VENGEFUL SLOWDOWN

An experimental investigation of the effects of punishment on worker effort

by

ELLEN M. STUART

Sarah Jacobson, Advisor

A thesis submitted in partial fulfillment
of the requirements for the
Degree of Bachelor of Arts with Honors
in Economics

WILLIAMS COLLEGE

Williamstown, Massachusetts

May 11, 2011

Table of Contents

List of Tables

List of Figures

Abstract

In this thesis I use a principal-agent experiment to examine how workers change their effort levels in response to two punishment schemes. Subjects are paired so that one subject is the worker (agent) and the other is the boss (principal). Workers avoid being punished by completing a task a specified number of times—the "punishment threshold." In the discretionary punishment treatment, the punishment threshold for worker *i* is selected by boss *i*. In the deterministic punishment treatment, the punishment threshold is pre-determined by the experimenter for all workers. Based on previous research in reciprocity and punishment I hypothesize that workers will reduce their effort level in response to both punishment treatments. Furthermore, I hypothesize that there will be a greater change in effort level in the discretionary treatment because intentional punishment may spur an urge to negatively reciprocate by vengefully slowing down. I find that workers in the discretionary treatment work harder before being punished than workers in the deterministic treatment. In both treatments workers increased their output after having been punished. The increase in effort in the discretionary treatment was greater than the increase in the deterministic treatment. In my conclusion I discuss various explanations for the speed-up, including a perverse reciprocity, guilt or shame, decreasing marginal utility of wealth, and reference points. These results, while surprising in the context of the reciprocity literature, imply that punishment can be successfully used to encourage worker effort in some circumstances.

1. Introduction

When someone is monetarily fined there are two main consequences: an income effect and an emotional effect. The income effect is clear. The punished individual has suffered reduced income as a result of the punishment. The emotional effect is more ambiguous and often ignored, but may also have economic consequences. Does the punished individual accept the punishment, or does he want to retaliate? Which of these effects dominates in determining how he responds to the punishment is an empirical question. This thesis uses a principal-agent experiment to investigate whether punishment chosen by a boss (principal) causes a worker (agent) to reduce his effort in order to take revenge on his boss. I explore this by comparing the slowdown after a punishment intentionally chosen by the boss to the slowdown after a pre-determined punishment holding the loss of income constant.

An event in New York City exemplifies the idea that workers may actively seek revenge against actions that they deem personal affronts. As 2010 came to a close, the Sanitation Department was charged with clearing the streets after an unanticipated blizzard covered the city in more than twenty inches of snow overnight. In the days after the storm, five city employees contacted City Council reporting that some supervisors were instructing their crews not to plow the roads in retaliation to a labor dispute with the city over the demotion of over 100 other Sanitation Department supervisors (Green, 2011). The matter was looked into by the Department of Investigation, an independent law-enforcement organization. “What we are looking at is whether there was intentional misconduct relating to the snow removal—whether or not there was a slowdown,” said Diane Struzzi, a spokeswoman for the Department of Investigation. Joe Mannion,

president of the Sanitation Officers Association, stated, "Absolutely not, there was no slowdown." The president of the union representing the workers under the supervisors in question reiterated this. "My workforce won't put their jobs on the line for their officers," said Harry Nespoli, president of the Uniformed Sanitationmen's Association. Whether or not supervisors from the Sanitation Department were responding to a demotion that was, in their eyes, undeserved and unanticipated, the fact that an organized slowdown was considered a real possibility raises an interesting question. Will workers stage a slowdown in revenge for punishment or another hurtful act by their bosses when such a slowdown is personally costly?

Revenge is an old and compelling plot driver. Sociologist and political theorist Jon Elster defines revenge as "the attempt, at some cost or risk to oneself, to impose suffering upon those who have made one suffer" (Elster, 1990). The impulse to seek revenge is instinctual. The same chemical is released in the brain when someone is plotting revenge as is when falling in love or on cocaine. In all three instances, dopamine fills the brain, causing great anticipation of pleasure in the near future (de Quervain et al., 2004).[1]

Imagine a scenario where a boss's pay is dependent upon her employee's effort. Suppose that, out of frustration that her employee is not working hard enough, she punishes him through a pay cut. If he believes that the punishment was unnecessarily harsh or undeserved, he could intentionally slow down, further reducing his boss's pay in retaliation. Work contracts are inevitably incomplete—the task of making a complete work contract, spelling out what would occur in every possible situation, is impossible.

[1] Dopamine is released when one is plotting revenge, not upon the execution of revenge, suggesting that carrying out revenge is not as satisfying as anticipated.

Workers who greatly underperform can be fired, but there is a range of underperformance that does not violate the work contract and cannot lead to being fired.

Why would a worker slow down, when that might result in additional punishment? This kind of behavior seems particularly inefficient in the case where shirking is directly costly to the worker. Imagine a scenario in which both a worker and his boss are paid based on how many times a worker completes a given task. In this case reduced effort directly affects the worker's pay as well as his boss's, and he would not be maximizing his own material payoff by enacting a costly revenge. In most basic economic models, not maximizing one's payoff is equivalent to not maximizing one's utility. Perhaps, however, utility is determined by more than just material payoff. Perhaps making his boss worse off gives the worker more utility than pursuing his own monetary gain. There are theories within behavioral economics that incorporate other-regarding preferences—the idea that someone's utility could be affected by someone else's monetary gain (or loss). Evidence for this is found in multiple literatures, including studies of reciprocity and gift exchange, punishment and counter-punishment, efficient contracting and intentionality (see Section 2). Countless studies in cooperation have demonstrated that individuals are willing to pay to punish people that do not cooperate. Punished workers may be just as willing to pay to punish their boss for punishing them.

In this paper I explore experimentally whether or not workers change their effort levels differently in response to two different punishment schemes: a discretionary punishment treatment, where bosses are allowed to choose what level of effort their worker must achieve to avoid being punished, and a deterministic punishment treatment, where bosses have no say in setting productivity standards. Instead the standards are

fixed by the experimenter, as if they were company policy. I am interested in the different effects of these different types of punishment on worker effort because the difference is the intentionality of the punishment.

By lowering their effort levels, workers with piece-rate contracts reduce their own material payoff. If the material payoff of the boss affects the worker's utility level, workers may experience increased emotional payoff for taking revenge. My hypothesis is that bosses will experience worker slowdown—that the emotional payoff for taking revenge will overrule the loss in material payoff—and that bosses will experience greater worker slowdown in response to discretionary punishment. This hypothesis is grounded in intuition as well as the basic tenets of research in reciprocity (see section 2.2.1.). If taking revenge provides utility to people who feel wronged, workers who feel they have been slighted personally will gain more emotional payoff by reducing their work efforts and taking revenge than workers who cannot blame their bosses specifically for being punished.

I use an experimental, principal-agent model involving a real effort task to test my hypothesis. As predicted, workers participating in treatments where bosses had a say in the punishment change their effort more than workers who participated in treatments where bosses had no say in the punishment. However, the change in behavior in both treatments is not a slowdown, but a speed-up. Subjects in both treatments work harder after having been monetarily punished and perhaps more so when punishment was discretionary.

There are several explanations for the speed-up. Workers may have been reacting to an increased marginal utility of wealth, or they may have had a priori expectations about

how much money they would earn in the lab. Perhaps they simply improved at the experimental task over the course of the experiment. However, none of these explain the difference between the treatments. It is possible that the workers were reacting to their bosses with some sort of perverse reciprocity, or that they viewed the thresholds as indicative of normative behavior. These explanations are all discussed further in Section 5.

My contribution is as follows: this is the first experiment using a real effort task over two periods controlling for the income effect of monetary punishment and looking for differences in behavioral responses as a result of different implementations of punishment. I believe this study is important because worker efficiency is an issue that workplace managers must face on a daily basis. If different punishment schemes result in different worker behavior, utilizing one or the other could be beneficial to the firm. Understanding what tools are available to principals to elicit maximally efficient behavior is of utmost important in many workplace environments.

The remainder of this thesis is organized as followed. Section 2 is split into subsections reviewing the major themes in reciprocity and gift exchange, punishment and counter-punishment, efficient contracting, and intentionality. I present the theory I will test, the experimental design, and a description of how the experiment was conducted in Section 3. In Section 4 I present and discuss the results. Section 5 concludes.

2. Background

My hypothesis is drawn from results in several major literatures within economics. First, from the literatures on reciprocity and gift exchange, there is a large body of evidence supporting the idea that individuals do in fact have "other-regarding preferences." That is, someone else's payoff may affect my utility level. More

importantly, there is reason to believe that individuals have conditional other-regarding preferences: how much someone else's payoff affects my utility level is conditional on my history with that individual. Second, the punishment literature demonstrates time and time again that people are willing to pay to punish those who do not cooperate or "play by the rules," while extensions of that literature show that the cooperation gained when a punishment option exists is lost when those punished have the opportunity to counter-punish, on net reducing efficiency. Third, literature in efficient contracting suggests that principals and agents prefer bonus contracts to incentive (piece-rate) contracts, but that bonus contracts result in higher payment for both parties. Finally, literature focusing on the importance of intentionality in how individuals interpret the kindness of an action suggests that people recognize a difference in being punished because someone else made a decision as opposed to being punished by protocol.

2.1. Reciprocity

Research in altruism and reciprocity has spanned many fields, including anthropology, sociology, economics, and psychology.[2] There is a difference between "unconditional altruism" and "conditional altruism." Unconditional altruism is when someone is nice to other people regardless of their history. Andrea could have done something mean to Bob, but Bob will be nice to Andrea if Bob is unconditionally altruistic. Conversely, conditional altruism (i.e., reciprocity) is when someone is nice to other people conditional on their past interactions. The conditionally altruistic Bob will only be nice to Andrea if the two have good history. It is also important to distinguish between "intrinsic reciprocity" and "instrumental reciprocity." Intrinsic reciprocity is the

[2] See Sobel (2005) for a survey of interdependent preferences and reciprocity.

willingness to give up some of one's own material consumption in order to increase someone else's material consumption in response to kind behavior, whereas instrumental reciprocity is behaving kindly in response to kind actions with the motive of maintaining a profitable long-run relationship or reputation. That is, instrumental reciprocity is purely strategic and not based on other-regarding preferences.

Theoretical models of reciprocity suggest that individuals will reciprocate both altruistic behavior and selfish or unfair behavior. Just as we are inclined to be nice to people who are nice to us, we have a gut reaction to be unkind to people who are unkind to us. The terms "positive reciprocity" and "negative reciprocity" have been used to distinguish between the mirror phenomena. Cox and Deck (2005) offer a clear and succinct distinction between how the two terms are used in experimental economics. Bob positively reciprocates Andrea when he acts in a way that is personally costly but is beneficial to Andrea because he believes that Andrea previously behaved toward him with kind intentions. Negative reciprocity is the opposite: Bob negatively reciprocates toward Andrea when he acts in a way that is personally costly but harmful to Andrea because he believes that Andrea previously behaved toward him with unkind intentions.

The history of reciprocity research in economics has yielded mixed results. Berg et al. (1995), Cox (2004), and others find evidence for intrinsic reciprocity while List (2006) casts doubts on those results by providing evidence that instrumental reciprocity dominates intrinsic reciprocity in some situations. Why would we care if preferences are reciprocal? Reciprocal preferences change how we think about utility functions in general. If the principal's payoff is an argument of an agent's utility function, we cannot assume that agents will behave solely with the intention of maximizing their own

material payoff. More specifically, reciprocal preferences are going to change what is the most efficient principal-agent contract possible.

2.1.1. Theories of Reciprocity

There are several popular theories of reciprocity. Falk and Fischbacher's (2006) theory of reciprocity is particularly well-suited for explaining my hypothesis that agents will react differently to deterministic punishment than to discretionary punishment. Unlike most other popular economic theories of reciprocity, Falk and Fischbacher use a simple model that evaluates how kind an action is both by what comes of the action but also the original intention behind the action. Falk and Fischbacher justify the inclusion of intentions in their model of reciprocity by pointing to the results of several experimental studies. Falk and Fischbacher apply their model to several standard games and are able to explain common results from many of the most common games (e.g., negative reciprocity in the ultimatum game, positive reciprocity with the gift-exchange/investment game).

Many other reciprocity models can be classified as either inequity aversion models or intentions models. Inequity aversion models are solely concerned with material outcome; intentions are not taken into account. Because of the importance of intentionality in my research question, models that only account for inequity aversion are not as well-suited to explain my hypothesis. Examples of inequity aversion models include Bolton and Ockenfels (2000) and Fehr and Schmidt (1999). Conversely, in intentions models (such as Rabin, 1993, and Dufwenberg and Kirchsteiger, 2004), reciprocity is assumed to be exclusively intention-driven and fairness evaluations are not based on how one's standing compares to those around him. What is most useful about

the Falk and Fischbacher model is that it allows for both inequity aversion and intentionality but does not require either.

The model of reciprocity put forth by Cox et al. (2007) focuses on the effect of one's emotional state on the marginal rate of substitution between his or her own and others' payoffs. This model was expanded in Cox et al. (2008), which focused on how players' choices respond to observable events and opportunities. Though this model also incorporates both inequity aversion and intentionality and could therefore be used to formulate my hypothesis, Falk and Fischbacher's model is simpler and more tractable for my research question. I will describe in more detail how a modified version of the Falk and Fischbacher model predicts my hypothesis in Section 3.

2.1.2. Gift Exchange

In labor economics, reciprocity in the workplace is termed "gift exchange." Gift exchange is not different than reciprocity but rather a way of framing reciprocity in the workplace. In his seminal 1982 paper, George Akerlof explains the existence of firms that consistently pay more than the market-clearing wage in terms of norms of reciprocity and gift exchange. These higher wages are known as "efficiency wages."

Firms have legal and logistical restrictions as to what they can and cannot do in their attempt to encourage workers to perform optimally. Firms rely on their workers not shirking to maximize profit. In order to achieve this high effort level, firms must obtain the willing cooperation of their workers. Akerlof explains that norms exist within groups of workers about what is a fair day's worth of work. Workers expect that, if they work hard, that they will be treated kindly by the firm in return. The worker's gift to the firm is the effort they put in above what is required. Workers give this gift in exchange for

higher wages than what they would receive if they chose to leave their current work in search of other employment. In other words, the most efficient wage may not be the market-clearing wage because firms pay more than the equilibrium wage in order to receive greater effort in return.

Akerlof formalized this idea in the fair wage-effort hypothesis (Akerlof and Yellen, 1990). The fair wage-effort hypothesis claims that workers only work as much as they believe they are being paid for. If workers believe they are not being paid enough for the work expected of them, they will not meet expectations. In formulating this hypothesis, Akerlof and Yellen were influenced by psychology and sociology in addition to data about correlations between wages, skills, and unemployment. Extending the fair-wage effort suggests that if workers feel that they are being slighted in some way, they will withdraw their effort (i.e., seek some form of revenge via negative reciprocity). Though this work focuses on hourly wage work, the same principles can be applied to other payment methods, such as piece-rate contracts.

Fehr et al. (1993) experimentally explored the fair wage-effort hypothesis. Their experiment involved buyers (employers) setting wage levels and sellers (workers) choosing effort levels. Their experimental design was such that there were more sellers than buyers. As a result, one might predict that all of the buyers would offer a price at the lowest amount they were allowed to offer in the experiment. In spite of the existence of involuntary unemployment, the majority of employers offered wages that were well above the market-clearing wage—in some cases more than 100% above the market

clearing wage. The employers provided higher-than-necessary wages in an attempt to elicit higher than average effort levels and were successful. [3]

In 2002 Anderhub, Gächter, and Köningstein used a finitely repeated game to further explore reciprocity in a principal-agent problem.[4] Their results suggest that principals recognize the inclination toward reciprocity and are fully prepared to make that inclination work to their advantage by offering generous contracts. Agents tended to reject contracts that left them with what could be considered an unfairly low fraction of the earnings, while they responded to what they perceived to be fair contracts with higher effort levels.

Gift exchange and reciprocal behavior are observed in the field as well as the lab. Bewley (1999) interviewed hundreds of managers and executives about why they do not lower wages during economic downturns. The response was overwhelming. Bewley's informants stressed the significance of morale. The mood of the employees is crucial to maximizing worker productivity. Using Akerlof's language, keeping morale high by not lowering wages during hard times is a gift given by the firm to the workers; the firm expects the workers to produce high effort levels in exchange. More recently, firms turned to furloughs as a way to deal with the 2009 economic crisis (Petrecca, 2009). Though firing workers may be a more efficient means of cutting costs in the short run, firms sacrificed that efficiency by requiring that many workers take mandatory, unpaid

[3] Evidence from the questionnaire used by the researchers after the experiment supported the idea that employers that offered high wages did so hoping to encourage sellers (of labor) to provide higher effort levels. The questionnaire results also made it clear that most workers chose their effort levels based on the wage received. Workers were more likely to supply high effort levels in response to high wages.

[4] The principal designed a contract and offered it to the agent. The agent then had two choices: to accept the contract or reject it. If the agent accepted the work contract, then the agent chose their effort level. Part of the contract included a recommended effort level, but because the contract was incomplete, the worker was not obligated to provide the effort level proposed by the principal.

leave instead of firing a smaller number of workers. The hope was that workers would also be willing to make some sacrifices in the short run in return for not losing their jobs.

