UC Santa Barbara
UC Santa Barbara Electronic Theses and Dissertations

Title
Essays in Behavioral Labor and Health Economics

Permalink
https://escholarship.org/uc/item/50g5h00t

Author
Cooper, Michael Thomas

Publication Date
2021

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

University of California
Santa Barbara

Essays in Behavioral Labor and Health Economics

A dissertation submitted in partial satisfaction
of the requirements for the degree

Doctor of Philosophy
in
Economics

by

Michael Thomas Cooper

Committee in charge:

Professor Peter Kuhn, Co-Chair
Professor Ryan Oprea, Co-Chair
Professor Gary Charness

March 2021

The Dissertation of Michael Thomas Cooper is approved.

Professor Gary Charness

Professor Ryan Oprea, Co-Chair

Professor Peter Kuhn, Co-Chair

March 2021

Acknowledgements

First and foremost I would like to thank my committee members for their invaluable guidance and support over the years. I am grateful to Peter Kuhn for his formative influence on my research interests – his passion for personnel economics directly led to the development of my own research agenda. I thank him for consistently providing detailed feedback on my projects which not only improved my papers but also taught me how to write better papers. I am thankful for his mentoring style which always left me feeling that the road to improvement was clearly laid out ahead.

I am grateful to Ryan Oprea for teaching me the big picture of conducting research. I am indebted to him for several breakthroughs which allowed me to see my projects through to the end. I thank him for shaping my taste in what makes a "good" paper, refining my writing style, and participating in countless discussions about my research which drastically improved my ability to design experiments and present results in a compelling way.

I am grateful to Gary Charness for his encouragement and guidance. Discussing ideas with him is truly a pleasure and his creativity and cleverness in designing experiments is remarkable. I also thank him for providing candid, practical advice about publishing and the job market in a supportive way.

I would also like to thank the faculty at UCSB. In particular, Erik Eyster and Sevgi Yuksel have provided valuable feedback on my projects. Sevgi was very helpful during the job market. I also thank Doug Steigerwald for teaching me econometrics in a way that finally clicked with a strong helping of warmth and encouragement.

I am thankful for my colleagues in the department and the members of my cohort for providing a supportive environment. I am particularly grateful for Max Lee, Sarah Bana, and Lucas Reddinger for stimulating and productive conversations about my research.

Lastly, this undertaking would have been impossible without the support and sacrifice of my parents and my girlfriend Eyvette. Their unconditional love and emotional support throughout my graduate education was indispensable for me to finish the degree. I am deeply grateful to have them in my life and thank them for helping me get through the bad times and cheering me through the good times.

Curriculum Vitæ
Michael Thomas Cooper

CONTACT INFORMATION

Address:	North Hall 2037, Department of Economics, University of California Santa Barbara, Santa Barbara, CA 93106
Email:	`michael_cooper@ucsb.edu`
Website:	`http://www.michaelthomascooper.com`

EDUCATION

2021	Ph.D. in Economics (Expected), University of California, Santa Barbara.
2014	M.A. in Economics, University of California, Santa Barbara.
2011	B.A. in Economics and Political Science, Loyola Marymount University

RESEARCH AND TEACHING INTERESTS

Labor Economics; Behavioral and Experimental Economics; Health Economics

RESEARCH

Job Market Paper

"Dismissal Threats as an Incentive Device."

Abstract: This paper explores dismissal threats as an incentive device in a laboratory experiment. Workers earn a fixed wage per period and complete real effort tasks to reduce their chance of being fired at the end of each period. Behavioral motivators are purposefully activated in addition to these monetary incentives. The design innovates on previous literature by implementing dismissal threats in a quantifiable way and by collapsing preference elicitation over incentives along with random assignment to incentives into the same round. The experiment produces two main results. First, dismissal threats are more motivating than comparable piece rates. Workers produce significantly more output under dismissal threats than they do under piece rates, even though the marginal benefit of output is lower. Second, the productivity gains from strengthening dismissal threats on the margin have a large self-selection component but significant heterogeneity in pure incentive effects. Workers who prefer higher pay with steeper dismissal threats appear to respond positively to this environment, but these high-pressure incentives backfire among workers who wanted to avoid them.

Refereed Journal Articles

Cooper, Michael, and Michael Pesko. 2017. "The Effect of E-Cigarette Indoor Vaping Restrictions on Adult Prenatal Smoking and Birth Outcomes." *Journal of Health Economics*, 56: 178-190.

Edited Volumes

Charness, Gary, Michael Cooper, and Lucas Reddinger. 2020. "Wage Policies, Incentive Schemes, and Motivation." In: Zimmerman K. (ed) *Handbook of Labor, Human Resources, and Population Economics*. Springer, Cham.

Cooper, Michael, and Peter Kuhn. 2020. "Behavioral Job Search." In: Zimmerman K. (ed) *Handbook of Labor, Human Resources, and Population Economics*. Springer, Cham.

Manuscripts in Submission

"The Effect of E-Cigarette Indoor Vaping Restrictions on Infant Mortality" with Michael Pesko. Revise and resubmit at *Southern Economic Journal*.

Other Manuscripts in Preparation

"Self-Image in Job Search: An Experimental Investigation of Biased Updating and Information Avoidance in Search Decisions."

EXTERNAL PRESENTATIONS

2021	(Scheduled) The Western Economics Association International. "Dismissal Threats as an Incentive Device."
2020	The Economic Science Association Job Market Candidate Series. "Dismissal Threats as an Incentive Device."
2019	The Economic Science Association North American Meeting. "Self-Image in Job Search: An Experimental Investigation of Biased Updating and Information Avoidance in Search Decisions."
2017	The Western Economics Association International, Session on Smoking, Health, Illicit Trade, and Public Policy. "The Effect of E-Cigarette Indoor Vaping Restrictions on Adult Prenatal Smoking and Birth Outcomes."
2017	The Choice Lab and Rady School of Management, Poster Presentation. "Dismissal Threats and Worker Effort: Sorting vs. Incentives."

HONORS AND AWARDS

2018	Economics Department "Grad Slam" Presentation Competition, 2nd Place

2018	Outstanding Teaching Assistant of Winter Quarter
2013-2014	Andron Fellowship, University of California, Santa Barbara

PROFESSIONAL EXPERIENCE

Teaching Assistant

Personnel Economics, Intermediate Microeconomic Theory (Head TA), Principles of Macroeconomics, Introduction to Economics

Teaching Credentials

Teaching Online Certificate from Quality Matters

Research Assistant

2016-2019	Experiment programming and running laboratory sessions for Gary Charness (University of California, Santa Barbara).
2019	Running laboratory sessions for Muriel Niederle (Stanford University).
2017-2018	Textbook editing services for Peter Kuhn (University of California, Santa Barbara).
2017	Running laboratory sessions for Chad Kendall (University of Southern California).

PROFESSIONAL SERVICE

Journal Referee

Journal of Labor Economics

Journal of Economic Behavior and Organization

Professional Memberships

American Economic Association

Western Economic Association International

Economic Science Association

COMMITTEE MEMBERS

Peter Kuhn (Co-Chair), Ryan Oprea (Co-Chair), Gary Charness

Permissions and Attributions

The content of Chapter 3 is the result of a collaboration with Michael F. Pesko and has previously appeared in *The Journal of Health Economics*. The journal allows the inclusion of previously published work in a dissertation with proper attribution. The published article can be found at: `https://doi.org/10.1016/j.jhealeco.2017.10.002`

Cooper, Michael T., and Michael F. Pesko. "The effect of e-cigarette indoor vaping restrictions on adult prenatal smoking and birth outcomes." *Journal of Health Economics* 56 (2017): 178-190.

Abstract

Essays in Behavioral Labor and Health Economics

by

Michael Thomas Cooper

This dissertation consists of three essays that analyze the impact of behavioral biases on labor market and health outcomes. The essays use tools from both experimental economics and applied econometrics. The common thread that runs through this research agenda is the goal of understanding how biases and suboptimal behaviors impact long-term outcomes crucial to well-being: labor market and health outcomes.

The first essay asks whether dismissal threats are more motivating than other types of incentives. In a laboratory experiment, workers earn a fixed wage per period and complete real effort tasks to reduce their chance of being fired at the end of each period. Behavioral motivators are purposefully activated in addition to monetary incentives. The design innovates on previous literature by implementing dismissal threats in a quantifiable way and by collapsing preference elicitation over incentive along with random assignment to incentives into the same round. The experiment produces two main results. First, workers produce significantly more output under dismissal threats than they do under piece rates, even though the marginal benefit of output is lower. Second, the productivity gains from strengthening dismissal threats on the margin have a large self-selection component but significant heterogeneity in pure incentive effects. Workers who prefer higher pay with steeper dismissal threats appear to respond positively to this environment, but these high-pressure incentives backfire among workers who want to avoid them.

The second essay implements a lab experiment to investigate the effects of self-image concerns on search behavior. Subjects play a simple sequential search game in which

they decide how many times to search for a wage offer before giving up. Feedback from search contains both instrumental information about search prospects and signals about subjects' relative performance on an intelligence test taken earlier in the experiment. Treatments isolate and shut down two mechanisms: biased belief updating and information avoidance. Despite replicating results from the literature on overconfidence in incentivized reporting of initial beliefs, subject search behavior does not differ between treatments with or without self-image concerns during search. These results seem to suggest that people are more likely to state overconfident beliefs when these beliefs are directly elicited, but that people act much closer to the rational Bayesian benchmark when actions only indirectly reveal self-relevant beliefs.

The third essay is joint work with Michael F. Pesko. We estimate the effect of county-level e-cigarette indoor vaping restrictions on adult prenatal smoking and birth outcomes using United States birth record data for 7 million pregnant women living in places already comprehensively banning the indoor use of traditional cigarettes. We use both cross-sectional and panel data to estimate our difference-in-difference models. Our panel model results suggest that adoption of a comprehensive indoor vaping restriction increased prenatal smoking by 2.0 percentage points, which is double the estimate obtained from a cross-sectional model. We also document heterogeneity in effect sizes along lines of age, education, and type of insurance.

Contents

Chapter 1

Dismissal Threats as an Incentive Device

1.1 Introduction

The threat of being fired for poor performance is a ubiquitous incentive throughout the modern economy. Workers who wish to retain their wages and benefits from employment must provide enough visible effort to avoid dismissal and a potentially costly period of involuntary unemployment. A recent poll conducted during a time of economic expansion and low unemployment found that 48% of employed Americans experienced anxiety over the possibility of losing their job, and 40% of Americans have been laid off or terminated from a job at least once (Harris Poll, 2019). Despite the prevalence of firing threats in the labor market, little is known about how strongly they motivate workers to provide effort compared to other monetary incentives.

This paper reports the results of an experiment which explores dismissal threats as a rigorously quantified incentive device in the lab. Subjects complete a real effort task for pay under various types and strengths of incentives. Under dismissal threats, subjects receive a fixed payment per period and complete tasks to reduce their chance of being fired at the end of each period. This implementation of dismissal threats in the lab as a known probabilistic relationship between output and the chance of being fired brings tractability to an incentive that is often ambiguous and not explicitly quantified in labor market contracts in the field. An important benefit of this approach is that the marginal benefit of output under dismissal threats is known to workers and observable to the researcher. This tractability allows direct comparisons of worker output under dismissal threats to output under simpler monetary incentives such as piece rates.

Few jobs are entirely devoid of firing threats, but in practice the strength of this incentive varies considerably. Some jobs, such as executive leadership of large public corporations, have high employer-initiated turnover rates – CEOs are often terminated based on poor performance (e.g., stock value). Other jobs seem to have much weaker dismissal threats, with a feature of strong job security as long as lenient baseline levels of performance are met. Jobs with unionized workforces have much higher levels of job security, as do jobs in the public sector (Freeman and Medoff, 1984; Farber, 2010). Variation in the strength of dismissal threats also occurs based on policies, including employment protection laws and court rulings (e.g., statewide "just cause" doctrines studied by Autor, Kerr, and Kugler, 2007).

An important benefit of studying dismissal threats in the lab is the ability to produce exogenous variation in the strength of the incentive. In the field, how motivating dismissal threats are to workers depends on a number of factors, including not only the relationship between output and the chance of being fired, but also considerations such as economic conditions. Indeed, research has shown that workers increase output when unemployment would be more financially painful due to a recession or local labor market conditions (Cappelli and Chauvin, 1991; Lazear, Shaw, and Stanton, 2016). The lab environment will abstract away from these concerns, simplifying the monetary incentives of dismissal threats to merely losing income from all remaining work periods after a once-and-for-all firing has occurred. In this design, the incentives under firing threats will be characterized by three parameters: the fixed wage per period and the slope and intercept of a linear function that relates output to the probability of being fired. Exogenous variation in these parameters allows analysis of worker behavior under steeper versus flatter dismissal threats.

Surprisingly, little research has been conducted on how firing threats operate to motivate workers. Instead, the literature has frequently analyzed financial incentives like piece rates, tournaments, and teams. Although these incentives serve as important tools for employers, they seem less common in practice than fixed wages (hourly wages or salaries) combined with dismissal threats. Piece rates could be less common than dismissal threats for a variety of reasons, including the difficulty of measuring individual worker output or the fear of workers neglecting unincentivized criteria at the cost of focusing only on measured, incentivized tasks. However, one potential reason stands out as particularly interesting to behavioral and personnel economists: the threat of being fired combines neoclassical and behavioral motivators in a way that could be uniquely powerful in motivating worker effort.

Consider the mixture of neoclassical and behavioral mechanisms involved when using firing threats to motivate workers. Most obviously, being fired means earning less money, causing a direct utility loss. Whether one will be fired is usually uncertain until it happens, thus bringing risk aversion into consideration. However, the widespread fear and anxiety over being fired (as mentioned in the polling data above) does not seem congruent with what would be caused under piece rates or even simple monetary incentives with risk involved. Instead, it seems likely that deeper behavioral and even evolutionary motivations are at play, such as the fear of exclusion from a social group. Fired workers may become jealous of retained coworkers who continue to earn money and perform interesting work. They may feel embarrassed about public announcements of their firing, or experience negative self-image impacts resulting from the signal that firing provides about their own value in the labor market. Loss aversion likely plays a role when keeping one's job is the status quo and being fired feels like a large negative change compared to prior expectations.

These behavioral motivators are purposefully activated in the experiment design because they seem likely to play a large role in why firing threats are motivating. The event of being fired is framed as a loss throughout the instructions and work interface, and public announcements of firings are made at the end of each work period.[1] Fired workers have their outside option tightly controlled; they must simply wait while other workers continue to earn money, further developing feelings of exclusion. Given that it would be impossible to activate these behavioral mechanisms in the lab at the full strength that they exist in the real labor market, the estimates presented in this paper of the motivating effect of dismissal threats are likely a lower bound on their effect in the field.

The results of the experiment shed light on the use of dismissal threats to motivate workers. The main results fall into two categories: the effect of dismissal threats on the extensive margin (the existence of credible dismissal threats) and on the intensive margin (intensifying existing dismissal threats). On the extensive margin, dismissal threats are shown to be highly motivating compared to piece rates. The experiment is designed such that the marginal benefit of output under dismissal threats is always lower than it was under the high piece rate work period. Despite the lower monetary benefit, mean productivity in the first dismissal threat work period is 23.6% higher than under the piece rate, a highly statistically significant difference. Parametric analysis shows that

[1] Subjects are assigned random worker identification numbers to retain their anonymity; these worker IDs are displayed in the public announcements.

this cannot be caused merely by the levels of risk aversion displayed in subject decisions over gambles; instead, the unique mixture of monetary and behavioral motivators invoked by the dismissal threats in the lab seems to be highly motivating in inducing worker effort.

The implications of this result are both prescriptive and descriptive. First, firms may experience large productivity gains by introducing credible dismissal threats that did not exist before (or were too weak to be salient to workers). Second, this may be an important reason why dismissal threats are so widespread in the modern labor market: they are more motivating than simpler monetary incentives like piece rates.

On the intensive margin, steeper dismissal threats combined with higher pay result in higher output than flatter dismissal threats with lower pay, as expected. However, the design allows further decomposition of this productivity increase into the pure incentive effect (steeper incentives result in higher optimal output) and the selection effect (more productive workers self-select into steeper incentives). The design elicits subject preferences over incentive schemes while retaining random assignment to those schemes, all collapsed into one round. This both saves time and avoids order contamination effects compared to previous approaches in the literature. With both subject preferences and random assignment, a clean decomposition of incentive effects from selection effects is possible.

The qualitative results of this decomposition of incentive versus selection effects are informative for both firms and policymakers. If the incentive effect dominates, then imposing stronger dismissal threats on an existing workforce would cause large short-term productivity gains; if the selection effect dominates, productivity would increase only gradually as the composition of workers changes over time. This decomposition is also relevant to policy debates over employment protection and productivity. Taking an extreme example, when all firms offer the same level of dismissal threats, the selection effect on productivity is entirely removed – there must be different levels of dismissal threats and pay for this effect to operate. Following this logic, widespread employment protection policies are predicted to have muted effects on productivity if the selection effect is a large component of the total effect, but would cause large declines in productivity if the incentive effect is large.

It turns out that the aggregate productivity gains on the intensive margin (strengthening existing dismissal threats) operate largely through selection effects. The workers who indicated a preference for the high-paying job with stronger dismissal threats were significantly more productive than workers with the opposite preference. Aggregate incentive effects were muted: random assignment to the steeper incentive scheme was not associated with significantly higher output.

However, these aggregate effects mask important heterogeneity in incentive effects. Among workers who preferred higher pay with stronger dismissal threats, there is suggestive evidence that they do increase output under these steeper incentives (although it is not statistically significant). On the other hand, workers who were willing to accept lower pay to avoid strong dismissal threats actually produce significantly *fewer* units of output when randomly assigned to the strong dismissal threats. This negative incentive effect of steeper incentives can be explained in line with Roy's (1951) model of worker self-selection: these are precisely the workers who expect to do the worst in this high-pressure environment. Indeed, 85% of this subset of subjects would end up with greater than a 50% chance of being fired even if they sustained their level of output from the high piece rate work period. The takeaway is that imposing steeper dismissal threats may be demotivating if they are too strong compared to the ability level of the worker – an important lesson for firms considering strengthening this form of incentive on a mixed-ability workforce.

The rest of the paper will be organized as follows. Section 1.2 will discuss the related literature and the unique contributions of this paper compared to previous work. Section

1.3 will provide details about the experimental design. Section 1.4 will present the results, and Section 1.5 will conclude.

1.2 Related Literature

This paper contributes to several different strands of economics literature, but it most directly relates to the personnel economics literature on motivating workers to provide effort. Early contributions to this literature established the principal-agent model and showed how providing piece rates can overcome the moral hazard issue facing employers trying to motivate their employees to provide costly effort (Ross, 1973; Stiglitz, 1975). Much of the subsequent experimental literature has focused on comparing piece rates to other payment schemes such as tournaments or team-based revenue sharing (Bull, Schotter, and Weigelt, 1987; Niederle and Vesterlund, 2007; Dohmen and Falk, 2011).

Applied work has also made many important contributions to understanding how motivating different types of incentives are for workers. In a seminal paper, Lazear (2000) studies the shift from hourly wages to piece rates in a large windshield repair company. Making this study even more relevant to the current design, Lazear also uses econometric methods (worker fixed effects) to try to decompose the productivity gains into pure incentive effects and selection effects. He finds that roughly half of the productivity gains come from the selection effect of hiring and retaining more productive workers. However, he cannot rule out that changes in management coinciding with the incentive changes may confound some of these measurements. Indeed, this is a perennial issue in applied work: identifying policy changes in a firm or government that are truly exogenous. It can also be difficult to find changes to employment contracts which only impact one part of the incentive structure.

