, I demonstrate that the controllability of a performance standard is affected by the interaction of the standard’s time horizon with the amount of noise in the worker’s effort-performance relationship. When uncertainty was low, short-horizon standards were perceived as more controllable than long-horizon standards. However, when uncertainty was high, the reverse was true. The pattern of effort (and therefore motivation) exhibited by participants was identical to the pattern of perceived controllability. My findings are therefore consistent with the controllability principle.

In the experiment, 140 participants performed a computer-based, real-effort task of counting the number of zeros in a sequence of tables of 1’s and 0’s. They were awarded points for each table counted correctly and paid by the point at the end of the experiment. Participants were assigned performance standards that were displayed prominently while they worked but that had no effect on their monetary compensation. The experiment employed a $2 \times 2$ between-subjects design, varying standard horizon (short, long) and environmental uncertainty (high, low). To vary standard horizon, half of the participants were assigned 8 standards and given 5 minutes to achieve each one (totaling 40 minutes). The other half were assigned 2 standards and given 20 minutes to achieve each one (also totaling 40 minutes). To vary environmental uncertainty, participants faced “shocks” that adjusted their point total every five minutes by a random number. For the
participants assigned to the low-uncertainty condition, this random number was between -2 and 2; for the participants assigned to the high-uncertainty condition, the number was between -12 and 12. Thus, the shocks had a mean of zero for all, but higher variance for some.

Results indicate a significant interaction in which the effect of horizon on effort (measured as the number of tables counted correctly) depends on the level of uncertainty. Consistent with the hypothesis that short-horizon standards are more motivating when uncertainty is low, participants in the short-horizon condition counted significantly more tables than their counterparts in the long-horizon condition when uncertainty was low. Consistent with the hypothesis that the motivating effect of short-horizon standards is reduced when uncertainty is increased, participants in the short-horizon condition counted no more tables than their counterparts in the long-horizon condition when uncertainty was high. Debriefing responses indicated that uncertainty had a different impact on participants’ perceptions of control over standard achievement. Specifically, uncertainty significantly reduced participants’ perception of control when they faced short-horizon standards, but had no impact when they faced long-horizon standards. Debriefing responses also indicated that participants with short-horizon standards enjoyed the task significantly more and were more willing to participate again, consistent with the robust finding in the goals literature that goals increase task enjoyment and can make a boring task seem fun (Harackiewicz et al. 1984; Locke and Latham 2002).

This paper makes three contributions. First, this paper shows that the horizon of a standard can have an impact on motivation, even if the level of the standard is held constant. In practice, variation in the horizon of standards is observed and can range from a day, or even an hour, to multiple years. The findings in this paper demonstrate that the horizon of a standard must be matched to the level of uncertainty in the environment, with shorter horizons preferred whenever uncertainty allows it. To the extent that investments in information systems can reduce uncertainty in measuring worker performance, shorter-horizon standards can be employed.
Second, the paper shows that the actual and perceived controllability of a performance standard varies with the interaction between the standard’s horizon and the uncertainty in the worker’s production function. Holding environmental variables constant, a performance standard with a short horizon may be far less controllable than a performance standard with a long horizon. This lack of controllability has the potential to reduce worker motivation, which could in turn reduce the efficiency of the worker's incentive scheme.

Finally, the paper demonstrates that performance standards can affect worker motivation even when they are not tied directly to an extrinsic incentive such as a bonus. That is, workers derive satisfaction from achieving the goals that the standard causes them to internalize, possibly allowing managers to supplement monetary pay with non-monetary utility.

The remainder of the paper is organized as follows. Section 2 discusses prior literature and develops hypotheses. Section 3 discusses my research design. Section 4 discusses my results, and Section 5 concludes.

**2. Background and hypotheses**

The U.S. Office of Personnel Management (OPM) defines a performance standard as “a management-approved expression of the performance thresholds, requirements, or expectations that must be met to be appraised at a particular level of performance” (U.S. Office of Personnel Management 1998). OPM further states that a key attribute of a performance standard is its timeliness (horizon), which specifies “when or by what date the work is produced.” In other words, a performance standard specifies a level of performance, but also requires specification of a time by which the level of performance must be met. Performance standards may be based on measures of inputs, such as hours worked, or outputs, such as profitability, production quantity or quality, uptime, and customer satisfaction. Managers use performance standards to communicate their desires and facilitate planning, to extract employees’ private information, and for compensation,
formal and informal performance evaluation, and promotion (Sprinkle 2003; Anderson et al. 2010; Murphy 2000; Indjejikian and Nanda 2002).

2.1 Performance standards and goals

Performance standards are distinct from the psychological construct of a “goal,” but the two are closely linked. A worker’s decision to attempt to achieve a performance standard represents an internalized desire, and this internalized desire meets Locke et al. (1981) definition of a goal: “what an individual is trying to accomplish; it is the object or aim of an action.” Locke (2001) argues that whenever a person makes a conscious choice to take an action that is directed towards an end, he has set a goal. Thus, the standard is an external stimulus, and the goal is the internal psychological state that results from it.

Architects of performance-standards systems appear to realize that standards create goals for workers. For example, Heckman et al. (1997) state that the primary function of performance-standards systems is to establish goals: “By defining goals and providing incentives for achieving these goals, performance-standards systems attempt to bring to public-sector agencies the type of discipline that markets bring to firms.” Sprinkle (2003) also acknowledges the goal-creating role of performance standards when he refers to the “goals contained in accounting budgets and standards.”

Goals are a critical part of the human cognitive subsystem through which people regulate their behavior (Bandura 1989). A large literature in psychology shows that achieving a goal is its own reward, providing non-monetary utility (Locke and Latham 1990; Locke and Latham 2002; Heath et al. 1999) and enhancing interest in the task (or equivalently, decreasing the cost of effort). Examples abound of cases where workers increased their performance simply because they were given performance standards. Latham (2004) reports that loggers who were assigned challenging standards cut down more trees and missed work less often than those told to “do your best” or those given no specific requirement. Kerr and Landauer (2004) describe General Electric’s system
of "stretch goals" (which would more-accurately be called "stretch standards") that are used as motivators but not always directly tied to pay. These anecdotes are representative of hundreds illustrating that performance standards can motivate workers even when not tied to extrinsic incentives. These studies also demonstrate that managers must carefully consider the motivational effects of performance standards, even if their own intent is to use the goal as a basis for communication, planning, or providing extrinsic incentives.

2.2 Choosing performance standard horizon

Managers must make a number of decisions when setting performance standards. One of the most studied is the level of the standard, which can range from a minimum performance standard to an easy-to-meet standard to a “stretch” standard (e.g. Merchant and Ferreira 1988; Anderson et al. 2010; Kerr and Landauer 2004). In this paper I focus on a far less-studied decision – the length of time that an employee is given to achieve the standard, called the horizon. In practice, there is considerable variation in the horizon of observed standards. For example, De Angelis and Grinstein (2011) analyze CEO compensation contracts and report that CEO performance standards range from a quarter to almost eight years.

2.2.1 Efficacy of short horizons

Since workers who commit to achieving performance standards set internal goals, I rely on the literature on goal setting to predict the effect of standard horizon on motivation. Numerous studies on goal horizon have been conducted over the last three decades yet the evidence on the motivating effect of goal horizon1 remains mixed (Locke and Latham 1990). Some authors argue that since goals become more motivating when one is closer to achieving them, shorter goal horizons are more effective (e.g. Heath et al. 1999). Heath et al. (1999) propose that goals serve as

---

1 The literature on goal setting has adopted the terms “proximal” and “distal” to refer to short-horizon and long-horizon goals, respectively. In the more recent behavioral economics literature, authors refer to the process of setting goal horizon as “goal bracketing.” An overarching goal is “bracketed” into multiple subgoals. In this paper, I use the term horizon to refer to the length of time available to achieve a goal.
reference points and that people experience sensations of gains or loss based on whether they exceed or fail to achieve their goals. Through numerous experiments, Heath et al. (1999) demonstrate that the gain/loss function of goals appears to be identical to the Prospect theory value function (Kahneman and Tversky 1979). The Prospect theory value function is convex in the loss domain and concave in the gain domain, and therefore exhibits the property of diminishing sensitivity. In the loss domain, where the value function is convex, the closer one gets to the reference point, the higher is the marginal utility. If goals function as Heath et al. (1999) propose, then the closer one is to achieving a goal, the more it will motivate them.

Other studies also argue that decomposing a goal into multiple subgoals means that individuals can achieve more goals and therefore attain more satisfaction from goal achievement (Locke and Latham 2002). Bandura and Simon (1977) argue that people intuitively realize the efficacy of short-horizon goals, because a number of experimental studies report that subjects assigned long-horizon goals set short-horizon goals for themselves.

Finally, another line of studies argues that short horizons prevent procrastination and thereby increase effort (e.g. Bandura 1989). Many recent papers in behavioral economics suggest that individuals set personal goals as commitment devices to help mitigate the effects of present bias, or self-control problems (e.g. Jain 2009; Hsiaw 2010; Koch and Nafziger 2009; Suvorov and van de Ven 2008). This stream is part of a growing literature suggesting that management control systems not only to align incentives of managers and workers, but also to help workers with their own self-control problems (Kaur et al. 2010a).

2.2.2 Inefficacy of short horizons

While the above literature provides evidence that short horizons are often effective in motivating workers and improving performance, other studies argue that long-horizon goals are preferable in at least some circumstances. One line of research argues that longer horizon goals afford more flexibility in allocating effort over time (Kirschenbaum 1985; Read et al. 1999). For
example, Kirschenbaum et al. (1981) conducted an experiment in which they assign study goals to undergraduates. They found that those assigned monthly study goals spent more time studying and improved grades more than those assigned daily goals or no goals. They conjecture but do not test the idea that daily goals were incompatible with the irregular schedules of their subjects, and that after failing to achieve initial daily goals, the subjects gave up.