Good work in response to good wages could very well be a form of positive reciprocity. The flip side, of course, is that anything that lowers morale could lead to subsequent retaliation—negative reciprocity—and a greater reduction in efficiency (Dreber et al., 2008). If gift exchange can be viewed as a form of positive reciprocity, where the workers increase effort in return for superior compensation and vice versa, then we would expect monetary punishment to induce a subsequent decrease in effort.

2.2. Punishment

If reciprocity matters to individual utility functions, then we would expect punishment to elicit acts of negative reciprocity from the punished individual. In 2000 Fehr and Gächter published one of the prominent papers on the effects of punishment on behavior, extending the seminal work of Ostrom et al. (1992). Both papers investigate the effect of an external agent in the maintenance of common pool resources. Fehr and Gächter used a common public goods game to explore the effect of punishment on cooperation levels. As expected, nearly complete free-riding was the long-term result in a "no-punishment" scenario. However, the opportunity to punish caused contribution rates to increase to over 50% of the total initial endowment. Despite the fact that it was costly to punish, those individuals that tended to contribute more were willing to pay in order to punish those individuals that thought they could get away with free riding.[5] These convincing results that punishment is effective in

[5] In their experiment, subjects were more likely to be punished when their contributions to the shared pool were below the average contribution, while subjects that contributed above the average contribution level

encouraging cooperative behavior have been replicated throughout the decade. Punishment has shown to be effective even through nonmonetary sanctions such as group disapproval ratings (Dugar, 2010). However, there are questions of whether punishment alone is the most effective way to induce cooperative behavior.[6]

Theories about payoff-maximization predict that subjects will not punish in a one-shot context, but Fehr and Gächter observed that subjects were willing to pay to punish free-riders in one-shot interactions. This suggests that subjects still find it worthwhile to punish even if they will not be able to reap the benefits of punished subjects changing their behavior; that is, there is some inherent utility gain to reciprocating negatively other than personal monetary gain, providing evidence for the idea that reciprocal behavior extends beyond instrumental reciprocity.

The evidence from the reciprocity and punishment literatures that people care about the outcomes of others supports the idea that the utility of bosses would affect the utility of workers. However, punishment is not a one-way street. Many people react to punishment with counter-punishment of their own.

2.2.1. Counter-punishment

Though the majority of punishment research does not include opportunities to counter-punish, there is increasing evidence that when counter-punishment is an option, the option to punish does not actually lead to increased efficiency (Nikiforakis, 2007;

---

were rarely punished. When groups were kept together for the entirety of the experiment, contributions levels increased almost to 100%, suggesting that people care about reputation and social history.

[6] Andreoni et al. (2003) experiment with punishment and rewards using a one-shot proposer-responder game with four treatments. The main result was that punishment and rewards seem to complement each other and are most effective when both are available.

Denant-Boemont, et al., 2007). Counter-punishment is having the ability to strike back at the person that did the original punishing.

Nikiforakis (2007) investigates the effect of opportunities to counter-punish on cooperation levels. Nikiforakis used a public goods game with three treatments: a treatment with no punishment, a treatment with punishment, and a treatment with a punishment stage followed by a counter-punishment stage.

Subjects had the highest earnings in the no-punishment treatment, with lower earnings in the punishment treatment and the lowest earnings in the punishment-counter-punishment treatment. The introduction of opportunities for counter-punishment had a significant effect on contribution levels. Subjects in the punishment-counter-punishment treatment were more likely to contribute to the public account than subjects in the no punishment treatment; subjects in the punishment treatment were even more likely. Most significantly, subjects in the punishment-counter-punishment treatment were significantly less likely to contribute to the public account than subjects in the punishment treatment.

Over three-fourths of the punishments could be classified as "pro-social punishments," meaning that the punished individual originally contributed less than the average contribution of the group. However, there was still some level of anti-social/unjustified punishments (where the target of the punishment was an individual who contributed more than the average of his peers). Those who were punished anti-socially were more likely to counter-punish. That is, individuals who were punished but believed that they did not deserve to be punished were more likely to seek revenge against the individual or individuals who punished them in the first place. This implies that individuals who felt they were intentionally punished for no good reason were more likely to retaliate.

Nikiforakis's results call into question the idea that cooperation and self-governance are achievable as put forth by Ostrom et al. (1992) and Fehr and Gächter (2000), claiming that fear of counter-punishment may make people believe that punishing in the first place is not worthwhile or would not be profitable.[7] His results suggest that punishment may not be a solution to the free rider problem since there are opportunities for counter-punishment in most situations.

Just because there is evidence for the existence of counter-punishment does not necessarily mean that counter-punishment is anticipated. In fact, there is mixed evidence as to whether or not people consider how others will respond to their actions. Research from Neelin et al. (1988) and Harrison and McCabe (1993) suggest that subjects do look one step ahead while other research suggests people behave without considering reciprocal behavior or the behavior of the responder (Andreoni et al., 2003; Cox and Deck, 2005). The importance of expecting counter-punishment is that the expected profitability of punishing someone depends on the expected rate of counter-punishment. This idea naturally leads to the question of how firms can best structure their contracts to maximize expected profitability.

2.3. Efficient Contracting

There is an extensive literature on how firms and employers design compensation contracts in order to encourage workers to perform in the best interests of the firm[8]. Principal-agent contracts that incorporate the intuition of gift exchange and punishment are among the most extensively studied. Anderhub et al. (2002) showed that agents

[7] However, the fact that Ostrom, et al. (1992) also utilize non-experimental data supports the idea that cooperation may be achievable.
[8] See Prendergast (1999) for an overview of this literature.

behaved in order to optimize their own payoff within the limits of their contracts while Fehr et al. (1997) used experimental evidence to argue that reciprocal preferences ought to be taken into account in the enforcement of contracts.

Utility depending solely on payoff maximization suggests that piece-rate contracts (payoffs monotonically increasing with output) would be the most effective at maximizing firm profit because worker pay is always increasing with effort. However, many workers find this kind of compensation alienating (Etzoni, 1971), and it is often not possible. Employers seem to prefer bonus contracts (contracts where workers are paid a base rate and are eligible to receive a bonus at the discretion of the employer). This agrees with the fair work-effort hypothesis. When workers care about reciprocity, they are more likely to provide higher effort levels in hopes of earning a bonus. Both Fehr and Schmidt (2004) and Fehr et al. (2007) found that principals overwhelmingly choose a bonus contract over an incentive contract, and that the bonus contract had the highest net payoff for both principal and agent (Fehr et al., 2007). There is evidence from the field that bonus contracts are frequently utilized, especially for tasks that are complex and difficult to evaluate (MacLeod and Parent, 2000).

The research mentioned previously focuses on piece-rate (linear payment) vs. bonus (nonlinear payment) contracts and finds that bonus contracts are preferred to piece-rate contracts. There are other kinds of nonlinear contracts—specifically, penalty contracts. Research on penalty contracts suggests that penalty contracts provide a stronger behavior-shaping incentive than bonus contracts. Hannan et al. (2005) explore workers' responses

to bonus and penalty contracts.[9] They found that subjects prefer the bonus contract but tend to exert more effort under a penalty contract. Hannan et al. explain this with the principles of loss aversion (Kahneman and Tversky, 1979). Workers anticipate the pain of a penalty more than the pleasure of a bonus and consequently work harder to avoid the punishment than to earn the bonus. This result is in line with Dugar's (2010) claim that people tend to be more concerned with minimizing bad feedback than with maximizing good feedback. Ragland et al. (2010) expand on Hannan et al. by including a treatment with a hybrid contract[10]. The researchers found that subjects exerted the same effort levels under hybrid contracts as they did penalty contracts and suggested that subjects may anchor on the penalty portion of a hybrid contract.

2.4. Intentionality

Intentionality is the idea that the intent of an actor matters to the person affected. This thesis is interested in the role of intentionality in how workers react to punishment. What makes it punishment is that someone else chose the action that resulted in the monetary loss. If Andrea monetarily punishes Bob, Bob must deal with both his lowered payoff and the fact that someone chose to punish him. Would Bob react the same way to Andrea's punishment as he would simply if the same amount of money was subtracted from his earnings?

Houser and Xiao (2010) utilize a similar distinction between types of punishment to determine whether individuals seek to create advantageous inequalities when they believe that they have been treated unfairly in a version of the ultimatum

[9] Subjects are assigned to either a bonus contract treatment or a penalty contract treatment. The experimenters record how much effort each subject intends to put forth in their assigned treatment and whether or not he or she would prefer the other treatment.
[10] Subjects could be rewarded or punished.

game. In one treatment, the amount of money offered to the second mover is stochastically decided by a machine; if the second mover accepts, he or she receives that amount of money while the first mover receives the remainder. The other treatment is more like the classic ultimatum game: the first mover decides upon an amount to offer the second mover, and if the second mover accepts he or she receives that amount while the first mover receives the remainder.

In both treatments, the second mover has the ability to pay a fixed cost in order to punish at whatever level they choose. Houser and Xiao find that individuals are significantly less likely to seek advantageous inequalities against individuals who had no choice in what portion of the original pool of money to offer the second player than against individuals who had discretion in the decision to strike back against unfair behavior. That is, individuals who received less in the randomly-determined offering treatment would still punish, but were significantly less likely to punish so much that their final payment would be higher than the first mover's. This differed from the treatment where the first mover had discretion in determining the offer to the second mover. In this case, second movers were more likely to punish to such an extent that their final payoff was actually greater than the first mover's.

Houser and Xiao's results fit in nicely with a history of research that indicates that intentions matter. Blount (1995) finds that subjects negatively reciprocate unkind, intentional actions more often than unkind but unintentional actions. Charness (1998) finds that subjects are more likely to give higher effort in response to high wages that were chosen by their employer rather than stochastically chosen. Not only is there is evidence that intentionality increases both negative and positive reciprocity, but intentions seem to

matter even more when the original action was perceived to be intentionally unkind. Offerman (2002) finds that not only are subjects more likely to reciprocate an intentional action over an unintentional action, the increase in likelihood is three times greater when the original action was hurtful. These results indicate that intentions matter—that we attribute the consequences of an act to the actor and his motivations when we know the actor took an active part in causing the consequences.

3. Research Method

I use an experiment to investigate whether workers change their effort levels differently in response to different ways of implementing punishment. This question is well suited for a lab experiment because of the control over the decision environment allowed by lab experiments. In the field there are numerous confounding variables that would make it impossible to isolate the effect of different types of punishment. Workers may be motivated to increase or decrease effort levels in response to other stimuli besides being punished, and responses to specific punishments may differ depending on industry, culture, etc. The control of the lab environment allows me to isolate the workers' reactions to the different types of punishment.

3.1. Experimental Design

I model a world in which there are two types of people: workers (agents) and bosses (principals). Each worker is paired with a boss. The worker exerts costly effort to engage in production in each of two periods (e_1 and e_2 respectively). Both worker pay (π_{w}) and boss pay (π_{b}) are functions of output (y). Because output is increasing in effort

(e), for the sake of simplicity I assume $y = e$.[11] The word effort is used from here on to mean both effort and output. If a worker does not meet a punishment threshold in the first period, he is monetarily punished— $p(e_1)$ is subtracted from his earnings. The effort in period 1 (e_1) can be considered a baseline effort for the worker ($\hat{e}$)—the amount of effort he gives when he has not just been punished.

There are two treatments. The difference between the two treatments is how the punishment threshold is determined. In the discretionary treatment the punishment threshold is chosen by the boss. In the deterministic treatment the threshold is set ahead of time—the bosses make no decisions.

The payment function for the worker is given in equation 1.

$$\pi_{\mathrm{w}}(e_1, e_2) = ae_1 + ae_2 - p(e_1) \quad (1)$$

where π_{w} is total payment, e_1 is effort in the first period, e_2 is effort in the second period, a is a parameter designating the piece-rate pay per task completed successfully, and $p(e_1)$ is the punishment as a function of effort as given in equation 2

$$p(e_1) = \begin{cases} 5 : e_1 < t, 0 \text{ otherwise (discretionary)} \\ 5 : e_1 < 180, 0 \text{ otherwise (deterministic)} \end{cases} \quad (2)$$

with t is the punishment threshold in the discretionary treatment chosen by the boss.

The payment function for the boss is given in equation 3.

$$\pi_{\mathrm{b}}(e_1, e_2) = f_1 + f_2 + ae_1 + ae_2 + s(e_1) + s(e_2) \quad (3)$$

where π_{b} is total payment, $f_1 = f_2$ = the base pay for each period, and $s(e)$ is the bonus as a function of effort, given in equation 4.

[11] Output is most likely concave with respect to effort. However, the important conclusions of this production shape are also predicted by the convex shape of the cost function I assume here. Thus assuming $y = e$ simplifies notation without altering the implications.

$s(e) =$ 5 if $e \geq 180$, 0 otherwise (4)

Suppose individuals have utility functions inspired by the Falk and Fischbacher (2006) model of reciprocity. The worker aims to maximize his utility over both work periods. Then the worker's utility is the sum of his individual material payoff minus the utility cost of effort plus what Falk and Fischbacher call reciprocity utility. Reciprocity utility is composed of a reciprocity parameter (ρ_{w}), an equality term (Δ_{wb}), an intentionality term (θ_{w}) and a reciprocation term (σ_{w}). Thus workers seek to maximize utility in period 1 as shown in equation 5 and in period 2 as shown in equation 6.

$$\max_{e_1} \; U_{\mathrm{w}}^{1}(e_1|E(t)) = ae_1 - \; c_{\mathrm{w}}(e_1) - p(e_1|E(t)) \tag{5}$$

$$\max_{e_2} \; U_{\mathrm{w}}^{2}(e_2, \sigma_{\mathrm{w}}) = ae_2 - \; c_{\mathrm{w}}(e_2) + \rho_{\mathrm{w}}\Delta_{\mathrm{wb}}\theta_{\mathrm{w}}\sigma_{\mathrm{w}} \tag{6}$$

The cost of choosing an effort level e is expressed by the function $c_{\mathrm{w}}(e)$. In the absence of punishment, and with no reciprocity, a simple utility maximization would yield a solution in which $c'_{\mathrm{w}}(e) = a$. It is reasonable to assume a convex cost of effort function (see Figure 1). Thus the solution to the utility maximization problem for subjects that do not care about punishment is where the marginal cost of effort is equal to the marginal benefit a.

$p(e_1|E(t))$ is the expected value of punishment. A worker's expected utility of punishment depends on how likely a worker believes it is that he is going to be punished—that is, it depends on his expected punishment threshold $E(t)$.

The reciprocity parameter (ρ_{w}) is a positive constant that measures how important the reciprocity utility is relative to the utility from the material payoff (the higher the constant, the more important the reciprocity utility). This is a parameter of a worker's

individual preferences. Note that if $\rho_{\mathrm{w}} = 0$, the worker has no other-regarding preferences and will never engage in costly reciprocity.

The equality term (Δ_{wb}) is a "state of the world" variable measuring the difference between the material payoffs of two players ($\Delta_{\mathrm{wb}} = \pi_{\mathrm{w}} - \pi_{\mathrm{b}}$) and is a measure of inequality. If $\pi_{\mathrm{w}} > \pi_{\mathrm{b}}$, Δ_{wb} is positive (the worker experiences an advantageous inequality). If $\pi_{\mathrm{w}} < \pi_{\mathrm{b}}$, Δ_{wb} is negative (the worker experiences a disadvantageous inequality). The experiment is designed so that a boss always makes more money her worker and therefore Δ_{wb} is always negative.

The intentionality term (θ_{w}) expresses how kind the worker perceives the last action of his boss toward him to be relative to the boss's action space. Let $\bar{t}_b$the kindest action the boss can take, and t_b the action the boss does take. Then θ_{w} is described by equation 7 [12]

$$\theta_{\mathrm{w}} = \begin{cases} 1 & : \ t_b \neq \bar{t}_b \\ \varepsilon_{\mathrm{w}} & : \ t_b = \bar{t}_b \end{cases} \tag{7}$$

where ε_{w} is an individual parameter measuring the worker's concern for an equitable outcome ($0 \leq \varepsilon_{\mathrm{w}} \leq 1$). A worker with $\varepsilon_{\mathrm{w}} = 1$ cares only about the consequences of his boss's action, while a worker with $\varepsilon_{\mathrm{w}} = 0$ is solely concerned with his boss's intentions.

The reciprocation term (σ_{i}) is the action the worker takes towards his boss. The term is positive when he enacts a rewarding action and negative when he enacts a punishing action. In terms of this experiment, σ_w is equal to $e_2 - e_1$ because the effort that the worker gives in the second period is the only action he can take in response to his boss and e_1 is assumed to be equal to some baseline effort level $\hat{e}$. If he reacts kindly

[12] θ_{w} is defined here for disadvantageous inequality from the perspective of the worker ($\pi_w < \pi_b$) because the worker always earns less than the boss by design. In the Falk and Fischbacher (2006) model, if it were the case that $\pi_w > \pi_b$, $\theta_{\mathrm{w}} = 1$ for $t_b = \bar{t}_b$ and $\theta_{\mathrm{w}} = \varepsilon_{\mathrm{w}}$ for $t_b \neq \bar{t}_b$.

toward his boss, he will work harder in the second period and $e_2 - e_1$ will be positive. If he responds unkindly toward his boss, he will work less hard in the second period and $e_2 - e_1$ will be negative. Therefore the worker's utility function in the section period can be rewritten as in equation 8.

$$U_{\mathrm{w}}^2(e_1, e_2) = ae_2 - c_{\mathrm{w}}(e_2) + \rho_{\mathrm{w}}\Delta_{\mathrm{wb}}\theta_{\mathrm{w}}(e_2 - e_1) \quad (8)$$

Now consider this model in terms of the investigation of discretionary versus deterministic punishment. If subjects in the two treatments are randomly drawn from the same population, we can assume that, on average, c_w, ρ_w, and Δ_{wb} are the same for workers between treatments, with Δ_{wb} being negative (because $\pi_w < \pi_b$). What is different between the treatments is θ_{w}, which may drive differences in e_1 and e_2.