Jacob (2013) produces the applied work most relevant to this experiment by studying a policy change in a Chicago public school district that made dismissing certain teachers with limited years of experience easier. He measures productivity in the number of days of teacher absences. Similar to Lazear (2000), Jacob uses worker fixed effects to try to isolate the selection effect on productivity. He finds that changes to the composition of teachers (i.e., the selection effect) is the predominant cause of the post-policy reductions in teacher absences, and finds much more limited evidence of pure incentive effects. In other work, Jacob (2011) shows that after this policy, principals are more likely to fire teachers with more absences and lower value-added scores. Although the policy he analyzes in Chicago provides a plausibly exogenous reduction in employment protection, it is impossible to cleanly quantify and manipulate the strength of dismissal threats in the field in the way one can in a controlled laboratory experiment. Other advantages of an experiment include the ability to observe the same workers under a variety of incentive schemes (e.g., both piece rates and dismissal threats), and the possibility of measuring many worker traits with incentivized tasks.

Kwon (2005) answers a similar but distinct question: When a company decides to fire a worker, was it done to incentivize other workers or because the fired worker was a poor match for the job? He develops a model that shows each explanation predicts an opposite relationship between tenure and dismissal probability. Using proprietary data from a large U.S. insurance company, he tests his theoretical predictions and finds evidence that workers are dismissed as a way of providing incentives for effort, rather than because they were a bad match for the job. This conclusion of course depends on various assumptions of the model, and differs from the primary questions of the present design: how strongly motivating dismissal threats are as an incentive device, and whether the productivity gains from exogenous changes in dismissal threats are primarily due to selection or incentive effects.

Other researchers have conducted important applied work on the effects of employment protection and productivity. Autor, Kerr, and Kugler (2007) exploit the variation in state court cases that established just cause doctrines and other exceptions to at-will employment. They find that this trend in employment law is associated with capital deepening and declines in total factor productivity in manufacturing plants, but caution that the findings are “suggestive but tentative” due to concurrent increases in employment. Acemoglu, Daron, and Angrist (2001) find a link between a major employment protection law for disabled people and subsequent reductions in their employment rate.

A number of experiments have alternatively allowed subjects to select into their preferred type of incentives with certainty, or randomly assigned subjects an incentive scheme. For example, Dohmen and Falk (2011) run an important experiment on multi-dimensional sorting of workers into fixed wages versus variable pay (i.e., payments that depend on output: piece rates, tournaments, or teams). However, because subjects are guaranteed the incentive scheme they choose, the authors are unable to precisely decompose the productivity gains under variable pay into incentive and selection effects. Instead, they are merely able to assert that selection on productivity exists, since subjects who selected the variable pay schemes were more productive on average in an earlier piece rate work period.

Other examples of experiments with guaranteed selection into incentive schemes include Cadsby, Song, and Tapon (2007) who examine piece rate contracts, and Leuven, Oosterbeek, Sonnemans, and van der Klaauw (2011) who examine selection into different tiers of tournaments. Fehrenbacher, Kaplan, and Pedell (2017) have two treatments: one with random assignment of incentive scheme, and one with guaranteed assignment to the selected incentive scheme. However, because they do not elicit the preferences of subjects who are randomly assigned, they cannot compare how well subjects with different job preferences would have performed under alternative incentive schemes.

The design reported in this paper innovates on this strand of literature on selection effects by eliciting subject preferences over incentives in an incentive-compatible way while retaining random assignment to incentive schemes, all within one round. This allows the observation of subjects in the incentive scheme they did not select, thus providing a clean decomposition of selection versus sorting effects. Although past work has both elicited subject preferences over incentives and imposed random assignment to those incentives in different rounds (e.g., Niederle and Vesterlund, 2007), this requires each subject to work under the same incentive scheme multiple times. The drawbacks to this approach include potential contamination by order effects and reducing the amount of time available for other experimental tasks. Importantly, in the context of dismissal threats, it would also remove the behavioral impact of the "once-and-for-all" nature of being fired.

Some experimental work has directly implemented dismissal threats in the lab. Corgnet, Hernan-Gonzalez, and Rassenti (2015) create groups of subjects who are randomly assigned to be bosses or employees, allowing the boss to dismiss one employee at the end of each work period. They do not examine selection effects, but find that employees increase their productivity and engage in "impression management" when dismissal threats are introduced (i.e., they spend less time on real leisure and more time on real effort tasks when both are visible to the boss). Their design differs from the one in this proposal because with human bosses and employee counterparts, the strength of the firing threat cannot be known or quantified by employees. Thus perceptions (or misperceptions) of the strength of dismissal threats may drive differences in worker behavior. Additionally, when examining the question of how motivating dismissal threats are in increasing worker output, they only compare dismissal threats to a situation with no monetary incentives at all, rather than to other kinds of monetary incentives such as piece rates.

Other experimental work seems to uncover similar behavioral forces to those analyzed here: the discontinuous effect of introducing a fear of exclusion or financial survival into an economic decision-making problem. For example, Kopányi-Peuker, Offerman, and

Sloof (2018) show that contributions greatly increase in a weakest link game when a designated manager is able to permanently exclude subjects with low contributions from the team. Oprea (2014) finds strong evidence that subjects suboptimally hoard excess cash to improve their odds of avoiding bankruptcy. It seems likely that similar behavioral forces are activated in the fear of being fired.

Finally, this paper is tangentially related to a few other literatures in labor economics. First, it is related to the literature on efficiency wages (Shapiro and Stiglitz, 1984). It is more costly to lose a job with higher pay, and this effect operates in this experiment as well. Even holding constant the probability of dismissal, increasing pay would be predicted to increase optimal output for this reason: there is more to lose by becoming unemployed. Second, it is related to the literature on deferred compensation (Lazear, 1979; Kuhn, 1986). The idea is that workers are paid below their marginal product when they are young, and then they collect rents above their marginal product when they are old. A similar force operates in the experiment because subjects can earn full job security in the final period and shirk without consequence – this is a motivating factor to provide higher effort in earlier periods.

Third, this paper is somewhat related to the literature on relational contracting in terms of the distinction between contracts which are locked in and repeated versus contracts which must be actively renewed after each transaction. The design in this paper has actively chosen a loss framing for dismissal (i.e., subjects are told they will "lose" their job and will not be able to earn more money). This stands in contrast to the alternative of temporary job contracts which are expected to be terminated unless actively renewed (more common in certain European countries with strong employment protection; see Cahuc and Postel-Vinay, 2002). Cromwell, Goerg, and Leszczynska (2018) show that this distinction can have an important impact on behavior in the lab even when it provides no tangible difference in flexibility or incentives. Therefore, the decision to frame dismissal as a loss of a contract compared to the status quo could have important implications for the power of dismissal threats as an incentive device in the present design.

1.3 Experiment Design

This section will explain the experimental design and the motivation behind various design decisions. The primary purpose of the experiment was to expose subjects to various types and strengths of incentives and observe their resulting productivity in a real effort task. The subject pool consisted primarily of students at the University of California, Santa Barbara. The experiment was run entirely online, with instructions read over a Zoom video chat conference and payments made through the mobile application Venmo.[2] The experiment software was coded in oTree (Chen, Schonger, and Wickens, 2016).

The real effort task involved counting shapes that appear in randomly generated images. This shape-counting task is largely the same real effort task used by Caplin, Csaba, Leahy, and Nov (2020). A primary strength of this particular geometric reasoning task is that Caplin et al. show that subject performance is responsive to different levels of monetary incentives. Additionally, it is culturally neutral and relatively simple to explain to subjects. The task was slightly modified from the implementation of Caplin et al. to disincentivize guessing and to provide more granular data on worker output. The difficulty level of the task did not vary throughout the experiment. Additionally, to increase variation in subject productivity, real leisure was included in the work interface

[2]Subjects initially presented identification in a private breakout room, but were muted with their cameras off for the remainder of the experiment. Thus the anonymity of subjects from one another was maintained. Venmo is ubiquitous among the subject pool; all subjects agreed to accept payment through this app before participating in the experiment.

in the form of interesting facts (e.g., about animals or astronomy).[3] A more detailed description of this task along with a screenshot of the work interface are provided in the appendix.

After the initial instructions, subjects proceeded through a series of work periods, each carefully designed to answer questions about dismissal threats as an incentive device. First, subjects completed a two-minute unpaid period to familiarize themselves with the task.[4] Next, subjects were told they would be paid per task completed in two successive five-minute work periods: the first period with a piece rate of $0.01 per task, and the second period with a $0.20 payment per task. These two pay rates were displayed saliently before either piece rate period began so subjects would know that their effort was significantly more valuable in one period than another.

Following the piece rate work periods, there were three work periods under dismissal threats, each lasting 10 minutes. Dismissal threats were implemented as a linear relationship between output and the probability of being fired at the end of each remaining work period. Subjects were paid a fixed wage up front at the start of each period. Given that there was no fourth dismissal threat period, the third period acted as a simulation of full job security; trivially, optimal output in the final period is zero. The monetary incentives of dismissal threats thus can be fully characterized by three parameters: the payment per period along with the intercept and slope of the linear firing function. Because these parameters are known to the experimenter and the subjects, the marginal benefit of output can be calculated in the first and second dismissal work periods, making direct comparisons with piece rates possible.

Workers are assigned to one of two jobs, displayed in Figure 1.1. In the Safe Job, the

[3]A silver lining of running the experiment online is that subjects could not be prevented from using their phones during the work periods – another form of real leisure that increases variation in output.

[4]Subjects were also shown a series of tutorial screens with screenshots of the work interface. The tutorial and unpaid practice period are important to reduce the potential effects of learning to do the task more efficiently.

pay per period is \$2.50, the firing-chance intercept is 50%, and the slope is a 1% lower chance of being fired per task completed. In the Risky Job, these parameters are \$5.00, 100%, and 2%, respectively.[5]

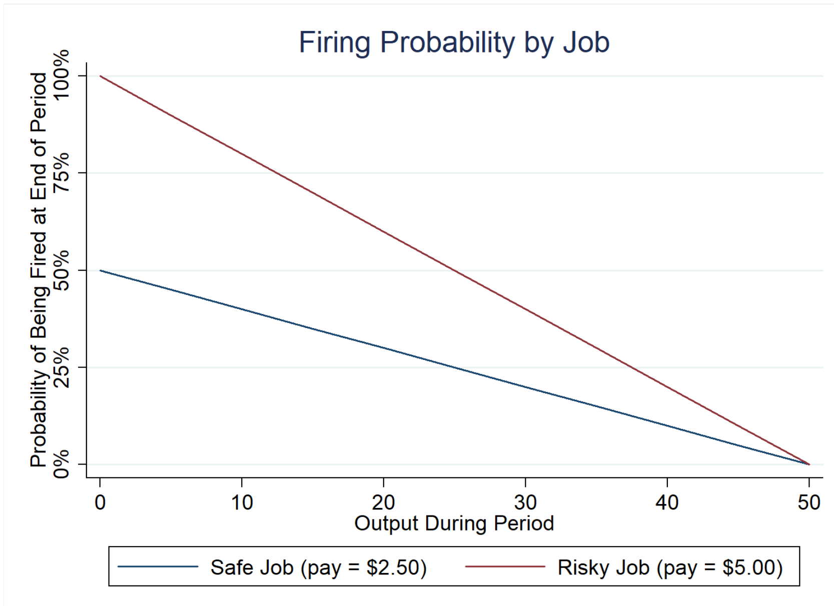


Figure 1.1: Firing Probability by Job

Crucially, the two piece rates of \$0.01 and \$0.20 were chosen as the lower and upper bounds on the marginal benefit of output under dismissal threats. This allows the design to answer the following research question:

Question 1: Extensive Margin. Are dismissal threats more motivating than piece rates with comparable monetary value?

Given that the marginal benefit of output under dismissal threats is always lower than \$0.20, one would expect that worker output under dismissal threats is lower than output under the piece rate of \$0.20. If output under dismissal threats is higher than

[5]Although the two parametrizations will be called the Safe Job and the Risky Job throughout the text, they were only referred to with neutral language ("Job 1" and "Job 2") in the experiment instructions and software.

output under this piece rate, it would be strong evidence that dismissal threats are a more motivating type of incentive than piece rates.

The marginal benefit of output in each period and job is calculated in Table 1.1 by taking the derivative of expected pay with respect to output in periods 1 and 2.[6]. Of course, the marginal benefit of output in period 1 depends on the level of output in period 2; but across the whole range of potential period 2 outputs (from 1 to 50), the marginal benefit of output in period 1 remains between \$0.01 and \$0.20. The table confirms that under any job-period combination, and while incorporating the option value of being retained after period 1, the marginal benefit of output under dismissal threats is always lower than \$0.20.

Table 1.1: Marginal Benefit of Output during Dismissal Periods

	Marginal Benefit of q_1	Marginal Benefit of q_2
Risky Job	$\frac{\partial \mathbb{E}[\text{Pay}]}{\partial q_1} = 0.01 + 0.002q_2$ if $q_2 = 0 \rightarrow$ **\$0.01** if $q_2 = 50 \rightarrow$ **\$0.11**	$\frac{\partial \mathbb{E}[\text{Pay}]}{\partial q_2} =$ **\$0.10**
Safe Job	$\frac{\partial \mathbb{E}[\text{Pay}]}{\partial q_1} = 0.0375 + 0.00025q_2$ if $q_2 = 0 \rightarrow$ **\$0.0375** if $q_2 = 50 \rightarrow$ **\$0.05**	$\frac{\partial \mathbb{E}[\text{Pay}]}{\partial q_2} =$ **\$0.025**

Of course, dismissal threats constitute more than simply risky payments, so behavioral motivators are purposefully activated by the design as well. The instructions framed retention as the status quo and being fired as an event which caused a subject to miss out on payments. Being retained was described in green text while being fired was always displayed in bold red text. To additionally activate social and self-image concerns,

[6]Note that the marginal benefit of output in period 2 is calculated from the perspective of period 2 (i.e., for a worker who has already been retained). Trivially, output in period 3 never impacts total pay, so that period is excluded from the analysis.

public announcements of firing decisions were made at the end of each work period. These announcements emphasized that retained workers would continue to earn money. Subjects were kept anonymous through the use of randomly assigned worker identification numbers, so examples of these notifications would be "Worker 3 was fired!" in bold red text, or "Worker 7 was not fired! They will continue to earn money next period." in green text.[7]

The outside option of workers was tightly controlled to prevent uncontrolled variation in the cost of being fired and to further induce feelings of exclusion among fired workers. Fired workers had to complete a boring filler task for any remaining work periods: clicking a large button when it occasionally turned red to avoid losing a small amount of money.[8] This control over the outside option increases the internal validity of the design by preventing subjects from attempting to get fired on purpose in order to engage in other work or leisure.

Further elements of the experiment were designed to analyze the intensive margin effects of strengthening dismissal threats:

Question 2.1: Intensive Margin. Do stronger dismissal threats result in higher output than weaker dismissal threats?

The parameters of the Risky Job are predicted to be significantly more motivating than the parameters of the Safe Job. Crucially, the slope of the firing function in the Risky Job is twice as steep, with each unit of output reducing the chance of being fired by 2% rather than 1% in the Safe Job. Also, since the pay per period is higher in the Risky Job, subjects have more to lose by being fired.

[7]Subjects were shown a screenshot of an example firing announcement ahead of time so that social/self-image concerns would be activated in the first dismissal work period.

[8]The button-clicking task was intentionally designed to be extremely easy and merely required subjects to stay attentive at their computer. Exceedingly few dismissed subjects ever missed a button click, and most of those subjects missed only a single click.

The answer to the above question is largely predictable, but the design seeks to further analyze these intensive margin productivity gains:

Question 2.2: Incentive and Selection Effects. Do stronger dismissal threats increase output primarily through the incentive effect or the selection effect?

The incentive effect is the straightforward impact of steeper incentives increasing the optimal level of output. This means that exogenous assignment to the Risky Job should result in higher levels of output. The selection effect is that more productive workers tend to choose the Risky Job while less productive workers tend to choose the Safe Job.[9] Although the precise quantitative decomposition of these effects will not be representative of the labor market, the qualitative results (i.e., which effect is larger) have important implications for firms and policymakers, as discussed in the introduction.

The design allows for a clean decomposition of the incentive and selection effects by eliciting worker preferences over jobs while still retaining a random element to job assignment. Specifically, subjects have a 60% chance of receiving the job they prefer. Because they are more likely to receive their preferred level of pay and dismissal threats, this mechanism retains incentive compatibility – subjects should honestly report their preference. The mechanism also imposes exogenous variation in the strength of dismissal threats, allowing the isolation of the effect of random assignment to a steeper or flatter firing function. This mechanism thus provides significant benefits over implementations in previous experiments, which have opted for either guaranteed assignment to one's preferred incentive scheme or for fully random assignment. Simple regression analysis can decompose the intensive margin productivity gains of steeper dismissal threats into

[9] Note that under the implemented parameters, the expected total pay in the Risky Job will be higher than in the Safe Job for any level of worker output. In the pilot session, the pay was $3.00 in the Safe Job to ensure that expected pay was higher for very low productivity subjects in the Safe Job. However, the majority of subjects chose the Safe Job under these parameters, leading to the final adjustment to $2.50 which caused a much more even split in job preference.

the effect of preferring the Risky Job and the effect of being randomly assigned the Risky Job.[10]

In the remainder of the experiment after the work periods, subjects completed a series of incentivized tasks and unincentivized survey questions designed to measure a wide variety of worker traits. Full descriptions of these measurement procedures are provided in the appendix. For the body of the paper, it will suffice to note that standard methods from the modern experimental literature are used to elicit the following traits: risk aversion, loss aversion, patience, confidence, altruism, trust, and reciprocity.

1.4 Results

A total of 8 sessions were run, and 101 subjects participated. Each subject participated in only one session. The average payment (including the show-up payment) was \$16.36, with a minimum of \$5.91 and a maximum of \$30.01. The typical session lasted about 1 hour and 20 minutes, and no session lasted longer than 1 hour and 30 minutes. Table 1.2 shows mean output in the first dismissal threats work period by job preference and job assignment.

This section will be divided into two subsections analyzing the extensive and intensive margin effects of dismissal threats on productivity, respectively. First, worker output under piece rates and dismissal threats will be compared. The results show a strong extensive margin effect of dismissal threats: this form of incentive is significantly more motivating than comparable piece rate incentives. The large difference in output cannot be explained by individual risk preferences.

[10]The experiment was not designed to decompose the extensive margin productivity gains into selection and incentive effects for two main reasons. First, it was unclear ex ante whether dismissal threats would actually result in productivity gains compared to higher-powered piece rates. Second, the mechanism requires subjects to be assigned to only one of the two incentive types (piece rates or dismissal threats). This would have prevented important within-subject analysis of whether the productivity gains are driven by individual risk preferences.

Second, the intensive margin effects of stronger dismissal threats will be analyzed. The broad picture of these results may be seen in Table 1.2, which shows mean output in the first dismissal work period by job assignment and job preference. Productivity is higher under strong dismissal threats than weaker ones (26.6 versus 25.7), but this is a small and statistically insignificant difference. It turns out that the selection effect is much larger, with workers who prefer the safe job completing 22.3 units of output on average compared to 28.8 units among those who prefer the risky job. Finally, there is interesting heterogeneity in incentive effects: strong incentives appear to backfire among workers who prefer to sacrifice pay for weaker dismissal threats (25.8 units when assigned to safe versus 19.7 units when assigned to risky).