Read et al. (1999) survey the literature on how to frame choices and reach a similar conjecture. They argue that when longer horizons are considered, people are better able to make tradeoffs across choices and that this leads to higher returns to effort. For example, they state that “artists and academics have days when the muse is resident, and days when she is painfully absent. In such situations, the most efficient way for workers to organize their time is to work long hours when the return to time spent is high, and take leisure when the return is low.” Consider what would happen if the artist/academic in this hypothetical scenario had daily output goals. On days when the muse was absent, the artist/academic would fail to achieve his goal, and on other days, he would far exceed his goal. If the artist/academic was loss-averse, the sensations of loss experienced on days when he failed to achieve his goals would outweigh the sensations of gain on the days when he exceeded his goals, and he would have been better off with a longer goal horizon. Under longer horizons, random variation in outcomes can cancel out through the power of large sample sizes, thereby giving people more control over goal achievement through their efforts.

2.3 A theory of optimal performance standard horizons

The choice of optimal performance standard horizon is based on the theory that performance standards lead workers to set internal goals, and that their responses to the properties of those goals will determine the observed responses to the standard. Therefore, in the following paragraphs, I assume that workers internalize the goal of achieving their assigned performance standards and discuss the effects of horizon and uncertainty on the motivating effect
of those goals. I also assume that achieving the performance standard has no implications for compensation, performance evaluation or other extrinsic incentives.

The evidence for and against short goal horizons implies a tradeoff between the greater motivation of having nearby goals and the loss of flexibility that accompanies short goal horizons. Goals with shorter horizons are more motivating when uncertainty in the worker’s effort-output relationship is low. However, when uncertainty is high, workers do not have time to respond to uncertainty in the environment and therefore cannot achieve short-horizon goals. I therefore propose that the motivating effect of performance standards is determined by an interaction between horizon and uncertainty because that interaction determines the controllability of the standard. This conjecture relies on the assumption that workers decide to achieve the performance standard assigned to them and therefore set a goal based on it.

I derive predictions about the interaction between horizon and uncertainty by incorporating one straightforward behavioral assumption into a traditional economic framework: that people derive utility from merely achieving a goal. There is support for this assumption in the literature. For example, several behavioral economic models explicitly assume a fixed increase in utility from goal attainment (e.g. Jain 2009), and in their review of goal setting theory, Locke and Latham (2002) explicitly state that “the more goal successes one has, the higher one’s total satisfaction.” My assumption is much simpler than that used in other recent behavioral economic models of goals. For example, Hsiaw (2010) and Koch and Nafziger (2009) both follow the Heath et al. (1999) model of goals as reference points and assume that people experience a sensation of gain or loss that is increasing in the amount by which they exceed or fall short of their goal. In contrast, I derive my hypotheses simply by assuming that simply achieving a goal results in a “burst” of utility that is greater than when the goal is missed.

If achieving more goals leads to higher utility, then in the absence of uncertainty, shorter goal horizons, which mean more goals, increase returns to effort. As a result, people will work
harder to achieve a series of short-horizon goals than they would to achieve a single long-horizon goal that is equivalent to the sum of the short horizon goals. This leads to the first hypothesis:

**Hypothesis 1**: Shorter goal horizons increase effort when uncertainty about the effort-output relationship is low.

Hypothesis 1 suggests that performance standard horizons should be as small as possible when uncertainty is low. However, uncertainty in the effort-output relationship imposes a lower bound on the optimal horizon of the performance standard. In a world with no uncertainty, workers know exactly what the results of their efforts will be and can decide whether the benefits of achieving a performance standard exceed the costs. When uncertainty is present, the relationship between effort and output becomes stochastic. A worker who exerts high effort might not achieve his performance standard, and as performance standard horizons are shortened, the control the worker has over achieving his standard through his efforts is diminished. The worker simply may not have enough time to exert the effort required to guarantee that he achieves his performance standard. By contrast, if the horizon is longer, the worker has the flexibility to respond to the realizations of uncertainty that he observes. It is also possible that uncertainty may cancel itself out over a longer horizon.

Consider a worker with a convex cost of effort who is paid a piece rate for his output. When there is no uncertainty in the environment, the relationship between effort and output is deterministic. The worker will choose a level of effort at which the marginal benefit (the piece rate) just equals the marginal cost of effort. Call this level of effort $e^*$. Since achieving a goal provides extra utility (by assumption), the worker can be induced to work a little harder if given a goal. He will work until the marginal benefit (piece rate plus goal utility) just equals the marginal cost. Call this level of effort $e^*_g$. If uncertainty is introduced into the effort-output relationship, where uncertainty has a mean of zero, a rational worker will reduce his effort. The reason is that, because of the uncertainty, he is no longer guaranteed to achieve the goal if he exerts $e^*_g$. At $e^*_g$, the
To see the differential effect on workers with short- and long-horizon goals, consider a setting where a worker faces two consecutive periods, each with a goal of $g$, and uncertainty that realizes at the end of each period. By the reasoning above, the worker will reduce effort as uncertainty (the variance in the distribution) is increased. Now consider the effect on a worker who faces a single goal of $2g$ for both periods. The worker can respond to the uncertainty realized after the first period. If the uncertainty is negative, he can increase effort. If it is positive, he can decrease effort. A rational worker will realize that the sum of the two draws might cancel each other out. His control over the attainment of his goal of $2g$ is therefore less affected by the uncertainty than his counterpart who faced two goals of $g$, and his reduction in effort in response to an increase in uncertainty will be less.

Thus, increasing uncertainty reduces the controllability of performance standards for workers facing short-horizon standards, but has a smaller effect on controllability for workers facing long-horizon standards. The reduction in controllability will reduce motivation (Bandura and Wood 1989). This leads to hypothesis 2.

**Hypothesis 2**: Greater uncertainty in the effort-output relationship reduces the motivational advantage of shorter goal horizons.

In this paper, I do not posit a main effect of uncertainty on effort. A series of analytical papers in neoclassical economics considered the problem of labor supply under uncertainty, but without the presence of a goal (e.g. Block and Heineke 1973; Tressler and Menezes 1980). One prediction of these papers is that when uncertainty is additive (e.g. a fixed salary plus a variable bonus), risk-averse workers increase effort to hedge against low realizations of uncertainty (i.e. an income effect). Some related empirical papers support this finding (e.g. Parker et al. 2005). However, because none of these papers considered the effects of goals, I do not rely on them for my
predictions. If an income effect arises because of uncertainty in this experiment, it would weaken the directional effects hypothesized above, and work against finding a result.

3. Method

3.1 Participants

Participants were 140 students at a large, private research university. Participants were recruited through the web site of the business school’s laboratory, which maintains a large pool of subjects for use in experimental research. Since the task required no specific skills or knowledge beyond basic computer skills, the ability to exert effort on a straightforward task, and the ability to respond to a typical pay-for-performance arrangement, this subject pool was an appropriate choice (Libby et al. 2002). Upon seeing advertisements for studies at the lab’s web site, members of the subject pool voluntarily sign up for experiments. This study’s advertisement stated “In this experiment you will be asked to perform a computer-based task for money. We expect that the study will take 75 minutes. This study is only for pay (no credit). Expect to earn about $16.”

3.2 Task overview and experimental design

The goal of the experiment was to examine the impact of performance standards on real effort, and how that impact varies with uncertainty and standard horizon. To elicit real effort, participants were asked to perform a straightforward, repetitive task for which they received monetary compensation. Participants were awarded points for each repetition of the task completed, and gained or lost points as a result of periodic random events. Points were redeemed for cash at the end of the experiment using a fixed, known exchange rate. Participants were also given performance standards but these were not tied to their pay – there was no bonus for achieving a standard or penalty for failing to do so.

---

2 A supplementary analysis in section 4.5.1 examines whether participants had a reference point of $16 when they arrived in the lab. It concludes that many did come with a goal of $16, but that the likelihood of arriving with this goal did not differ across treatments.
Participants’ effortful task was to count the number of zeros in tables comprised of 54 randomly generated zeros and ones. The task is adapted from Abeler et al. (2011), who argued that it was useful for assessing real effort because it “does not require any prior knowledge, performance is easily measurable, and there is little learning possibility; at the same time, the task is boring and pointless, and we can thus be confident that the task entailed a positive cost of effort for subjects.” The positive cost of effort is desirable for this study since the theory employs an assumption common to most agency models – that effort is costly (Lambert 2001).

The experiment used a 2 × 2 between-subjects design. Performance standard horizon was manipulated at two levels, short and long. Participants in the short-horizon condition were given 8 standards to achieve and 5 minutes to achieve each one. Participants in the long-horizon condition were given 2 standards to achieve and 20 minutes to achieve each one. Uncertainty was manipulated at two levels, low and high. After every 5 minutes, a random number was added to each participant’s point total. For participants in the low-uncertainty condition, this random number was drawn from a discrete uniform distribution with support [-2, 2]. For participants in the high-uncertainty condition, this random number was drawn from a discrete uniform distribution with support [-12, 12]. To ensure that realizations of random events were the same across the four cells of the design, each cell used the same 35 sequences of random numbers (a different one for each participant in the cell). The dependent variable, effort, was measured by the number of tables counted correctly in the main round. A timeline of the experiment in the short- and long-horizon conditions is shown in Figure 2.

<< INSERT FIGURE 2 ABOUT HERE >>

---

3 Each sequence was drawn from a continuous, uniform distribution with support [-1, 1], and then scaled up to either [-2, 2] or [-12, 12], depending on the uncertainty level for that participant. The number was also rounded down to the nearest integer.
3.3 Task and procedures

Upon entering the lab, each participant was seated at a partitioned computer station, assigned to a treatment using a predetermined random-assignment scheme, and verbally instructed to “follow the instructions on the screen.” The experimental session was divided into two rounds: a 10-minute preliminary round and a 40-minute main round. The preliminary round required participants to count tables without any performance standards or random events. The main round was identical to the preliminary round, except that it included the performance standards and random events described above. The preliminary round reduces noise in the analysis of effort by giving participants an opportunity to learn the task before the main round, and by providing a baseline measure of individual variation in performance before performance standards are introduced.