In the discretionary punishment treatment, the boss has the option of setting the punishment threshold at numerous levels of t, including 0 ("never punish"). This means that for any level of punishment above "never punish," there is always an action in the boss's action space that would have been kinder to the worker. Because $t_b \neq \bar{t}_b$, $\theta_{\mathrm{w}} = 1$ in the discretionary setting.

In the deterministic treatment, however, the boss cannot do any better by the worker because the boss has no say in the punishment decision. That means that θ_{w} is ε_w in the deterministic setting. With each individual ε_w being $0 \leq \varepsilon_w \leq 1$, it is safe to assume that, on average, $0 < \varepsilon_w < 1$ and therefore $\theta_{\mathrm{w(disc)}} > \theta_{\mathrm{w(det)}}$.

Since Δ_{wb} is negative, workers should act unkindly toward their boss in order for the reciprocation utility term to be positive. An unkind action will cause $e_2 < e_1$, which means that σ_w will be negative and the whole term will be positive. This should be true in both treatments. Because $\theta_{\mathrm{w(disc)}} > \theta_{\mathrm{w(det)}}$, workers in the discretionary treatment will

suffer a greater loss of utility if they act kindly toward their boss, but will gain more utility if they respond unkindly as demonstrated in equation 9.

$$\frac{\partial U_{\mathrm{w}}^{2}}{\partial e_{2}} = a - c'(e_{2}) + \begin{cases} \rho_{\mathrm{w}}\Delta_{\mathrm{wb}} \text{ (discretionary)} \\ \rho_{\mathrm{w}}\Delta_{\mathrm{wb}}\varepsilon_{\mathrm{w}} \text{ (deterministic)} \end{cases} \quad (9)$$

Because $\rho_{\mathrm{w}}\Delta_{\mathrm{wb}} > \rho_{\mathrm{w}}\Delta_{\mathrm{wb}}\varepsilon_{\mathrm{w}}$, $\frac{\partial U_{\mathrm{w}}^{2}}{\partial e_{2}}$ (disc) $>$ $\frac{\partial U_{\mathrm{w}}^{2}}{\partial e_{2}}$ (det). Therefore, if agents maximize a utility function of this form, there should be a greater slowdown in the discretionary treatment. In words, workers in the discretionary treatment will experience a greater change in utility by counter-punishing than workers in the deterministic treatment, and in both cases that change would be greater than 0. Based on this theory, we would expect punished workers to take revenge in both treatments, and to a greater degree in the discretionary treatment.[13]

3.2. Hypotheses

I experimentally investigate the null hypothesis that workers will not change their effort levels in response to punishment ($e_2 = e_1$), with the alternative hypothesis that they will ($e_2 \neq e_1$). More specifically, I hypothesize that workers will reduce their effort level in response to punishment (($e_2 < e_1$); $\sigma_w$would be negative) and that the difference will be greater in the discretionary treatment ($|\Delta e_{\mathrm{disc}}| > |\Delta e_{\mathrm{det}}|$). This hypothesis follows naturally from Falk and Fischbacher (2006) if $\rho_{\mathrm{w}} > 0$ (an individual's utility from reciprocating is greater than 0) and $\varepsilon_w > 0$ (individuals care about intentionality). In other words, I hypothesize that workers will negatively reciprocate in response to punishment, and that that response will be stronger in a situation where punishment is chosen intentionally.

[13] If a worker were to not care about reciprocity ($\rho = 0$, perhaps), he would not retaliate and we would see $e_1 = e_2 = \hat{e}$. In this case $\sigma_w = e_2 - e_1 = 0$ and the whole reciprocity term of equation 8 drops out.

These hypotheses are made with the following two assumptions. First, $E(t)_{\text{disc}} = E(t)_{\text{det}} = 180$. That is, the expected punishment threshold is the same in both treatments. Second, a subject devoid of emotions ($\rho_{\text{w}} > 0$) would give the same effort in both periods ($e_2 = e_1$). This is an implication of my model; alternatives are explored in the conclusion.

3.3. Implementation

To test my hypothesis I designed a double-blind principal-agent game with an incomplete contract that utilizes a real-effort task and the strategy method. All participants sign the consent form given in Appendix 3. Subjects are randomly assigned the role of worker or boss by picking a card from a shuffled deck. The subjects are paired so that there is one boss associated with one worker. Each subject has their own individual, randomly assigned ID number so that the experimenters cannot associate decisions with any specific individual. Subjects are also unaware of who their partner is.[14] Each subject type has its own task: workers fill out bubble sheets (e.g., Scantron®) forms according to a specific pattern and bosses complete a survey. In the model, e_1 and e_2 are the number of bubbles completed by the worker in each period respectively. Both worker pay and boss pay are functions of the number of bubbles that the worker successfully completes.

Unlike many experiments where the effort level is chosen from a list with the price of each effort level given, I chose to use a real effort task. I believe there is a real difference between the decision of choosing a specific effort level without actually having to put forth the effort and the decision necessary regarding effort level when

[14] Cox and Deck (2005) found that they had significantly different results depending on whether they used a single blind or double blind procedure (more specifically, they found that reciprocal behavior is sensitive to whether the experimenters do or do not know the decisions that a particular individual makes). Single blind means that subjects do not know with whom they are interacting, whereas double blind means that the experimenters are also unaware of which subjects interact with each other.

actually performing a task. I believe this because in most situations, people do not take a step back and think "On a scale of 1 to 10, I am going to give this task an effort level of 6." In their day to day lives, people give effort levels based on their energy levels, their mood, what has happened to them recently, and so on. Because my experiment focused on a specific emotional response, using a real effort task was more appropriate because it takes those factors into account.

A worker's productivity in any real effort task is a function of his ability and effort. It was important to pick a task where effort was by far the more important factor in a subject's success with the task. Filling in the bubble sheets according to a specific pattern had several nice properties. First, because the subjects in the experiment were students at an elite college, it was reasonable to assume that the majority of them had previous experience filling in bubble sheets as a result of standardized testing. Therefore most subjects' final output ought to be a function mainly of their effort level, mitigating learning effects. Second, it is relatively easy for the experimenter to measure task completion and hard for subjects to cheat. Finally, because each bubble completion takes very little time, subjects ought to be able to complete the task numerous times in each work period so the measurement of output is nearly continuous. For these reasons, I assume a fairly direct mapping of effort to output.

The sequence of events is given in Table 1 with detailed explanations below. For subject instructions, see Appendices 4.1 and 4.2.

1. *Practice*: After subjects are aware of their role and are assigned experimental IDs, there is a brief practice period so that workers may practice filling in the bubble sheets according to the specified pattern (see Appendix 4.3 for an example of the pattern).

Workers are instructed to fill in 15-20 bubbles and are told that they will receive feedback on their work so that they know how strictly their work will be graded. The goal is to reduce learning effects so that e_1 can be an accurate proxy for $\hat{e}$. In the discretionary period, bosses use this time to set a punishment threshold, t (see Appendix 5 for the punishment threshold decision form). Bosses are allowed to choose anything from 0 ("never punish") and then continuing in increments of 20 until 300. Bosses can also choose "always punish." A worker is not informed of the threshold set by his boss until after the first work period. In the deterministic treatment, bosses do not make a decision about the punishment threshold. The punishment threshold is set at 180 bubbles for all subjects—workers and bosses are informed of this before the practice period. This procedure differs from those mentioned in Section 2.4 which utilize a stochastic treatment as the control group rather than a pre-determined treatment. During the practice period, bosses in the deterministic treatment are asked to wait patiently.

Bosses earn a bonus if their worker reaches 180 bubbles ($s(e)$). The bonus was meant encourage bosses in the discretionary treatment to set the punishment threshold at 180 so that the thresholds faced by workers in the discretionary treatment would be the same as that faced by workers in the deterministic treatment.

The method by which the boss chooses the punishment threshold is an implementation of the strategy method. The strategy method is a means of conducting an experiment where subjects must state how they would act in any given situation before they are placed in a situation. In the context of my experiment, bosses in the discretionary setting decide before the first work period what number of bubbles their worker must complete in order to not be punished. Workers are then punished after the first work

period according to the decision made by their boss. The alternative would have been to use the *game method*, in which the boss sets t after learning e_1. There are several advantages and disadvantages to the strategy method which Casari and Cason (2009) outline nicely. The main advantage is being able to obtain data from bosses who choose a punishment threshold but whose worker meets that threshold and therefore gain information about a continuous measure of boss kindness that the worker could also be responding to. The biggest disadvantage is the trade-off between a considered choice and the heat-of-the-moment decision. In other words, the strategy method is a "cold" decision, and there is some evidence that people punish less when making cold decisions (Casari and Cason, 2009). However, my research question was concerned with how a worker responds to his boss's decision ($e_2 - e_1$), not the boss's decision itself. Use of the strategy method should not have biased the measurement in which I was interested. Consequently the benefit of having a continuous measure of boss decisions outweighed the cost of the cold decision.

2. *Period 1*: After the practice period, workers receive feedback from the experimenters on their performance in the practice round so that they are aware of exactly what is considered a successfully completed bubble. Once any questions have been answered, the first work period begins. Period 1 lasts for five minutes. In both treatments workers try to complete as many bubbles successfully as possible while bosses work on a survey (see Appendix 6 for the survey completed by the bosses). The survey was preliminary data gathering for a different experiment and completely unrelated to the experiment taking place. The survey had no effect on boss pay.

3. *Information*: After period 1, the experimenters grade the workers' period 1 efforts. Workers and bosses are given earnings sheets that inform them of a few pieces of information. In the discretionary treatment, workers learn how many of their period 1 bubbles were graded as "completed successfully." They also learn what punishment threshold was set by their boss and whether or not they were punished (i.e., whether or not they reached the threshold; see Appendix 7.1 for the workers' earning sheet in the discretionary treatment). In the deterministic treatment, workers also learn how many of their period 1 bubbles were graded as "completed successfully." Unlike in the discretionary treatment, workers in the deterministic treatment are already aware of the punishment threshold (set at 180). Therefore workers in the deterministic treatment are only informed whether or not they were punished (see Appendix 7.2 for the workers' earning sheet in the deterministic treatment). Workers receive \$0.08 per bubble completed successfully and lose \$5.00 if they were punished. Therefore worker payment in period 1 is given in equation 10.

$$\pi_{w1} = 0.08e_1 - p(e_1) \tag{10}$$

with p being a function of e_1, the number of bubbles that worker successfully completes (equation 11).

$$\begin{aligned} &\text{discretionary treatment: } p(e_1) = 5 \text{ if } e_1 < t, 0 \text{ otherwise} \\ &\text{deterministic treatment: } p(e_1) = 5 \text{ if } e_1 < 180, 0 \text{ otherwise} \end{aligned} \tag{11}$$

with t = the punishment threshold in the discretionary treatment chosen by the boss.

In both treatments, bosses are informed if they received a bonus (see Appendix 7.3 for the boss' earning sheet). Bosses receive an additional \$5.00 if their worker successfully completes 180 or more bubbles. This incentive should have increased the

focus on 180 as a punishment threshold. Like workers, bosses earn $0.08 per bubble. Bosses also receive a $3.00 base pay each period. Therefore boss payment in period 1 is given in equation 12.

$$\pi_{b1} = 3 + 0.08e_1 + s(e_1) \quad (12)$$

where $s(e_1)$ is the bonus as a function of e_1, the number of bubbles that worker successfully completes (equation 13). [15]

$$s(e) = 5 \text{ if } e \geq 180, 0 \text{ otherwise} \quad (13)$$

And the end of period 1, all subjects are reminded that workers cannot be punished after period 2, while bosses are still eligible to earn a bonus if their worker complete 180 or more bubbles.

4: *Period 2*: Like period 1, period 2 lasts for five minutes. In both treatments workers complete as many bubbles successfully as possible while bosses continue to work on the survey.

5: *Information*: After period 2, the experimenters grade the workers' period 1 efforts and return the earnings sheets to all workers and bosses. Workers learn how many bubbles they completed successfully, their period 2 earnings, and their total payment. Worker payment in period 2 is shown in equation 14.

$$\pi_{w2} = 0.08e_2 \quad (14)$$

and worker total payment is given in equation 15.

$$\pi_w = \pi_{w1} + \pi_{w2} = 0.08(e_1 + e_2) - p(e_1) \quad (15)$$

[15] The values for the monetary parameters were chosen so that subjects would earn $20.00 on average per hour and a half session. This is based on the expectation that workers complete an average of 130 bubbles per work period.

Bosses learn how many bubbles their workers completed successfully, whether or not they received a bonus, their period 2 earnings, and their total payment. Boss period 2 payment is shown in equation 16

$$\pi_{b2} = 3 + 0.08e_2 + s(e_2) \quad (16)$$

where $s(e_2)$ defined as above for $s(e_1)$, and boss total payment is given in equation 17.[16]

$$\pi_b = \pi_{b1} + \pi_{b2} = 6 + 0.08(e_1 + e_2) + s(e_1) + s(e_2) \quad (17)$$

6: *Questionnaire*: While experimenters assemble payment for the subjects, all subjects complete a questionnaire (see Appendix 8 for questionnaire).

7: *Pay & leave*: Once experimenters have assembled and double-checked payments and all subjects have completed the questionnaire, payments are distributed to subjects and subjects exit the lab. The payment method preserves anonymity of decisions by allowing subjects to privately choose an envelope containing their payment that is labeled with their experimental ID number.

There are three main differences between the treatments, in bold in Table 1. The first is that the punishment threshold is decided by the bosses in the discretionary treatment, whereas in the deterministic treatment the punishment threshold is set at 180. The second is that workers in the discretionary treatment are unaware of what the punishment threshold is until after the first work period, whereas workers in the deterministic treatment are aware that the punishment threshold is the same as the bonus threshold and is therefore 180. This decision was made because we anticipated that workers in the discretionary treatment would expect a punishment threshold of 180 because of the bonus, and therefore would face the same work environment as workers in the deterministic treatment. The third is that, due to variation in the choices made by the

[16] Remember that boss pay does not depend on the quality or completion of their task (the survey).

bosses in the discretionary treatment, the punishment thresholds faced by workers in the discretionary treatment were not always the same as that faced by workers in the deterministic treatment.

For logistical ease and to ensure that all instructions were common knowledge and that there was no deception of subjects, I decided to run the experiment in one room with both types of subjects in the room. This ensured that all participants knew the same information about the sequence of events, how each type of subject would be paid, and so on.

To avoid biasing subjects during the experiment, I purposefully used neutral framing, which meant avoided using words like "fairness" and "equity" as well as words directly associated with the context of a workplace. Some loaded words were used (e.g., punishment, bonus, boss, and worker), but only those words that made easiest to explain the experiment to subjects. It is possible that these words carry enough weight to give the experiment enough context that some control of the decision environment was lost. However, the context was held constant across treatments.

I hypothesized that workers would reduce their effort level in response to punishment. In terms of this experiment, worker effort (e_1 and e_2) is translated into output (y_1 and y_2). Though it is an imperfect measure of effort, the output of the task is more or less continuous and worker output should be more dependent on effort and less on ability. If my hypotheses are correct, the number of bubbles completed in period 1 should be greater than the number of bubbles completed in period 2 ($e_1 > e_2$) and that difference should be greater in the discretionary treatment than in the deterministic treatment ($\Delta e_{\text{disc}} < \Delta e_{\text{det}}$) and ($|\Delta e_{\text{disc}}| > |\Delta e_{\text{det}}|$).

4. Results

As predicted, workers in the discretionary treatment changed their effort more between the two work periods than workers in the deterministic treatment ($|\Delta e_{\text{disc}}| > |\Delta e_{\text{det}}|$), suggesting that either intentionality matters ($\varepsilon > 0$) or some other factor that differs between the treatments matters. Unlike my hypothesis, the change in effort was not a slowdown but a speed-up. Subjects in both treatments completed significantly more bubbles in the second period than the first ($e_2 > e_1$) and this speed-up was greater in the discretionary treatment.

There are several explanations for the speed-up. If the Falk and Fischbacher (2006) model of reciprocity is an appropriate method for modeling this behavior, the only explanation for the speed-up is that subjects were expressing some kind of perverse reciprocity by responding to unkindness with kindness. If the Falk and Fischbacher model is not a good model for this behavior, there are three distinct possibilities. First, subjects might have responded to seeing the thresholds set by their bosses, believing the threshold to be an indication of others' expectations. Second, if subjects have decreasing marginal utility from wealth, being punished should increase the marginal utility of wealth, prompting subjects to work harder. Finally, subjects may have been reacting to their own expectations about how much money they would earn during the experimental. I expand on these possibilities in Section 5.