Table 1.2: Mean Output as a Function of Job Preference and Assignment

	Safe Assignment	Risky Assignment	All Assignments
Safe Preference	25.8	19.7	22.3
Risky Preference	25.7	30.9	28.8
All Preferences	25.7	26.6	26.2

1.4.1 Extensive Margin: Dismissal Threats versus Piece Rates

The real effort task successfully induced large differences in output between subjects. During the 5-minute work period with the high \$0.20 piece rate, there was significant dispersion in output levels. The mean level of output was 10.6 tasks with a standard deviation of 4.9. Figure 1.2 displays a histogram of subject output during this work period, visually confirming large differences in productivity between subjects facing the same level of incentives. This dispersion in output under identical incentives is evidence that the real effort task was sensitive to pre-existing differences in subject ability – a necessary prerequisite for detecting selection on worker productivity into different incentive schemes.

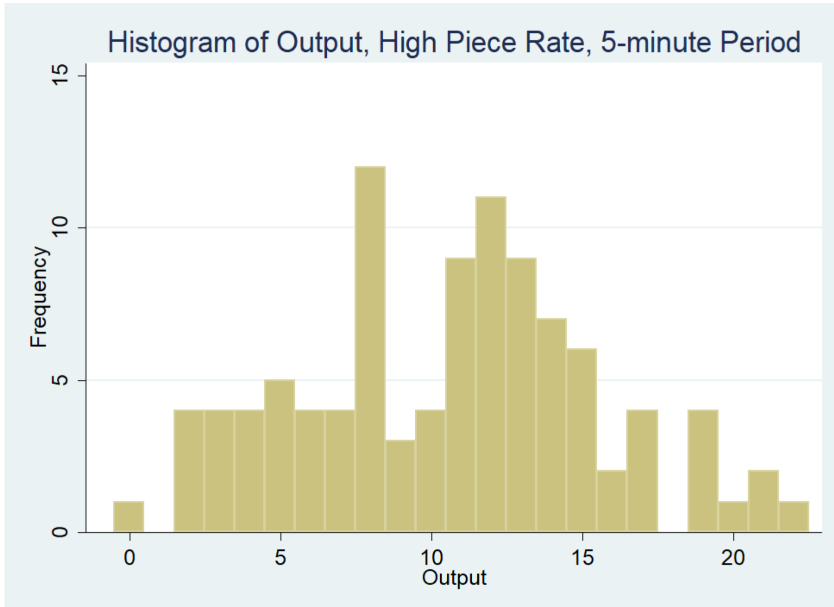


Figure 1.2: Histogram of Output in High Piece Rate Period

Additionally, there is strong evidence that output changes in response to financial incentives. When the piece rate increased from \$0.01 to \$0.20, 80.2% of subjects increased output. The mean increase in tasks completed was 2.89 tasks (with a median of 3), and the mean percentage increase in tasks completed was 82.4% (although this metric is inflated by certain subjects who heavily shirked under the low piece rate – the median percentage increase was 44.2%). Of the 20 subjects who did *not* increase output in response to the piece rate increase, 6 subjects produced the same amount in both periods and 16 subjects showed a decline in output of 2 units or fewer. A Wilcoxon signed-rank test rejects the null hypothesis that the distribution of output was equivalent in both piece rate periods ($p < 0.0001$). Taken together, the between-subject dispersion in output while facing the same incentives along with the within-subject dispersion in output while facing different incentives implies that the real effort task was sensitive to both pre-existing ability and financial incentives.

Having established that subjects change output in response to variation in simple piece rate incentives, the analysis will now turn to output under dismissal threats. Differences in output between the Risky Job and the Safe Job will be primarily reserved for the following subsection which analyzes the incentive and selection effects; the remainder of this subsection will focus on comparing piece rates to dismissal threats in general.

The following analysis will present compelling evidence that the unique mixture of monetary and behavioral motivators invoked by dismissal threats were highly motivating – indeed, much more motivating than would be expected under piece rates combined with uncertainty and risk aversion. First, simple nonparametric analysis will be presented, showing that worker productivity during the first two dismissal work periods is significantly higher than productivity during the high piece rate work period. This occurs even though the monetary marginal benefit of output is lower during the dismissal work periods than during the high piece rate period. Second, parametric analysis is conducted in which a simple quadratic functional form for the cost of effort function is assumed. The cost of effort function is calibrated using output from both the low and high piece rate work periods. The resulting utility function is used to predict each worker's output during the dismissal periods. Observed output is significantly higher than predicted output. This result holds true even when incorporating risk preferences into the output predictions, using workers' incentivized decisions over gambles to recover their CRRA parameters. Finally, analysis of output over time within the same work period under the same incentives is presented to show that learning to perform the task more efficiently is not driving this conclusion.

Recall that the marginal benefit of output under all job-period combinations under dismissal threats is below the high piece rate of \$0.20 (as shown in Table 1.1 above). This calculation takes into account the option value of being retained. This implies that under reasonable theoretical assumptions, the optimal level of output for workers under dismissal threats is less than the optimal level of output under the high piece rate. This is not the pattern of behavior that is observed.

Result 1.1: Extensive Margin. Dismissal threats are more motivating than piece rates: Subjects are more productive under dismissal threats even though the monetary benefit of output is lower.

The vast majority of subjects in both dismissal periods 1 and 2 increase their productivity compared to the high piece rate period. Doing a simple within-subject comparison, 76.2% of subjects increase productivity between the high piece rate period and the first dismissal period, while 79.7% of retained subjects do the same in the second dismissal period. In terms of magnitude, the mean change in productivity between the high piece rate period and first dismissal period is a 23.6% increase in tasks per minute (the equivalent increase in productivity is 32.9% in the second dismissal period among retained workers).

Figures 1.3 and 1.4 display CDFs of output showing this comparison.[11] Figure 1.3 displays the CDFs of high piece rate output side-by-side with the first 10-minute dismissal period; Figure 1.4 shows the same comparison with the second dismissal period. In both cases, the CDF of output in the dismissal periods is clearly to the right of the CDF of piece rate output. Hypothesis tests confirm this visual evidence. Signed-rank tests and t-tests reject the null hypotheses that output in the first and second dismissal periods are equal to output in the high piece rate period (all with $p < 0.0001$).[12]

Parametric analysis was also performed to predict output during the dismissal periods, with the primary purpose being to show that risk aversion alone cannot explain the increased output. This requires making assumptions about the functional form of the cost of effort. The cost of effort function was assumed to be increasing and convex, as is

[11] Note that to make the distributions of output comparable between 5-minute periods and 10-minute periods, the output numbers were doubled for all 5-minute periods in the displayed CDF graphs.

[12] Again, piece rate output is doubled for this hypothesis test to make a fair comparison between a 5-minute period and a 10-minute period.

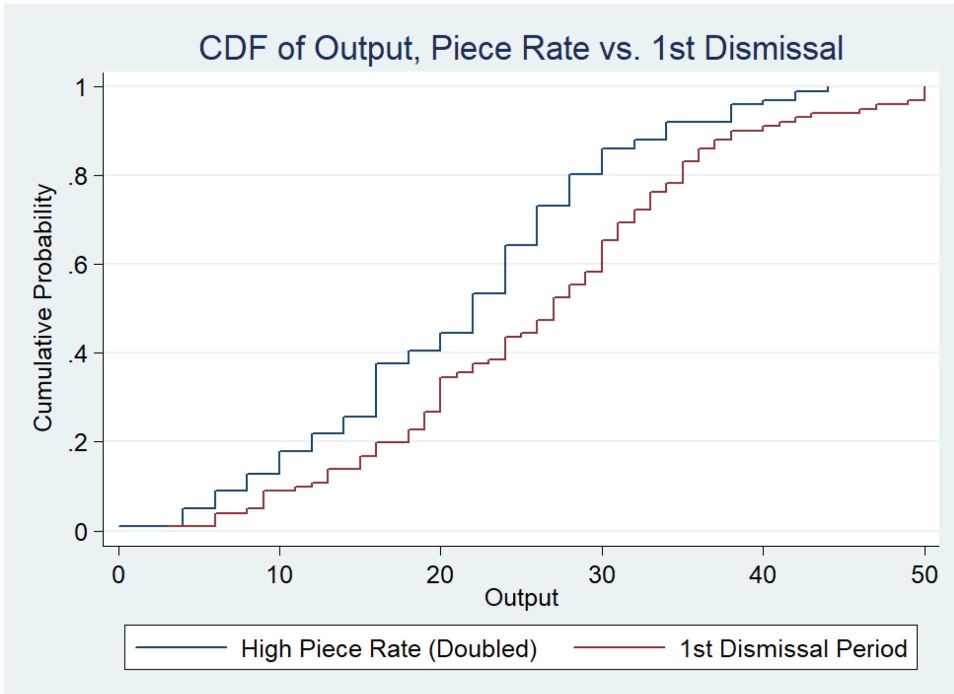


Figure 1.3: CDF of Output in High Piece Rate vs. First Dismissal Periods

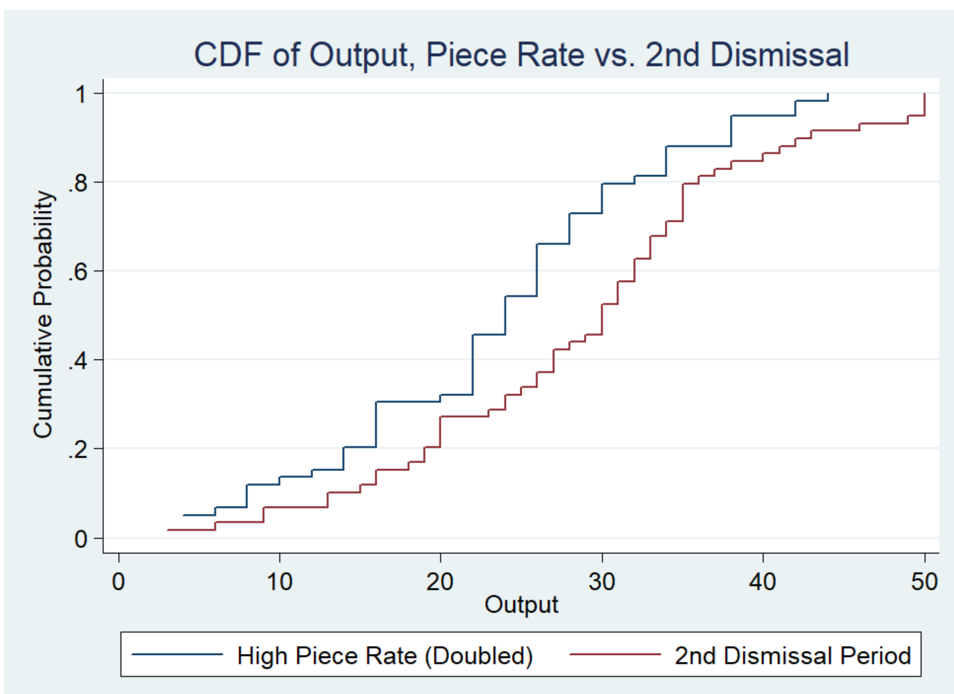


Figure 1.4: CDF of Output in High Piece Rate vs. Second Dismissal Periods

the convention in theoretical literature on principal-agent models. It was also assumed that the cost of producing zero units of effort is zero. (Although effort is unobservable, one unit of effort is normalized as the amount required to create one unit of output, without loss of generality.) To satisfy these assumptions with the simplest functional form possible, a quadratic form is assumed: $C(e) = ae^2 + be$. Additionally, utility is assumed to be additively separable in money and the cost of effort: $U = y - C(e)$, where y is total pay from that work period.

The parameters of the cost of effort function are calibrated using observed output from both piece rate work periods along with the known marginal benefit of output in each period. This results in a system of two equations with two unknown effort cost parameters for each subject: a and b. This setup is only rationalizable for subjects who increase output in response to a higher piece rate, so only the 80.2% of subjects who do so are included in the following analysis.

To account for the difference in the length of the periods (5 minutes under piece rates versus 10 minutes under dismissal threats), the amount of output observed in the piece rate periods was doubled before calibrating the cost of effort function, giving the best possible chance to the null hypothesis that predicted dismissal output is greater than or equal to observed output.[13]

The calibrated cost of effort function was then used to predict the output of each worker in the first and second dismissal work periods, given the job they were assigned. This exercise was first conducted assuming risk neutrality, and the results are displayed in Figures 1.5 and 1.6. Also note that only subjects who were retained after period 1 are

[13]Doubling piece rate output before calibrating the cost of effort function results in larger dismissal output predictions than the alternative of doubling the output predictions after they are made. This is due to the assumed convexity of the cost of effort function. However, both versions of this prediction exercise were run, and the results were qualitatively similar (the distribution of predicted output is always to the left of the distribution of observed output, and no hypothesis tests change significance levels).

included in Figure 1.6, further restricting the sample to 47.5% of subjects.

Incorporating risk aversion in the model predicts higher effort than assuming risk neutrality. The data exercise incorporates risk aversion by using the decisions over gambles made by subjects in the second part of the experiment. These decisions elicited a certainty equivalent for a 50/50 gamble between \$0 and \$6. This certainty equivalent is used to calibrate a CRRA parameter for each subject. These risk preferences are incorporated into the expected utility function of each subject, and the optimal levels of output in dismissal periods 1 and 2 are recalculated. To give the null hypothesis that predicted output equals observed output the best possible chance, the following analysis only includes subjects who exhibited risk averse preferences over the gamble options (i.e., risk-loving subjects are excluded). Combined with the other sample restrictions, 51.5% of subjects are included in Figure 1.7 and 25.7% of subjects are included in Figure 1.8.[14]

Whether one assumes risk neutrality or incorporates the measured risk aversion of subjects, the results of the parametric analysis are the same: subjects are significantly more productive under dismissal threats than neoclassical models would predict given their behavior throughout the rest of the experiment. Hypothesis tests strongly echo the visual evidence displayed in Figures 1.5-1.8. Signed-rank tests and t-tests reject the null hypotheses that first- or second-period dismissal outputs are equal to predicted output from either the risk neutral or the risk averse models (all p-values < 0.001).

Result 1.2: Risk Aversion. Parametric analysis suggests that risk aversion does not explain the high productivity under dismissal threats.

An important potential criticism to address is whether the above results are caused merely by subjects learning to complete the task more efficiently over time. For example,

[14] Figure 1.7 includes subjects who are risk averse and increased output in response to the piece rate increase. Figure 1.8 includes subjects who are risk averse, increased output in response to the piece rate increase, and were retained after dismissal period 1.

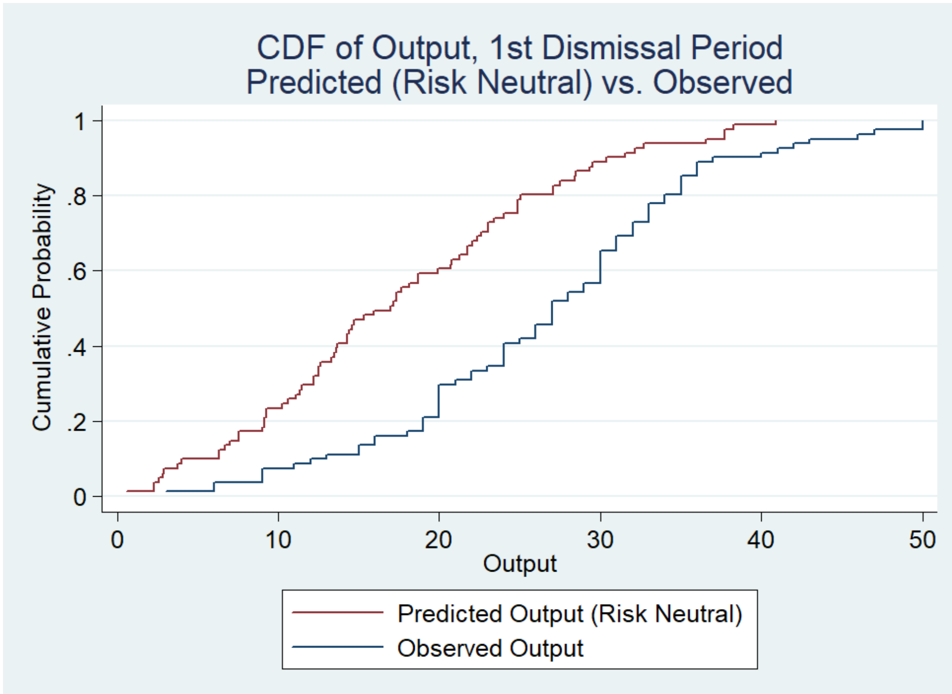


Figure 1.5: CDF of Output in First Dismissal Period, Predicted (Risk Neutral) vs. Observed

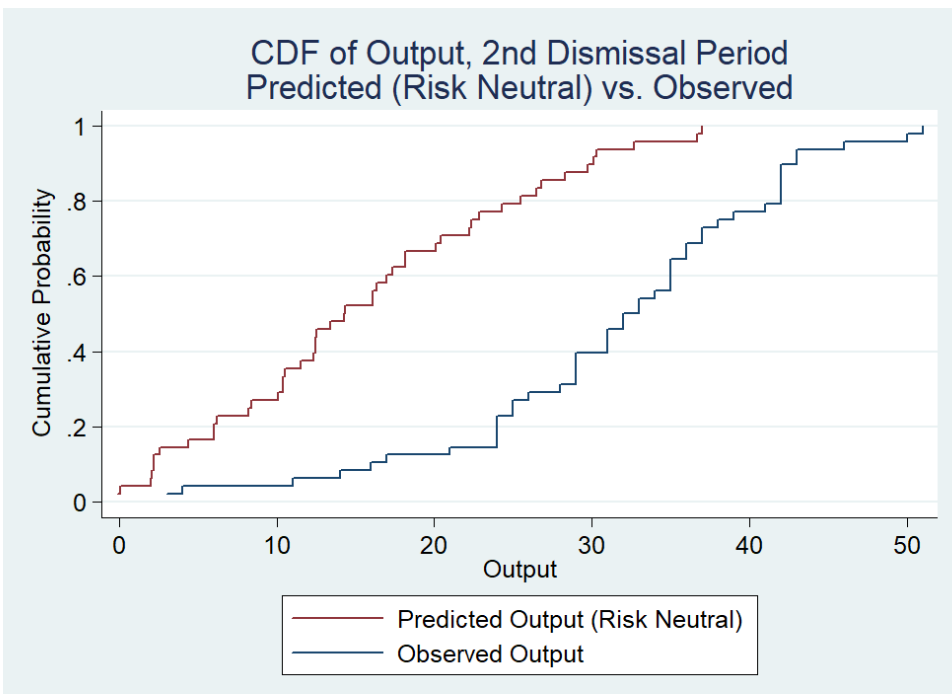


Figure 1.6: CDF of Output in Second Dismissal Period, Predicted (Risk Neutral) vs. Observed