3.3.1 Preliminary round

Instructions for the preliminary round (see appendix in section 9) described the rules governing the task and incentives. The instructions informed participants that, after counting the zeros in a table, they should enter their count into a textbox (see screen snapshot in Figure 1). If the participant entered the correct count, the computer displayed the next table and incremented the participant’s point total by 2 points. If the participant entered an incorrect count, he was given two more tries to count the same table before a penalty of 2 points was assessed. The penalty scheme, adopted from Abeler et al. (2011), discourages effort avoidance through guessing and skipping tables that look difficult to count. Participants were also informed that they would be paid a show-up fee of $10 plus an additional 15¢ for every 10 points earned.

After the preliminary round ended, the software informed participants of the number of tables they had counted correctly and their pay for the preliminary round. They then progressed to the main round of the experiment.
3.3.2 Main round

After completing the preliminary round, participants were shown instructions for the main round, which are reproduced in the appendix in section 10. These instructions informed participants that they would continue to perform the same task of counting tables of 1’s and 0’s, that each table would continue to be worth 2 points, that a penalty of 2 points would be assessed after 3 incorrect answers, and that the pay rate would remain at 15¢ for every 10 points received. The instructions then informed participants of changes from the preliminary round.

3.3.2.1 Performance standards

All participants were informed that they would now be given performance standards to achieve. Participants in the short-horizon conditions were told that their standard was to earn 42 points every 5 minutes for the next 40 minutes (8 standards). Participants in the long-horizon conditions were told that their standard was to earn 168 points every 20 minutes for the next 40 minutes (2 standards). They were told that each standard was independent of the others – points earned on an early standard would not carry over to future standards. Finally, participants were informed that there would be no bonus for achieving a standard or penalty for failing to achieve a standard. It was emphasized that pay would be determined solely by total points accumulated.

3.3.2.2 Random events

Participants were then informed that since “unexpected events can affect work performance ... a random number will be added to – or subtracted from – your point total after every 5 minutes.” Participants in the low-uncertainty condition were informed that this random number would be between -2 and 2, and those in the high-uncertainty condition were informed that the random number would be between -12 and 12. These point additions were referred to as “random events.”

---

4 Instructions to subjects described performance standards as targets. See page 2 of the main round instructions in section 10.
The mean of small samples of random numbers often deviates significantly from the population mean, and such deviation occurred for many participants. Twenty-five percent of participants (35 / 140) experienced a significant net loss to their points when their random numbers turned out to be mostly negative. Approximately 30% (41 / 140) experienced a significant net addition to their points. The remaining participants experienced a mix of positive and negative additions to their points that summed to approximately zero. The distribution of the sum of random events for all participants is shown in Figure 6.

3.3.2.3 Real-time feedback

While counting tables, participants were given real-time feedback about progress towards their current performance standard. As depicted in the screen snapshot in Figure 1, on-screen feedback displayed the current standard, the current point level, and the distribution of points after the next random event. For example, if a participant in the low-uncertainty condition had 4 points, the screen reminded him that after the next random event, he would have between 2 and 6 points. Finally, the screen displayed time remaining to achieve the current standard, and the timer was updated every minute.

3.3.2.4 Notification of random events and standard achievement

After every 5 minutes of the main round, a window appeared notifying participants of the random event that occurred; it showed the amount of the point gain or loss and the new point total (see Figure 3). If time to achieve a standard had elapsed, this window also told participants whether or not they had achieved their standard. Standard achievement was announced in bold, green letters, and animated fireworks were displayed on the screen (see Figure 4). Failure to achieve a performance standard was announced in dark red letters with no fireworks (see Figure 5). A timeline of the main round that indicates the relative timing of notifications of random events and standard achievement is shown in Figure 2.
3.3.2.5 The Leisure Zone

Participants were given the choice to either count tables or engage in an alternative task that I named the Leisure Zone. The purpose of the Leisure Zone was to obtain a clear measure of time spent working. For example, if participants wished to temporarily stop working to reward themselves for achieving a standard or to procrastinate on a current standard, they could enter the Leisure Zone. Participants were told that they could take a break from counting tables at any time during the main round by clicking on a button titled “Leisure Zone” on the main screen shown in Figure 1. Clicking the button opened a window that displayed a sequence of content from popular websites (see Figure 7). The content consisted of comics from two strips (XKCD and Calvin and Hobbes), celebrity gossip from TMZ.com, whimsical photos of animals, recent news stories, and pictures from a fashion website (The Sartorialist). To entice participants to enter the Leisure Zone, they were shown sample content while reading the instructions. They were told that content would appear in random order\(^5\) and that they could stay in the Leisure Zone as long as they liked. There was no penalty for time spent in the Leisure Zone, but it was not possible to earn points while in there.

3.3.2.6 Debriefing and final payment

After completing the 40-minute main round, participants were shown a screen that informed them of the total points they had accumulated from counting tables and from random events. The screen informed them of their total pay, which was equal to the $10 show-up fee plus the total of accumulated points multiplied by the point-cent exchange rate. Participants then answered 14 debriefing questions. After completing these questions, the software instructed the participants to...

---

\(^5\) The actual sequence of content was predetermined and was the same for all participants.
participant to go to the lab’s office, where the experimenter paid the participant in cash\textsuperscript{6}. The participant then left the lab.

3.3.3 Comprehension and manipulation checks

Participants were asked nine questions to assess their understanding of the main round. These questions asked participants about their assigned standards (level of each standard, number of standards, time allotted for each standard, independence of each standard from the others), determinants of pay, random events (distribution, frequency, possible impact on standard achievement), and the Leisure Zone (when it would be available and the effect on their pay). On every question, more than 85\% of participants provided a correct answer on the first try; on most questions only one or two participants out of 140 provided an incorrect answer on the first try. The responses to the comprehension check questions indicate that participants understood the instructions. Most importantly, participants understood that their pay was determined only by the total points they accumulated and not by standard achievement. Ninety-four percent of participants (131/140) answered this question correctly on the first try. Identical inferences can be drawn from the results reported below if these nine participants are excluded.

3.4 Software

The task, instructions, comprehension check questions, and debriefing questions were administered using custom software written in the C# programming language. The software was designed to allow participants to perform the experiment with minimal involvement from the experimenter. The software captured every keystroke and mouse movement made by each participant, and disabled the Windows system keys to prevent participants from using their assigned computer for any purpose but the experiment.

\textsuperscript{6} After each participant completed debriefing questions, the software sent the experimenter an email notifying him of the participant’s computer number and final pay. This allowed the experimenter to remain in the office in case many participants finished at the same time and had to line up to be paid.
4. Results

4.1 Cell means of dependent variables

Panel A of Table 1 reports cell means for raw performance, which is defined as the number of tables counted correctly by each participant during the main round of the experiment. The panel also reports means for baseline-adjusted performance, which is defined as raw performance less ten times the number of tables counted by each participant during the last four minutes of the preliminary round. Baseline-adjusted performance reduces the substantial between-subject variation in productivity by accounting for the number of tables the participant would have counted in the 40-minute main round had he continued to work with the same effort intensity as in the preliminary round. Figure 8 presents raw and baseline-adjusted performance across the four cells of the design.

Mean baseline-adjusted performance is different from zero in every cell at the 1% level, indicating that participants’ rate of counting tables in the main round was higher than that in the preliminary round. Table 1, panel B reports the results of an ANOVA for baseline-adjusted performance. Horizon is significant (F = 4.47, p = 0.0364), indicating that participants worked harder when provided with short-horizon goals. The interaction between Horizon and Uncertainty is also significant (F = 4.93, p = 0.0280), indicating that the effect of Horizon changes at different levels of Uncertainty.

4.2 Hypotheses tests

Table 1, panel C reports the results of the hypothesis tests. Hypothesis 1 states that when uncertainty is low, short-horizon performance will exceed long-horizon performance. I test hypothesis 1 with a planned contrast that $\mu_{\text{ShortLow}} > \mu_{\text{LongLow}}$, where $\mu_{XY}$ represents the mean of

---

7 The choice of 4 minutes reflects a tradeoff. Performance continues to improve throughout the preliminary round, but shorter windows provide noisier measures of preliminary-round performance since participants were interrupted while counting their last table. Identical inferences can be drawn from the results reported below if the last 5, 3, or 2 minutes are used as the basis for measuring baseline performance.
baseline-adjusted performance with horizon of X and uncertainty of Y. Participants in the short-horizon/low-uncertainty condition did indeed work harder than their counterparts in the long-horizon/low-uncertainty condition. Their adjusted performance was higher by 26.64 tables (F = 9.53, p = 0.0013, one-tailed\(^8\)). This result strongly supports hypothesis 1 and suggests that short-horizon goals are motivating when there is low task uncertainty.