In the remainder of the results section I elaborate on the results stated above by examining whether or not punishment thresholds were the same between the two treatments. I continue by exploring the number of bubbles completed in period 1 and

period 2 separately, whether there was an overall speed-up or slowdown, and whether the magnitude speed-up/slowdown differed between the two treatments.

Subject 2009 has been excluded from all of the following analyses and tables. Subject 2009 was an outlier in the discretionary treatment because he did not complete any bubbles in the second work period. On his questionnaire, subject 2009 indicated that he was "suffering from severe body aches due to fever." Removing an outlier is a delicate decision and could introduce bias due to selective data trimming. If he had chosen to leave the lab as a result of his fever rather than sit through the experiment, I would not have included his data. Because that is essentially what he chose to do, less bias will result from his exclusion than from his inclusion.

Most of the analysis that follows focuses on comparing workers in the deterministic treatment with workers in the discretionary treatment that were punished. While all workers in the deterministic treatment were punished, only 10 workers in the discretionary treatment were punished. However, since comparing punished workers between treatments holds the change in income constant, it is the comparison in which I am most interested and the comparison that receives the focus of this analysis. As a result of low sample sizes and high variance, none of the following tests have very much power—even if an effect exists it may not show up as a result of this study.

4.1. Demographics and summary statistics

The experiment was run at Williams College. 140 students were recruited to participate via ORSEE (Greiner, 2004). Twenty subjects participated in a pilot session and twenty more in each of six sessions. Each session ran approximately one and a half hours.

The pilot session was not included in the analysis.[17] There was no demographic difference across treatments and roles. A breakdown of subject demographics is given in Table 2.

Summary statistics (mean and standard deviation) are given in Table 3 for the five major variables for subjects that participated as "workers" in the deterministic treatment, the discretionary treatment, the discretionary treatment (punished subjects only), and the discretionary treatment (non-punished subjects only).

4.2. Punishment Decisions

One difference between the deterministic treatment and the discretionary treatment was that workers in the deterministic treatment knew the punishment threshold going into the first work period, whereas workers in the discretionary treatment learned their punishment threshold after the first work period. Workers' expectations about t could have influenced behavior in the first period. Workers in the discretionary treatment may have expected that bosses would set the punishment threshold at 180 because of the bonus. If that were the case, the difference in behavior resulting from discrepancy in decision environments would present itself in the second period, not the first. The alternative is that workers maintained "rational expectations" of t, where what they anticipated was what actually happened.

Workers may have behaved differently in period 2 depending on what punishment threshold they faced. t could be viewed as a continuous measure of boss kindness (as compared to the discrete measure of "punished" or "not punished") that workers can reciprocate against through their actions in period 2. Alternatively t could be seen as

[17] The pilot session data was not included in the analysis of the results due to several design changes that were implemented after the pilot session. Average earnings in the pilot session were low; the parameters were changed to account this in subsequent sessions. The bonus threshold was lowered to make it more attainable. Various minor changes were made to make the procedure more efficient.

informative of normative effort and others' expectations. In either case, being informed of t (and of whether you were punished) could affect e_2.

Ideally the punishment threshold would have been the same across treatments. Only 16 out of the 30 subjects that participated as bosses in the discretionary treatment chose to punish at any level. Of those 16, six chose a punishment threshold of 180—equal to the punishment threshold in the deterministic treatment. Figure 2 is a histogram displaying the threshold decisions of bosses in the discretionary treatment. Though 180 was the most popular option other than "0", this means that there was more variation between the treatments in the decision environments faced by workers besides whether or not the punishment had an element of intentionality. The median thresholds faced by workers in the two treatments were significantly different (Mann-Whitney p = 0.000). See Appendix 9 for more details on the punishment thresholds in the discretionary treatment.

4.3. Baseline effort in period 1

As mentioned previously, the measure of effort in this experiment is the number of bubbles successfully completed by a worker in a given work period. Period 1 output (e_1) was intended to act as a measure of a worker's baseline effort ($\hat{e}$).

I ran the Mann-Whitney test to determine whether there was a significant difference in the effort level in the first period between the two treatments. That is, I tested to see if there was a significant difference between the number of bubbles completed in the first period by subjects in the discretionary treatment and by subjects in the deterministic treatment. The Mann-Whitney test looks for significant differences in medians. It is considered to be a more robust test than the traditional Student's *t*-test

because it is less affected by outliers and non-normal populations. It tests the null hypothesis that the distributions of the two populations are equal (that is, the probability of selecting a larger observation from either sample is the same).[18]

In the case of period 1 effort, it makes sense to include all of the subjects in the discretionary treatment because they have not yet been punished. Figure 3 is a CDF showing the number of bubbles completed in period 1 for each treatment. The CDF of e_1 in the discretionary treatment is to the right of the CDF of e_1 for the deterministic treatment for all levels of effort, suggesting that effort is higher in the discretionary treatment. A simple univariate test supports this idea. On average, workers completed more bubbles in the first period of the discretionary treatment (mean 98.7) than in the first period of the deterministic treatment (mean 83.3). This difference is weakly significant (Mann-Whitney $p = 0.085$), with more bubbles completed on average in the discretionary treatment than in the deterministic treatment.

Looking at Table 4, where I estimate OLS regression models of period 1 output, there is additional evidence that being in a discretionary treatment did yield a significant difference in the number of bubbles completed in the first period. Specification (1) simply controls for treatment and finds that subjects in the discretionary treatment completed, on average, 15.42 more bubbles that workers in the deterministic treatment. Specification (2) controls for each session and finds that there was no significant different associated with any given session.[19] Specification (3) controls for treatment again as well as several demographic variables and finds that treatment is no longer significant. The significance of treatment is absorbed by class year and GPA. I believe this is due to a loss

[18] Bartlett, Levene, and Brown-Forsythe tests of variance yield no significant or interesting results.
[19] Hereafter I assume no session effects and consequently do not include these dummies.

of power due to a small sample size (note that class year and GPA are not significantly different between discretionary and deterministic treatments). Underclass students and students with higher GPAs tended toward higher e_1. The influence of GPA suggests that students who perform better in class may do so in part because they are more efficient workers while class year suggest that students that are closer to their high school years (and thus when they were taking standardized tests and using bubble sheets on a more regular basis) may have an advantage.

One plausible explanation for workers in the discretionary treatment completing more bubbles in period 1 than workers in the deterministic treatment is a slight difference in knowledge between the two treatments. In the deterministic treatment, workers knew that the punishment threshold was set at 180 bubbles, whereas in the discretionary treatment the punishment threshold was different for each worker, and each worker did not know what punishment threshold they faced until after the end of period 1. This difference in knowledge affected the decision environment faced by workers in the different treatments. In the deterministic treatment, workers may have realized whether they were going to be able to meet the punishment threshold or not at some point during the work period. If they believed that they would not be able to meet the threshold, the cost of continuing to exert a high effort level may have outweighed the constant benefit of an additional $0.08 per bubble, in which case they may have chosen a lower e_1. In the discretionary treatment, on the other hand, workers did not know whether each additional bubble they completed would be the bubble that kept them from being punished. If workers expected t=180, the expectations of the marginal benefit from each bubble would be the same in the discretionary and deterministic treatments. If the workers held rational

expectations (i.e., expected what actually happened) or made some other predictions, the expectation of the marginal benefit from each bubble would be greater in the discretionary treatment for all bubbles up to and including the 179th bubble. In other words, the expected marginal benefit of each additional bubble completed in the discretionary treatment could have been higher than in the deterministic treatment, because each additional bubble could be the bubble that saved the subject from monetary punishment. A discussion of possible worker expectations of the punishment threshold in the discretionary treatment is given in Section 5.

4.4. Post-punishment effort

Figure 4 is a CDF showing the number of bubbles completed during period 2 by each treatment. The CDF of e_2 for the discretionary treatment is to the right of the CDF of e_1 for the deterministic treatment for almost all effort levels, suggesting that effort was higher in the discretionary treatment during period 2 as well. Though this result is not supported by a simple univariate test, the OLS regression models described below suggest that period 2 effort was significantly greater in the discretionary treatment.

The difference between the number of bubbles completed in the second period of the discretionary treatment (mean 127.40) and the second period of the deterministic treatment (mean 96.20) was not significant using a simple univariate comparison, (Mann-Whitney $p = 0.146$). However, since only punished subjects are being considered, *N* is quite small for the discretionary treatment (10 subjects), it is not a surprise that this test was unable to detect the difference that appears in the CDF.

I estimate a series of OLS regression models of period 2 output in Table 5. [20] Specification (1) suggests that being in the discretionary treatment alone did not cause a significant difference in the number of bubbles completed in period 2. Specifications (2), (5), and (6) show that regardless of what else is controlled for, just being in the discretionary treatment is not predictive of a significant difference in period 2. Specification (3), however, suggests that being in the discretionary treatment and being punished did result in a significant difference in the number of bubbles completed in period 2, with punished discretionary subjects completing 31.20 more bubbles on average in period 2 than deterministic subjects (all of whom were punished), whereas being in the discretionary treatment and not being punished did not yield a significant difference. The significance of being punished in the discretionary treatment holds even when e_1 is controlled for (specifications (4) and (8)) and when a variety of demographics are controlled for (specifications (7) and (8)). e_1 is a positive significant predictor in every specification in which it is controlled for, as are class year and GPA. Athletics is also a significant predictor in each specification in which it is controlled for. On average, the less a subject cared about athletics, the fewer bubbles he completed in period 2.

Table 6 presents a second set of OLS regression models of period 2 output. These estimates only consider subjects in the discretionary treatment and examine the effect of the threshold set by the boss. Specifications (1) and (4) suggest that the threshold set by the boss was a positive predictor of period 2 effort (the higher the threshold set, the more bubbles a worker would complete). Specifications (2) and (5) suggest that whether or not a subject was punished is also a positive predictor of period 2 effort (being punished

[20] A regression of e_2 on e_1 may seem redundant to a regression of $e_2 - e_1$ (Section 4.5). This would be true if the coefficient on e_1 is 1 and if there are no relationships between e_1 and the other variables. Both of these are false, and therefore e_1 is included in some e_2 regressions.

resulted in greater period 2 effort). The significance of both the threshold and being punished disappears when these are controlled for simultaneously in specifications (3) and (6), which indicates that we cannot separate an effect of the threshold from an effect of having been punished. Specifications (4), (5), and (6) control for a variety of demographics and find again that GPA and athletics are significant predictors. Each GPA point is associated with approximately 18 more bubbles while the importance of athletics is again a negative predictor.

4.5. Slowdown/Speed-up

Instead of the anticipated slowdown in worker effort, there was a speed-up. The difference between $e_1$and e_2 was significantly positive across both treatments (*t*-test, p = 0.000).[21] Figure 5 is a histogram of $e_2 - e_1$ including subjects in both treatments. It is clear looking at Figure 5 that the mean value of $e_2 - e_1$ is positive. There are a few explanations for why there would have been a speed-up. It is possible that subjects improved at the task. Perhaps subjects did not actually learn what quality of bubble would be accepted in the practice period, even though that was the point of the practice period. More substantial possibilities are discussed in Section 6.

The existence of a speed-up does not mean that we are not seeing negative reciprocation of punishment. It is possible that the intentionality of the punishment in the discretionary setting had an effect in the size of the speed-up: that is, the effect of negative reciprocity as a result of intentionality in the discretionary treatment could be exhibited as a smaller speed-up. Figure 6 is a CDF displaying $e_2 - e_1$ by treatment, and it is clear that punished workers in the discretionary treatment increased their effort even more than

[21] A *t*-test is used here because the subjects were pooled and therefore $N > 30$. The distribution was also approximately normal, which can be observed in Figure 5.

punished workers in the deterministic treatment. In fact, the speed-up in the discretionary treatment (mean 29.70) was almost significantly more than the speed-up in the deterministic treatment (mean 12.90; Mann-Whitney test $p = 0.104$). Given the high variance and low sample size, I ran a bootstrapped *t*-test comparing the difference between effort in period 1 and period 2 between the two treatments. The results of this were almost significant as well (*t*-test $p = 0.101$).

OLS regression estimates support the idea that punished workers in the discretionary treatment increased their effort more than punished workers in the deterministic treatment. Specifications (1) and (2) in Table 7 show that there is not a difference between subjects in the discretionary treatment and subjects in the deterministic treatment. Specifications (5) and (6) suggest the same result even with various demographics controlled for. However, specification (3) shows that punished workers in the discretionary treatment sped up, on average, 16.80 bubbles more than workers in the deterministic treatment, while workers in the discretionary treatment that were not punished did not have a significantly different speed-up from workers in the deterministic treatment. Specification (4) suggests that this significance holds when baseline effort (e_1) is controlled for, and specifications (7) and (8) suggest that it is a significant difference even when various demographics are controlled for. This is an important result: regardless of what is being controlled for, on average being punished in the discretionary treatment leads to a speed-up that is 15.91 – 17.15 bubbles larger than a speed-up caused by a simple reduction of earnings of the same size. GPA and class year are again positive predictors, while the less important athletics are, the smaller $e_2 - e_1$ is.

In Table 8 I estimate a second set of OLS regression models of $e_2 - e_1$. These estimates focus on subjects in the discretionary treatment. The threshold set by the boss is significant in specification (1) and even when demographics are controlled for in (4), with higher thresholds predicting larger $e_2 - e_1$. Whether or not the subject was punished is significant in specification (2) and with demographics controlled for in specification (5), with punished subjects having significantly larger $e_2 - e_1$. However, threshold and whether or not the subject was punished lose their significance when they are both controlled for in specifications (3) and (6). As discussed with Section 4.4, this indicates that we are unable to disentangle an effect of the threshold from an effect of having been punished. GPA and the importance of athletics are similarly significant as they were previously.

The fact that punished workers in the discretionary treatment had a significantly greater speed-up than punished workers in the deterministic treatment is an important result. It means that intentional punishment causes a greater speed-up as compared to a commensurate reduction in income.

5. Conclusion

This thesis used a lab experiment to investigate whether or not there would be a difference in response to two different punishment schemes: a discretionary punishment, where punishment thresholds were set by each individual boss, and a deterministic punishment, where the bosses made no decision about the punishment threshold. I hypothesized that people care about reciprocity and therefore would reduce effort levels in response to any form of punishment, and that people care about intentionality and

therefore the reduction in effort would be greater in the discretionary treatment than in the deterministic treatment. While the change in effort was significantly greater in the discretionary treatment, the change was a positive change: workers increased their effort level in response to punishment.

There are several candidate explanations for this deviation from my hypothesis. One is the idea that subjects were acting with some kind of perverse reciprocity by acting kindly in response to unkindness. That is, if the model I proposed is appropriate for this experiment, $\rho > 0$ but $\varepsilon < 0$. There are three possibilities if the model is not appropriate. The first is that workers viewed the thresholds set by the bosses as a statement of what was a reasonable expectation of effort, and were responding to their bosses' expectations. The second is that subjects experienced an increase in the marginal utility of wealth after being punished since they had reduced wealth. The third is that subjects had a priori expectations about how much money they would making during the experiment and were adjusting their effort accordingly. Finally, it is possible that subjects became better at the task, as discussed briefly in Section 4.

If the model was appropriate, then what could have caused that sort of perverse reciprocity? If there was less social distance between the worker and the boss in the discretionary treatment, being punished may have led to a feeling of shame or guilt which might have spurred workers in the discretionary treatment to increase their effort even more. This would imply that the reciprocity parameter ρ is partially a function of an individual's own past actions. Workers may have worked harder out of concern for causing disappointment in an individual with whom they felt they had a relationship. Pride could have been a factor. Subjects may have feared some sort of disapproval from

their peers—despite the fact that their decisions were blind to all other subjects and to the experimenters—and were worried about being judged.

If the model was not appropriate for modeling this behavior, it is still possible that guilt and/or shame played a role in how workers responded to being punished. In the discretionary treatment, workers learned what punishment threshold they faced after the first work period. If workers viewed the threshold as a statement of the level of effort that is "normal", they might have increased their effort more in order to meet a social norm.

Three other possible explanations for the speed-up are increased marginal utility of wealth, reference points, and learning. If individuals have decreasing marginal utility functions, then the marginal utility of wealth would have increased after being punished due to the income loss. This would have resulted in subjects increasing their effort in the second period. Alternatively, subjects may have priors as to how much money they wanted to leave with but found they were not on track after being punished. There is evidence that subjects behave differently when they have different expectations about how much money they are going to make during the experiment (Abeler et al., 2011). The subjects use their expectations as a reference point and may change their behavior in order to meet their expectations, as in prospect theory (Kahneman and Tversky, 1979). Therefore if subjects entered the lab with an idea of how much money they were going to make, they may have worked harder in the second period if they were punished as a result of being further from that expectation. Subjects may have improved at the task during period 1, so that e_1 was not a good estimate of $\hat{e}$. However, punished workers in the deterministic treatment would also have experienced an increase in the marginal utility of wealth, would be just as likely to have a priori expectations about

their earnings in the lab, and would also have improved at the task between periods. Because these effects should have been constant across punished subjects, they are not a compelling explanation for the difference between treatments, even if they provide some explanation for the speed-up.