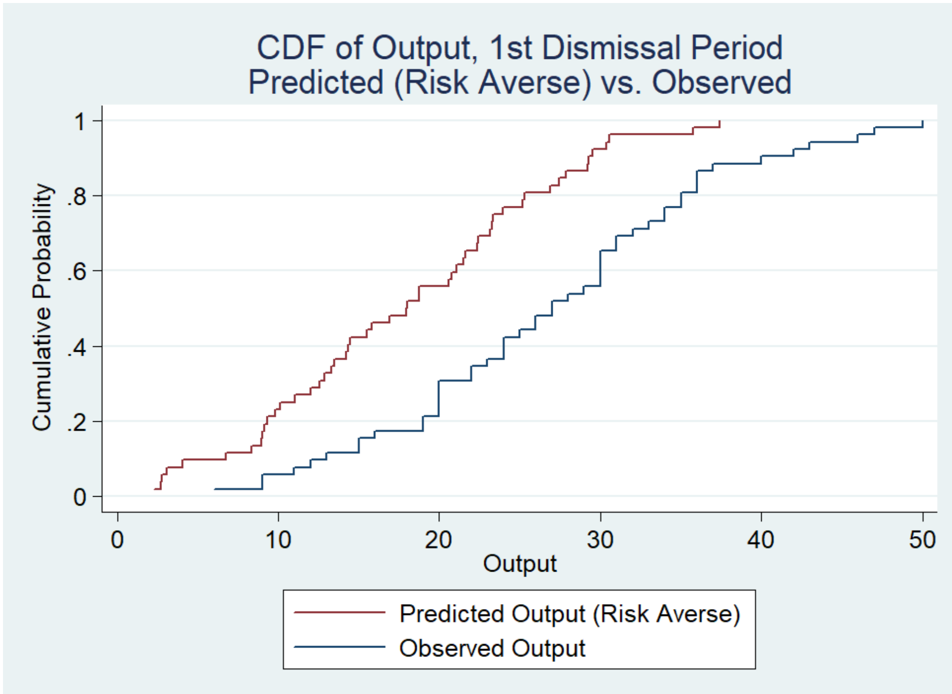


Figure 1.7: CDF of Output in First Dismissal Period, Predicted (Risk Averse) vs. Observed

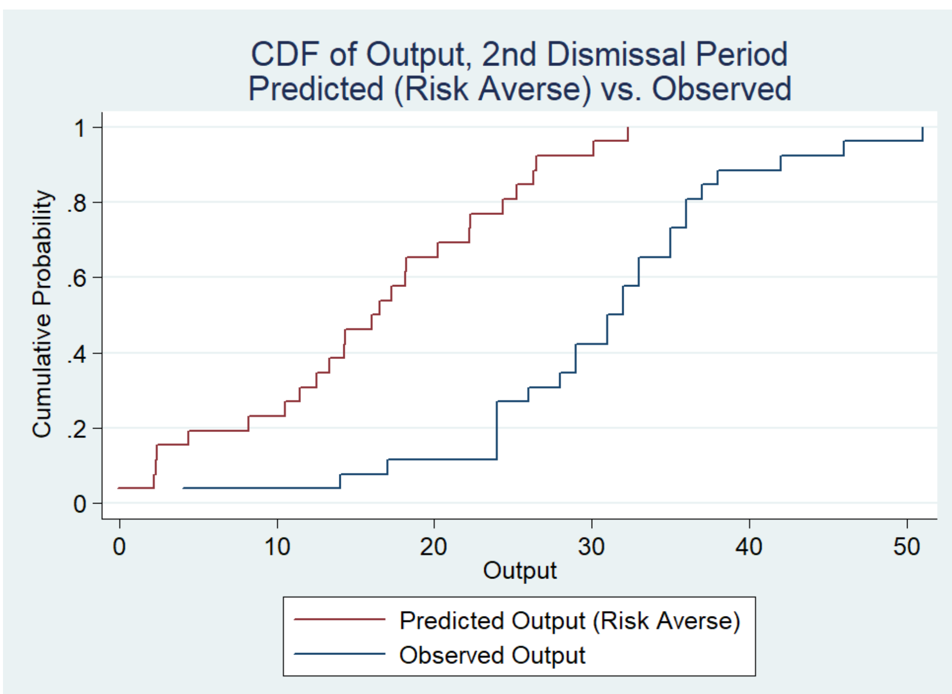


Figure 1.8: CDF of Output in Second Dismissal Period, Predicted (Risk Averse) vs. Observed

the first dismissal period comes after the high piece rate period, so it could be the case that subjects are producing more output because they have more experience rather than because dismissal threats are highly motivating.[15] Comparing worker output between discrete periods with different levels and types of incentives will not shed light on how much learning occurred because incentives are changing along with worker experience.

The best way to analyze worker learning during the real effort task is to compare output over time within the same work period. If workers are truly improving their ability during the work task, a 10-minute period would likely be enough time to pick up a difference in productivity. The experiment software stored the exact time at which each work task page was submitted, allowing analysis of how many tasks were submitted in each minute of every work period. Figures 1.9 and 1.10 plot mean worker output over time for both 5-minute piece rate periods and for the first two 10-minute dismissal work periods. (The third dismissal period is less informative about learning because a large proportion of retained subjects shirk.)

The figures show that apart from the first minute, output is relatively constant over time within each work period. However, there is a strong mechanical reason to expect output to increase from the first to the second minute. In the first minute, workers must start a fresh task; in all subsequent minutes, there is typically a task in progress that carries over from the previous minute. Another trend observed in the data explains why output slightly increases in the final minute: many subjects try to submit a final answer just before the timer ticks down to zero, even if that answer is incorrect.

Keeping these mechanical trends in mind relating to the first and last minutes of the period, hypothesis tests were run on the null hypothesis that output in the first half of the work period (omitting the first minute) is equal to output in the second half of

[15]Alternatively, exhaustion could reduce output over time, but this is not a plausible potential explanation for the results shown above because output is higher in later periods as dismissal threats are introduced.

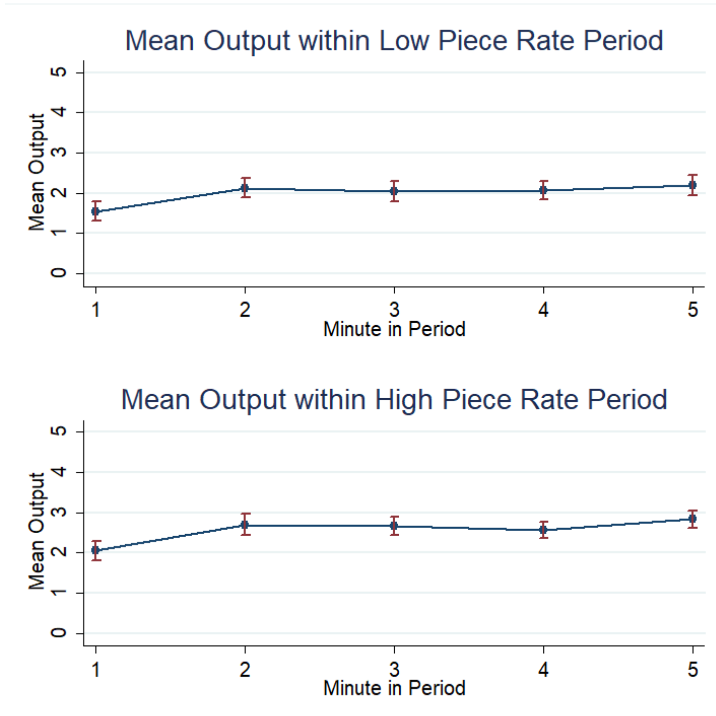


Figure 1.9: Mean Output over Time in Piece Rate Periods

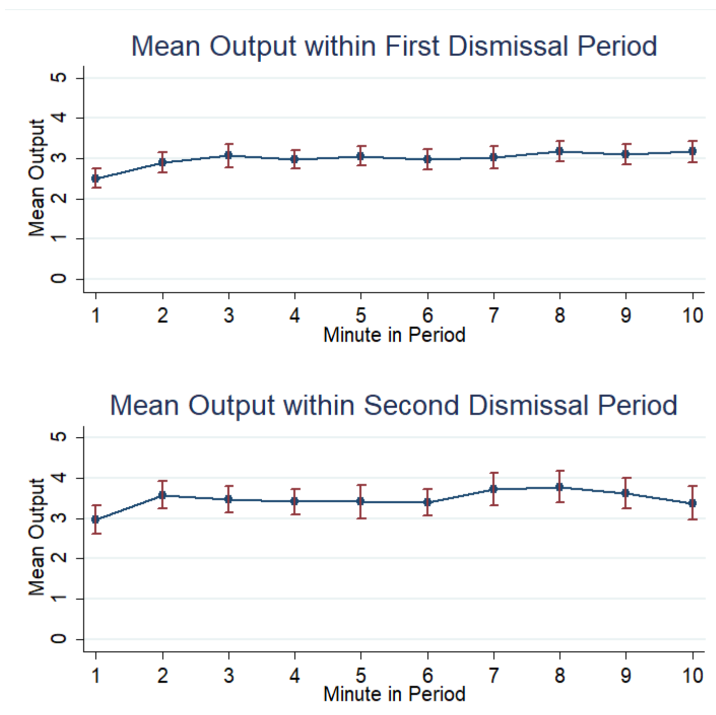


Figure 1.10: Mean Output over Time in Dismissal Periods

the work period (omitting the final minute). Rank-sum tests and t-tests were run on all four work periods, and every test failed to reject the null at a 95% confidence level. The evidence from these figures and hypothesis tests suggests that learning to be more efficient at the task did not contribute to higher worker output under dismissal threats.

Result 1.3: Learning. Subjects do not increase productivity over time, suggesting that learning does not explain the high productivity under dismissal threats.

Another alternative hypothesis is that subjects worked very hard under dismissal threats either due to intrinsic motivation or experimenter demand effects, but this explanation does not hold up to scrutiny after examining worker output during the final dismissal period with full job security. Because all fixed payments are up-front and there is no fourth dismissal period, there is no monetary incentive to work at all during the third period. Thus, the possible motivations to produce output during this period were only intrinsic motivation (i.e., enjoyment of the work task), experimenter demand effects, or a desire to avoid a public announcement of one's dismissal at the end of the period.

Most workers who made it to the final tenured work period produced lower output than in previous periods. Of the 39 workers (out of 101 total) who made it to the final period, 71.8% of them produced fewer units of output than in period 2. Mean output in period 2 among workers who would become tenured was 34.5 units, and mean output in period 3 dropped to 20.6 units. This decline in output is highly statistically significant (signed-rank test, $p = 0.0002$). Some of the tenured workers displayed severe levels of shirking: 25.6% of tenured workers produced 0 units of output, while 43.6% of then produced 5 units or fewer. Keeping in mind that these are systematically the most productive subjects in the experiment, these levels of output are quite low.

The final period turned off most but not all of the behavioral motivations involved in firing threats. There was no reason to work to avoid further social exclusion, jealousy of coworkers who would continue to earn money, or any potential self-image impacts of dismissal (the signal of being fired is uninformative if the worker was not exerting effort in the first place). Because the public announcement of firing remained, if anything the productivity decline is an overestimate of any remaining demand effects or intrinsic motivation that might have been driving the high productivity in periods 1 and 2. The reasonable conclusion is that the high productivity when all behavioral and monetary mechanisms were active was due to the combined strength of these motivators, and not merely due to intrinsic motivation or experimenter demand effects.

Result 1.4: Intrinsic Motivation. Subjects are significantly less productive in the final dismissal period, suggesting that intrinsic motivation does not explain the high productivity under dismissal threats.

The above analysis suggests that dismissal threats produce a large extensive margin effect on productivity which cannot be rationalized by subject responses to piece rate changes or risk preferences. Although the vast majority of subjects produce more output under dismissal threats than the parametric model with risk aversion predicts, the magnitude of this difference varies considerably by subject. An interesting follow-up question is which if any of the measured behavioral traits of subjects predicts the size of this residual. In other words, do the behavioral traits of workers predict how strongly they will react to the existence of dismissal threats?

Table 1.3 addresses this question by regressing the output residual on worker behavioral traits. The output residual for each worker is defined as observed output in the first dismissal period minus output predicted by the parametric model with risk aversion. Given the parametric model's restrictions, only subjects who increase output between the low piece rate and high piece rate work periods are included. Additionally, only risk averse subjects are included. The variables for worker behavioral traits include incentivized measures of loss aversion, patience, overconfidence, altruism, trust, and reciprocity.

Table 1.3: Linear Regression of Output Residual on Worker Traits

	(1)
	Output Residual
Loss Choice	0.725
	(0.612)
Patience Choice	2.620*
	(0.041)
Confidence: Actual - Reported Rank	0.339
	(0.331)
Social Prefs: Altruism	-0.734
	(0.124)
Social Prefs: Trust	0.639
	(0.540)
Social Prefs: Reciprocity	0.362
	(0.800)
Constant	-1.491
	(0.826)
Observations	52

Notes: The table shows the results of an OLS regression. The dependent variable is observed output in the first dismissal threat work period minus output predicted by the parametric model including risk aversion. The subsample only includes subjects who increase output under a higher piece rate and only subjects who display risk aversion in decisions over gambles. Standard errors are clustered at the session level; p-values are shown in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The results of this regression show that patience is a significant predictor of the extensive margin effect of dismissal threats. The patience choice variable represents how much money a subject would be willing to accept in one week instead of receiving $3 today, so subjects with a higher value are more impatient. The coefficient implies that a subject who requires $1 additional dollar received in one week will produce 2.6 additional units of output under dismissal threats than the parametric model predicts.

Result 1.5: Behavioral Predictors of Extensive Margin. Subjects who are more impatient display larger extensive margin productivity gains under dismissal threats.

This subsection has shown evidence suggesting that the mixture of monetary and behavioral motivators invoked with dismissal threats in this experiment were highly motivating. Output under dismissal threats was significantly higher than would be predicted under a variety of assumptions and methods, all tilted to give the alternative hypothesis the best chance whenever possible. The results cannot be explained by the level of risk aversion observed in decisions over gambles in the experiment. They also cannot be explained by learning because subjects do not improve output-per-minute even over the course of long 10-minute work periods while incentives remain constant. Instead, it appears that dismissal threats are a much stronger motivating incentive device than piece rates.

1.4.2 Intensive Margin: The Incentive and Selection Effects of Stronger Dismissal Threats

This subsection decomposes the productivity gains from stronger dismissal threats into the pure incentive effect from a higher marginal benefit of output and the selection effect from attracting more productive workers.[16] Recall that the firing chance reduction per unit of output is twice as high in the Risky Job: a 2% decline in the chance of being fired per unit of output, compared to only a 1% decline in the Safe Job. Additionally, subjects in the Risky Job have more to lose by being fired because the wage per period

[16]Note that the following analysis will only involve output from the first work period with dismissal threats unless stated otherwise.

is twice as high. These steeper incentives suggest that output should be higher in the Risky Job than in the Safe Job, which is indeed the pattern of behavior observed.

Result 2.1: Intensive Margin. Workers are more productive under stronger dismissal threats with higher pay than under weaker dismissal threats with lower pay.

Mean output of workers who preferred the Risky Job and were assigned the Risky Job was 30.9 units, while mean output of workers who preferred the Safe Job and were assigned the Safe Job was only 25.8 units. A Wilcoxon rank-sum test rejects the null hypothesis that these samples of output levels come from the same distribution (p = 0.028). However, comparing output between these two groups only reveals the combined effect of both the incentive effect and the selection effect for the subset of subjects who were assigned their preferred job.

To decompose the selection and incentive effects, each of the four job preference-assignment combinations will be exploited in data analysis. In terms of job preference, there was a fairly even split: 60.4% of subjects preferred the Risky Job and 39.6% of subjects preferred the Safe Job. Seventeen subjects preferred the Safe Job and were assigned the Safe Job; 23 subjects preferred the Safe Job and were assigned the Risky Job; 25 subjects preferred the Risky Job and were assigned the Safe Job; and 36 subjects preferred the Risky Job and were assigned the Risky Job.

Result 2.2: Incentive and Selection Effects. The aggregate productivity gain from stronger dismissal threats is caused largely by the selection effect, not the incentive effect.

The story is clear whichever method of analysis is used: the aggregate incentive effect of stronger dismissal threats is small and statistically insignificant, but the aggregate selection effect on output is large and significant. Mean output is 25.7 among workers who are assigned the Safe Job, while it is 26.6 among workers who are assigned the Risky

Job (statistically insignificant with a t-test). Thus there appears to be little difference in mean output indicating a pure incentive effect from stronger dismissal threats. Mean output is 22.3 among workers who prefer the Safe Job and 28.8 among workers who prefer the Risky Job (with a t-test p-value of 0.003). Therefore, the selection effect on mean output is both statistically significant and large in magnitude (roughly a 30% increase in output between job preferences).

Figures 1.11 and 1.12 visually display the output differences caused by the incentive and selection effects by showing the CDFs of output by either job preference or job assignment. Figure 1.11 shows roughly overlapping distributions of output between job assignments, indicating little or no incentive effect. A Wilcoxon rank-sum test does not indicate a statistically significant difference between these two distributions of output. On the other hand, Figure 1.12 shows a clear separation of the distribution of output between job preferences: the distribution of output among subjects who prefer the Risky Job is clearly to the right of the distribution for subjects who prefer the Safe Job. This difference is highly statistically significant (rank-sum test: $p = 0.001$).

These aggregate results hide substantial heterogeneity in subject responses to stronger dismissal threats. It seems that subjects who prefer the Risky Job do produce more output when randomly assigned the steeper incentives, although this result is only marginally significant. On the other hand, subjects who prefer the Safe Job but are thrust into the steep incentives of the Risky Job actually perform *worse* than their counterparts who prefer the Safe Job and are assigned the Safe Job. This suggests that imposing strong dismissal threats on workers who do not prefer them can backfire.

Result 2.3: Heterogeneity in Incentive Effects. Strong dismissal threats backfire among subjects who were willing to sacrifice pay in exchange for weaker dismissal threats, producing a negative incentive effect among this subset of subjects.