Hypothesis 2 predicts that the beneficial effect of short horizon standards will diminish as task uncertainty is increased. This hypothesis was tested by using a planned contrast of

\[
(\mu_{\text{ShortLow}} - \mu_{\text{LongLow}}) > (\mu_{\text{ShortHigh}} - \mu_{\text{LongHigh}}).
\]

This hypothesis is also supported strongly. The performance benefit of short-horizon goals is an additional 27.3 tables when uncertainty is low compared to when it is high (F = 4.93, p = 0.0140, one-tailed). This result suggests that uncertainty provides a bound on the usefulness of shorter performance standard horizons. Support for the hypothesis is driven primarily because uncertainty reduces performance when horizons are short (untabulated but see Figure 8: F = 3.45, p = 0.0327, one-tailed). Performance under long-horizon goals increases as uncertainty increases, but the increase is not statistically significant (untabulated but see Figure 8: F = 1.64, p = 0.2030, two-tailed\(^9\)).

\[4.2.1\] Robustness of hypotheses

An alternative to analyzing baseline-adjusted performance is to use an ANCOVA in which raw performance is the dependent variable and baseline performance is a covariate to account for between-subject variation in performance when goals are absent. Table 1, panel D reports the results of this ANCOVA. There is a main effect of horizon (F = 4.67, p = 0.0325) and a statistically

\[8\] MacDonald (2011) states that the F-distribution with 1 numerator degree of freedom and d denominator degrees of freedom is equivalent to the square of the Student t-distribution with d degrees of freedom. Therefore, “as long as the F test has 1 numerator degree of freedom, the square root of the F statistic is the absolute value of the t statistic for the one-sided test.” Since all contrast tests reported in this paper are Wald tests with 1 numerator degree of freedom, this method is appropriate, and is equivalent to dividing the two-tailed p-value in half.

\[9\] I use a one-tailed test for the simple effect of uncertainty when horizon is short because the theory suggests that short-horizon goals become less effective when uncertainty increases. Since I offer no prediction for long-horizon goals, a two-tailed test is more appropriate for them.
significant interaction between horizon and uncertainty ($F = 4.06, p = 0.0460$). Panel E shows that hypothesis 1 is supported and reveals a similar effect size as with ANOVA: performance in the short-horizon/low-uncertainty cell was higher than that in the long-horizon/low-uncertainty cell by 24.53 tables ($F = 8.83, p = 0.0018$, one-tailed), compared to 26.64 with ANOVA. Panel E shows that hypothesis 2 is also supported and reveals a similar effect size: the benefit of short-horizon standards over long-horizon standards decreases by 23.71 tables as uncertainty increases ($F = 4.06$, $p = 0.0230$, one-tailed), compared to 27.30 with ANOVA. Thus both hypotheses are supported and the effect sizes are similar to their counterparts in Table 1, panel C. I conclude that the use of baseline-adjusted performance as my dependent variable yields identical inferences and is easier to interpret10.

4.3 Perceived controllability of standards

The theory underlying hypotheses 1 and 2 suggests that the controllability of a standard is determined by the interaction between standard horizon and uncertainty in the environment. Uncertainty reduces the controllability of short-horizon standards by much more than it reduces that of long-horizon standards, because when the horizon is long, there is time to respond to uncertainty as it is realized.

I test this notion of controllability by analyzing the responses to a debriefing question on perceived controllability of performance standards. Participants were asked to rate agreement with the statement "I had control over whether I could achieve the targets" on a 7-point Likert scale that ranged from strong disagreement to strong agreement. Because some participants experienced mostly negative or mostly positive realizations of uncertainty (see Figure 6), I group the realizations of net (total) realizations into terciles and analyze the responses to this question separately for each tercile, as well as for the entire sample.

10 In an untabulated ANCOVA, I test for, but do not find, a 3-way interaction between horizon, uncertainty, and baseline performance. Thus, variations in baseline performance do not have different effects in the different experimental conditions.
Table 2, panel A shows the mean response to this question for the entire sample and for each tercile of net shocks. Results from the entire sample show that participants in the low-uncertainty conditions clearly perceived the task as more controllable than those in the high-uncertainty conditions. The ANOVA in Table 2, panel B shows a significant main effect of uncertainty ($F = 13.23, p = 0.0004$), and the top-left graph of Figure 9 shows this pattern graphically. The means of the low-uncertainty cells are positive and significantly different from zero, indicating that these participants felt they had control over standard achievement. By contrast, the means of the high-uncertainty cells are not significantly different from zero, indicating a neutral perception of controllability.

The ANOVA for the entire sample (Table 2, panel B) does not show an interaction between uncertainty and horizon ($F = 0.00, p = 0.9611$), due largely to the great variation in perceptions of controllability when shocks consistently hinder or aid standard attainment. To get a better look at perceived controllability in the absence of such effects, I examine the tercile of participants who experience neutral net shocks. The top-right graph of Figure 9 shows that perceived controllability in the short-horizon, low-uncertainty cell is significantly higher than that in the short-horizon, high-uncertainty cell (untabulated but see Figure 9: $F = 6.79, p=0.0059$, one-tailed). However, when standards have long horizons, there is no simple effect of uncertainty (untabulated but see Figure 9: $F = 0.21, p=0.6524$, two-tailed). A planned contrast (Table 2, panel C, neutral net shocks) shows that short-horizon standards are perceived as more controllable than long-horizon standards when uncertainty is low, but that the difference in perceived controllability declines as uncertainty is increased ($F = 2.21, p = 0.0715$, one-tailed). Thus, consistent with the theory underlying hypotheses 1 and 2, controllability of short-horizon standards is strongly affected by uncertainty, but controllability of long-horizon standards is not affected.

When uncertainty consistently hinders standard achievement (negative net shocks), participants in the high-uncertainty conditions unsurprisingly report that the standards are less
controllable. Table 2, panel C, negative net shocks, shows a significant main effect of uncertainty ($F = 16.40, p = 0.0003$). Surprisingly, participants in the long-horizon/high-uncertainty condition found the standards the least controllable. The bottom-left graph of Figure 9 shows the pattern, and an ANOVA in Table 2, panel C, negative net shocks, confirms a significant interaction between horizon and uncertainty when in the most negative tercile of net shock ($F = 3.21, p = 0.0830$).

Furthermore, the mean response for the long-horizon/high-uncertainty condition is -1.33 (Table 2, panel A) and this is significantly different from zero ($t = -2.53, p = 0.0353$, two-tailed). I conjecture that the effect of uncertainty on perceived controllability is strongest among participants with long-horizon standards because they received consistently negative shocks that repeatedly pulled them back from each standard, making it difficult for them to achieve their goals. By contrast, those with short horizon standards were able to achieve many of their goals because some shocks were positive, even if the sum of all their shocks was negative.

Participants who received consistently positive shocks report that controllability is the same, regardless of experimental condition. The ANOVA in Table 2, panel C, positive net shocks, shows no significant main effects of horizon and uncertainty, and no interaction. The bottom-right graph of Figure 9 shows the pattern of responses graphically. The responses for three of the four cells do not differ significantly from zero. This suggests that participants recognized that uncertainty weakened the link between effort and standard achievement, even though the uncertainty was helpful.

**4.4 Task enjoyment and willingness to participate again**

The theory underlying hypotheses 1 and 2 assumes that people enjoy having goals to work towards, and that achieving more goals leads to more satisfaction (i.e. utility) than achieving fewer goals. I address this aspect of the theory by analyzing data from two debriefing questions. The first question asked participants to rate their agreement with the statement: “*I enjoyed my task of counting tables and trying to achieve targets.*” The second asked them to rate their agreement with
the statement: "I would participate in this experiment again." Both questions were answered on a 7-point Likert scale that ranged from strong disagreement (-3) to strong agreement (+3), with zero indicating a neutral response. Given that the task was the same for all participants, differences in responses between experimental conditions must be due to the effects of horizon and uncertainty.

Table 3, panel A shows the mean responses to the question on task enjoyment by experimental condition, and Figure 10 displays the pattern of responses. Consistent with the theory, there is a significant main effect of horizon (F = 4.09; p = 0.0452); participants in the short-horizon condition reported significantly higher enjoyment of the task than participants in the long-horizon condition. In fact, the responses of participants in the short-horizon condition do not differ significantly from zero (untabulated: \( \mu = 0.21, t = 0.5756, p = 0.2844 \)), indicating neither enjoyment nor dislike of the task, while participants in the long-horizon condition gave responses that were negative and significantly different from zero (untabulated: \( \mu = -0.80, t = -2.5026, p = 0.0173 \)), indicating dislike of the task. The first contrast test in Table 3, panel A shows that participants in the short-horizon/low-uncertainty cell enjoyed the task significantly more than participants in the long-horizon/low-uncertainty cell (F = 4.40, p = 0.0189, one-tailed), providing support for the theory that more goals provide more utility, and that goals increase task enjoyment.

Table 3, panel B shows the mean responses to the question on willingness to participate again by experimental condition, and Figure 10 displays the pattern of responses. The pattern of results is identical to that for the question on task enjoyment. There is a main effect of uncertainty (F = 4.83; p = 0.0297), with participants in the short-horizon condition indicating a greater (and positive) willingness to participate again than those in the long-horizon condition. There is a significant simple effect of horizon when uncertainty was low, with participants in the short-horizon/low-uncertainty cell indicating a significantly higher willingness to participate again than participants in the long-horizon/low uncertainty cell (F = 2.47, p = 0.0594, one-tailed).
The pattern of responses for the questions on task enjoyment and willingness to participate again provide support for the theory that achieving more goals provides more utility. They also confirm a robust finding in the goal setting literature: that working towards and achieving goals enhances interest in the task.

4.5 Supplementary results

4.5.1 Other goals

This paper asserts that participant behavior was affected by goals. It is therefore possible that goals other than those assigned by the experimenter affected behavior. Other likely sources of goals are participants’ predetermined goals (set before coming to the lab) and goals set by participants during the experiment. Debriefing questions were used to capture these extraneous goals.

To measure participants’ predetermined goals, a debriefing question asked “Did you have a goal of earning a certain amount of money before you came to the lab today?” Fifty-two percent of participants answered yes. The percentage varied from 50-54% across experimental conditions, but the differences were not significant (in an untabulated ANOVA, neither the main effects of horizon and uncertainty nor their interaction were significant). Of those participants who came to the lab with a goal, 84% said their goal was to earn $16; the others gave responses that ranged from $10-20. Since the predetermined goals did not vary across experimental condition, it is unlikely they affected behavior systematically during the experiment.

To measure goals that participants self-set during the experiment, a debriefing question asked “Besides the targets that you were given, did you set any goals for yourself during the experiment?” This question was followed by one that asked participants to describe their goals. Sixty-eight percent of participants reported setting an additional goal during the experiment. The percentages ranged from 61-73% across cells, but an untabulated ANOVA reveals no statistically significant main effects or interaction between horizon and uncertainty. The open-text responses
reveal that many participants attempted to increase their table-counting accuracy and a few set higher or lower standards for themselves. The open-text responses will be analyzed in a future version of this paper.