Are these results generalizable, since they were generated in a lab? There are advantages and disadvantages to using a lab experiment. The advantages are substantial. Experimenters have more control over the decision environments that their subjects are faced with and are able to test specific hypotheses (Falk and Gächter, 2000; Falk and Heckman, 2009). Experimenters can also observe behavior at the level of individual agents and results ought to be replicable (Falk and Gächter, 2008). Because most experiments involve payments, participants are believed to take their decisions seriously (Falk and Heckman, 2009).

There are a few common complaints about experimental evidence. There is concern about how realistic experiments are, and therefore how close to reality any results may be (Falk and Heckman, 2009). Though lab experiments may not be able to exactly mimic real situations, they themselves are real situations, and therefore any significant, replicable results are interesting. The results of experiments may not be totally generalizable, but this is also an issue with field research (Falk and Gächter, 2000; Falk and Heckman, 2009). Falk and Heckman address several other major concerns regarding lab experiments. In response to the statement that the money involved in decisions is trivial and therefore people do not take the decisions seriously, the authors ask how often people make decisions involving nontrivial amounts of money. Participants may act different because of being observed, but being observed is not a

feature unique to the lab. Finally, there might be a self-selection bias amongst people that participate in experiments, but again this can also be a problem in field research and is not unique to lab experiments. Despite the arguments against lab experiments, there are relative downsides to field research. Ultimately the choice is not between the lab or the field, but what evidence could be best discovered by which method and what conclusions can be drawn from the results of both methods. Because of this, the research question I ask may benefit both from further experimentation and from field research.

Though the results were not what was hypothesized, obtaining significant results despite the small subject pool and high variance leads me to believe that this is a question worth pursuing further. Though subjects reported no confusion about the experimental procedures, there are several changes to the experimental design that I believe are worth considering. The first would be to utilize a stochastic method for the control treatment. Rather than setting all of the thresholds in the control group at a certain amount, the thresholds in the control group could be randomly assigned. In this case, workers in both treatments would learn what threshold they faced after the first work period. If the discretionary treatments were run first, the distribution of thresholds used in the stochastic treatment could be chosen particularly to mirror what was selected in the discretionary treatment. This change seems to be the most efficient way of making the two treatments more similar, but it would not necessarily result in increased levels of punishment.

Another potential change would be increasing the number of interactions between the boss and the worker. The reason there were only two work periods in my experimental design was due to the time it took to process and return materials. In order

to increase the number of interactions between the boss-worker pairs the task would have to be changed to something that was faster and more objective to grade since there was significant time spent organizing and evaluating subject materials between periods. Increasing the number of interactions between workers and bosses has two main advantages. Subjects would become more identified with their specific role. Workers may begin to resent the fact that their bosses are paid more than they are for doing less work, aggravating the effect of the parameter Δ_{wb} (disadvantageous inequality). Bosses may become more frustrated that they are not earning a bonus and thus more likely to punish. It is clear that there was an emotional effect in the experiment in both directions. Subject 7019 wrote in her questionnaire, "My worker filled in 179! What do YOU think that means? There was no reason not to cheat and fill one more bubble! I am UPSET." However, increased interaction may not have the desired effect of increasing how much people care about reciprocity. Workers may be less willing to take revenge if they know they will have repeated interactions with their boss (Fehr and Gächter, 2000).

Workers may not have been willing to take revenge because the cost of revenge was so high relative to the damage it did to the boss. As the design stands is that the cost of revenge is very high. A possible alternative to this would be to offer the worker two possible tasks in the second period: one that paid them less per bubble, but paid the boss nothing per bubble, and a second task that offered the same rates as before. This design allows workers to reduce their boss' pay at lower cost to themselves and simultaneously makes the option to shirk more visible.

The reason why there was such a small subject pool of workers that were punished in the discretionary treatment was that there was less than anticipated punishment in the

discretionary treatment—especially at 180 bubbles, the punishment threshold for the deterministic treatment. This is important because all workers in the deterministic treatment faced a punishment threshold of 180, and the fact that many workers in the discretionary treatment faced no punishment at all is a significant discrepancy between the treatments.

The fact that workers in the discretionary treatment were not aware of the punishment threshold they faced ahead of time, whereas workers in the deterministic setting knew that the punishment threshold was set at 180 bubbles, brings up an interesting question: how would punishment decisions change and output e_1 if workers saw t before starting period 1? This design choice was made with hopes that workers in both treatments would act as if they were facing a punishment threshold of 180, but the design could also have been structured so that workers in the discretionary treatment were told what punishment threshold they were facing before the first work period began. One idea would be to ask workers in the discretionary treatment on the questionnaire what they believed the punishment threshold was going to be before starting to see if workers in the discretionary treatment actually believed that the threshold they were facing was 180 or not.

If workers knew that they would not be punished, then they would not necessarily be motivated to give a certain effort level. Bosses may actually make less money in a situation where their workers knew what the punishment threshold was ahead of time (we saw this happen in the deterministic treatment) and may ultimately be more inclined to set a punishment threshold higher than 0. If that were the case, bosses would be pressured to set punishment thresholds as high as possible but still

achievable. This would help keep workers from believing the threshold was out of reach and therefore become discouraged.

Along with changes to the current experimental design, there are several possible extensions of the research question. One is would the results be different if bosses participated in the work task to some degree in order to understand just what is demanded of the workers? As boss 3019 stated, "I selected 'no punishment' because I didn't know how hard the task was."

Another extension would be to allow bosses the option of offering some percentage of their bonus to workers as another means of incentive, using the carrot/stick approach. This thesis finds that, holding the monetary effect of punishment constant, a punished worker whose punishment was chosen by someone whom the worker's actions will later affect will increase their effort more than a punished worker whose punishment was pre-determined. Both workers will increase their effort in response to being punished rather than reciprocating with a vengeful slowdown. This experiment cannot separate out the possible explanations for this behavior, but it is clear that intentionality plays a role in behavioral response to monetary punishment. These results imply that punishment can be successfully used to encourage worker effort in specific circumstances. Continued research in this area can help us understand more thoroughly how and when punishment is an effective behavior-shaping tool.

Acknowledgments

I wish to thank Nathan Stuart for the conversation that first sparked the idea for this project, Nick Arnosti for listening to me talk about it endlessly, Bob Gazzale for his input in developing the initial idea and continued investment. I especially wish to thank Sarah Jacobson for her unwavering support, encouragement, friendship, and endless cups of tea. I gratefully acknowledge the Van Duyne Prize for funding that made this project possible.

Appendices

Appendix 1: Tables

Table 1: Sequence of events

		Discretionary		Deterministic	
Step	Description	Worker	Boss	Worker	Boss
1	Practice	Practices task	**Decides punishment threshold**	Practices task	Waits
2	Period 1	Works	Survey	Works	Survey
3	Information	Learns: -punished? -**threshold** -period 1 pay	Learns: -got bonus? -period 1 pay	Learns: -punished? -period 1 pay	Learns: -got bonus? -period 1 pay
4	Period 2	Works	Survey	Works	Survey
5	Information	Learns: -period 2 pay -total earnings	Learns: -got bonus? -period 2 pay -total earnings	Learns: -period 2 pay -total earning	Learns: -got bonus? -period 2 pay -total earnings
6	Questionnaire	Completes questionnaire			
7	Pay & Leave	Receives payment and leaves			

Differences between treatments in bold

Table 2: Subject demographics

	Worker (discretionary)	Worker (deterministic)	Boss (discretionary)	Boss (deterministic)	Total
Male	14	16	17	15	62
Female	16	14	13	15	58
Senior	6	6	4	10	26
Junior	6	6	5	3	20
Sophomore	6	6	9	6	27
Freshman	12	12	12	11	47
GPA	3.62	3.60	3.67	3.60	3.62
Economics	15	15	15	14	59
Athletics	2.90	2.60	2.63	2.53	2.66
Earnings	$14.76	$9.36	$23.09	$20.36	$16.89
N	30	30	30	30	120

"GPA" is the mean GPA of subjects; "Economics" is the number of subjects who have taken at least one economics course; "Athletics" is the mean importance of athletics in the subjects' lives (on a scale of 1-4, 1 being more important); "Earnings" is the mean earnings of subjects

Table 3: Effort and earnings, by treatment

	Deterministic	Discretionary	Discretionary—punished	Discretionary—not punished
e_1 (number of bubbles completed in period 1)	83.30 (30.81)	98.70 (36.81)	97.70 (41.93)	99.26 (35.02)
e_2 (number of bubbles completed in period 2)	96.20 (39.21)	112.60 (41.62)	127.40 (56.41)	104.84 (30.30)
e_2 - e_1	12.90 (20.76)	13.89 (25.91)	29.70 (29.61)	5.58 (19.86)
Total earnings	$9.36 (5.39)	$15.18 (6.09)	$13.01 (7.59)	$16.32 (4.99)
Punishment threshold	180 (0)	73.10 (79.51)	164.00 (35.02)	25.26 (47.07)
N	30	29	10	19

Means listed. Standard deviations in parentheses.
Note: all workers in the deterministic treatment were punished.

Table 4: OLS regressions on e_1 (the number of bubbles completed in period 1)

	(1)	(2)	(3)
Discretionary	15.42* (8.83)		12.92 (8.28)
Session 3		-17.82 (15.70)	
Session 4		7.68 (15.70)	
Session 5		-19.42 (15.70)	
Session 6		-21.62 (15.70)	
Session 7		-15.72 (15.70)	
Female			6.36 (8.44)
Class year			6.91* (3.70)
GPA			26.64** (10.83)
Economics			-6.21 (8.69)
Athletics			-1.94 (3.14)
Constant	83.30*** (6.19)	102.22*** (11.39)	-26.76 (42.72)
N	59	59	58†
R^2	0.05	0.10	0.21

Discretionary = 1 if discretionary treatment, 0 if deterministic treatment. Session dummies: sessions 2 (dropped), 3, and 4 were discretionary, sessions 5, 6, 7 were deterministic. Female=1 if subject was female, 0 if male. Class year =1 for senior, 2 for junior, 3 for sophomore, and 4 for freshman. Economics = 1 if subject had taken any economics course, 0 if not. Athletics is the importance of athletics in the subjects' lives (on a scale of 1-4, 1 being more important).

† One subject dropped from specification (3) because "Athletics" empty for one subject

Standard errors in parentheses
** significant at 10%; ** significant at 5%; *** significant at 1%*

Table 5: OLS regressions on e_2 (the number of bubbles completed in period 2)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Discretionary	16.42	1.43			14.50	3.75		
	(10.53)	(6.32)			(9.00)	(5.99)		
Punished, discretionary			31.20**	17.15**			22.77*	16.92**
			(14.62)	(8.26)			(13.27)	(8.32)
Not punished, discretionary			8.64	-6.93			10.38	-3.20
			(11.74)	(6.70)			(10.25)	(6.58)
e_1		0.97***		0.98***		0.83***		0.85***
		(0.09)		(0.09)		(0.10)		(0.10)
Female					0.75	-4.54	3.36	-0.42
					(9.17)	(5.99)	(9.70)	(6.07)
Class					11.04***	5.29*	10.82***	4.79*
					(4.02)	(2.70)	(4.04)	(2.61)
GPA					40.42***	18.25**	39.14***	15.64*
					(11.77)	(8.09)	(11.90)	(7.89)
Economics					-3.08	2.09	-3.13	2.12
					(9.45)	(6.17)	(9.47)	(5.95)
Athletics					-6.56*	-4.95**	-6.68*	-5.10**
					(3.42)	(2.23)	(3.43)	(2.15)
Constant	96.20***	15.23*	96.20***	14.92*	-61.95	-39.69	-57.63	-32.12
	(7.38)	(8.82)	(7.31)	(8.34)	(46.45)	(30.30)	(46.85)	(29.41)
N	59	59	59	59	58†	58†	58†	58†
R^2	0.04	0.68	0.08	0.72	0.31	0.71	0.32	0.74

All subjects included. Discretionary = 1 if discretionary treatment, 0 if deterministic treatment. Punished, discretionary = 1 for subjects that were punished in the discretionary treatment. Not punished, discretionary = 1 for subjects that were not punished in the discretionary treatment. Female=1 if subject was female, 0 if male. Class year =1 for senior, 2 for junior, 3 for sophomore, and 4 for freshman. Economics = 1 if subject had taken any economics course, 0 if not. Athletics is the importance of athletics in the subjects' lives (on a scale of 1-4, 1 being more important).

† One subject dropped from specifications (5), (6), (7), and (8) because "Athletics" empty for one subject

Standard errors in parentheses
** significant at 10%; ** significant at 5%; *** significant at 1%*

Table 6: OLS regressions on e_2 in discretionary treatment, including threshold

	(1)	(2)	(3)	(4)	(5)	(6)
e_1	0.844***	0.898***	0.878***	0.550***	0.616***	0.607***
	(0.125)	(0.122)	(0.131)	(0.140)	(0.141)	(0.152)
Threshold	0.141**		0.055	0.104*		0.020
	(0.058)		(0.113)	(0.050)		(0.101)
Punished		23.962**	16.318		21.338**	18.134
		(9.259)	(18.291)		(9.160)	(18.962)
Female				-0.988	5.790	4.938
				(8.209)	(9.095)	(10.299)
Class year				6.276	5.214	5.319
				(3.803)	(3.809)	(3.941)
GPA				19.457*	17.426*	17.578*
				(9.351)	(9.311)	(9.575)
Economics				3.197	3.061	3.115
				(7.976)	(7.794)	(7.993)
Athletics				-10.102***	-10.402***	-10.360***
				(3.017)	(2.954)	(3.035)
Constant	18.975	15.692	16.347	-10.627	-9.091	-9.065
	(13.135)	(13.246)	(13.512)	(34.758)	(34.021)	(34.871)
N	29	29	29	28†	28†	28†
R^2	0.69	0.70	0.70	0.80	0.81	0.81

Discretionary treatment only. Punished=1 if subject was punished. Female=1 if subject was female, 0 if male. Class year =1 for senior, 2 for junior, 3 for sophomore, and 4 for freshman. Economics = 1 if subject had taken any economics course, 0 if not. Athletics is the importance of athletics in the subjects' lives (on a scale of 1-4, 1 being more important). Religion is the importance of athletics in the subjects' lives (on a scale of 1-4, 1 being more important);Min. job=1 if subject had ever worked a minimum wage job, 0 if not. Sibling =1 if subject had at least one sibling, 0 if not.

† One subject dropped from specifications (4), (5), and (6) because "Athletics" empty for one subject

Standard errors in parentheses
** significant at 10%; ** significant at 5%; *** significant at 1%*

Table 7: OLS regressions on e_2-e_1

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Discretionary	0.10	1.43			1.57	3.75		
	(6.10)	(6.32)			(5.96)	(5.99)		
Punished, discretionary			16.80**	17.15**			15.91*	16.92**
			(8.09)	(8.26)			(8.41)	(8.32)
Not punished, discretionary			-7.32	-6.93			-5.56	-3.20
			(6.50)	(6.70)			(6.49)	(6.58)
e_1		-0.03		-0.02		-0.17*		-0.15
		(0.09)		(0.09)		(0.10)		(0.10)
Female					-5.61	-4.54	-1.08	-0.42
					(6.07)	(5.99)	(6.14)	(6.07)
Class year					4.13	5.29*	3.75	4.79*
					(2.66)	(2.70)	(2.56)	(2.61)
GPA					13.78*	18.25**	11.57	15.64*
					(7.79)	(8.09)	(7.54)	(7.89)
Economics					3.13	2.09	3.04	2.12
					(6.25)	(6.17)	(6.00)	(5.95)
Athletics					-4.62**	-4.95**	-4.83**	-5.10**
					(2.26)	(2.23)	(2.17)	(2.15)
Constant	12.90***	15.23*	12.90***	14.92*	-35.20	-39.69	-27.70	-32.12
	(4.28)	(8.82)	(4.05)	(8.34)	(30.74)	(30.30)	(29.67)	(29.41)
N	59	59	59	59	58	58	58	58
R^2	0.00	0.00	0.12	0.12	0.15	0.20	0.23	0.27

All subjects included. Discretionary = 1 if discretionary treatment, 0 if deterministic treatment. Punished, discretionary = 1 for subjects that were punished in the discretionary treatment. Not punished, discretionary = 1 for subjects that were not punished in the discretionary treatment. Female=1 if subject was female, 0 if male. Class year =1 for senior, 2 for junior, 3 for sophomore, and 4 for freshman. Economics = 1 if subject had taken any economics course, 0 if not. Athletics is the importance of athletics in the subjects' lives (on a scale of 1-4, 1 being more important).