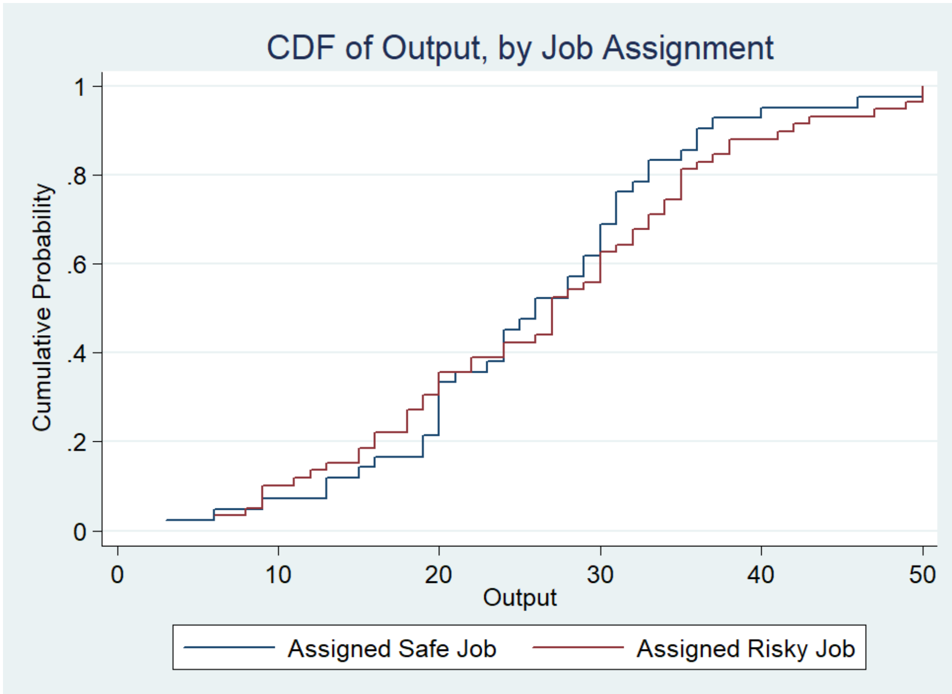


Figure 1.11: CDF of Output in Dismissal Periods, by Job Assignment

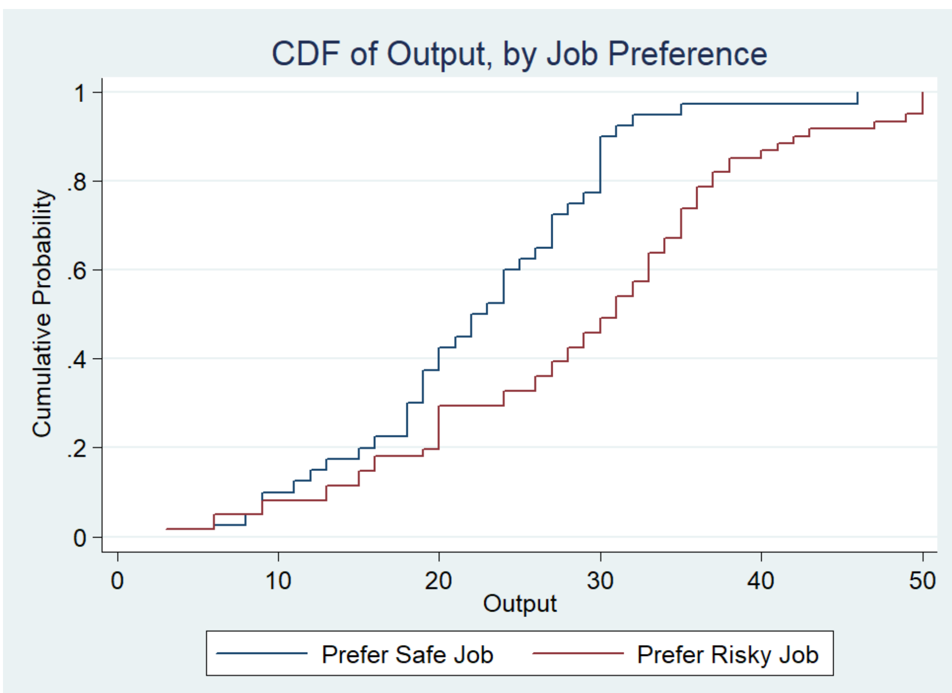


Figure 1.12: CDF of Output in Dismissal Periods, by Job Preference

Figures 1.13 and 1.14 illustrate this point by plotting the distribution of output between job assignment while only including a subset of subjects with the same job preference. Figure 1.13 only includes subjects who prefer the Safe Job and plots the distribution of output separately by job assignment. Notice that the expected placement of the distributions is reversed: the distribution of output under the steeper incentives in the Risky Job is actually further to the left among this subset of subjects. This difference in output is statistically significant using both a t-test ($p = 0.02$) and a rank-sum test ($p = 0.025$).

This negative incentive effect may seem surprising at first, but many of these workers were placed into a job in which they were not expected to perform well. The median output of workers who preferred the Safe Job during the high piece rate period was only 8 units in 5 minutes. A worker who maintained this same level of output per minute during an entire work period in the Risky Job would end up with a dismissal probability of 68%. In fact, fully 85% of workers who preferred the Safe Job would end up with a dismissal probability of greater than 50% even when maintaining their output-per-minute from the high piece rate period. A logical conclusion is that placing workers under dismissal threats that are too strong compared to their ability level can actually be demotivating.

Figure 1.14 shows suggestive evidence that the incentive effect of stronger dismissal threats was positive among the subset of subjects who preferred the Risky Job. The distribution of output under the Risky Job incentives is clearly to the right of the distribution of output under the Safe Job incentives, but this difference is only marginally significant (t-test: $p = 0.078$; rank-sum test: $p = 0.079$).

Although these graphs and hypothesis tests are visually useful and do tell most of the story, one must keep in mind that they do not fully isolate the selection and incentive effects because the random assignment to incentives did not occur with 50% probability. For example, in the plot of output distributions illustrating the incentive effect (Figure

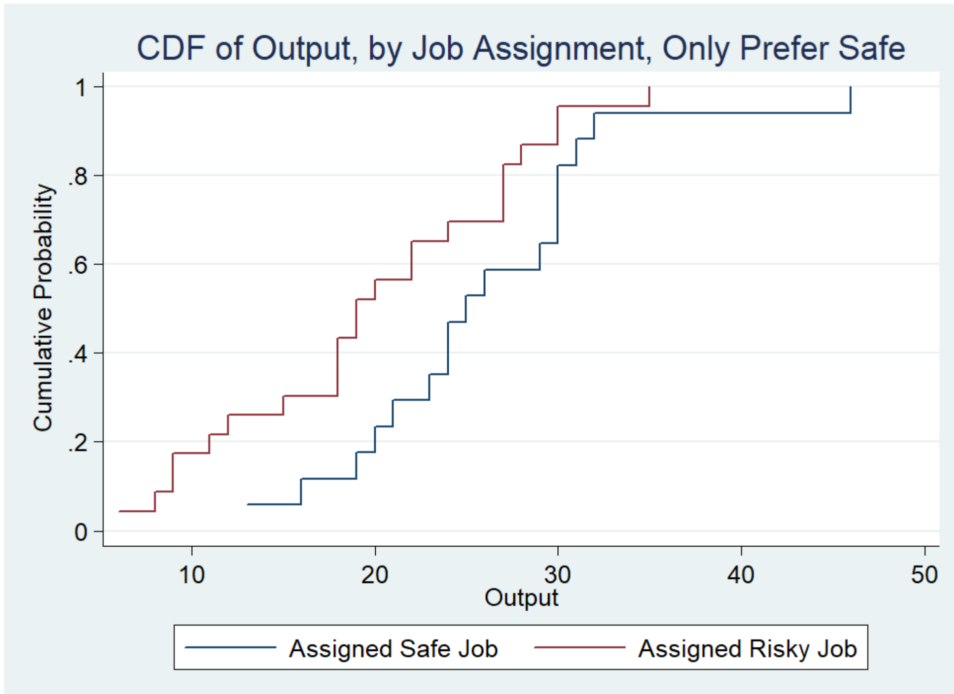


Figure 1.13: CDF of Output by Job Assignment, Only Prefer Safe

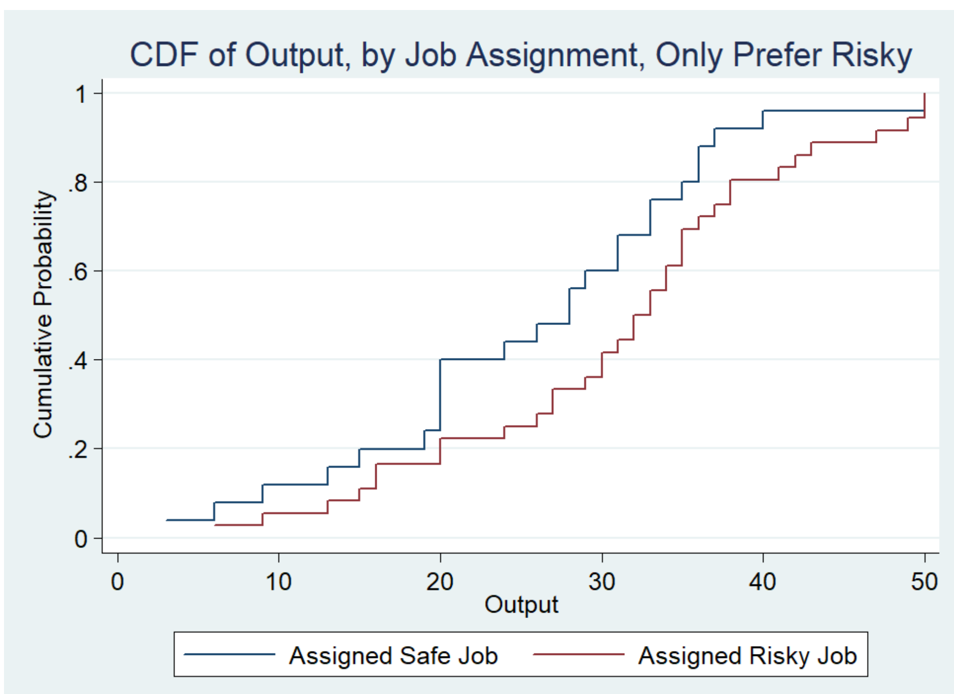


Figure 1.14: CDF of Output by Job Assignment, Only Prefer Risky

1.11), roughly 60% of the subjects working in the Risky Job will also have preferred the Risky Job due to the random job assignment design mechanism. As a result, regressions are run to fully isolate the impact of job preference versus job assignment.

Table 1.4 displays the results of regressions which fully isolate the incentive and selection effects on output. These specifications are simple OLS regressions with output during the first 10-minute dismissal work period as the dependent variable. Standard errors are clustered at the session level throughout. Column 1 includes all subjects and includes two indicator variables as covariates: whether the subject was assigned the Risky Job, and whether the subject preferred the Risky Job. The coefficient on job assignment is small and statistically insignificant, indicating that there were not strong incentive effects despite facing a much larger marginal benefit of output in the Risky Job compared to the Safe Job. In contrast, the selection effect on output is large and statistically significant. After controlling for job assignment, indicating a preference for the Risky Job is associated with a 6.4-unit increase in output ($p < 0.01$). This magnitude is large enough to be economically impactful, given that mean output in the first dismissal period was 26.2 units.

Regressions were also run to understand the heterogeneity in incentive effects between subjects with different job preferences. Columns 2 and 3 of Table 1.4 show the results of regressing output on job assignment among only subjects who preferred the Safe Job and only subjects who preferred the Risky Job, respectively. These columns confirm the visual evidence presented in the CDF plots above: among subjects who preferred the Safe Job, being assigned the Risky Job is associated with about a 6-unit reduction in output ($p = 0.018$). There is again suggestive evidence that the incentive effect of stronger dismissal threats is positive among subjects who preferred the Risky Job, although the coefficient is not statistically significant at the 5% level ($p = 0.103$).

The analysis so far has focused on productivity during the first period under dismissal

Table 1.4: Linear Regression of Output under Dismissal Threats on Job Preference and Job Assignment

	(1) All Subjects	(2) Only Prefer Safe	(3) Only Prefer Risky
Assigned Risky Job=1	0.725	-6.084*	5.237
	(0.729)	(0.018)	(0.103)
Prefer Risky Job=1	6.434**		
	(0.001)		
Constant	21.91***	25.82***	25.68***
	(0.000)	(0.000)	(0.000)
Observations	101	40	61

Notes: The table shows the results of OLS regressions. The dependent variable is output during the first 10-minute work period with dismissal threats. Column 1 includes all subjects. Columns 2 and 3 include only subjects who preferred the Safe Job or the Risky Job, respectively. Standard errors are clustered at the session level; p-values are shown in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

threats, but retention over time also plays an important role in the selection effect in the real labor market. More productive worker are systematically retained in jobs with strong dismissal threats, and this of course occurs in this experiment as well. The mean output in period 1 was 26.2 units, while mean output among retained workers in period 2 was 31.1 units. Additionally, more workers are fired from the Risky Job due to the higher chance of being fired at all levels of output: 58.4% of workers start out in the Risky Job, but by period 2 only 49.2% of retained workers are in the Risky Job. This confirms that the selection out of lower-productivity workers is an important element of the long-term impact of strong dismissal threats on productivity.

The lack of aggregate pure incentive effects from being randomly assigned stronger dismissal threats is an interesting behavioral puzzle. It appears that once the mixture of monetary and behavioral motivators of dismissal threats is activated, subjects generally work their hardest regardless of adjustments on the margin to the dismissal probability function. A noteable exception to this is that when the dismissal threats are too strong for the ability level of a worker, they can become demotivating.

Analyzing the differences in output between periods 1 and 2 also speak to this lack of responsiveness to changes in dismissal threats on the margin. Any model of rational worker behavior predicts that workers should reduce output from period 1 to period 2. This is because in the first period, there are two remaining payments to earn, while in the second period, only one potential payment remains. However, this does not occur. Restricting to the sample of workers who were be retained after period 1, their mean output was 28.7 in period 1 and 31.1 in period 2. It appears that at the parameters explored in this experimental design, workers were far more sensitive to the mere presence of credible dismissal threats than they were to the marginal strength of dismissal threats.

Additional analysis was conducted to determine whether selection into dismissal threats occurred on worker traits other than productivity. Although some univariate differences were detected between workers who preferred the Safe Job and the Risky Job, these traits were all correlated with productivity. For example, workers who preferred the Risky Job tended to be more confident but also more accurate in those beliefs. However, both confidence and accuracy of beliefs were highly correlated with productivity in the high piece rate work period. A probit regression of preference for the Risky Job on all incentivized worker traits results in productivity being the only trait which is significantly predictive of job preference. The full explanation of this analysis including tables and regression specifications may be found in the appendix.

1.5 Discussion

This paper makes two significant contributions to the literature on employee motivation and selection. First, the experiment design imports the threat of being fired to the lab in a clearly defined probabilistic fashion, linking worker output to the chance of being fired. This brings tractability to an incentive that is ubiquitous in the modern labor market, but rarely explicitly quantified in labor contracts and thus difficult to study in the field. This tractability allows for clean comparisons between dismissal threats and other monetary incentives such as piece rates, along with the ability to measure worker responses to exogenous, quantifiable changes in the strength of dismissal threats. Further, the design purposefully activates many of the behavioral mechanisms concomitant with the threat of being fired: the loss framing, the fear of exclusion from a social group, the embarrassment and self-image implications of public announcements of firings, and the jealousy of retained coworkers who will continue to work and earn money. These design elements produce a rich implementation of dismissal threats in the lab, enabling detailed analysis of this unique and understudied type of incentive.

Second, the design elicits worker preferences over incentives while retaining random assignment to those incentives, all within the same round. Previous approaches in the literature typically opt for either fully random assignment or guaranteed selection into incentive schemes, neither of which allows for a clean decomposition of incentive and selection effects. Another approach has been to include both random assignment and guaranteed selection over multiple rounds, but this is vulnerable to order contamination effects and takes up scarce time during the experiment. Collapsing preference elicitation and random assignment into one round also retains the once-and-for-all nature of dismissal threats, likely a key component of the power of the incentive. The mechanism is quite simple to understand for both researchers and subjects: workers select an incentive scheme to improve the odds that they are randomly assigned to that scheme. This mechanism is ready to import into other experiments on selection into a wide variety of incentive schemes such as competitions, teams, goal-setting, commitment contracts, and more.

The results paint a clear picture: on the extensive margin, dismissal threats are highly motivating as an incentive device. Introducing credible dismissal threats results in large productivity gains over the most basic monetary incentive of piece rates. The incentive schemes in the experiment were designed such that the monetary benefit of output under dismissal threats was always lower than the high piece rate. However, the vast majority of subjects increase productivity under dismissal threats compared to their own productivity under the high piece rate. The mean increase in productivity is 23.6% in the first dismissal work period, an economically important magnitude. Additionally, parametric analysis shows that this result is unlikely to be caused purely by risk aversion. Although incorporating subject risk preferences does increase predicted output under dismissal threats compared to a risk neutral benchmark, risk aversion alone does not explain the large increase in productivity induced by dismissal threats.

Turning to the intensive margin effects of steeper incentives, stronger dismissal threats result in higher output than weaker ones, as predicted. The interesting results appear in the decomposition of these productivity gains into the pure incentive effect versus the self-selection effect. Analysis of the data clearly shows that the aggregate productivity gains from stronger dismissal threats come largely from selection effects. Workers who prefer higher pay with a steeper firing probability function are significantly more productive than other workers. However, being randomly assigned to these steeper incentives appears to have little effect on productivity on average.[17]

A likely explanation for this result is that the mixture of financial and behavioral motivators invoked under firing threats was so strongly motivating that effort was not very sensitive to marginal changes in the pure financial incentive. In other words, as long

[17]This does not contradict the results on the extensive margin. Even the weaker dismissal threats implemented in this experiment resulted in significantly higher productivity than the high piece rate. This was true among both the workers who preferred weaker dismissal threats and the ones who preferred stronger dismissal threats.

at least a moderate threat of being fired is present, along with the social, self-image, and loss-framing motivators, subjects will put forth high levels of effort. If this is the case, introducing credible, salient firing threats is more powerful than making modifications to existing firing threats on the margin. Further evidence for this interpretation is the fact that retained workers did not show declines in productivity in the second dismissal period, despite the decline in the marginal benefit of output due to fewer payments remaining to be collected.

However, the results revealed an important exception to this finding of weak incentive effects: strong dismissal threats can backfire among subjects who preferred to avoid them. Subjects who were willing to sacrifice pay in order to work under weaker dismissal threats actually produced lower output when randomly assigned to strong dismissal threats. Clearly, these workers were trying to self-select into the incentive scheme under which they expected to perform better, as in Roy's (1951) model of worker selection. Indeed, most of these workers were very likely to be dismissed even while maintaining their effort levels from the high piece rate period.

These results contain important implications for firms and policymakers. Firms that rely largely on monetary incentives other than dismissal threats may experience large productivity gains by shifting to moderate-strength dismissal threats. Firms with existing credible dismissal threats may consider strengthening them to improve productivity, but they should realize that these effects are likely to occur over time as the composition of workers changes through hiring and retention decisions. It is important to avoid imposing dismissal threats that are too strong for the ability level of a worker given the evidence that this can backfire and cause negative incentive effects. This is an important consideration when imposing stronger uniform dismissal threats on an existing workforce of mixed ability levels. Policymakers considering implementing widespread employment protection may expect to experience muted productivity losses given that the aggregate effect operates largely through selection, but this hides differential productivity impacts on workers with different preferences over dismissal threats.

There are a number of potentially fruitful avenues for future research. In this experiment design, only two levels of pay and dismissal threats were offered. This was necessary to maintain incentive compatibility of the job preference elicitation and allow a simple decomposition of the selection effect from the incentive effect. However, it seems quite possible that a more complex design that offered many different levels of pay, baseline chance of firing, and relationship between output and firing chance could uncover much richer evidence on worker preferences over dismissal threats. Other modifications to the framing of firing threats might be considered, such as displaying the chance of retaining one's job rather than the chance of losing it.

This paper has intentionally activated many of the behavioral motivators of firing threats in the lab, yet still more mechanisms could be explored in future research. The present design included a strong loss framing, public notifications of firing, the fear of exclusion from a group of coworkers, and jealousy of retained workers who continue to earn money. Firing threats in the real labor market may be even more motivating when the rejection comes from a human boss with discretion compared to an impersonal mass layoff.[18] Social motivators may be stronger when worker identities are revealed and social connections are developed, rather than the anonymous setting implemented in this design. It also seems possible that a sense of ambiguity plays an important role in the field as well – workers never quite know precisely what it takes to avoid being fired. This ambiguity might produce even stronger motivation for workers than a known, quantified link between productivity and the probability of being fired.

The results of this experiment suggest that the threat of being fired is a highly mo-

[18]However, without a clever design solution, using a human subject as a boss involves losing control over the ability to quantify and manipulate the strength of dismissal threats.

tivating incentive device, and this may partially explain why this form of incentive is so ubiquitous in the modern labor market compared to simpler incentives such as piece rates. The behavioral motivators activated in this lab experiment did not operate at the full strength at which they exist in the real labor market. This suggests that the productivity gains from dismissal threats estimated in this paper are a lower bound on the motivating effect they may have in the field. A fruitful avenue for future research would be to explore which of the various behavioral components of firing threats are the most powerful and why. This sort of decomposition of the motivating mechanisms of firing threats would be fascinating for labor and personnel economists and informative for management teams trying to motivate their employees.