5. Concluding remarks

This paper examines how the horizon of performance standards affects worker motivation. I conduct an experiment that shows that performance standards with short horizons can be especially motivating for workers, but that the beneficial effect of short horizons is reduced by uncertainty in the worker’s effort-output relationship. The interaction between horizon and uncertainty occurs because workers with short-horizon standards have (and perceive) little control over standard achievement. Despite their effort, negative realizations of uncertainty might prevent them from attaining their standards, and given the short horizon, they do not have time to adjust their effort level in response to random shocks as they occur. Participants who faced short horizons and high uncertainty understood that they had little control over standard achievement. The results are consistent with the theory that workers internalize the performance standards assigned to them in the form of personal goals, and that psychological models of the motivating effect of goals can therefore be applied to performance standards.

My study has several implications. First and foremost, I demonstrate that the horizon of a performance standard can have a measureable impact on worker motivation, even if the level of the standard is held constant. By demonstrating that worker motivation increases when performance standard horizons are properly matched to the uncertainty in production functions, I provide a potential explanation for the variation observed in performance standard horizons in practice.

Next, I show that consistent with past research, workers are likely to internalize performance standards as goals. Practitioners should therefore realize that goals have motivational effects on workers, and that these effects will operate independently of links to extrinsic incentives
that are based on standards. Two robust findings in the goal setting literature are that people work harder when working towards a goal, and that goals can make boring tasks fun (Locke and Latham 2002). Therefore, simply providing a standard to workers can motivate them, and managers can augment monetary compensation with satisfaction from achieving goals. This occurs routinely in nonprofit organizations, who are able to attract workers who are willing to accept less compensation in order to pursue their social goals (Hallock 2000).

The motivational benefits of goals are not a free lunch. Since people experience a sensation of loss if they fail to achieve a goal, they will be less motivated if they do not believe they can achieve their assigned goal (Locke and Latham 1990; Heath et al. 1999). To avoid this, managers must choose performance standards carefully. Many previous studies have shown that the level of a performance standard is important. The present results show that the horizon of a standard can have positive effects on motivation, but only when properly matched to the level of uncertainty in the environment. An implication of the interaction between horizon and uncertainty is that investments in accounting systems that reduce uncertainty can give managers more flexibility in choosing the horizon of performance standards.

This study contributes to the managerial accounting literature by demonstrating that the controllability of a performance standard varies with the interaction between the standard’s horizon and the uncertainty in the worker’s production function, and that controllability affects worker effort. Thus, my findings are consistent with the controllability principle.

My study contributes to the goal setting literature by clarifying earlier mixed findings. Proponents of short-horizon goals had argued that proximity to a goal increases motivation, while opponents argued that short-horizon goals reduce flexibility, lead to failures to achieve goals, and therefore reduce motivation. My study is the first to provide direct evidence of this tradeoff. I demonstrate that uncertainty is the moderating factor that determines whether short-horizon goals are more effective than long-horizon goals.
This study is subject to several limitations. First, the experiment is predicated on the assumption that the performance standards assigned to participants would be internalized and would cause participants to set personal goals. This assumption appears to be valid, as debriefing responses indicate that participants paid attention to their assigned standards. However, some participants set additional goals that need to be analyzed further to determine whether the goal of achieving the external standard (and not another goal) truly drove behavior. Another limitation is that the results may vary when overall performance and achievement of standards have strong effects on extrinsic incentives. Larger financial incentives may swamp the non-monetary psychological effects of internalized goals. More generally, different methods of linking standard achievement to incentives may alter how goals are internalized, and how subjects react to hitting or missing standards. Incentives might also alter the degree of the intensity of a worker’s desire to achieve his goal, which Locke and Latham (2002) refer to as goal commitment.

I envision two extensions of this work. This experiment took place in a non-strategic setting and was an out-of-equilibrium test of worker behavior. Future research might examine whether principals understand the motivating effects of performance standards and choose the levels and horizons of performance standards to match the uncertainty in the agent’s effort-output relationship. Future research might also investigate whether agents respond differently to performance standards that were chosen strategically by the principal to maximize his profits. In such a setting, fairness concerns such as reciprocity and inequity aversion could easily arise and distort agent behavior from what was observed in the non-strategic setting studied here (Rabin 1993; Bolton and Ockenfels 2000).

Other extensions could explore the use of performance standards to mitigate workers’ self-control problems. An emerging view in economics is that the modern organization has evolved to help workers deal with their desires to enjoy utility in the present moment, even though doing so
entails a greater cost to future utility (Kaur et al. 2010a). Under this view, penalties for misbehavior and low output are written into contracts to mitigate such effects; workers supply low effort not because they are effort averse, as is commonly assumed in agency models (Lambert 2001), but because the costs and benefits of effort are not contemporaneous. A worker with a self-control problem will overweight the immediate cost relative to the long-term benefit. By providing an immediate penalty for poor present performance, the firm can offset the effects of the self-control problem by providing a commitment device to the worker. Future studies could alter the timing of the costs and benefits of effort and leisure in an experimental setting, and see if agents and principals choose performance standards to mitigate the effects of self-control problems.

---

11 The tendency to overweight the present relative to the future is known as present-bias. Present-bias can cause time-inconsistent preferences, such as self-control problems. See Frederick et al. (2002) for a review of time-inconsistency and present-bias, and O'Donoghue and Rabin (1999) for a model of self-control problems.  
12 Kaur et al. (2010b) provide compelling evidence that workers realize they are afflicted by present bias. In a field experiment, they show that workers voluntarily choose contracts with penalties for low output, even when they could have chosen a standard piece rate contract that dominates the penalty contract.
6. References


7. Tables

Table 1: Effect of performance standard horizon and uncertainty on effort

Participants counted tables of 1’s and 0’s in two rounds, a preliminary round and a main round. **Raw performance** refers to the number of tables counted correctly during the 40-minute main round. **Baseline performance** refers to the number of tables counted correctly during the last 4 minutes of the 10-minute preliminary round, multiplied by 10; this represents the number of tables the participant would have counted correctly in the main round if he had continued working with the same effort intensity as during the last 4 minutes of the preliminary round. **Baseline-adjusted performance** equals **raw performance** minus **baseline performance**. Each participant earned two points for every correctly counted table. During the main round, a random number was added to each participant’s point total after every 5 minutes. In the **low** (high) **uncertainty** condition, this random number was uniformly distributed between -2 and 2 (-12 and 12). Participants were also given performance standards that did not affect their pay. **Short** (long) **horizon** refers to the experimental condition in which participants were given eight 5-minute (two 20-minute) standards of earning 42 (168) points.

Panel A shows performance in each of the experimental conditions. One, two, or three asterisks (⁎) next to a baseline-adjusted cell mean indicate that the mean is different from 0 at the 10%, 5%, or 1% level, respectively (two-tailed t-test). Panel B shows an ANOVA of **baseline-adjusted performance** on **horizon** and **uncertainty**. Panel C shows the results of the planned contrasts that were used to test hypotheses 1 and 2. Panel D shows an ANCOVA of **raw performance** on **horizon** and **uncertainty**, with **raw performance** as a covariate. Panel E shows tests of hypotheses 1 and 2 based on the ANCOVA in panel D.

Panel A: Cell means

<table>
<thead>
<tr>
<th>Goal horizon</th>
<th>N</th>
<th>Low uncertainty</th>
<th>High uncertainty</th>
<th>Diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short</td>
<td>69</td>
<td>173</td>
<td>168</td>
<td>5</td>
</tr>
<tr>
<td>Long</td>
<td>71</td>
<td>155</td>
<td>163</td>
<td>-8</td>
</tr>
<tr>
<td>Difference</td>
<td>18</td>
<td>5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Panel B: ANOVA of baseline-adjusted performance on horizon and uncertainty

<table>
<thead>
<tr>
<th>Source</th>
<th>Partial SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Prob &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>12,807.86</td>
<td>3</td>
<td>4,269.29</td>
<td>3.23</td>
<td>0.0245</td>
</tr>
<tr>
<td>Horizon</td>
<td>5,903.08</td>
<td>1</td>
<td>5,903.08</td>
<td>4.47</td>
<td>0.0364</td>
</tr>
<tr>
<td>Uncertainty</td>
<td>238.41</td>
<td>1</td>
<td>238.41</td>
<td>0.18</td>
<td>0.6717</td>
</tr>
<tr>
<td>Horizon x Uncertainty</td>
<td>6,520.27</td>
<td>1</td>
<td>6,520.27</td>
<td>4.93</td>
<td>0.0280</td>
</tr>
<tr>
<td>Residual</td>
<td>179,767.32</td>
<td>136</td>
<td>1,321.82</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>192,575.17</td>
<td>139</td>
<td>1,385.43</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Panel C: Tests of hypotheses using planned contrasts on baseline-adjusted performance

<table>
<thead>
<tr>
<th>Hypothesis 1:</th>
<th>Effect size</th>
<th>F</th>
<th>p-value (1-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short-horizon/low-uncertainty performance &gt; long-horizon/low-uncertainty performance</td>
<td>26.64</td>
<td>9.53</td>
<td>0.0013</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hypothesis 2:</th>
<th>Effect size</th>
<th>F</th>
<th>p-value (1-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Short-horizon minus long-horizon performance at low uncertainty) &gt; (Short-horizon minus long-horizon performance at high uncertainty)</td>
<td>27.30</td>
<td>4.93</td>
<td>0.0140</td>
</tr>
</tbody>
</table>

Panel D: ANCOVA of raw performance on horizon and uncertainty using baseline-adjusted performance as a covariate

<table>
<thead>
<tr>
<th>Source</th>
<th>Partial SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Prob &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>145,276.69</td>
<td>4</td>
<td>36,319.17</td>
<td>30.17</td>
<td>0.0000</td>
</tr>
<tr>
<td>Baseline performance</td>
<td>138,671.60</td>
<td>1</td>
<td>138,671.60</td>
<td>115.20</td>
<td>0.0000</td>
</tr>
<tr>
<td>Horizon</td>
<td>5,619.29</td>
<td>1</td>
<td>5,619.29</td>
<td>4.67</td>
<td>0.0325</td>
</tr>
<tr>
<td>Uncertainty</td>
<td>87.59</td>
<td>1</td>
<td>87.59</td>
<td>0.07</td>
<td>0.7878</td>
</tr>
<tr>
<td>Horizon x Uncertainty</td>
<td>4,883.38</td>
<td>1</td>
<td>4,883.38</td>
<td>4.06</td>
<td>0.0460</td>
</tr>
<tr>
<td>Residual</td>
<td>162,504.20</td>
<td>135</td>
<td>1,203.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>307,780.89</td>
<td>139</td>
<td>2,214.25</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Panel E: Tests of hypotheses using planned contrasts on covariate-adjusted raw performance

<table>
<thead>
<tr>
<th>Hypothesis 1:</th>
<th>Effect size</th>
<th>F</th>
<th>p-value (1-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short-horizon/low-uncertainty performance &gt; long-horizon/low-uncertainty performance</td>
<td>24.53</td>
<td>8.83</td>
<td>0.0018</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hypothesis 2:</th>
<th>Effect size</th>
<th>F</th>
<th>p-value (1-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Short-horizon minus long-horizon performance at low uncertainty) &gt; (Short-horizon minus long-horizon performance at high uncertainty)</td>
<td>23.71</td>
<td>4.06</td>
<td>0.0230</td>
</tr>
</tbody>
</table>
Table 2: Perceived controllability of performance standards

After completing the main round of the experiment, participants were asked a series of debriefing questions. This table shows the response to the question on whether participants believed they had control over achievement of their assigned performance standards (“targets”). Participants were asked, “How strongly do you agree with the following statement: ‘I had control over whether I could achieve the targets.’” Responses were solicited on a 7-point Likert scale in which -3 indicated strong disagreement, 0 indicated a neutral response, and 3 indicated strong agreement.