† One subject dropped from specifications (5), (6), (7), and (8) because "Athletics" empty for one subject

Standard errors in parentheses
** significant at 10%; ** significant at 5%; *** significant at 1%*

Table 8: *OLS regressions on e_2-e_1 in discretionary treatment, including threshold*

	(1)	(2)	(3)	(4)	(5)	(6)
e_1	-0.16	-0.10	-0.12	-0.45***	-0.38**	-0.39**
	(0.13)	(0.12)	(0.13)	(0.14)	(0.14)	(0.15)
Threshold	0.14**		0.06	0.10*		0.02
	(0.06)		(0.11)	(0.05)		(0.10)
Punished		23.96**	16.32		21.34**	18.13
		(9.26)	(18.29)		(9.16)	(18.96)
Female				-0.99	5.79	4.94
				(8.21)	(9.10)	(10.30)
Class year				6.28	5.21	5.32
				(3.80)	(3.81)	(3.94)
GPA				19.46*	17.43*	17.58*
				(9.35)	(9.31)	(9.58)
Economics				3.20	3.06	3.12
				(7.98)	(7.79)	(7.99)
Athletics				-10.10***	-10.40***	-10.36***
				(3.02)	(2.95)	(3.04)
Constant	18.98	15.69	16.35	-10.63	-9.09	-9.07
	(13.14)	(13.25)	(13.51)	(34.76)	(34.02)	(34.87)
N	29	29	29	28	28	28
R^2	0.21	0.22	0.23	0.58	0.60	0.60

Discretionary treatment only. Punished=1 if subject was punished. Female=1 if subject was female, 0 if male. Class year =1 for senior, 2 for junior, 3 for sophomore, and 4 for freshman. Economics = 1 if subject had taken any economics course, 0 if not. Athletics is the importance of athletics in the subjects' lives (on a scale of 1-4, 1 being more important).

† One subject dropped from specifications (4), (5), and (6) because "Athletics" empty for one subject

Standard errors in parentheses
** significant at 10%; ** significant at 5%; *** significant at 1%*

Table 9: OLS regressions on thresholds in discretionary treatment

	(1)	(2)
Female	-22.21	-10.014
	(35.22)	(26.404)
Class year	20.93	35.978**
	(17.37)	(13.364)
GPA	-41.97	-115.549**
	(59.32)	(47.574)
Economics	31.03	53.481
	(41.10)	(31.236)
Athletics	14.06	4.316
	(11.08)	(9.681)
Session 3		-107.421***
		(36.337)
Session 4		-143.560***
		(31.358)
Constant	152.878	445.505**
	(209.160)	(175.143)
N	30	30
R^2	0.08	0.56

Female=1 if subject was female, 0 if male. Class year =1 for senior, 2 for junior, 3 for sophomore, and 4 for freshman. Economics = 1 if subject had taken any economics course, 0 if not. Athletics is the importance of athletics in the subjects' lives (on a scale of 1-4, 1 being more important). Session 3=1 if data from session 3 (discretionary, 0 if not. Session 4 = 1 if data from session 4 (discretionary), 0 if not.

Standard errors in parentheses
** significant at 10%; ** significant at 5%; *** significant at 1%*

Appendix 2: Figures

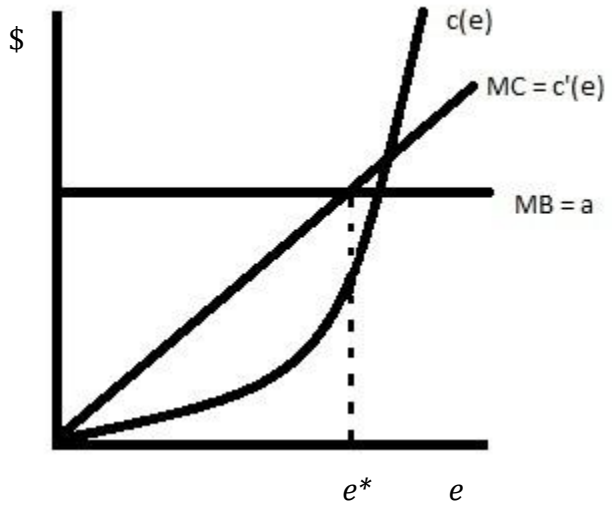
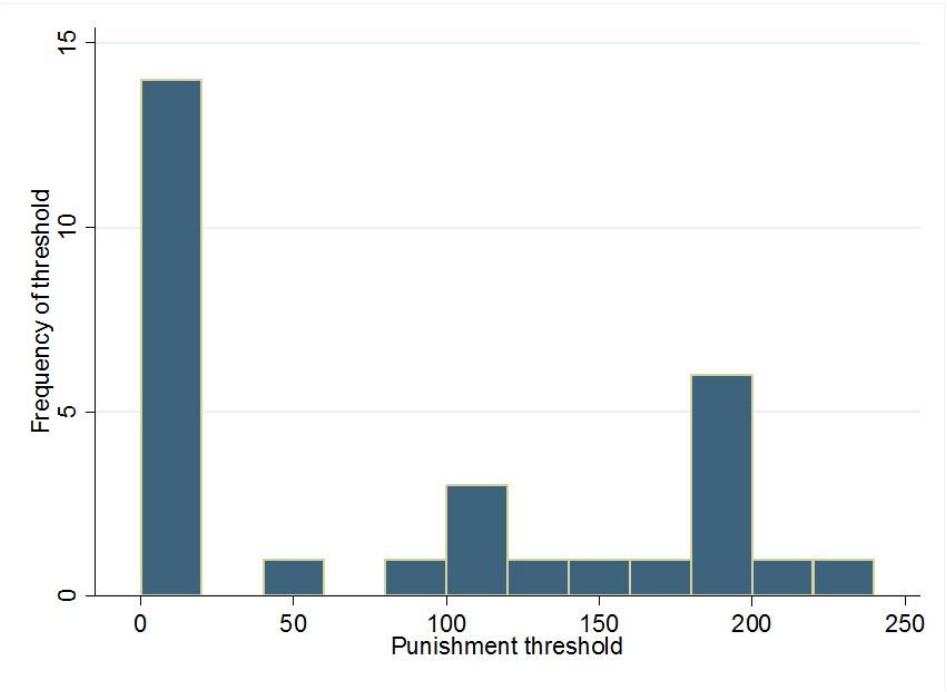
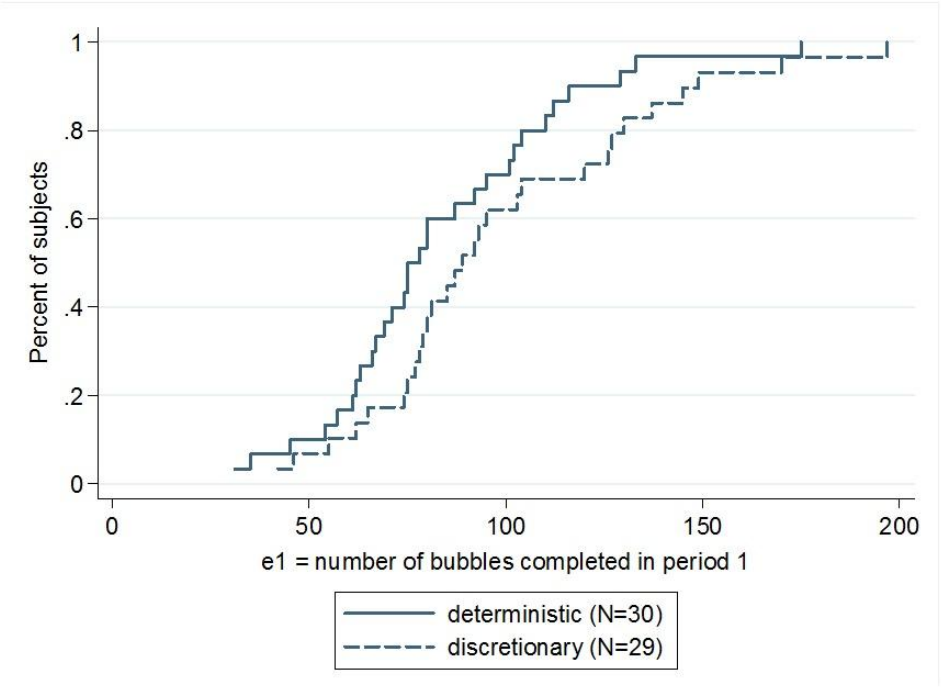


Figure 1: Utility maximization condition with no punishment or reciprocity



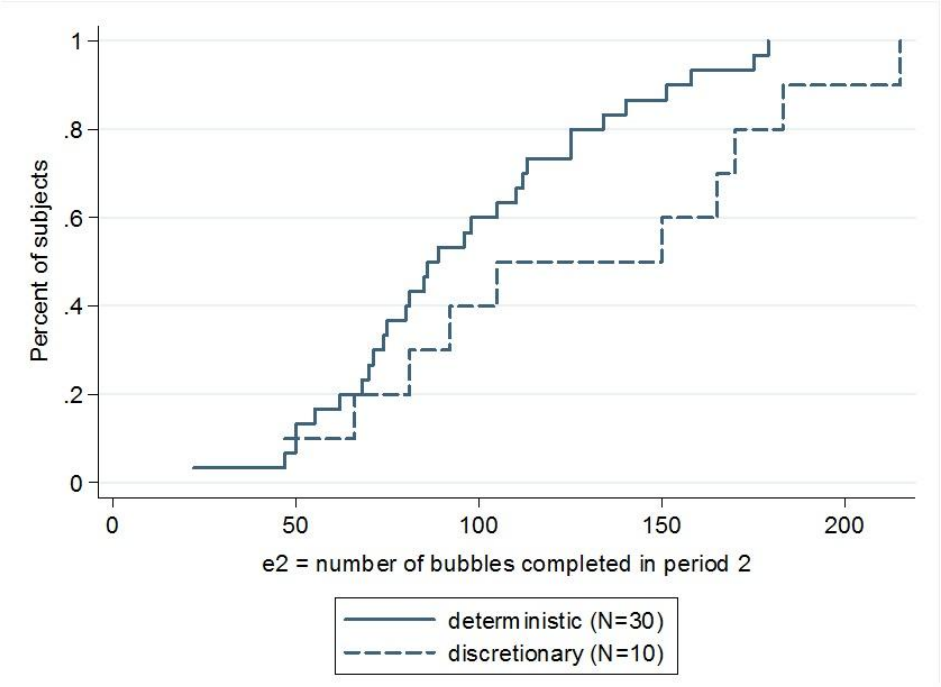
Note: 0 had the highest frequency (14) followed by 180 (6).

Figure 2: Punishment thresholds in the discretionary treatment



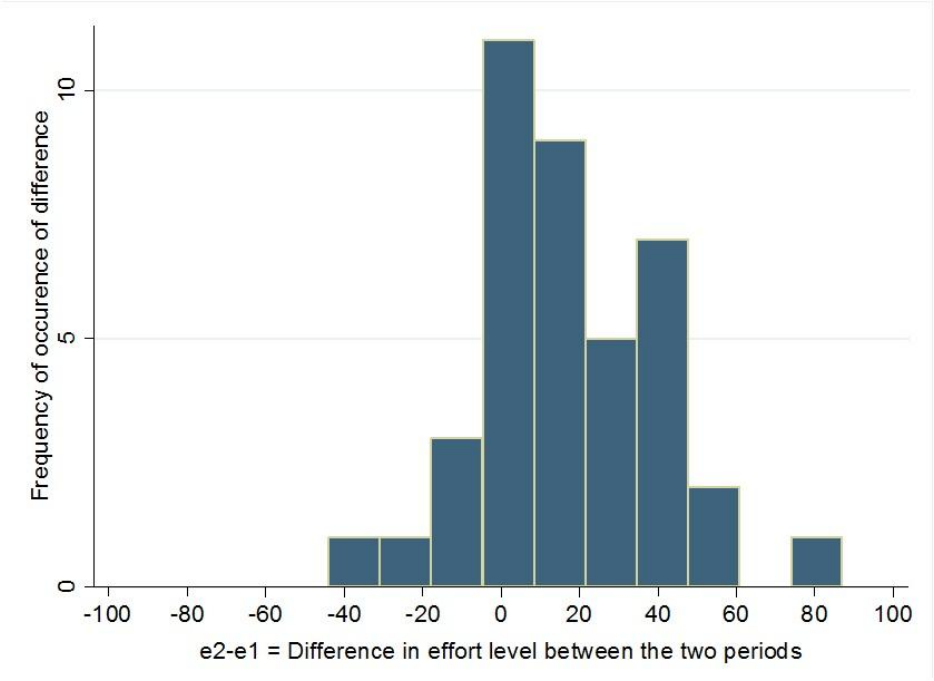
All subjects

Figure 3: CDF plot showing e_1, by treatment



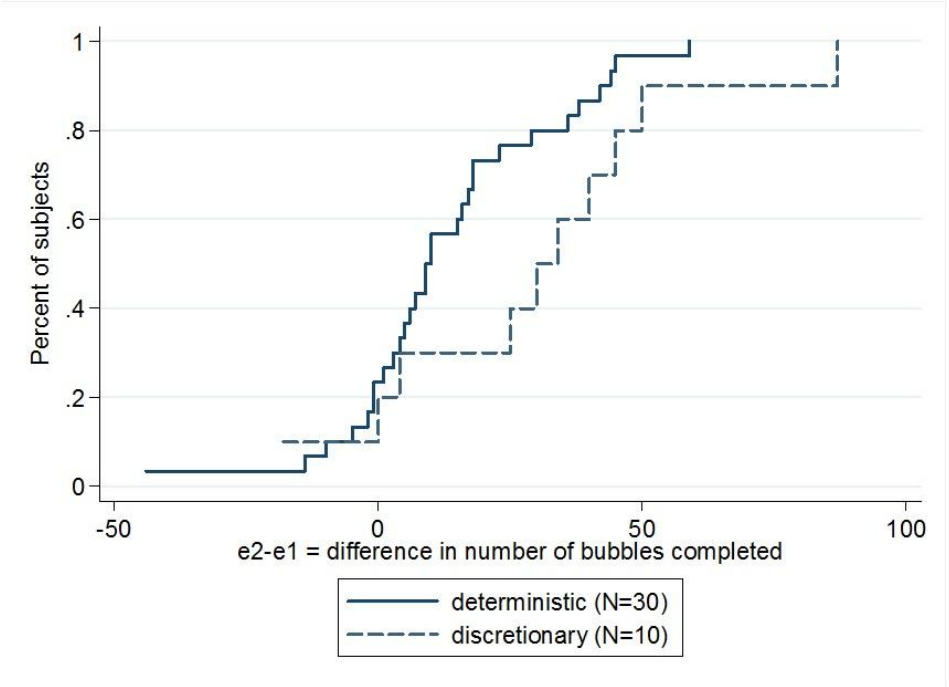
Punished subjects only

Figure 4: CDF plot showing e_2, by treatment



N=40; discretionary and deterministic punished subjects only (subjects are pooled)

Figure 5: e_2-e_1



Punished subjects only

Figure 6: CDF plot showing e_2-e_1, by treatment

Appendix 3: Consent Form

Vengeful Slowdown
Primary Investigator: Ellen Stuart
Faculty Adviser: Sarah Jacobson, Economics Department

You are invited to participate in a research study. The purpose of the study is to investigate individual decision-making. You are invited to participate because you are a student at Williams College.

Please read this form carefully and ask any questions you may have before agreeing to take part in the study.

I. Procedures

If you agree to participate in this study you will perform tasks and/or make decisions. At the end of the experiment you will fill out a short questionnaire. At that point, your participation will be complete. The study will take up to an hour and a half. You will receive instructions in the lab that explain how monetary compensation will be determined.

II. Risks

This experiment involves intensive writing for short periods of time. If you are unable to do this please inform one of the researchers. Otherwise you will not be faced with any more risk than you would in a normal day.

III. Benefits

Participation in this study may not benefit you personally. Overall, we hope to gain information about individual decision-making.

IV. Compensation

You will receive instructions in the lab that explain how monetary compensation will be determined.

V. Voluntary Participation and Withdrawal

Participation in this research is completely voluntary. If you decide not to take part in this study, it will not affect your current or future relationship with Williams College. During the course of the study you have the right to drop out at any time. If you do not finish the study, we will not use any of your data.

VI. Confidentiality

The records of this study will be kept private to the extent allowed by law. Public reports will not include information that will make it possible to identify you. No one will know the decisions that you personally make. The findings will be summarized and reported in group form. Research records will be kept in a locked file and on a password- and firewall protected computer; only the researchers will have access to the records. Information may also be shared with those who make sure the study is done correctly (Williams College Institutional Review Board and the Office for Human Research Protections (OHRP)).

VII. Contact Persons

Contact Ellen Stuart at 352-226-7187 or Ellen.M.Stuart@williams.edu, or Sarah Jacobson at 413-597-4766 or Sarah.A.Jacobson@williams.edu, if you have questions about this study. If you have questions or concerns about your rights as a participant in this research study, you may contact the Institutional Review Board (IRB) at 413-597-2240 (Kenneth Savitsky, committee chair) or access their website at http://wiki.williams.edu/display/IRB/.

You will be given a copy of this consent form to keep for your records.

VIII. Statement of Consent

I have read the above information, and have received answers to any questions I asked. By signing below I consent to take part in the study.

_____ _____

Participant Date

_____ _____

Principal Investigator or Researcher Obtaining Consent Date

This consent form will be kept by the researcher for at least three years beyond the end of the study.