Chapter 2

Self-Image in Job Search

2.1 Introduction

The process of searching for a job involves receiving repeated signals that are inherently self-relevant, suggesting that biases related to self-image may play an important role in the behavior of job seekers. Many people take pride in their career, deriving an important sense of self-worth from their work. Applying to a job puts this source of self-worth under scrutiny, inviting an employer to compare one's skills, education, and experience to that of other competing job seekers. The results of job applications are self-relevant signals for job seekers, who must form and continuously update their beliefs about job prospects over the course of a search spell.

Previous work has consistently shown that self-image concerns alter how people interpret signals and update beliefs, but it's not clear whether and how these updating biases might directly translate into actions in various contexts. Stating overconfident beliefs about oneself may come naturally and feel good, but acting congruent with that overconfidence is not assured when the stakes are high. Additionally, directly stating beliefs may activate a different mental frame than choosing an action that only indirectly reveals those beliefs. Although there is evidence from the field that job seekers are typically overconfident about their reemployment prospects, applied work has not yet directly linked overconfidence, biased signal processing, or other biases related to self-image to search effort decisions. This is important research because searching for a job is one of the most consequential activities undertaken by the working-age population, and search outcomes have large long-term impacts on the well-being of workers. If job seekers alter their search behavior due to self-image concerns, it could lead to inefficiencies such as skill mismatch among the employed, unused and deteriorating human capital among the unemployed, and serious mental health and financial consequences for the long-term unemployed.

Self-image concerns are predicted expected to affect search behavior through two primary mechanisms: biased beliefs and information avoidance. Past work has shown that people process signals differently when the new information is relevant to their self-image. This biased updating often occurs in a self-serving manner, leading to more optimistic beliefs about oneself than would occur otherwise. In labor market search, the primary feedback job seekers receive comes in the form of ignored or explicitly rejected job applications, causing a rational job seeker to update beliefs about the probability of receiving a high quality offer downwards. Agents with self-image concerns might interpret these signals through a self-serving lens, discounting negative signals to avoid painful belief revisions. Positive feedback also occurs during job search in the form of call-backs, interview requests, and job offers, meaning that job seekers may also overweight positive signals in a self-serving way. If job seekers do introduce bias into the belief updating process, it could lead to search behaviors and outcomes that would otherwise be suboptimal.

In addition to the concern that job seekers may alter their beliefs in a self-serving way, thus impacting search behavior, the opposite concern also exists: that job seekers might alter their behavior to manipulate the signals they receive from search. Information avoidance is the second mechanism through which self-image concerns could lead to suboptimal search behavior. The most obvious and damaging case of this would be a job seeker who becomes discouraged from repeated rejections and decides to stop search altogether to avoid further painful belief revisions. It could also occur in more nuanced ways if job seekers apply to openings for which they are overqualified, thus rendering signals less informative about their true labor market value. Or it could occur in a partial way if job seekers apply to fewer openings to reduce the number of rejections, but do not halt search altogether.

These two mechanisms are predicted to exert forces on search duration in opposite directions, and are not easily isolated in the field; thus experimental methods are particularly useful to answer questions about the role of self-image in search. Consider the extensive margin of search effort in which an agent decides each period whether to apply to a job opening or to quit search and become a discouraged worker. (Analogously, an agent might decide whether to continue applying to high quality openings or to switch to lower quality openings.) An agent who protects self-image through biased updating would update beliefs too little after receiving negative signals, thus searching for too long compared to an unbiased agent. An agent who protects self-image through information avoidance would quit search too early to avoid further negative signals.

This paper reports the results of a laboratory experiment that compares search behavior between environments with and without self-image concerns. In the treatments involving self-image, feedback from search is informative about subjects' relative performance on an intelligence test. Other treatments decisively shut down this self-image channel while holding all other aspects of the search environment constant, including prior beliefs before beginning search. Strikingly, subject search behavior does not differ significantly between treatments with and without self-image. This holds across numerous methods of data analysis, and while comparing various subsets of the sample to test whether specific biases impact search behavior. The design uses standard methods from the literature to induce self-image concerns: providing signals about subject test scores on a series Raven's Matrices problems, which is directly referred to as an "intelligence test" in the instructions. Additionally, subject behavior replicates one of the most common results in the literature on self-image: subjects are significantly overconfident on average about their relative performance on the test. However, once the context switches from directly reporting beliefs about oneself to taking actions in a search environment that only indirectly reveal beliefs about oneself, all of the predicted differences in search behavior stemming from self-image biases vanish.

The structure of the experiment consists of three stages. In the first stage, subjects take a test of cognitive ability without feedback about whether their answers were correct or not. Subjects are ranked as "High Scorers" or "Low Scorers" based on whether they performed above or below the median score in a group of 10 subjects, but they do not know their own scoring type, their score, or the score of any other subjects. Instead, they must form beliefs about the probability that they are a High Scorer and report those beliefs through an incentive compatible mechanism prior to the second stage. In the second stage, subjects engage in a sequential search task where only one scoring type can receive a wage offer with positive probability. Because subjects know which scoring type can win, but not their own scoring type, rational subjects will engage in Bayesian updating about their own scoring type after each failed search attempt. Theory suggests that subjects should quit search once a threshold belief has been reached (or equivalently, after a threshold number of search attempts, given a prior belief). The third and final stage uses standard methods from the literature to measure subject risk aversion, loss aversion, and demographics.

A between-subjects design with three primary treatments sequentially shuts down information avoidance and then biased updating. In the *Full Self-Image* treatment, both mechanisms may impact search behavior. Signals from search are informative about one's relative performance on the intelligence test, allowing self-serving biased updating to play a role. Additionally, subjects do not find out their scoring type when they quit search; they only learn their type if they win. Therefore, subjects may engage in information avoiding or seeking behaviors by altering their search durations to manipulate the number of informative signals they receive from search. An intermediate *Biased Updating Only* treatment shuts down information avoidance by informing subjects in advance that they will learn their type immediately upon the termination of the search game, regardless of whether they win or decide to quit. This treatment still allows for biased updating because signals from search remain self-relevant. Finally, the *No Self-Image* treatment shuts down both mechanisms by removing the self-relevance of signals altogether. All references to High Scorers and Low Scorers are removed and replaced by explanations that only "Bonus Types" may win additional money from search. Additional design elements hold constant prior beliefs and all other aspects of the search environment, allowing for cognitive errors in Bayesian updating but removing self-image motivations for changing search behavior. This means that other influences on search behavior, such as risk aversion, loss aversion, base-rate neglect, and other cognitive errors or preferences cannot explain the lack of difference in search durations between treatments.

The results of this experiment show that overconfident beliefs do not necessarily lead to overconfident actions. Across many methods of data analysis, the result is strikingly consistent: shutting down channels for self-image to influence search behavior does not impact observed search durations. There are a few ways to interpret these results, and future research will certainly be necessary to disentangle the potential explanations. It could be that self-image does play an important role in real search behavior, but these biases were not strongly activated in this experimental environment. For example, perhaps pride in one's occupation is a stronger motivation to change search behavior than one's relative performance on a short intelligence test. However, this explanation is less credible because the intelligence test employed is the exact one used in previous experiments that did identify differential biased updating between self-relevant and non-self-relevant signals. Still, if it is the case that the biases matter but were simply not activated here, it will be important for future research to identify exactly which aspects of a decision problem cause self-image biases to be strongly activated.

Alternatively, it could be that changing the framing of the decision problem mutes the effects of motivated reasoning. Most previous research on motivated belief updating has used decision problems that remain solely in the domain of reporting beliefs. Changing the context to one in which beliefs are only indirectly revealed through payoff-relevant actions could allow people to take less confident actions without fully internalizing the implications for self-relevant beliefs. In other words, it may be easier for people to act as if they are below average than it is to directly report to someone else that they are below average. Simulation analysis reported in the results section implies that in many cases, subject do search as if they hold prior beliefs closer to the truth, even though they often report overconfident beliefs before the search environment is introduced.

If motivated reasoning is truly muted by contextual aspects of decision problems, there are important implications for both behavioral scientists and policymakers. Behavioral scientists should not assume that reported beliefs will directly map to actions across many decision problems. Instead, they should take into account whether the decision problem causes the link between self-relevant beliefs and actions to become indirect. Self-image concerns will be activated most strongly when directly eliciting beliefs about oneself. The findings of this paper also imply that perhaps policymakers need not be overly concerned about biased updating directly causing suboptimal search behavior. Instead, the indirect nature of search effort decision may mute the impact of motivated reasoning on behavior.

Policymakers may focus more on other behavioral aspects of improving search behaviors and outcomes.

The paper will proceed as follows. Section 2.2 briefly covers the related literature. Section 2.3 presents a model which incorporates self-image concerns into search decisions and derives hypotheses about how these concerns are predicted to impact search duration. Section 2.4 describes the experimental design in detail. Section 2.5 presents the results, including duration analysis that shows subjects do not search differently when self-image-related mechanisms are shut down. This section also uses simulations of subject search to show that observed behavior matches more closely with search from accurate priors than search from reported overconfident priors. Section 2.6 concludes by discussing the implications of these findings for behavioral science and labor market policy.

2.2 Related Literature

This research aims to contribute to the economics literature on labor market search by analyzing the effects of self-image concerns, a topic more prevalent in the psychology literature. The economics literature has recently devoted much attention to the role of various behavioral biases in job search. In experimental work, economists have explained suboptimal search behavior with risk aversion (Cox and Oaxaca, 1989; Miura, Inukai, and Sasaki, 2017), reference dependence (Schunk, 2009; Schunk and Winter, 2009), satisficing heuristics (Caplan, Dean, and Martin, 2011), and subjective wait time costs (Brown, Flinn, and Schotter, 2011). In applied work, economists have also explored the role of present bias (DellaVigna and Paserman, 2005; Paserman, 2008), reference dependence (DellaVigna, Lindner, Reizer, and Schmieder, 2017), and locus of control (Caliendo, Cobb-Clark, and Uhlendorff, 2015). The purpose of this paper is not to deny the important role of these other biases, but to add to this list an emotional concern that has long been salient in the psychology literature on job search: self-image and emotional resilience to the difficult process of search. Barber, Daly, Giannantonio, and Phillips (1994) develop an emotional model of search behavior and argue that accumulating emotional distress can cause job seekers to withdraw from the labor market altogether as a defense mechanism. Chen and Lim (2012) show that optimism and resilience are associated with higher perceived probability of success for job applications. Wanberg, Zhu, and Hooft (2010) show that short-term emotional fluctuations can heavily affect search effort. Numerous papers in psychology have studied mental health degradation among the long-term unemployed (see McKee-Ryan, Song, Wanberg, and Kinicki, 2005, and Paul and Moser, 2009, for literature reviews).

This paper is most directly related to a working paper by Falk, Huffman, and Sunde (2006). They run an experiment in which subjects are sorted into types based on their ability to perform simple multiplication problems within a given time limit. Subjects then choose between a gamble or a safe payoff, with high performers having a higher probability of winning money from the gamble than low performers. Subjects can repeatedly switch between the gamble and the certain payoff over the course of eight periods, and they always learn their type after the eighth period. Their experiment does not incentivize belief elicitation, and as a result the authors have a limited ability to draw conclusions from the belief data. Additionally, the experiment includes no parallel search task devoid of self-image, so the design does not allow one to isolate the role of self-image from other obviously relevant characteristics such as risk aversion, loss aversion, or simple cognitive errors. The experimental results reported in this paper differ from those reported by Falk et al. by enriching the design to isolate multiple self-image biases and answer new questions about the role of self-image in search, in addition to updating experimental procedures with modern incentive-compatible elicitation methods.

A small number of economics papers have directly examined the roles of confidence and learning in search. A recent job market paper by Potter (2018) analyzes the role of learning in intensive margin search effort among unemployed job seekers in New Jersey during the Great Recession. He calibrates a structural model using this data to show that job seekers' effort responds to both positive feedback (i.e., receiving an offer but rejecting it) and negative feedback (i.e., cumulative search effort without offers) as expected. Similarly, Kudlyak, Lkhagvasuren, and Sysuyev (2013) analyze online job board data to show that job seekers tend to apply to high-wage openings first, and over time apply to lower-quality openings. Although it is not their preferred explanation, this behavior is consistent with slow learning about job prospects. Survey evidence has shown that unemployed workers greatly overestimate how quickly they will find new work (Spinnewijn, 2015). This strand of applied literature is an important contribution to our understanding of self-image in job search, but the approach of these papers tends to use reduced form regression analysis or structural models to recover parameters from the data, whereas the approach in this paper is to use experimental methods to causally isolate the impact of search behaviors influenced by self-image concerns.

Finally, this paper relates to the behavioral economics literature on biased belief updating and information avoidance (distinct from the search literature). Three recent papers have shown that self-image concerns cause conservatism and asymmetry in belief updating (Eil and Rao, 2011; Ertac, 2011; Möbius, Niederle, Niehaus, and Rosenblat, 2014). Conservatism refers to updating beliefs less in response to signals that are ego-relevant than equivalent ego-irrelevant signals. Asymmetry refers to weighting negative ego-relevant signals differently from positive ones; Ertac (2011) finds that subjects react more to negative signals, while Eil and Rao (2011) and Möbius et al. (2014) find that subjects react more to positive signals. There is a longer history of interest in information avoidance in the behavioral theory literature (e.g., Bénabou and Tirole, 2002), but this has rarely been applied directly to job search. In a notable exception, Andolfatto, Mongrain, and Myers (2005) discuss job search as one of the potential applications of their theory. Golman, Hagmann, and Loewenstein (2017) provide a more general recent literature review of information avoidance in economics.

2.3 Theory and Hypotheses

This section develops a discrete-time, sequential search model to formally develop hypotheses about the role of self-image concerns in search behavior. The search environment is intended to be simple enough for human subjects to understand in the lab, but rich enough to provide insight about how information avoidance and biased updating would change the search behavior of an agent with self-image concerns.

2.3.1 Basic Model

The model takes place in discrete time over a finite horizon: denote the current period as t and the final period as T. In each period, the agent may choose one of two actions: pay a fixed cost c to search, or quit search permanently.

Before beginning the search problem, agents are classified as either Type A or Type B, but they do not know their own type. Instead, they hold a prior belief about the probability of being Type A. Crucially, Type A agents have a positive probability p of receiving a wage offer in each period of search, but Type B agents have zero probability of receiving an offer.

The wage offer distribution is degenerate, with all offers resulting in a one-time payoff of w, and the game ends when an offer is accepted. Trivially, all wage offers should be accepted because a better wage offer will never arrive. Thus, the strategy of an agent can be fully characterized by the number of periods in which to search before quitting. This is an important simplification compared to typical search models that focus on solving for the optimal reservation wage. The purpose of this simplification is to reduce the search choice to only one dimension: the extensive margin of search effort. Once self-image concerns are introduced, it will be easy to interpret how they affect search behavior because they will either increase or decrease the optimal number of searches.

After each round of search, the agent will perform Bayesian updating on the belief in being Type A. After enough failed search attempts, the belief in being Type A will decrease below a threshold value and the agent will decide to quit search. Formally, the agent should search if the marginal benefit of searching net of the marginal cost is greater than the value of stopping search (normalized to zero):

$$(Belief_A \cdot p) \cdot w + (1 - Belief_A \cdot p) \cdot V - c > 0, \tag{2.1}$$

where $Belief_A$ is the agent's current belief in being Type A, p is the probability of Type A receiving a wage offer of w, c is the cost of search, and V is the option value of being able to search in the next period if no offer is received.

However, this problem can be significantly simplified by using backwards induction. Note that in the final period T, there is no continuation value of search: $V = 0$. The agent should search in the final period if the following reduced inequality holds:

$$(Belief_A \cdot p) \cdot w - c > 0. \tag{2.2}$$

Further, if the agent chooses *not* to search in the final period, then there is also no continuation value in the prior period ($V = 0$ in the period $T - 1$). Thus, the same inequality determines whether it is optimal to search in the prior period. By backwards induction, continuing this line of reasoning means that this simpler inequality identifies the last period of search for the agent. In other words, if failing to receive a wage offer in this period would reduce the agent's $Belief_A$ such that the agent will not search in the next period, then $V = 0$ this period. In the period that determines the agent's strategy – the last period of voluntary search – the above inequality holds true. The intuition here is that because the agent never receives "good news" about the value of search relative to the value of stopping, and there is a finite time horizon, the continuation value never matters for the agent's decision of how many periods to search. [1]

Therefore, the optimal behavior in this model can be characterized by a simple belief threshold. The agent should search as long as his belief in being Type A is high enough.

$$Belief_A > \frac{c}{p \cdot w} \tag{2.3}$$

Using Bayes' rule to decompose $Belief_A$, this threshold could alternatively be expressed as the number of rejections the agent should endure before giving up search:

$$\frac{(1-p)^r \cdot Prior_A}{(1-p)^r \cdot Prior_A + (1 - Prior_A)} > \frac{c}{p \cdot w}, \tag{2.4}$$

where r is the total number of rejections so far (i.e., search attempts without wage offers) and $Prior_A$ is the agent's belief in being Type A before receiving any feedback from the search process.

Rearranging and taking logs provides a closed form solution for the optimal number of rejections before giving up search, as a function of prior beliefs and parameters of the search environment (c, w, and p).

$$r < \frac{\ln(c) + \ln(1 - Prior_A) - \ln(p \cdot w - c) - \ln(Prior_A)}{\ln(1-p)} \tag{2.5}$$

[1] A similar result exists in the literature on search with unknown wage offer distributions – the reservation wage remains static as long as searching never reveals good news about the wage distribution. See Rothschild (1978).

2.3.2 Biased Updating

Previous literature suggests that people update too little when signals are relevant to self-image (conservatism), and that they update even less when the feedback about themselves is negative rather than positive (asymmetry). These biases have been found to exist in addition to other cognitive errors in Bayesian updating such as base rate neglect. To add this ego-related bias to the model in the simplest way possible, the model includes a single new bias parameter, β, which scales up or down the amount of belief updating compared to the rational Bayesian benchmark. In the above expressions in the baseline model, $Belief_A$ may be replaced with $BiasedBelief_A$:

$$BiasedBelief_A = Prior_A + \beta(BayesBelief_A - Prior_A). \tag{2.6}$$

As above, $Prior_A$ is the agent's belief in being Type A prior to receiving any feedback from search. $BayesBelief_A$ is the correct posterior belief calculated through Bayes' rule. The bias parameter β increases or decreases the amount of belief updating that should have occurred. For example, if a rational Bayesian agent would have updated the belief from 80% to 60% after receiving a rejection, then a biased agent with $\beta = 0.5$ would only update half as much and the new $BiasedBelief_A$ would be 70%.

Previous literature justifies the hypothesis that $\beta < 1$.[2] Although this does not change the model solution in terms of a belief threshold at which the agent should quit search (Equation 2.3), it does change the model solution in terms of the optimal number of rejections endured before quitting search (Equation 2.5). An agent who engages in biased updating with $\beta < 1$ will search for more periods (i.e., endure more rejections) before giving up search. The following two hypotheses are direct implications importing the

[2] For example, Mobius et al. (2014) find that subjects update only 35% as much as rational Bayesians would when the signals are relevant to self-image, implying that $\beta = 0.35$ in terms of this model.

behavioral updating biases of conservatism and asymmetry into the search environment.

Hypothesis 1. *Subjects will search for longer when search feedback is relevant to self-image than when it is not relevant to self-image.*

Hypothesis 2. *Subjects will search for longer when search feedback is relevant to self-image and negative than when it is positive.*

2.3.3 Information Avoidance

In a standard rational agent model, the only value of beliefs is to inform decisions that lead to better monetary payoffs. Models of information avoidance instead suppose that agents care about their self-image, such that beliefs about oneself enter directly into the utility function and matter beyond their pure instrumental value in informing payoff-relevant decisions. This is incorporated into the model by adding an additional term called belief utility (or ego utility). As in the previous literature on information avoidance, assume that the belief utility function, μ, is additively separable to utility from wealth. Additionally, assume that μ is monotonically increasing in beliefs, meaning that agents prefer to have higher beliefs about their own ability.