Panel A shows the mean response to the question on perceived controllability for the entire sample and for each tercile of net shock. An asterisk (*) next to a cell mean indicates that the mean is different from 0 at the 10% level or better. Panel B displays an ANOVA of the response to the question using horizon and uncertainty to explain the variation; it also shows planned contrasts.

Eight random numbers (“shocks”) were added to each participant’s points total. Some participants experienced mostly negative shocks, others experienced mostly positive shocks, and others received a mix. Each participant’s cumulative, or net, shock was computed and these were ranked into terciles. The distribution of net shocks is shown in Figure 6. Four subjects were dropped from this analysis because they experienced extreme values of net shocks (-13 in low uncertainty, and -72 in high uncertainty). Panel C displays a separate ANOVA and planned contrasts for each net shock tercile.

During the main round, a random number was added to each participant’s point total after every 5 minutes. In the low (high) uncertainty condition, this random number was uniformly distributed between -2 and 2 (-12 and 12). Participants were also given performance standards that did not affect their pay. Short (long) horizon refers to the experimental condition in which participants were given eight 5-minute (two 20-minute) standards of earning 42 (168) points.

Panel A: Cell means of perceived controllability

<table>
<thead>
<tr>
<th>Goal horizon</th>
<th>Entire sample</th>
<th>Negative net shocks</th>
<th>Neutral net shocks</th>
<th>Positive net shocks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Low uncertainty</td>
<td>High uncertainty</td>
<td>Diff</td>
</tr>
<tr>
<td>Short</td>
<td>67</td>
<td>1.53*</td>
<td>0.42</td>
<td>1.11</td>
</tr>
<tr>
<td>Long</td>
<td>69</td>
<td>1.40*</td>
<td>0.26</td>
<td>1.14</td>
</tr>
<tr>
<td>Difference</td>
<td></td>
<td>0.13</td>
<td>0.16</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Goal horizon</th>
<th>N</th>
<th>Low uncertainty</th>
<th>High uncertainty</th>
<th>Diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short</td>
<td>31</td>
<td>1.87*</td>
<td>0.25</td>
<td>1.62</td>
</tr>
<tr>
<td>Long</td>
<td>29</td>
<td>1.36*</td>
<td>1.07*</td>
<td>0.29</td>
</tr>
<tr>
<td>Difference</td>
<td></td>
<td>0.51</td>
<td>-0.82</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N</th>
<th>Low uncertainty</th>
<th>High uncertainty</th>
<th>Diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short</td>
<td>20</td>
<td>0.91</td>
<td>0.67</td>
</tr>
<tr>
<td>Long</td>
<td>21</td>
<td>1.00*</td>
<td>0.50</td>
</tr>
<tr>
<td>Difference</td>
<td></td>
<td>-0.09</td>
<td>0.17</td>
</tr>
</tbody>
</table>
Panel B: ANOVA and contrasts of perceived controllability for entire sample

<table>
<thead>
<tr>
<th>Source</th>
<th>Partial SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Prob &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>43.39</td>
<td>3</td>
<td>14.46</td>
<td>4.49</td>
<td>0.0049</td>
</tr>
<tr>
<td>Horizon</td>
<td>0.71</td>
<td>1</td>
<td>0.71</td>
<td>0.22</td>
<td>0.6398</td>
</tr>
<tr>
<td>Uncertainty</td>
<td>42.65</td>
<td>1</td>
<td>42.65</td>
<td>13.23</td>
<td>0.0004</td>
</tr>
<tr>
<td>Horizon x Uncertainty</td>
<td>0.01</td>
<td>1</td>
<td>0.01</td>
<td>0.00</td>
<td>0.9611</td>
</tr>
<tr>
<td>Residual</td>
<td>425.55</td>
<td>132</td>
<td>3.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>468.94</td>
<td>135</td>
<td>3.47</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Effect size:  
Contrast 1:  
(Short-horizon minus long-horizon response at low uncertainty) >  
(Short-horizon minus long-horizon response at high uncertainty) = -0.03  
F = 0.00  
p-value (1-tailed) = 0.4806

Panel C: ANOVA and contrasts of perceived controllability by net shock tercile

ANOVA and contrasts of perceived controllability – Neutral net shocks

<table>
<thead>
<tr>
<th>Source</th>
<th>Partial SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Prob &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>21.30</td>
<td>3</td>
<td>7.10</td>
<td>2.38</td>
<td>0.0790</td>
</tr>
<tr>
<td>Horizon</td>
<td>0.35</td>
<td>1</td>
<td>0.35</td>
<td>0.12</td>
<td>0.7320</td>
</tr>
<tr>
<td>Uncertainty</td>
<td>13.61</td>
<td>1</td>
<td>13.61</td>
<td>4.57</td>
<td>0.0370</td>
</tr>
<tr>
<td>Horizon x Uncertainty</td>
<td>6.58</td>
<td>1</td>
<td>6.58</td>
<td>2.21</td>
<td>0.1429</td>
</tr>
<tr>
<td>Residual</td>
<td>166.88</td>
<td>56</td>
<td>2.98</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>188.18</td>
<td>59</td>
<td>3.19</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Effect size:  
Contrast 1:  
(Short-horizon minus long-horizon response at low uncertainty) >  
(Short-horizon minus long-horizon response at high uncertainty) = 1.33  
F = 2.21  
p-value (1-tailed) = 0.0715
### ANOVA and contrasts of perceived controllability – Negative net shocks

<table>
<thead>
<tr>
<th>Source</th>
<th>Partial SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Prob &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>60.74</td>
<td>3</td>
<td>20.25</td>
<td>7.62</td>
<td>0.0006</td>
</tr>
<tr>
<td>Horizon</td>
<td>6.15</td>
<td>1</td>
<td>6.15</td>
<td>2.31</td>
<td>0.1385</td>
</tr>
<tr>
<td>Uncertainty</td>
<td>43.59</td>
<td>1</td>
<td>43.59</td>
<td>16.40</td>
<td>0.0003</td>
</tr>
<tr>
<td>Horizon x Uncertainty</td>
<td>8.53</td>
<td>1</td>
<td>8.53</td>
<td>3.21</td>
<td>0.0830</td>
</tr>
<tr>
<td>Residual</td>
<td>82.40</td>
<td>31</td>
<td>2.66</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>143.14</td>
<td>34</td>
<td>4.21</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Effect size

<table>
<thead>
<tr>
<th>Contrast 1:</th>
<th>Effect size</th>
<th>F</th>
<th>p-value (1-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Short-horizon minus long-horizon response at low uncertainty) &gt; (Short-horizon minus long-horizon response at high uncertainty)</td>
<td>-1.98</td>
<td>3.21</td>
<td>0.0415</td>
</tr>
</tbody>
</table>

### ANOVA and contrasts of perceived controllability – Positive net shocks

<table>
<thead>
<tr>
<th>Source</th>
<th>Partial SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Prob &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1.62</td>
<td>3</td>
<td>0.54</td>
<td>0.15</td>
<td>0.9280</td>
</tr>
<tr>
<td>Horizon</td>
<td>0.01</td>
<td>1</td>
<td>0.01</td>
<td>0.00</td>
<td>0.9492</td>
</tr>
<tr>
<td>Uncertainty</td>
<td>1.40</td>
<td>1</td>
<td>1.40</td>
<td>0.39</td>
<td>0.5336</td>
</tr>
<tr>
<td>Horizon x Uncertainty</td>
<td>0.17</td>
<td>1</td>
<td>0.17</td>
<td>0.05</td>
<td>0.8286</td>
</tr>
<tr>
<td>Residual</td>
<td>131.41</td>
<td>37</td>
<td>3.55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>133.02</td>
<td>40</td>
<td>3.33</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Effect size

<table>
<thead>
<tr>
<th>Contrast 1:</th>
<th>Effect size</th>
<th>F</th>
<th>p-value (1-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Short-horizon minus long-horizon response at low uncertainty) &gt; (Short-horizon minus long-horizon response at high uncertainty)</td>
<td>-0.26</td>
<td>0.05</td>
<td>0.4143</td>
</tr>
</tbody>
</table>
Table 3: Task enjoyment and willingness to participate again

After completing the main round of the experiment, participants were asked a series of debriefing questions. Panel A shows the response to the question on task enjoyment. Participants were asked, "How strongly do you agree with the following statement: ‘I enjoyed my task of counting tables and trying to achieve targets.’” Panel B shows the response to the question on participants’ willingness to participate again. Participants were asked, “How strongly do you agree with the following statement: ‘I would participate in this experiment again.’” Responses to both questions were solicited on a 7-point Likert scale in which -3 indicated strong disagreement, 0 indicated a neutral response, and 3 indicated strong agreement.