Appendix 4.1: Subject Instructions (front page - discretionary)

- Two roles: worker and boss
- Two work periods
    - Separate tasks for each role

Period I

- Boss task
    - Complete survey
    - Paid base pay ($3.00)

- Worker task
    - Fill bubbles on bubble sheet according to a pattern
        - Examples on other side of this sheet
    - Paid per bubble filled in correctly
        - $0.08/bubble → worker
        - $0.08/bubble → boss
    - Boss gets a bonus (+$5.00) if worker fills 180 or more bubbles

- Punishment threshold
    - Boss chooses number of bubbles worker must complete in order to NOT be punished
    - May choose any level between 20 and 300 or "always punish" or "never punish"
    - If applicable, punishment applied to worker's earnings at the end of period I
        - Punishment subtracts $5.00 from worker's earnings (this money is NOT added to the boss's pay)

- Payment (Period I)

Worker	Boss
_____ bubbles x $0.08/bubble = _____	base pay = $3.00
− punishment = _____	+_____ bubbles x $0.08/bubble = _____
net period I pay = _____	+ bonus = _____
	net period I pay = _____

Period II is like period I with one change

- Worker, boss still complete their tasks
- Boss still eligible for bonus
- No punishment for worker
- Payment (Period II)

Worker	Boss
_____ bubbles x $0.08/bubble = _____	base pay = $3.00
net period II pay = _____	+_____ bubbles x $0.08/bubble = _____
	+ bonus = _____
	net period II pay = _____

Net payment = period I pay + period II pay

Appendix 4.2: Subject Instructions (front page - deterministic)

- Two roles: worker and boss
- Two work periods
    - Separate tasks for each role

Period I

- Boss task
    - Complete survey
    - Paid base pay ($3.00)

- Worker task
    - Fill bubbles on bubble sheet according to a pattern
        - Examples on other side of this sheet
    - Paid per bubble filled in correctly
        - $0.08/bubble → worker
        - $0.08/bubble → boss
    - Boss gets a bonus (+$5.00) if worker fills 180 or more bubbles

- Punishment threshold
    - If worker completes fewer than 180 bubbles, worker will be punished
    - If applicable, punishment is applied to worker's earnings at the end of period
        - Punishment subtracts $5.00 from worker's earnings (this money is NOT added to the boss's pay)

- Payment (Period I)

Worker

_____ bubbles x $0.08/bubble = _____
− punishment = _____
net period I pay = _____

Boss

base pay = $3.00
+_____ bubbles x $0.08/bubble = _____
+ bonus = _____
net period I pay = _____

Period II is like period I with one change

- Worker, boss still complete their tasks
- Boss still eligible for bonus
- No punishment for worker
- Payment (Period II)

Worker

_____ bubbles x $0.08/bubble = _____
net period II pay = _____

Boss

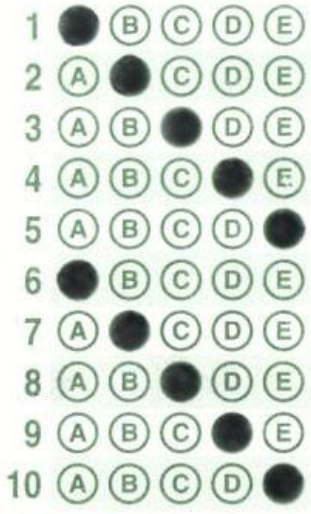
base pay = $3.00
+_____ bubbles x $0.08/bubble = _____
+ bonus = _____
net period II pay = _____

Net payment = period I pay + period II pay

Appendix 4.3: Subject Instructions (back page)

Examples – Worker Task

The bubbles must be filled in neatly and completely in the following pattern:



The worker will not be paid for bubbles that are not filled in properly:

1. Bubble not filled in within the lines
2. Bubble not completely filled in
3. Bubble not completely filled in
4. Bubble not completely filled in; also too light
5. Bubble doesn't follow pattern

Appendix 5: Punishment Decision Form

Circle the number of bubbles your worker must complete to NOT be punished. **Circle only one.**

Punishment means that $5.00 will be subtracted from your worker's earnings.

ALWAYS PUNISH (I will punish my worker regardless of how many bubbles s/he fills)

300

280

260

240

220

200

180 (Reminder: this is your bonus threshold)

160

140

120

100

80

60

40

20

NEVER PUNISH (I will not punish my worker regardless of how many bubbles s/he fills)

Appendix 6: Boss Survey

Please complete the following survey. For each question, there is no right or wrong answer. We simply want to learn what you feel to be true about yourself and what you prefer.

If you are unsure about how to interpret a question or a suggested answer, do not raise your hand to clarify. Simply interpret it in the way that seems most sensible and meaningful to you.

There is no penalty for failing to complete the survey, but please try to complete it as fully and thoughtfully as you can. If you feel that a question does not apply to you, you may skip it.

For each statement in the table below, check the box that corresponds to the extent to which you agree with the statement.

	1 – Strongly Disagree	2 – Somewhat Disagree	3 – Slightly Disagree	4 – Slightly Agree	5 – Somewhat Agree	6 – Strongly Agree
1. Luck plays an important part in everyone's life.	☐	☐	☐	☐	☐	☐
2. Some people are consistently lucky, and others are unlucky.	☐	☐	☐	☐	☐	☐
3. I consider myself to be a lucky person.	☐	☐	☐	☐	☐	☐
4. I believe in luck.	☐	☐	☐	☐	☐	☐
5. I often feel it's my lucky day.	☐	☐	☐	☐	☐	☐
6. I consistently have good luck.	☐	☐	☐	☐	☐	☐
7. It's a mistake to base any decision on how lucky you feel.	☐	☐	☐	☐	☐	☐
8. Luck works in my favor.	☐	☐	☐	☐	☐	☐
9. I don't mind leaving things to chance because I'm a lucky person.	☐	☐	☐	☐	☐	☐
10. Even the things in life I can't control tend to go my way because I'm lucky.	☐	☐	☐	☐	☐	☐

11. There is such a thing as luck that favors some people but not others	☐	☐	☐	☐	☐	☐
12. Luck is nothing more than random chance.	☐	☐	☐	☐	☐	☐

13. How would you define "luck"? _____

_____

_____

14. When you have had a lucky (good) thing happen to you, do you more often feel:
    a. I have good luck right now; I expect more lucky things to happen to me today
    b. I've used up my luck; I can't expect any more lucky things to happen to me today
    c. Neither

15. Do you feel that luck or chance has played a large role in your life?
    a. Yes, a large role
    b. No, not a large role

16. In general, do you think that most people's fortunes and happiness depend more on their own choices or on chance? (You may consider chance to include all circumstances outside of the person's control.)
    a. More on their own choices than on chance
    b. More on chance than on their own choices
    c. Equally on choice and chance

17. Have you ever decided NOT TO take a risk because of a hunch that you'd be unlucky?
    a. Never
    b. Sometimes
    c. Often

18. Have you ever decided TO take a risk because of a hunch that you'd be lucky?
    a. Never
    b. Sometimes
    c. Often

19. Do you consider yourself to be: (circle one)
    a. A very unlucky person
    b. A somewhat unlucky person
    c. Neither a lucky nor unlucky person
    d. A somewhat lucky person
    e. A very lucky person

The following questions ask you to choose a number from 0 to 10 to indicate how willing you are to take risks. 0 means "unwilling to take risks" and 10 means "fully prepared to take risks."

20. Do you see yourself GENERALLY as a person who is fully prepared to take risks or do you try to avoid taking risks?

0	1	2	3	4	5	6	7	8	9	10
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Unwilling to take risks										Fully prepared to take risks

21. In the context of DRIVING (e.g., car), how willing are you to take risks?

0	1	2	3	4	5	6	7	8	9	10
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Unwilling to take risks										Fully prepared to take risks

22. In the context of FINANCES (e.g., savings, investments, credit cards), how willing are you to take risks?

0	1	2	3	4	5	6	7	8	9	10
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Unwilling to take risks										Fully prepared to take risks

23. In the context of COMPETITIVE SPORTS or games (e.g., choice of sports, moves/plays within a sport), how willing are you to take risks?

0	1	2	3	4	5	6	7	8	9	10
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Unwilling to take risks										Fully prepared to take risks

24. In the context of LEISURE ACTIVITIES (e.g., snowboarding, mountain climbing), how willing are you to take risks?

0	1	2	3	4	5	6	7	8	9	10
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Unwilling to take risks										Fully prepared to take risks

25. In the context of YOUR CAREER (e.g., your choice of internships, career paths you are considering), how willing are you to take risks?

0	1	2	3	4	5	6	7	8	9	10
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Unwilling to take risks										Fully prepared to take risks

26. In the context of YOUR ACADEMICS (e.g., your choice of courses or projects), how willing are you to take risks?

0	1	2	3	4	5	6	7	8	9	10
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Unwilling to take risks										Fully prepared to take risks

27. In the context of YOUR HEALTH (e.g., smoking, other behaviors that pose risks to your health), how willing are you to take risks?

0	1	2	3	4	5	6	7	8	9	10
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Unwilling to take risks										Fully prepared to take risks

28. In the context of PERSONAL RELATIONSHIPS (e.g., choice of friends / partners, behavior with them), how willing are you to take risks?

0	1	2	3	4	5	6	7	8	9	10
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Unwilling to take risks										Fully prepared to take risks

Imagine that there is a lottery that operates as follows. You choose how much to "invest" in the lottery, and that amount may be doubled or lost with equal likelihoods. That is, for any amount you put in, you have a 50-50 chance of winning twice that amount or losing all of it. For example, if you invest $100, you are just as likely to end up with $200 as $0.

29. How much money would you invest if you were offered the chance to enter the lottery today? (Write 0 if you would not invest anything.) _____

30. Imagine you receive a surprise gift of $10,000 today and are also offered the chance to enter the same lottery. How much would you invest in this case? _____

31. What do you expect your annual income to be when you are 40 years old? _____

32. Imagine yourself at age 40, earning the income you now expect for that age. If you were offered the lottery in that situation, how much do you think you would invest? _____

Imagine instead that the money you invest may be tripled or lost with equal likelihoods. That is, for any amount you put in, you have a 50-50 chance of winning three times that amount or losing all of it. For example, if you invest $100, you are just as likely to end up with $300 as $0.

33. How much money would you invest if you were offered the chance to enter the lottery today? (Write 0 if you would not invest anything.) _____

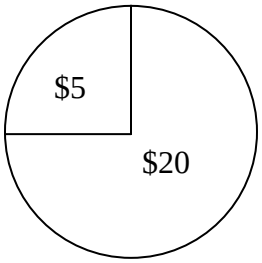
34. How much would you invest if you got a surprise gift of $10,000 today? _____

35. How much would you be likely to invest if you were 40 and earning the income you currently expect to earn at that age? _____

Imagine now that you're facing the pairs of lotteries (you can also think of them as gambles or bets) on the next page.

In each pair, your task is to choose the ONE lottery of the two that you prefer (like best).

Each lottery is represented by a pie chart like the following:

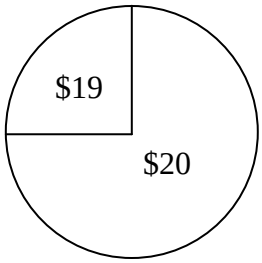
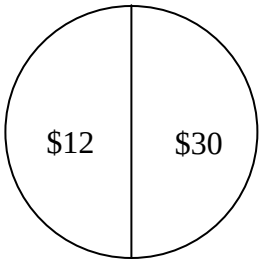
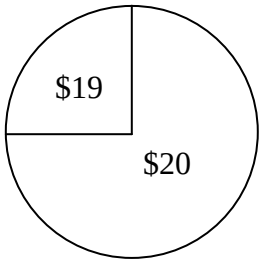
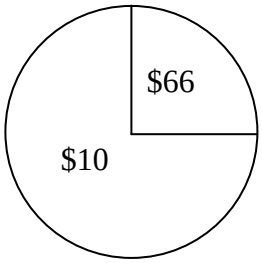
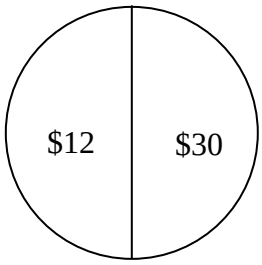
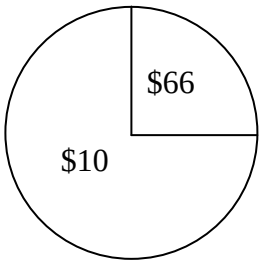
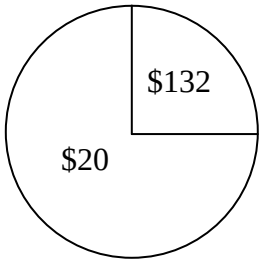
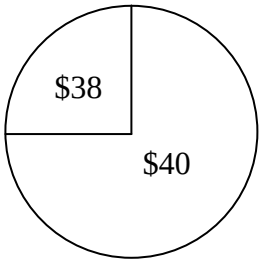


You'd read this lottery as: a ¼ (25%) chance of $5, and a ¾ (75%) chance of $20.

In each pair, you have two different lotteries like this with different chances of winning (always 25%, 50%, or 75%) and with different amounts that could be won.

In each pair, some people will prefer the lottery on the left and others will prefer the lottery on the right. We want to know which YOU prefer.

Choose ONE LOTTERY for each PAIR.

Pair #	LEFT Lottery	RIGHT Lottery	Circle YOUR Choice
1			**LEFT** or **RIGHT**
2			**LEFT** or **RIGHT**
3			**LEFT** or **RIGHT**
4			**LEFT** or **RIGHT**

Imagine now that you face the following set of pairs of bets. Your task is to choose which bet you prefer in each pair.

Each pair consists of a SAFE bet and a RISKY bet. In a given row, the SAFE and RISKY bets have the same likelihood of winning. The SAFE bet always pays $20 if you win and $16 if you lose. The RISKY bet always pays $38.50 if you win and $1 if you lose. As you move down the table, the chances of winning each bet increase.

Here's an example of how to read the table. Look at Row 3. The SAFE bet offers a 2 in 10 (20%) chance of winning $20 and an 8 in 10 (80%) chance of winning $16. The RISKY bet offers a 2 in 10 (20%) chance of winning $38.50 and an 8 in 10 (80%) chance of winning $1. Some people prefer the SAFE bet and others the RISKY bet. We want to know what YOU prefer.

Note that for Row 1, in both bets, you are assured of "losing" (you earn the smaller amount for sure). Thus, in Row 1 everyone probably prefers the SAFE bet because it gives $16 instead of $1 for sure. Similarly, for Row 11, you are assured of "winning" so everyone probably prefers the RISKY bet for Row 11 because it gives $38.50 instead of $20 for sure.

For each pair, indicate whether you prefer the SAFE (left column) or RISKY (right column) bet.

Row	**SAFE Bet**	**RISKY Bet**	**Circle your choice**
1	0/10 of $20, 10/10 of $16	0/10 of $38.50, 10/10 of $1	SAFE or RISKY
2	1/10 of $20, 9/10 of $16	1/10 of $38.50, 9/10 of $1	SAFE or RISKY
3	2/10 of $20, 8/10 of $16	2/10 of $38.50, 8/10 of $1	SAFE or RISKY
4	3/10 of $20, 7/10 of $16	3/10 of $38.50, 7/10 of $1	SAFE or RISKY
5	4/10 of $20, 6/10 of $16	4/10 of $38.50, 6/10 of $1	SAFE or RISKY
6	5/10 of $20, 5/10 of $16	5/10 of $38.50, 5/10 of $1	SAFE or RISKY
7	6/10 of $20, 4/10 of $16	6/10 of $38.50, 4/10 of $1	SAFE or RISKY
8	7/10 of $20, 3/10 of $16	7/10 of $38.50, 3/10 of $1	SAFE or RISKY
9	8/10 of $20, 2/10 of $16	8/10 of $38.50, 2/10 of $1	SAFE or RISKY
10	9/10 of $20, 1/10 of $16	9/10 of $38.50, 1/10 of $1	SAFE or RISKY
11	10/10 of $20, 0/10 of $16	10/10 of $38.50, 0/10 of $1	SAFE or RISKY

Appendix 7.1: Earning form (worker, discretionary)

Period I

Punishment threshold set by your boss: _____

Number of bubbles completed: _____

Punished?

YES	NO
PUNISHMENT (-$5.00)	NO PUNISHMENT ($0.00)

_____ bubbles x $0.08/bubble = **YOUR EARNINGS**= _____

PUNISHMENT = _____

PAYMENT (period I) = _____

Your boss's pay (per. I) =_____ = $3.00 base + _____ bubbles x $0.08/bubble + _____ (bonus)

Period II

_____ bubbles x $0.08/bubble = **YOUR EARNINGS**= _____

PAYMENT (period II) = _____

Your boss's pay (per. II) =_____ = $3.00 base + _____ bubbles x $0.08/bubble + _____ (bonus)

Total Payment

Payment in period I = _____

Payment in period II = _____

YOUR NET PAYMENT = _____

Your boss's net payment = _____

Appendix 7.2: Earning form (worker, deterministic)

Period I

Pre-determined punishment threshold: 180

Number of bubbles completed: _____

Punished?