An agent with belief utility will incorporate the potential change in belief as a result of search into the expected marginal benefits and costs of search. An agent with belief utility should search if:

$$\begin{aligned} & Prior_A \cdot p[w + \mu(1) - \mu(Prior_A)] \\ & +(1 - Prior_A \cdot p)[\mu(Posterior_A) - \mu(Prior_A)] - c > 0. \end{aligned} \tag{2.7}$$

In the above expression, $Prior_A$ refers to the agent's belief in being Type A after incorporating all information prior to this round, while $Posterior_A$ is the Bayesian-updated posterior belief after including one additional rejection. Note that the posterior belief in being Type A must be 1 if a wage is offered, since only Type A agents ever receive wage offers. While in previous expressions, the expected value of failing search dropped out because the monetary payoff was zero, in this extension there is always a belief utility cost from failed search due to the downward revision in belief.

How these changes impact search behavior depends on the curvature of the belief utility function. Previous literature typically proposes that belief utility is concave, which leads to a sort of risk aversion with respect to signals about oneself: information avoidance. This loosely means that agents prefer to keep their current beliefs about themselves rather than take the risk of receiving a signal that might revise their beliefs either upwards or downwards. If the concavity is strong enough, agents may avoid free information that might be useful to make decisions that lead to better monetary payoffs. In the search model described here, the concavity of the belief utility function implies that agents prefer to search less than a rational agent would.

Hypothesis 3. *Subjects will search for fewer periods when search feedback is relevant to self-image and when quitting search allows them to avoid revealing further information about themselves.*

The final hypothesis involves how strongly agents engage in information avoidance depending on whether the search environment has a chance of providing a large positive belief revision versus a large negative belief revision. Consider the different implications for an agent winning search and learning they are Type A when being classified as Type A means the agent is especially smart, attractive, charismatic, or hardworking, compared to a situation where being Type A means they are below average on these qualities. In the situation in which Type A is a positive self-image trait, searching involves a small chance of a large upward belief revision upon winning search, but a more likely chance of a small downward belief revision. If the ego utility function is concave, this signal environment will cause weaker information avoidance behavior than the converse environment, where there is a small chance of a large downward belief revision.

This prediction is akin to using risk aversion to explain preferences over monetary gambles. Under standard risk aversion, agents are more averse to gambles with low probability large losses and high probability small gains compared to gambles with low probability large gains and high probability small losses. The same intuition applies to concavity over ego utility and informs Hypothesis 4.

Hypothesis 4. *In a search environment that allows information avoidance of self-relevant signals, subjects will search fewer times when winning search reveals they are a Low Scorer than when winning search reveals they are a High Scorer.*

Of course, subject search behavior should differ based on their prior beliefs in being the winning type. The above hypotheses refer to predicted differences in search behavior even after controlling for prior beliefs and other observable characteristics of subjects.

2.4 Experiment Design

The hypotheses about self-image concerns in job search are tested using a between-subjects design with three treatments: *Full Self-Image*, *Biased Updating Only*, and *No Self-Image*. These treatments should be compared sequentially because they are designed to "turn off" one self-image-related behavior at a time. The *Full Self-Image* treatment allows for both biased belief updating and information avoidance. Signals from search are informative about subjects' relative performance on the intelligence test, providing motivation for biased updating. Subjects who quit search before winning will never learn whether they were a "High Scorer," allowing for information avoidance. The *Biased Updating Only* treatment retains the self-relevance of signals, but removes the possibility of information avoidance. This is accomplished by informing subjects that their scoring type will be revealed immediately after search ends, regardless of whether they win or quit. The *No Self-Image* treatment is a control treatment designed to turn off self-image concerns altogether, so that neither biased belief updating nor information avoidance can be driving search behavior. Because the control treatment allows for all remaining biases and preferences to impact behavior besides self-image, pairwise comparisons of search durations between these treatments separately identify each mechanism. Differences in search duration between *Full Self-Image* and *Biased Updating Only* should be caused by information avoidance, and differences between *Biased Updating Only* and *No Self-Image* should be caused by self-serving updating biases.

In addition to the three treatments, two signal strength conditions varied the search environment parameters: *Weak Signals* and *Strong Signals*. Upon analyzing the data from an initial set of sessions which showed no statistically significant differences in search duration between treatments, it was hypothesized that the self-relevant signals were too weak to induce self-image concerns that were detectable in search behavior. Therefore, the *Biased Updating Only* and *No Self-Image* treatments were rerun with much stronger self-relevant signals, such that a rational agent engaging in Bayesian updating would update beliefs about the intelligence test score much more after each search attempt. Details of the parameters in the *Weak Signals* and *Strong Signals* search environments are explained below.

In all treatments, the design consists of three stages. In the first stage, subjects answer a subset of questions from the Raven's Advanced Progressive Matrices test, which is a culturally neutral intelligence test based on complex pattern recognition.[3] To provoke

[3]For the purposes of this design, it is not crucial whether the IQ test is truly an accurate measure

self-image concerns, subjects are informed that the questions are commonly used as an intelligence test, and it is explicitly labeled as such in the interface heading. Subjects have 15 minutes to answer as many of 20 questions correctly as they can, and they are paid \$0.10 per correct answer. However, they receive no feedback whatsoever about whether their submitted answers are correct, or how well other subjects scored. Additionally, they do not learn their payments until the end of the experiment, and even then, the payment information is displayed as a total and is not itemized by task. Subjects are informed that their score will also matter later in the experiment, but are not given details about future tasks at this point.

The intelligence test is used to classify subjects as either "High Scorers" or "Low Scorers." Specifically, this means that the subject performed either above or below the median score (i.e., total correct answers) in a group of 10 randomly selected subjects. Immediately upon the conclusion of the intelligence test, these scoring types are explained and prior beliefs in being a High Scorer are elicited. Subjects are asked to report the probability that they think they are a High Scorer, and they are incentivized using the binarized scoring rule, a standard belief elicitation method that is robust to risk aversion (Hossain and Okui, 2013). It is emphasized in the instructions that reporting beliefs accurately maximizes their expected earnings. Additionally, full details of the binarized scoring rule payment mechanism are provided in a drop-down text box, for subjects who decide to review the details. In this drop-down box, the mechanism is explained based on the language recommended by Wilson and Vespa (2018).

The second stage consists of the search task. The task is thoroughly explained on an instructions screen, and subjects are not allowed to proceed to the next screen until after the instructions have additionally been read out loud by the experimenter. Before the

of cognitive ability, whether it is culturally neutral, or other related issues. What is important is that subjects care about their performance relative to others, so that self-image becomes salient.

subjects proceed, there is also a period for subjects to ask private questions to ensure they understand the upcoming task.

The parameters of the search environment differ by the signal strength condition, but the structure of the task closely mirrors the model laid out in the theory section above. The *Weak Signals* search environment used the following parameters. Subjects may spend \$0.10 for a chance to win a one-time bonus of \$4.00. If they are the scoring type that can win the bonus, there is a 10% chance of winning the bonus on each try. If they are not the type that can win the bonus, there is a 0% chance of winning. Subjects are given an additional balance of \$2.00 to spend on search, meaning that they can search at most 20 times before exhausting these funds. The *Strong Signals* environment had the following differences. If subjects are the scoring type that can win the bonus, there is a 25% chance of winning the bonus on each try. To hold the expected value of search constant, the cost of one search is \$0.25. As a result, subjects in the *Strong Signals* condition can search at most 8 times before exhausting their balance.

The task is explained as a chance to try to win a bonus, and does not include any references to search that might cause framing or demand effects. Two large buttons appear on the screen during the task: Stop and Try. When subjects click Try, they spend the search cost for a chance to win the bonus. If they do not win the bonus, the screen updates to display their new balance of funds and the number of tries so far. If they do win the bonus, the task immediately ends and they are informed that they won. Finally, when subjects click Stop, the task immediately ends. Subjects keep any leftover search funds whether the game ends due to winning or quitting.

The search instructions and task differ by treatment. In the *Full Self-Image* treatment that allows for both biased updating and information avoidance, subjects were told "If you click Stop, we will not tell you your type, but later on there may be a chance to pay to learn your type. If you win the bonus, you will know your type." In the *Biased*

Updating Only treatment, motivations for information avoidance are removed with the following statement: "You will immediately learn whether you were a High Scorer or Low Scorer when the task ends, whether you win the bonus or stop trying to win the bonus." In the *No Self-Image* treatment, there are no references to High Scorers or Low Scorers. Instead, the relevant portions of the instructions are replaced with "Bonus Type" or "not Bonus Type." In this treatment, subjects are also told whether they are a Bonus Type or not after the task, whether they quit or win, in order to hold constant any motivations for search such as pure curiosity or regret aversion between the latter two treatments.

Additionally, the design randomizes at the subject level whether it is High Scorers or Low Scorers who are able to win the bonus, and subjects know which type can win. In the two treatments with self-image concerns, the software randomizes which type of scorer can win, with equal chance for each type. Thus, subjects know which type can win but not their own type, and must update beliefs about their own type as they search but fail to win the bonus.

In the *No Self-Image* treatment, the software randomizes whether subjects are the Bonus Type or not using the subject's own reported prior belief in being a High Scorer. However, to construct a mathematically identical search environment, there is an equal chance of the software using the reported prior or one minus the reported prior (i.e., the opposite belief) as the chance of being the Bonus Type. The case of using one minus the reported prior as the chance of being the Bonus Type is parallel to the case in the self-image treatments in which Low Scorers can win. Subjects are told the randomization probability used in determining whether they are a Bonus Type or not.

The purpose of this mechanism is two-fold. First, it enables testing hypotheses about biased updating and information avoidance relating to receiving positive versus negative signals from search. Second, it allows an analogue control treatment to hold prior beliefs about being the type that can win constant between treatments, without using honestly reported beliefs against the subjects (which borders on deception). To see how omitting this mechanism would be problematic, consider that a subject who dishonestly reported a 100% prior belief would be guaranteed to be the Bonus Type, while a subject who honestly reported a 25% prior belief would be unlikely to be the Bonus Type. The mechanism avoids disincentivizing subjects from reporting beliefs honestly in future experiments because their expected earnings in the search task are not systematically impacted by the beliefs they report.

The third and final stage of the experiment consists of a posterior belief elicitation, additional preference elicitations, and a demographics survey. Posterior beliefs in being the High Scorer type (or the Bonus Type in the *No Self-Image* treatment) are elicited only for subjects who did not win search, since subjects who did win already know which type they are. The posterior belief is incentivized using the same method described above, and only one of the two beliefs is randomly selected for payment (with equal likelihood of the prior and posterior being selected for payment). Additionally, in only the *Full Self-Image* treatment, subjects who do not win search are offered a price list for learning their scoring type. The prices range from positive to negative $2.00 in increments of $0.25 (i.e., they are offered both to pay and to receive money to learn their type). One row of the price list is randomly selected for implementation. This was added to provide some direct evidence on information avoidance or information seeking behaviors. The preference elicitations include standard price lists for evaluating risk and loss aversion, which are used as control variables in regression analysis, but cannot be driving any treatment differences because parameters relating to risk and loss are not varied between treatments. Finally, there is a demographics survey and a text box for free-form feedback. All subjects are paid at the same time at the end of the experiment and not allowed to leave early if they finish the tasks quicker, to avoid introducing waiting time costs into the search decision.

Two other differences exist between the *Weak Signals* and *Strong Signals* conditions. First, in the *Strong Signals* condition, subjects are asked their intended number of searches before they begin receiving signals. Subjects still proceed through the normal search task, but the software implements either the initial strategy or the observed actions during search with equal chance at the end of the search task. The purpose of this mechanism is to collect at least some uncensored data on intended search duration. Second, the *Strong Signals* condition includes repeated belief elicitation during the search process. After each failed search attempt, subjects are asked to report their new belief in being a High Scorer. This allows the analysis of posterior beliefs of all subjects in the condition, rather than only subjects who quit search before winning, as in the *Weak Signals* condition.

2.5 Results

This section will analyze data from 15 experimental sessions, with 3 sessions each for the following 5 treatment-condition combinations: *Full Self-Image (Weak Signals)*, *Biased Updating Only (Weak Signals)*, *Biased Updating Only (Strong Signals)*, *No Self-Image (Weak Signals)*, *No Self-Image (Strong Signals).* In total, 275 subjects participated, with between 50-60 subjects in each treatment-condition combination. Each subject participated in only one session. All sessions were run at the University of California, Santa Barbara Experimental and Behavioral Economics Laboratory. The experiment software was coded in oTree (Chen, Schonger, and Wickens, 2016). The average subject payment (including the show-up payment of $5) was $12.48, with a minimum of $7.25 and a maximum of $18.25. The typical session took about 1 hour, and no session took longer than 1 hour and 30 minutes.

Tables 2.1-2.3 present subject characteristics by treatment. These treatment balance tables include Raven's Matrices score, reported prior belief in being a High Scorer, risk aversion and loss aversion measures, and a variety of demographic characteristics. Table 2.1 compares subject characteristics between the *Full Self-Image* and *Biased Updating Only* treatment, while Table 2.2 compares *Biased Updating Only* to *No Self-Image.* Table 2.3 compares subjects in both *Strong Signal* treatments to both *Weak Signal* treatments. Across these comparisons, the only significant differences are in demographic characteristics between *Biased Updating Only* and *Full Self-Image.* The difference in age is driven by a small number of older subjects in the *Full Self-Image* treatment, and the differences in race distribution are not expected to impact behavior (the coefficients on race are insignificant in all regressions).

Turning to subject search behavior, Figures 2.1 and 2.2 show histograms of the total number of searches in the *Weak Signals* and *Strong Signals* conditions, respectively. Approximately 11% of subjects choose to search 0 times, while about 10% of subjects choose to search the maximum number of times. There are spikes in all treatments at half of the maximum possible number of searches – apparently a focal point for subjects (about 17% of subjects choose to search half the maximum number of times). The number of searches appears more evenly spread in the *Strong Signals* treatments, where the maximum number of searches was only 8. The discrete time logistic hazards model will be used to control for these spikes in certain periods, as explained below. Also note that these figures display the observed number of searches, which may be a censored measure if the subject won the bonus before deciding to quit searching. Standard methods from duration analysis will be used to handle the censored data.

Figures 2.3 and 2.4 show the empirical CDF of observed searches separately by self-image treatment. Figure 2.3 shows the three self-image treatments with *Weak Signals*, and Figure 2.4 shows the two self-image treatments with *Strong Signals.* Again, these figures display all observed search durations, regardless of whether search ends due to the subject's decision to quit or the subject winning the bonus and thus having a censored intended search duration. There do not appear to be large differences in search durations between treatments. This visual appearance is confirmed with simple t-Tests on the null hypothesis that observed mean search durations are equal between treatments: *Full Self-Image, Weak Signals* vs. *Biased Updating Only, Weak Signals* (p = 0.42), *Biased Updating Only, Weak Signals* vs. *No Self-Image, Weak Signals* (p = 0.30), and *Biased Updating Only, Strong Signals* vs. *No Self-Image, Strong Signals* (p = 0.11). Further statistical tests which properly account for censoring will be reported in the Duration Analysis subsection below.

One search measure which is not subject to censoring is the proportion of subjects who choose to search zero times. Two-sample tests of proportions were run to detect differences in this proportion between treatments or between which scoring type can win within treatment, but no statistically significant differences were detected among any of these comparisons. This implies that manipulating the self-image mechanisms in play does not impact the decision to participate in search.

2.5.1 Search Duration Analysis

When a subject wins the bonus, the latent strategy of the intended number of searches is unobservable. To account for this right-censoring of the search data, standard methods of duration analysis are used. First, Kaplan-Meier estimated survival functions are plotted for visual comparisons, and log-rank tests are performed to test for equality of these functions. Second, Cox proportional hazards regression models are run to test for equality of search durations between treatments while controlling for various subject characteristics such as intelligence test score and risk aversion. Third, the discrete time logistic hazards model is used to control for the spikes in quit rates in certain focal periods.

Figures 2.5 and 2.6 display the Kaplan-Meier estimated survival functions by treatment for the *Weak Signals* and *Strong Signals* versions, respectively. The only apparent visual difference is that subjects search for slightly longer in the *No Self-Image (Weak Signals)* treatment compared to the other *Biased Updating Only (Weak Signals)* treatment, which would indicate self-relevant belief updating may cause subjects to update beliefs more (and search for shorter durations) compared to the self-irrelevant context. However, this comparison is not significant in a log-rank test of equality of survival distributions (p = 0.1538). Additionally, the other two relevant comparisons of estimated survival functions between treatments are not significantly different with a log-rank test: *Full Self-Image (Weak Signals)* vs. *Biased Updating Only (Weak Signals)* (p = 0.8547), *Biased Updating Only (Strong Signals)* vs. *No Self-Image (Strong Signals)* (p = 0.6434). It would not be informative to compare survival distributions between High Scorers Win and Low Scorers Win conditions within treatment due to the systematically overconfident prior beliefs driving the differences in search behavior between those groups. Instead, these comparisons designed to reveal asymmetric updating and asymmetric information avoidance are reserved for multivariate analysis that controls for prior beliefs.

The Cox proportional hazards model is used to test for treatment differences in search duration while controlling for various subject characteristics. Tables 2.4-2.9 display the results of these regressions. These tables show the estimated hazard rates, with hazard rates above one indicating a higher likelihood of quitting search earlier. P-values are displayed in parentheses below each hazard rate, indicating whether the rate is statistically significantly different from one. Each column in the tables introduces additional control variables. Column 1 includes an indicator variable for treatment, an indicator variable for whether High Scorers or Low Scorers can win the bonus, a continuous integer ranging from 0-100 for the subject's prior belief in being a High Scorer, the interaction between treatment and whether High Scorers can win, and an interaction between prior belief in being a High Scorer and whether High Scorers can win. Column 2 additionally includes the subject's intelligence test score (ranging from 0-20), a proxy for risk aversion ranging from 1-6 (from a standard multiple price list offering sequentially riskier gambles), and a proxy for loss aversion ranging from 1-7 (again from a standard multiple price list decision). Column 3 additionally includes demographic controls for gender, race, and household income (prior to university attendance for students), although these coefficients are suppressed to improve the legibility of the tables.

Table 2.4 tests whether subjects searched different durations between the *Full Self-Image (Weak Signals)* and *Biased Updating Only (Weak Signals)* treatments. Only observations from these two treatments are included in the regression sample, so a significant coefficient on the treatment indicator variable would provide evidence of information seeking or information avoidance behavior, depending on whether the hazard rate is greater than 1 (indicating a higher likelihood of quitting search earlier) or less than 1 (indicating a lower likelihood of quitting search earlier). Across the specifications, the hazard rate on the treatment indicator variable is not significantly different from 1.

Tables 2.5 and 2.6 test whether subjects searched differently in the *Biased Updating Only* and *No Self-Image* treatments, in the *Weak Signals* and *Strong Signals* conditions, respectively. If the hazard rate on the *Biased Updating Only* indicator was significantly greater than 1, it would indicate that conservatism in self-relevant belief updating was causing shorter search durations in the *Biased Updating Only* treatments. However, there is no evidence of conservatism impacting search durations apparent in these results.