The first table in each panel shows the mean response in each experimental condition. An asterisk (*) next to a cell mean indicates that the mean is different from 0 at the 10% level or better. The second table in each panel shows an ANOVA of responses on horizon and uncertainty. The third table in each panel shows the results of two planned contrasts on the responses, based on the preceding ANOVA.

During the main round, a random number was added to each participant’s point total after every 5 minutes. In the low (high) uncertainty condition, this random number was uniformly distributed between -2 and 2 (-12 and 12). Participants were also given performance standards that did not affect their pay. Short (long) horizon refers to the experimental condition in which participants were given eight 5-minute (two 20-minute) standards of earning 42 (168) points. Four participants were excluded from this table because they experienced extreme negative random numbers. Inferences are identical if these participants are included.

Panel A: Analysis of debriefing question on task enjoyment

<table>
<thead>
<tr>
<th>Goal horizon</th>
<th>N</th>
<th>Low uncertainty</th>
<th>High uncertainty</th>
<th>Diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short</td>
<td>67</td>
<td>0.21</td>
<td>-0.24</td>
<td>0.45</td>
</tr>
<tr>
<td>Long</td>
<td>69</td>
<td>-0.80*</td>
<td>-0.62*</td>
<td>-0.18</td>
</tr>
<tr>
<td>Difference</td>
<td></td>
<td>1.01</td>
<td>0.38</td>
<td></td>
</tr>
</tbody>
</table>

ANOVA and contrasts of task enjoyment

<table>
<thead>
<tr>
<th>Source</th>
<th>Partial SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Prob &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>20.37</td>
<td>3</td>
<td>6.79</td>
<td>1.71</td>
<td>0.1675</td>
</tr>
<tr>
<td>Horizon</td>
<td>16.21</td>
<td>1</td>
<td>16.21</td>
<td>4.09</td>
<td>0.0452</td>
</tr>
<tr>
<td>Uncertainty</td>
<td>0.60</td>
<td>1</td>
<td>0.60</td>
<td>0.15</td>
<td>0.6976</td>
</tr>
<tr>
<td>Horizon x Uncertainty</td>
<td>3.38</td>
<td>1</td>
<td>3.38</td>
<td>0.85</td>
<td>0.3575</td>
</tr>
<tr>
<td>Residual</td>
<td>523.25</td>
<td>132</td>
<td>3.96</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>543.62</td>
<td>135</td>
<td>4.03</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Panel B: Analysis of debriefing question on willingness to participate again

Cell means of willingness to participate again

<table>
<thead>
<tr>
<th>Goal horizon</th>
<th>N</th>
<th>Low uncertainty</th>
<th>High uncertainty</th>
<th>Diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short</td>
<td>67</td>
<td>1.18*</td>
<td>1.27*</td>
<td>-0.10</td>
</tr>
<tr>
<td>Long</td>
<td>69</td>
<td>0.43</td>
<td>0.53</td>
<td>-0.10</td>
</tr>
<tr>
<td>Difference</td>
<td></td>
<td>0.75</td>
<td>0.74</td>
<td></td>
</tr>
</tbody>
</table>

ANOVA and contrasts of willingness to participate again

<table>
<thead>
<tr>
<th>Source</th>
<th>Partial SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Prob &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>19.23</td>
<td>3</td>
<td>6.41</td>
<td>1.64</td>
<td>0.1837</td>
</tr>
<tr>
<td>Horizon</td>
<td>18.89</td>
<td>1</td>
<td>18.89</td>
<td>4.83</td>
<td>0.0297</td>
</tr>
<tr>
<td>Uncertainty</td>
<td>0.33</td>
<td>1</td>
<td>0.33</td>
<td>0.08</td>
<td>0.7719</td>
</tr>
<tr>
<td>Horizon x Uncertainty</td>
<td>0.00</td>
<td>1</td>
<td>0.00</td>
<td>0.00</td>
<td>0.9946</td>
</tr>
<tr>
<td>Residual</td>
<td>516.53</td>
<td>132</td>
<td>3.91</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>535.76</td>
<td>135</td>
<td>3.97</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Effect size</th>
<th>F</th>
<th>p-value (1-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contrast 1:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short-horizon/low-uncertainty response &gt;</td>
<td>0.75</td>
<td>2.47</td>
<td>0.0594</td>
</tr>
<tr>
<td>long-horizon/low-uncertainty response</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contrast 2:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Short-horizon minus long-horizon response at low uncertainty) &gt;</td>
<td>0.00</td>
<td>0.00</td>
<td>0.4973</td>
</tr>
<tr>
<td>(Short-horizon minus long-horizon response at high uncertainty)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
8. Figures

Figure 1: The screen on which participants performed their task during the main round of the experiment. The table is the box with 1’s and 0’s. Participants had to count the number of zeros in the table and enter it into the textbox labeled “How many zeros are in the table?” If they entered a correct answer, the word *Correct* appeared under the textbox, the table flashed, and a new table was displayed. If they entered an incorrect answer, the word *Wrong* appeared along with “Tries left: x”, where x was 2 or 1 (a total of 3 tries were allowed for each table). The timer in the lower right corner of the screen updated every minute, as did the box with “Remember!” The screen for the preliminary round differed in that there was no *Leisure Zone* button, no reminder of an upcoming random event, no timer in the bottom right hand corner of the screen, and the only feedback text was “Your current points”.

<table>
<thead>
<tr>
<th>Target</th>
<th>42</th>
</tr>
</thead>
<tbody>
<tr>
<td>Your current points</td>
<td>4</td>
</tr>
<tr>
<td>After the next random event you will have between</td>
<td>2 and 6</td>
</tr>
<tr>
<td><strong>Remember! The next random event will occur in less than 5 minutes.</strong></td>
<td></td>
</tr>
</tbody>
</table>

Time remaining for this target: 5 min
Figure 2: Timeline of the task. Participants counted tables of 1’s and 0’s for 40 minutes. Each table was worth 2 points. At the end of every 5-minute period, participants experienced a “shock” – a random number was added to their points; this is represented by a lightning bolt symbol in the timelines above. Shocks in the low (high) uncertainty condition were drawn from discrete U[-2, 2] (U[-12, 12]). In the short-horizon conditions, participants were given a performance standard (or goal) of earning 42 points every 5 minutes. In the long-horizon conditions, participants were given a standard of earning 168 points every 20 minutes. Participants in the long-horizon conditions experienced 4 shocks per goal, whereas those in the short-horizon conditions experienced 1 shock per goal and had no opportunity to react to the shock before goal evaluation.

Figure 3: The Notification window appeared every 5 minutes during the 40-minute main round. This window informed participants of the value of the random event, or “shock” that occurred, as well as their new point total.
Figure 4: The Notification window appeared every 5 minutes during the 40-minute main round. This window informed participants of the value of the random event, or "shock" that occurred, as well as their new point total. When the time for a goal had elapsed, this window also informed participants of whether or not they achieved their goal. This example shows performance standard achievement. When standard achievement occurred, animated fireworks were displayed on the screen for 15 seconds.

Figure 5: The Notification window appeared every 5 minutes during the 40-minute main round. This window informed participants of the value of the random event, or "shock" that occurred, as well as their new point total. When the time for a performance standard ("target") had elapsed, this window also informed participants of whether or not they achieved their goal. This example shows a failure to achieve the standard.
Figure 6: Distribution of the net, or total, shock experienced by participants. During the main round, a random number was added to each participant’s point total after every 5 minutes. In the low (high) uncertainty condition, this random number was uniformly distributed between -2 and 2 (-12 and 12). Each participant experienced 8 random changes (“shocks”) to their points, and this graph shows the distribution of the sum of these shocks across all participants.

All shocks were initially drawn from $U[-1, 1]$ and then scaled to $[-2, 2]$ or $[-12, 12]$ depending on the experimental condition. This figure shows a histogram of the sum of the normalized shocks (i.e. the sum of the $U[-1, 1]$ numbers) experienced by participants. In the low uncertainty condition, the left (right) cluster corresponds to a net point addition of approximately -4 (+5), and the middle cluster corresponds to a net point addition of zero. In the high uncertainty condition, the left (right) cluster corresponds to a net point addition of approximately -20 (+26), and the middle cluster corresponds to a net point addition of approximately +4.

Four participants were excluded from this graph because they had extreme realizations of net shock (normalized net shock of -5.8).
Figure 7: Leisure Zone screen displaying a comic with statistical humor. This screen completely
covered the main work screen (Figure 1) so participants could not earn points while in the
Leisure Zone. The sequence of content appeared random from participants’ perspective, but
was predetermined and was the same for all participants. Once participants clicked the Next
button, they could not return to the previous content. Participants could resume work at any
time by clicking the Show task screen button.
Figure 8: Participants counted tables of 1’s and 0’s in two rounds, a preliminary round and a main round. 

*Raw performance* refers to the number of tables counted correctly during the 40-minute main round.  
*Baseline performance* refers to the number of tables counted correctly during the last 4 minutes of the 10-minute preliminary round, multiplied by 10; this represents the number of tables the participant would have counted correctly in the main round had he continued to work with the same effort intensity as during the last 4 minutes of the preliminary round. *Baseline-adjusted performance* equals *raw performance* minus *baseline performance*. Each participant earned two points for every correctly counted table. During the main round, a random number was added to each participant’s point total after every 5 minutes. In the low (high) uncertainty condition, this random number was uniformly distributed between -2 and 2 (-12 and 12). Participants were also given performance standards that did not affect their pay. *Short (long) horizon* refers to the experimental condition in which participants were given eight 5-minute (two 20-minute) standards of earning 42 (168) points. 