YES	NO
PUNISHMENT (-$5.00)	NO PUNISHMENT ($0.00)

_____ bubbles x $0.08/bubble = **YOUR EARNINGS**= _____

PUNISHMENT = _____

PAYMENT (period I) = _____

Your boss's pay (per. I) =_____ = $3.00 base + _____ bubbles x $0.08/bubble + _____ (bonus)

Period II

_____ bubbles x $0.08/bubble = **YOUR EARNINGS**= _____

PAYMENT (period II) = _____

Your boss's pay (per. II) =_____ = $3.00 base + _____ bubbles x $0.08/bubble + _____ (bonus)

Total Payment

Payment in period I = _____

Payment in period II = _____

YOUR NET PAYMENT = _____

Your boss's net payment = _____

Appendix 7.3: Earning form (boss)

Period I
Number of bubbles completed by your worker: _____ (bonus threshold: 180 bubbles)

Met threshold?

YES	NO
BONUS (+$5.00)	NO BONUS ($0.00)

BASE PAY = $3.00

_____ bubbles x $0.08/bubble = **YOUR EARNINGS** = _____

BONUS = _____

PAYMENT (period I) = _____

Your worker's pay (per. I) =_____ = _____ bubbles x $0.08/bubble - _____ (punishment)

Period II
Number of bubbles completed by your worker: _____ (bonus threshold: 180 bubbles)

Met threshold?

YES	NO
BONUS (+$5.00)	NO BONUS ($0.00)

BASE PAY = $3.00

_____ bubbles x $0.08/bubble = **YOUR EARNINGS** = _____

BONUS = _____

PAYMENT (period II) = _____

Your worker's pay (per. II) =_____ = _____ bubbles x $0.08/bubble

---

Total Payment

Payment in period I = _____

Payment in period II = _____

YOUR NET PAYMENT = _____

Your worker's net payment = _____

Appendix 8: Questionnaire

Demographics

Gender

- ☐ Male
- ☐ Female

Year

- ☐ 2011
- ☐ 2012
- ☐ 2013
- ☐ 2014

Major (check all that apply)

- ☐ Division I
- ☐ Division II
- ☐ Division III
- ☐ Undeclared

Current GPA

- ☐ < 1.5
- ☐ 1.5-2.0
- ☐ 2.0-2.5
- ☐ 2.5-3.0
- ☐ 3.0-3.5
- ☐ > 3 .5

Have you taken any economics courses?

- ☐ Yes
- ☐ No

If yes, which ones? Check all that apply.

- ☐ 110
- ☐ 120
- ☐ 251

☐ 252

☐ 255

☐ Other: _____

☐ Does not apply

Personal history

Parental marital status

☐ Married

☐ Divorced

☐ Separated

☐ Never married

☐ Deceased

What is the highest level of education of your father?

☐ Did not complete high school

☐ High school

☐ Some college

☐ Bachelor's degree

☐ Graduate degree

☐ Don't know

What is the highest level of education of your mother?

☐ Did not complete high school

☐ High school

☐ Some college

☐ Bachelor's degree

☐ Graduate degree

☐ Don't know

How many older brothers do you have (including half- and step-siblings)? _____

How many older sisters do you have (including half- and step-siblings)? _____

How many younger brothers do you have (including half- and step-siblings)? _____

How many younger sisters do you have (including half- and step-siblings)? _____

What was your family's annual income growing up?

- ☐ Less than $30,000
- ☐ $30,000-$60,000
- ☐ $60,000-$100,000
- ☐ More than $100,000
- ☐ Don't know

Have you ever worked a minimum wage job?

- ☐ Yes
- ☐ No

What is your religious affiliation?

- ☐ Catholic
- ☐ Protestant
- ☐ Eastern/Orthodox
- ☐ Non-Trinitarian (e.g., Jehovah's Witness, Latter Day Saints)
- ☐ Other Christian
- ☐ Buddhist
- ☐ Hindu
- ☐ Jewish
- ☐ Muslim
- ☐ Other: _____
- ☐ No religion

How often to you attend religious services or participate in religious ritual?

- ☐ Every day
- ☐ Every week
- ☐ 1-2 times a month
- ☐ 2-11 times a year
- ☐ Once per 1-3 years
- ☐ Never

How important is your religion/religious affiliation to your identity?

- ☐ Very important
- ☐ Somewhat important
- ☐ A little important
- ☐ Not important

Are you an athlete?

- ☐ Yes
- ☐ No

How important is your sport to your identity?

- ☐ Very important
- ☐ Somewhat important
- ☐ A little important
- ☐ Not important

Questions about the experiment

Was there anything in the written or verbal instructions for this experiment that you found confusing or did not understand?

- ☐ Yes:

_____

_____

_____

- ☐ No

What do you think the experimenters will learn from this experiment?

_____

_____

_____

[The following questions were only asked of the workers]

Did you enjoy the work task?

☐ Yes

☐ No

At the beginning of the experiment, did you believe that the bonus threshold was achievable?

☐ Yes

☐ No

Looking back, do you believe that the bonus threshold was achievable?

☐ Yes

☐ No

Did you feel discouraged about your ability to reach the bonus threshold?

☐ Yes

☐ No

If you were punished, did you feel as though the punishment was deserved?

☐ Yes

☐ No

☐ I was not punished

If you were punished, did you feel as though the punishment was unnecessarily harsh?

☐ Yes

☐ No

☐ I was not punished

Did you feel as though your boss understood the effort level necessary to reach the bonus threshold?

☐ Yes

☐ No

Did you feel as though your ability to perform the task noticeably changed over the course of the experiment?

☐ Yes, I felt as though my ability to perform the task noticeably improved

☐ Yes, I felt as though my ability to perform the task noticeably declined

☐ I felt as though my ability to perform the task neither improved nor declined

Did the possibility of punishment cause you to consciously change your effort level in the first period to avoid being punished?

☐ Yes

☐ No

If you were punished, did being punished cause you to consciously change your effort level in the second work period in an attempt to "take revenge" on your boss?

☐ Yes

☐ No

☐ I was not punished

If you were not punished, did *not* being punished cause you to consciously change your effort level in the second work period?

☐ Yes

☐ No

☐ I was punished

Do you have any problems with your hands that interfered with your ability to do the task?

☐ Yes

☐ No

Was there any other reason for a decline in performance throughout the experiment?

☐ Yes:

_____

_____

_____

☐ No

☐ I did not experience a decline in performance

[The following questions were only asked of bosses – deterministic]

At the beginning of the experiment, did you think it was likely that your worker would reach the bonus threshold?

☐ Yes

☐ No

Looking back, do you believe that your worker was capable of achieving the bonus threshold?

☐ Yes

☐ No

If you didn't receive a bonus, were you upset when your worker *did not* reach the bonus threshold?

☐ Yes

☐ No

☐ I received a bonus

If you received a bonus, were you pleased when your worker *did* reach the bonus threshold?

☐ Yes

☐ No

☐ I didn't receive a bonus

Did you feel satisfaction from the fact that your worker was monetarily punished for not meeting the punishment threshold?

☐ Yes

☐ No

☐ My worker reached the punishment threshold

Were you frustrated that you had no say in the punishment decision?

☐ Yes

☐ No

[The following questions were only asked of bosses – discretionary]

At the beginning of the experiment, did you think it was likely that your worker would reach the bonus threshold?

☐ Yes

☐ No

Did this belief effect your punishment cut-off decision?

☐ Yes

☐ No

Looking back, do you believe that your worker was capable of achieving the bonus threshold?

☐ Yes

☐ No

What factors influenced your punishment cut-off decision?

_____

_____

_____

If you did not receive a bonus, were you upset when your worker did *not* reach the bonus threshold?

☐ Yes

☐ No

☐ I received a bonus

If you received a bonus, were you pleased when your worker *did* reach the bonus threshold?

☐ Yes

☐ No

☐ I didn't receive a bonus

Did you feel uncomfortable making the punishment decision?

☐ Yes

☐ No

Would you rather that someone else had set the punishment threshold for you?

☐ Yes

☐ No

Did you believe that punishment would increase your worker's effort level in the next period?

☐ Yes

☐ No

Did you feel satisfaction from the fact that your worker was monetarily punished for not meeting the punishment threshold?

☐ Yes

☐ No

☐ My worker reached the punishment threshold

[The following questions were asked of all participants]

Of the other participants in this session, how many do you know by name? _____

How many of their cell phone numbers do you have? _____

Appendix 9: Thresholds in the discretionary treatment

Though not directly related to the research question, one of the benefits of utilizing the strategy method is the ability to gather data about the punishment decision from all bosses in the discretionary setting regardless of whether or not their workers were punished. Even within one session, the reasons given by bosses on the questionnaire for the thresholds chosen varied greatly:

- 2014: "The worker, I assumed, would want the money enough to try to fill in as many bubbles as possible. I felt imposing a punishment would not be fair to them."
- 2017: "I just set the cut-off at 180 without thinking about whether or not it was actually attainable."
- 2019: "Not very rational. Just a random choice."
- 2020: "Punishing might hurt me in the second round, and had no benefit in the first round, so I chose not to punish."
- 2022: "I didn't want to punish the worker but if they couldn't achieve the bonus, I felt they should pay for my loss of compensation."

I estimate OLS regression models of threshold decisions and boss demographics in Table 9. Specification (1) suggests that boss characteristics have little to do with threshold decisions. However, specification (2) controls for session effects and indicates that some factors are predictive, with higher GPA and higher class year associated with higher thresholds. Sessions 3 and 4 were also significant, suggesting that the thresholds set were not consistent over the three discretionary sessions.

There were two more interesting relationships between boss demographics in the discretionary treatment and questionnaire responses regarding punishment. First, female bosses were more likely to believe that the bonus threshold was achievable than male bosses (OLS regression coefficient $p = 0.043$). Second, female bosses were also more likely to feel uncomfortable setting a punishment threshold (OLS regression

coefficient $p = 0.000$). Though gender was never a significant predictor for the punishment threshold (gender did not influence actions), it was an influence on boss attitudes toward the actions they had to make.

Works Cited

J. Abeler, A. Falk, L. Goette, and D. Huffman. Reference points and effort provision. *American Economic Review*, 101(4):470-492, 2011.

G.A. Akerlof. Labor contracts as partial gift exchange. *The Quarterly Journal of Economics*, pages 534-569, 1982.

G.A. Akerlof and J.L. Yellen. The fair wage-effort hypothesis and unemployment. *The Quarterly Journal of Economics*, 105(2):255-283, 1990.

V. Anderhub, S. Gächter, and M. Königstein. Efficient contracting and fair play in a simple principal-agent experiment. *Experimental Economics*, 5(1):5-27, 2002.

J. Andreoni, W. Harbaugh, and L. Vesterlund. The carrot or the stick: Rewards, punishments, and cooperation. *American Economic Review*, 93(3):893-902, 2003.

J. Berg, J. Dickhaut, and K. McCabe. Trust, reciprocity, and social history. *Games and Economic Behavior*, 10(1):122-142, 1995.

T.F. Bewley. *Why wages don't fall during a recession.* Harvard University Press, 2002.

S. Blount. When social outcomes aren't fair: The effect of causal attributions on preferences. *Organizational Behavior and Human Decision Processes*, 63(2):131-144. 1995.

G.E. Bolton and A. Ockenfels. ERC: A theory of equity, reciprocity, and competition. *American Economic Review*, 90(1):166-193, 2000.

M. Casari and T.N. Cason. The strategy method lowers measured trustworthy behavior. *Economics Letters*, 103(3):157-159, 2009.

G. Charness. Attribution and reciprocity in a simulated labor market: An experimental investigation. *Journal of Labor Economics*, 22(3):665-688, 2004.

J.C. Cox. How to identify trust and reciprocity. *Games and Economic Behavior*, 46(2): 260-281, 2004.

J.C. Cox and C.A. Deck. On the nature of reciprocal motives. *Economic Inquiry*, 43(3): 623-635, 2005.

J.C. Cox, D. Friedman, and S. Gjerstad. A tractable model of reciprocity and fairness. *Games and Economic Behavior*, 59(1):17-45, 2007.

J.C. Cox, K. Sadiraj, and V. Sadiraj. Implications of trust, fear, and reciprocity for modeling economic behavior. *Experimental Economics*, 11(1):1-24, 2008.

D.J.F. de Quervain, U. Fischbacher, V. Treyer, M. Schellhammer, U. Schnyder, A. Buck, and E. Fehr. The neural basis of altruistic punishment. *Science*, 305(5688):1254-1258, 2004.

L. Denant-Boemont, D. Masclet, and C. Noussair. Punishment, counterpunishment, and sanction enforcement in a social dilemma experiment. *Economic Theory*, 33:145-167, 2007.

A. Dreber, D.G. Rand, D. Fudenberg, and M.A. Nowak. Winners don't punish. *Nature*, 452:348-351, 2008.

M. Dufwenberg and G. Kirchsteiger. A theory of sequential reciprocity. Games and Economic Behavior, 47(2):268-298, 2004.

S. Dugar. Nonmonetary sanctions and rewards in an experimental coordination game. *Journal of Economic Behavior & Organization*, 73(3):377-386, 2010.

J. Elster. Norms of revenge. *Ethics*, 100(4):862-885, 1990, July.

A.W. Etzioni. *Modern Organizations*. Englewood Cliffs, NJ: Prentice-Hall, 1971.

A. Falk and U. Fischbacher. A theory of reciprocity. *Games and Economic Behavior*, 54(2):293-315, 2006.

A. Falk and S. Gächter. Experimental Labour Economics. *The New Palgrave Dictionary of Economics, Basingstoke: Palgrave Macmillan*, 2008.

A. Falk and J. Heckman. Lab experiments are a major source of knowledge in the social sciences. *Science* 326:535-538, 2009.

E. Fehr, S. Gächter, and G. Kirchsteiger. Reciprocity as a contract enforcement device: Experimental evidence. *Econometrica*: 65(4):833-860, 1997.

E. Fehr and S. Gächter. Cooperation and punishment in public goods experiments. *American Economic Review*, 90(4):980-994, 2000.

E. Fehr and S. Gächter. Fairness and retaliation: The economics of reciprocity. *The Journal of Economic Perspectives*, 14(3):159-181, 2000.

E. Fehr, G. Kirchsteiger, and A. Riedl. Does fairness prevent market clearing? An experimental investigation. *The Quarterly Journal of Economics*, 108(2):437-459, 1993.

E. Fehr, A. Klein, and K.M. Schmidt. Fairness and contract design. *Econometrica*, 75(1):121-154, 2007.

E. Fehr and K.M. Schmidt. A theory of fairness, competition, and cooperation*. *Quarterly Journal of Economics*, 114(3):817-868, 1999.

E. Fehr and K.M. Schmidt. Fairness and Incentive in a Multi-task Principal-Agent Model. *Scandinavian Journal of Economics*, 106(3):453-474, 2004.

P.S. Green. "NYC Investigating Reports of Work Slowdown During Snow Removal." *Business Week*, January 3, 2011. http://www.businessweek.com/news/2011-01-03/nyc-investigating-reports-of-work-slowdown-during-snow-removal.html

B. Greiner. An online recruitment system for economic experiments. In Kurt Kremer and Volker Macho, editors, *Forschung und wissenschaftliches Rechnen 2003*, volume 63 of *GWDG-Bericht*, pages 79-93. Gesellschaft für wissenschaftliche Datenverarbeitung mbh, Göttingen, Germany, 2004.

L. Hannan, V. Hoffman, A. Rapoport. Bonus versus reality: Does the contract frame affect employee effort? *Experimental Business Research*, 3:151-169, 2005.

G. Harrison and K. McCabe. The role of experience for testing bargaining theory in experiments. *Research in Experimental Economics*, 5:127-169, 1993.

D. Houser and E. Xiao. Inequality seeking punishment. *Economics Letters*, 109(1):20-23, 2010.

D. Kahneman and A. Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263-292, 1979.

J.A. List. The behavioralist meets the market: Measuring social preferences and reputation effects in actual transactions. *Journal of Political Economy*, 114(1):1-36, 2006.

W. Bentley MacLeod and D. Parent. Job characteristics and the form of compensation. *Research in Labor Economics, Volume 18* (eds. Professor Solomon Polachek, Dr Konstantinos Tatsiramos). Emerald Group Publishing Limited, 177-242, 2000.

J. Neelin, H. Sonnenschein, and M. Spiegel. A further test of noncooperative bargaining theory: Comment. *American Economic Review*, 78:824-836, 1988.

N. Nikiforakis. Punishment and counter-punishment in public good games: Can we really govern ourselves? *Journal of Public Economics*, 92(1-2):91-112, 2008.

T. Offerman. Hurting hurts more than helping helps. *European Economic Review*, 46(8):1423-1437, 2002.

E. Ostrom, J. Walker, and R. Gardner. Covenants with and without a sword: Self-governance is possible. *American Political Science Review*, 86(2): 404-417, 1992.

L. Petrecca. "More companies turn to furloughs to save money, jobs." *USA Today*, March 6, 2009. http://www.usatoday.com/money/economy/employment/2009-03-04-furloughs-workers-companies_N.htm

C. Prendergast. The provision of incentives in firms. *Journal of Economic Literature*, 37(1):7-63, 1999.

M. Rabin. Incorporating fairness into game theory and economics. *The American Economic Review*, 83)5:1281-1302, 1993.

L. Ragland, L. Kersting, and N. Stuart. Contract preferences and effort choice under bonus, penalty, and hybrid incentive contracts. University of Wisconsin-Oshkosh Working Paper, 2010.

J. Sobel. Interdependent preferences and reciprocity. *Journal of Economic Literature*, 43(2):392-436, 2005.