The left graph shows the mean of *raw performance* in each experimental condition. The right graph shows the mean of *baseline-adjusted performance* in each experimental condition. P-values are simple effects of uncertainty at each level of horizon (i.e. whether the slope of the line differs from zero).
Figure 9: Participants counted tables of 1’s and 0’s in two rounds, a preliminary round and a main round. Participants earned two points for every correctly counted table. During the main round, a random number was added to each participant’s point total after every 5 minutes. In the low (high) uncertainty condition, this random number was uniformly distributed between -2 and 2 (-12 and 12). Participants were also given performance standards that did not affect their pay. Short (long) horizon refers to the experimental condition in which participants were given eight 5-minute (two 20-minute) standards of earning 42 (168) points.

After completing the main round of the experiment, participants were asked a series of debriefing questions. The figures show the response to the question on whether they believed they had control over performance standard achievement. Participants were asked, “How strongly do you agree with the following statement: ‘I had control over whether I could achieve the targets.’” Responses were solicited on a 7-point Likert scale in which -3 indicated strong disagreement, 0 indicated a neutral response, and 3 indicated strong agreement.

The top left graph shows the mean response by experimental condition using the entire sample. Going clockwise, the other graphs are for participants who experienced neutral, positive, and negative net shocks, respectively. P-values are simple effects of uncertainty at each level of horizon (i.e. whether the slope of the line differs from zero).
Figure 10: Participants counted tables of 1’s and 0’s in two rounds, a preliminary round and a main round. Participants earned two points for every correctly counted table. During the main round, a random number was added to each participant’s point total after every 5 minutes. In the low (high) uncertainty condition, this random number was uniformly distributed between -2 and 2 (-12 and 12). Participants were also given performance standards that did not affect their pay. Short (long) horizon refers to the experimental condition in which participants were given eight 5-minute (two 20-minute) standards of earning 42 (168) points.

After completing the main round of the experiment, participants were asked a series of debriefing questions. The left figure shows the mean responses by experimental condition to the question on task enjoyment. Participants were asked, “How strongly do you agree with the following statement: ‘I enjoyed my task of counting tables and trying to achieve targets.’” The right figure shows the mean responses by experimental condition to the question on participants’ willingness to participate again. Participants were asked, “How strongly do you agree with the following statement: ‘I would participate in this experiment again.’” Responses to both questions were solicited on a 7-point Likert scale in which -3 indicated strong disagreement, 0 indicated a neutral response, and 3 indicated strong agreement. P-values are simple effects of uncertainty at each level of horizon (i.e. whether the slope of the line differs from zero).
9. Appendix – Instructions for preliminary round

This appendix contains instructions for the preliminary round. These instructions were displayed one page at a time using a viewer application written in the C# programming language. A snapshot of the viewer application is shown in Figure 11 below. The instructions were displayed in a panel on the left, and a sample screen, if available for that page of the instructions, was shown in a separate panel on the right. The panel at right was also used to administer comprehension check questions for the preliminary round.

Figure 11: Snapshot of viewer application for reading instructions and comprehension check questions. Participants read instructions in the panel on the left, and viewed sample screens, if available for that page of instructions, in the panel on the right.
Overview
We need your undivided attention for the next 75 minutes. If you have other commitments, please contact the lab administrator so that you can reschedule.

Reading these instructions
Use the Next page and Previous page buttons below to read these instructions.

Restrictions
No electronic devices
You may not use any electronic devices, such as cell phones, PDAs, iPhones, iPads, and laptops. Please turn off or mute your electronic devices at this time. If you cannot agree to this restriction, then you may not participate.

Computer use
You have been assigned a lab computer. You may only use it for the task described below. If you use your assigned computer for any other purpose, you will be asked to leave and will not receive any payment.

Your task
You will be shown a series of tables. Your task is to count the number of zeros in each table.

A sample work screen is shown at right with the table highlighted. Don’t count the zeros just yet.

Points
You will receive 2 points for each table you count correctly.

If you submit three incorrect answers for a table, 2 points will be deducted from your point total and a new table will be generated.

Example
You count six tables correctly, and miscount one table three times. You will be awarded points as follows:

Correctly counted tables: 6 x 2 points = 12 points
Penalty: 1 x -2 points = -2 points
Total points: 10 points

Tracking your points
The work screen will show you the total number of points accumulated. See highlighting at right.
Instructions - page 5 of 7

Your pay
• You will be paid a fixed fee of $10.
• In addition, you will be paid 15¢ for every ten points that you earn.

Instructions - page 6 of 7

How to enter your answers
After you have counted the number of zeros in a table, enter your answer into the blue box and press the Enter key on the keyboard.

If you enter the correct answer, Correct will appear under the blue box and a new table will be generated. If your input was wrong, Wrong will appear under the blue box, and you will have two additional tries to enter the correct number.

Try it now
Try entering different numbers into the blue box on the sample screen and pressing the Enter key. The correct answer is 24.
Click the Next page button below when you are done with the sample screen.

Instructions - page 7 of 7

If you have questions...
please ask the lab administrator at this time.

When you are comfortable with these instructions...
click the green button below that says I'm done with the instructions. You will be asked some questions that test your understanding of these instructions. Once you answer them correctly, your task will begin.

I'm done with the instructions

Feel free to use the Previous page and Next page buttons to review the instructions while you are answering the questions.
10. Appendix – Instructions for main round

This appendix contains instructions for the main round. These instructions were displayed one page at a time using a viewer application written in the C# programming language. Comprehension check questions for the main round were administered using the viewer application. A snapshot of the viewer application is provided in Figure 11 in the previous appendix. Sample screens were shown to participants as part of the main round instructions. The sample screens for pages 4-8 of these instructions are shown in Figure 1, Figure 3, Figure 5, Figure 4, and Figure 7, respectively.

Different main round instructions were required for each experimental condition. To implement this, I created a master document with embedded tags, such as the “TT” and “MM” tags shown in page 2 of the instructions below. At runtime, the viewer application parsed the master document and replaced each tag with the appropriate value for the experimental condition for that participant. The table below shows the value of each tag for each experimental condition:

<table>
<thead>
<tr>
<th>Tag</th>
<th>Short-horizon Low uncertainty</th>
<th>Short-horizon High uncertainty</th>
<th>Long-horizon Low uncertainty</th>
<th>Long-horizon High uncertainty</th>
</tr>
</thead>
<tbody>
<tr>
<td>TT</td>
<td>42</td>
<td>42</td>
<td>168</td>
<td>168</td>
</tr>
<tr>
<td>MM</td>
<td>5</td>
<td>5</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>NT</td>
<td>8</td>
<td>8</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>MAX</td>
<td>2</td>
<td>12</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>NS</td>
<td>1 random event</td>
<td>1 random event</td>
<td>4 random events</td>
<td>4 random events</td>
</tr>
<tr>
<td>OLDRAND</td>
<td>“”</td>
<td>“”</td>
<td>and any past random events</td>
<td>and any past random events</td>
</tr>
<tr>
<td>LONGTEXT</td>
<td>“”</td>
<td>“”</td>
<td>The window will also let you know how many minutes remain to achieve your target.</td>
<td>The window will also let you know how many minutes remain to achieve your target.</td>
</tr>
</tbody>
</table>

### Instructions - page 1 of 9

Now that you have some experience with the task, there are some changes.

**What has not changed**

- Your task remains the same – counting zeros.
- You will earn 2 points for every table that you count correctly.
- You will lose 2 points if you enter three incorrect answers for a table.
- You will be paid 15¢ for every ten points that you receive.

Click the **next page>>** button to find out what has changed.

### Instructions - page 2 of 9

**Targets**

- People sometimes find it useful to have targets to work towards as they perform tasks like this one. Therefore, to help you perform better...

- Your target is to earn (TT) points every (MM) minutes.

- The rest of the experiment will last for 40 minutes. That means that you will have (NT) targets to achieve.

- Each target is independent.
  - If you exceed a target, those extra points will not count towards the next target.
  - However, all points earned will count towards your pay.

- There is no bonus or penalty for achieving a target or failing to achieve it. Your pay depends only on the total points that you accumulate.
Random Events

- Unexpected events can affect work performance. To depict that here, a random number will be added to – or subtracted from – your point total after every 5 minutes.

- You will gain or lose up to (MAX) points every time a random event occurs.

- Since you have (MM) minutes to achieve each target, you will experience (NS) per target.

- The random events can affect target achievement. For example, say that you have exactly (TT) points at the end of a (MM)-minute period. If you then lose 1 point from a random event, you will not achieve your target.

Revised task screen

The screen has been revised to show progress toward your targets.

- **Target** is your current target.

- **Your current points** shows progress towards your current target. This number includes points earned by counting tables (OLDRAND).

- **After the next random event** shows how your point total could change after the next random event.

- **Time remaining for this target** is shown in the lower right of the screen. It will be updated every minute.

Example

- Type 24 into the blue box and press *Enter* to see how the numbers change.

Notification of random events

When a random event occurs, a window will appear to let you know the amount of the gain or loss, as well as your new point total. (LONG-TEXT)

The **continue** button will appear after the notification window has been open for ten seconds. You should click this button when you are ready to resume work.

If you fail to achieve a target

When time for a target runs out, the notifications window will let you know if you failed to achieve your target.

After you have finished working on one target, you will start working on the next target. The points and timer shown on your task screen will be reset. Don't worry, all points earned towards previous targets will count towards your pay.
Instructions - page 7 of 9

If you achieve a target
When time for a target runs out, the notifications window will let you know if you achieved your target.
The fireworks will appear every time you achieve a target.

Instructions - page 8 of 9

The Leisure Zone
At any time you need a break, you may click the Leisure Zone button on the task screen.

When you do, you will be presented with content from a variety of popular websites, such as funny pictures, comic strips, and celebrity gossip. You may view as much content as you wish, but you won’t earn points while doing so.

Use the Next button to see new content. When you are ready to resume counting tables, click the Show task screen button.

A sample Leisure Zone screen is shown at right with some sample content displayed. Click the Next button a few times to see more sample content.

Instructions - page 9 of 9

If you have questions...
please ask the lab administrator at this time.

When you are comfortable with these instructions...
click the green button below that says I’m done with the instructions. You will be asked some questions that test your understanding of these instructions. Once you answer them correctly, your task will begin.

I’m done with the Instructions

Feel free to use the Previous page and Next page buttons to review the instructions while you are answering the questions.