Table 2.7 tests whether subjects change their information avoidance or seeking behavior in response to the signal type. Recall that when High Scorers can win the bonus, there is high probability of a minor belief adjustment downward about oneself, but a low probability of a large belief adjustment upward. When Low Scorers can win, there is a chance of a large belief adjustment downwards – a possibility that would strengthen information avoidance if subjects are risk averse with respect to signals about themselves. Table 2.7 runs the same Cox regression specifications but restricts the sample to only the *Full Self-Image, Weak Signals* treatment. Because the hazard rate on the indicator variable for High Scorers Win is not statistically significantly different from 1, there is no evidence to support different levels of information avoidance in response to signal type.

Tables 2.8 and 2.9 test whether asymmetry in self-relevant belief updating affects search durations. Table 2.8 restricts the sample to only the *Biased Updating Only, Weak Signals* treatment, while Table 2.9 includes only the *Biased Updating Only, Strong Signals* treatment. If subjects engage in self-serving positive asymmetric updating, they update beliefs less in response to negative signals than positive ones. This means that beliefs will be resistant to updating when High Scorers win, where failing to win the bonus provides negative signals about intelligence. This positive asymmetric updating would be reflected in estimated hazard rates less than 1 on the indicator for High Scorers Win in these tables (i.e., subjects are less likely to quit because they are resistant to update beliefs about their chance of winning). On the other hand, a hazard rate greater than 1 would indicate negative asymmetric updating: subjects who receive negative signals update beliefs more and are more likely to quit earlier. In the Weak Signals condition in Table 2.8, there are no significant differences in hazard rates between High Scorers Win and Low Scorers Win conditions, providing no evidence of asymmetric updating impacting search behavior. In the Strong Signals condition in Table 2.9, there is initially evidence of negative asymmetric updating in the first column; however the hazard rate is unreasonably high and the significance vanishes once risk and loss aversion are added as controls, indicating these characteristics are important in explaining subject behavior. Overall, these regressions do not support the hypothesis that asymmetric updating affects search behavior.

These regression models may be sensitive to the impacts of large quit rates in specific periods (e.g., the halfway point to the maximum number of searches), so a discrete time logistic hazards model is run to allow the inclusion of a full set of period fixed effects. Although the coefficients on these period fixed effects are suppressed to improve legibility of the tables, the fact that certain periods have consistently significant impacts on hazard rates indicates that these fixed effect regressions capture an important phenomenon: subjects are more likely to quit in certain focal periods. The results are included in Tables 2.10-2.15, and their structure exactly mirrors the hypotheses tested in order in Tables 2.4-2.9. Table 2.10 tests for information avoidance or information seeking behavior by comparing the *Full Self-Image* treatment to the *Biased Updating Only* treatment. Tables 2.11 and 2.12 test for biased belief updating affecting search behavior, separately in the Weak Signals and Strong Signals environments, respectively. Table 2.13 tests for differential information avoidance behavior between High Scorers Win and Low Scorers Win signal environments within the *Full Self-Image* treatment. Tables 2.14 and 2.15 test for asymmetric updating impacting search behavior within the *Biased Updating Only* treatment under Weak Signals and Strong Signals, respectively.

In these discrete time logistic hazards model results, there are again no significant differences in search durations between treatments, indicating that shutting down the channels of information avoidance and biased updating does not significantly change search behavior (see Tables 2.10, 2.11, and 2.12). Additionally, there is no evidence that the winning scoring type impacts search durations when information avoidance or seeking behavior is possible (i.e., no evidence of risk aversion with respect to signals about oneself – see Table 2.13).

However, within the *Biased Updating Only (Weak Signals)* treatment in Table 2.14, there is a statistically significant reduction in the log odds of quitting among subjects in the High Scorers Win regime compared to the Low Scorers Win regime. Because the specification already controls for subject prior beliefs in being the winning type, this is suggestive evidence of self-serving, positive asymmetric updating: subjects update less in response to negative signals about themselves than positive ones. In other words, subjects under the High Scorers Win condition are reluctant to update their beliefs about their intelligence score downwards, so they are less likely to quit in a given period than subjects under the Low Scorers Win condition, who are updating beliefs about their intelligence score upwards. This difference is only detected under the *Weak Signals* comparison in Table 2.14 and not under the *Strong Signals* comparison in Table 2.15, suggesting that this difference may be sensitive to signal strength or the maximum possible duration of search (8 versus 20 periods). Of course, these results should be interpreted cautiously – it may be a spurious finding given the many regression models run in this analysis. Additional evidence on asymmetry and other hypotheses about belief updating will be gathered by examining reported subject beliefs directly in the next subsection below.

Tables 2.16 and 2.17 pool the data from the *Weak Signals* and *Strong Signals* sessions while controlling for an interaction between period fixed effects and the signal environment. This interaction effect allows pooling the data despite the difference in the maximum possible number of periods (8 versus 20 periods). Even with the additional power from pooling these sessions, no treatment difference in search duration is found between *Biased Updating Only* and *No Self-Image* (Table 2.16). Additionally, evidence of asymmetric updating within the *Biased Updating Only* treatment is absent in the pooled regressions, suggesting that the previous finding of positive asymmetric updating is either specific to *Weak Signals* or a spurious result.

One additional set of discrete time logistic hazards models is run to check for differences in search durations between treatments (although the results are not displayed in tables here). Rather than retaining subjects under both the High Scorers Win and Low Scorers Win conditions and controlling for the condition, these regressions compare treatments while only retaining subjects under the same winning scoring type. For example, one regression compares subjects under *Full Self-Image* High Scorers Win and *Biased Updating Only* High Scorers Win while omitting all subjects under the Low Scorers Win condition. This is repeated for each treatment and winning condition combination: *Full Self-Image* Low Scorers Win and *Biased Updating Only* Low Scorers Win; *Biased Updating Only* High Scorers Win and *No Self-Image* High Scorers Win; *Biased Updating Only* Low Scorers Win and *No Self-Image* Low Scorers Win. The specifications include the full set of control variables (period fixed effects, prior beliefs, risk and loss preferences, intelligence test scores, and demographics). Across all of these results, no significant differences in search durations between treatments is detected, again implying that information avoidance and conservatism updating are not impacting search durations.

2.5.2 Beliefs

On average, subjects are overconfident in being a High Scorer in their reported prior beliefs. The median prior belief in being a High Scorer is 75%, and the mean is 71.9%. Figure 2.7 displays the empirical CDF of reported subject prior beliefs.

Subject beliefs are qualitatively accurate in the sense that more extreme beliefs are more likely to match the truth. For example, among the small number of subjects who reported a prior belief in being a High Scorer of less than 50%, about three-quarters of them were actually low scorers. Among subjects who reported a prior belief greater than 50%, 54% of them were actually High Scorers; for beliefs greater than or equal to 75%, 64% were actually High Scorers; and for beliefs greater than or equal to 90%, 70% were actually High Scorers.

Next, this analysis will turn to subject belief updating. Some subjects updated incorrectly in the sense that either their beliefs did not change after receiving signals from search, or their beliefs changed in the incorrect direction. About 27% of subjects updated incorrectly, and this proportion does not significantly differ between self-image or non-self-image treatments (two-sided test of proportions, p = 0.80). These subjects will be omitted from the following analysis, along with 36 subjects who reported prior beliefs of 100% since a rational Bayesian will not update from this prior. (Despite the Bayesian implication that a prior belief of 100% will never be updated, more than half of these subjects did revise their reported beliefs after feedback.)

Tables 2.18 shows the magnitude of error in belief updating. The first row displays the mean signed difference between the rational Bayesian posterior (updating from the subject's reported prior) and the reported posterior. The value of this difference is the reported posterior minus the Bayesian posterior, in decimal form (e.g., a value of -0.09 indicates that the mean reported posterior belief was 9 percentage points below the mean Bayesian posterior belief). Standard deviations are displayed in parentheses. The first column indicates that on average, in the *No Self-Image* treatment, subject posterior beliefs ended up below the rational Bayesian posterior beliefs. Since all subjects in this treatment update beliefs downwards, this implies that they updated too much. However, a signed-rank test fails to reject the null that the difference is zero (p = 0.73). Column 2 shows the signed difference for pooled self-image treatments, but this masks the fact that High Scorers Win subjects are updating beliefs downwards while Low Scorers Win subjects are updating beliefs upwards (so the sign is less informative in this column). Columns 3 and 4 indicate that subjects updating beliefs downwards about their intelligence tend to update too much (although this is insignificant with a signed-rank test of p = 0.12), and that subjects updating beliefs upwards about their intelligence also tend to update too much (p = 0.001). This does not provide strong evidence of either conservatism or asymmetric updating, at least when looking at subjects' reported posterior beliefs compared to their reported prior beliefs.

The second row of Table 2.18 shows the absolute value of the difference between reported and Bayesian posteriors. All four differences in this row are significantly difference from zero with signed-rank tests ($p < 0.001$ in all cases). This is evidence that subjects update significantly differently from rational Bayesian updating under all treatment conditions. Interestingly, subjects tend to update closer to the Bayesian benchmark when feedback is positive about themselves (Low Scorers Win, shown in column 4). A Wilcoxon rank-sum test finds that the absolute value of the difference between Bayesian and reported posterior beliefs is significantly different between the High Scorers Win and Low Scorers Win conditions ($p = 0.009$), indicating that subjects update significantly closer to Bayesian when self-relevant feedback is positive than when it is negative.

Table 2.19 shows the proportion of subjects who update too little versus too much in each subgroup. All subjects are classified as updating either too much or too little in this table because no subject exactly matched the Bayesian belief. It appears that most subjects generally updated too much, but that subjects in the Self-Image treatments with Low Scorers Win were more evenly split between updating too much or too little. Tests of proportions do not indicate differences at the 5% confidence level between *No Self-Image* and *Self-Image* treatments ($p = 0.06$) or between High Scorers Win and Low Scorers Win within the self-image treatments ($p = 0.0503$). Still, the table shows the heterogeneity in updating styles: most subjects appear to overreact to negative signals about themselves when updating beliefs, but a sizable minority does the opposite.

To better understand this heterogeneity in subject belief updating, another measure of updating bias is constructed: the difference between the reported posterior and the reported prior, divided by the difference between the Bayesian posterior and the reported prior. This is the simple bias parameter that scales updating, as introduced to the model in equation 2.6 of the theory section above. The value is informative about the proportion of updating a subject performed compared to how much a Bayesian would have updated.

For example, if it takes a value of 2, then the subject updated twice as much as Bayesian updating implied. If it takes a value of 0.5, then the subject updated half as much as a Bayesian would.

Figures 2.8 through 2.10 display empirical CDFs of this bias parameter, split by various subgroups. Figure 2.8 shows the belief bias parameter by Strong Signals versus Weak Signals environments. Because the CDFs nearly overlap, this provides evidence that subjects generally understand that stronger signals call for larger belief updates. Confirming previous analysis, most subjects update too much because most of the mass is to the right of the bias parameter value of 1. (Note that outlier subjects who update more than 3 times as much as Bayesian updating implied are top-coded at a value of 3 to improve the display of these figures.) Figure 2.9 shows the bias parameter split by self-image treatments versus the *No Self-Image* treatment. Across both contexts, the majority of subjects still update too much compared to the Bayesian benchmark, but the mass is shifted slightly to the left when self-image is involved, suggesting that subjects update slightly less with self-relevant signals than without (i.e., conservatism in updating). However, this visible difference is not statistically significant (rank-sum test, $p = 0.14$). Figure 2.10 shows the bias parameter split by High Scorers Win versus Low Scorers Win, only keeping subjects in the self-image treatments. This figure shows that subjects update less in response to positive signals than in response to negative signals because the mass of the bias parameter is shifted to the left for Low Scorers Win subjects. However, this difference is not statistically significant using a rank-sum test ($p = 0.16$).

2.5.3 Information Avoidance and Information Seeking

Subjects in the *Full Self-Image* treatment who did not learn their type by winning search were given the opportunity to pay or receive some money to learn whether they were a High Scorer. Figure 2.11 displays a histogram of the willingness to pay or accept money to learn one's scoring type, where the dollar values range from negative to positive \$2.00 in increments of \$0.25. Negative values mean subjects are willing to pay that amount to learn their type, implying information seeking behavior. Positive values mean subjects are unwilling to receive that amount of money while learning their type, implying information avoidance behavior (i.e., they would rather leave the money on the table than learn their type). Strikingly, only six subjects displayed information avoidance, while 17 subjects displayed (mostly mild) information seeking behavior and 16 subjects were indifferent between learning or not learning their type. This willingness to pay measure does not vary by whether High Scorers or Low Scorers can win the bonus: splitting out these groups for a visual comparison does not change the qualitative results, and a t-Test on the equality of willingness to pay between which scoring type can win does not indicate a significant difference. Additionally, even subjects who displayed information seeking or avoiding behaviors were only willing to accept or pay very small dollar amounts compared to the money at stake in the rest of the experiment – typically much less than \$1. Overall, direct evidence for either information avoidance or information seeking behavior from the willingness to pay measure is weak.

2.5.4 Strategy Method Search

In the *Strong Signals* treatments, the strategy method was implemented for use in supplemental analysis. Subjects reported how many times they would like to search at most without winning the bonus before quitting, and this choice was implemented with 50% chance. The purpose was to gather data on subject search intentions that was not vulnerable to any data censoring issues. Table 2.20 runs an ordinary least squares regression on the search strategy, and the indicator variable for treatment is insignificant across all specifications. However, this may be expected because the feedback from a hypothetical situation is less visceral than actually receiving real-time feedback relevant to self-image. Additionally, Table 2.21 shows a similar regression on the *Biased Updating Only* treatment to test for asymmetric belief updating affecting search durations. The first specification shows a significant negative effect on search duration when High Scorers can win, after controlling for the interaction between prior beliefs and which type can win. However, this effect disappears in richer specifications that control for risk and loss aversion. In sum, the strategy method echoed previous results that self-image does not affect search durations.

2.5.5 Simulated Belief Formation and Search

The data analysis so far has taken subjects' reported initial beliefs in being a High Scorer at face value and compared subject behavior to the rational Bayesian benchmark. However, it is possible that self-image biases are activated differently in different contexts. For example, subjects might adjust their prior beliefs even before viewing any feedback from the search process, just because the new context has activated a different frame of mind.

This subsection presents additional evidence on initial belief formation and the resulting search behavior. What is the magnitude of error in subjects' reported prior beliefs in being a High Scorer, given their performance? Do subjects select search durations as if they are using their initial reported beliefs, or as if they are using more accurate beliefs? Simulations are run to address these questions. In summary, the following steps are taken: a proxy for correct prior beliefs is generated for subjects based on their intelligence test scores; subjects engage in simulated rational search with Bayesian updating, starting from either the "correct" prior or the reported prior; simulated search durations from each starting prior are compared to observed actual search behavior from the experiment.

The first part of this analysis is generating a proxy for correct prior beliefs in being the High Scorer type for each subject. One iteration for one subject involves the following steps. First, twenty draws are taken from a binomial distribution with a success rate equal to their percent of correct answers on the actual Raven's Progressive Matrices test. This counted as the subject's number correct for this iteration. Second, 9 test scores from other subjects are randomly drawn to form the subject's group of counterparts. The simulated test score is compared to the median of the group to determine whether the subject was a High Scorer for this iteration of the simulation. This algorithm was then repeated for 1,000 iterations for each subject, and the proportion of times they were classified as a High Scorer was used as a proxy for correct prior beliefs in further analysis. The mean of this proxy for correct beliefs is close to 50%, and there is substantial variation in simulated beliefs, indicating that this procedure has captured a proxy that is much closer to the truth than reported prior beliefs were.

The goal of the analysis is to compare two potential search strategies to observed behavior: first, search with Bayesian updating from the "correct" prior belief, and second, search with Bayesian updating from the reported prior belief. However, subject prior beliefs imply an optimal *maximum* number of search attempts before quitting. This is an unobserved strategy in the experiment due to censoring: some subjects win the wage and drop out of the sample, so their full strategy is never observed. To make the two uncensored search strategies directly comparable to censored observed search durations in the experiment, the strategies must be run through the experimental environment and subject to censoring under the right conditions.

The next step of the analysis is to run the subject strategies through 1,000 iterations of the experimental search environment, subject to censoring at the parameters from the experiment (but only if subjects were the type that could win). For example, suppose that the reported prior belief of a subject implied an optimal strategy of 15 searches at most, and this subject happened to be a type that could win search in the Weak Signals environment. This subject is run through the search environment with a 10% chance of winning the bonus on each sequential try for up to 15 tries, and the resulting number of observed searches was stored in the data over 1,000 iterations. The mean number of censored observed search tries is stored. Thus, three final numbers exist for each subject: the actual number of searches observed in the experiment, the simulated observed number of searches with Bayesian updating starting from the reported prior belief, and the simulated observed number of searches with Bayesian updating starting from the proxy for the correct prior belief.

The results are displayed in Tables 2.22 and 2.23. Each cell in the table displays a mean number of observed searches for the given subgroup of that row. The first column shows the mean number of searches observed in simulations of subjects starting from their simulated "correct" prior belief. The second column shows the actual mean searches observed in the experiment. The third column shows the mean number of searches observed in simulations of subjects starting from their reported prior belief. Table 2.22 shows these results for the *Weak Signals* treatments, where the action space ranged from 0 to 20 searches; Table 2.23 shows the results for the *Strong Signals* treatments, where the action space ranges from 0 to 8 searches. Stars in between columns indicate statistically significant differences between the means (using a two-sided t-Test of equality).

The results show that in many cases, subjects search closer to as if they are using correct prior beliefs than as if they are using their reported beliefs. This is most apparent in the *Weak Signals* environment in the first row of Table 2.22, where all subjects searched significantly differently from optimal search using their reported prior, but not significantly differently from optimal search using the simulated "correct" prior. There is also evidence for this in the *Strong Signals* environment - there are no statistically significant differences between optimal search with the correct prior and actual search among any subgroups, while there are some significant differences between optimal search with reported priors and actual search among certain subgroups.

Turning to the treatment subgroups isolated in each row, the most consequential difference is among the subjects searching where High Scorers Win in the *Self-Image* treatments. Across both the *Weak Signals* and *Strong Signals* search environments, subjects search significantly less than their overconfident stated prior beliefs suggest they should. In the *Weak Signals* environment, they adjust actual search partly downwards towards optimal search under a correct prior, but they still search significantly more than correct priors would suggest. In the *Strong Signals* environment, subjects adjust search downwards enough that there is little difference between actual search and optimal search under correct prior beliefs. Although there are some other differences apparent in the tables, this is the most consistent and important one: subjects state highly confident beliefs in being a High Scorer before being introduced to the search task, but then behave in the search task as if they have much more accurate beliefs about their scoring type.

2.6 Conclusion

Labor economists have recently devoted much attention to incorporating behavioral biases into models of labor market search. The innate self-relevance of signals received during job search is highly suggestive that biases related to self-image should play a role in behavior. However, the results of this experiment are broadly unable to identify any significant impact of self-image on search behavior. This finding occurs despite the obvious relationship between search outcomes and beliefs about one's relative intelligence in the experimental search environment. If biases stemming from self-image concerns do not manifest themselves in search behavior in such a simplified search environment, what are the implications for real labor market search?

On one hand, it is possible that self-image *does* influence the behavior of job seekers in important ways, but these biases were simply not activated in a strong way in this particular experimental environment. This could have happened for a number of reasons. Subjects may not have believed that the pattern recognition test was a true measure of intelligence, despite its framing as such in the instructions and the interface. Real job-seekers may feel a deeper sense of pride and identity with respect to their occupation than subjects felt in this short experiment. Perhaps self-image has a greater impact on search behavior when search takes place over the course of months rather than hours. Of course, a challenge for this explanation is that bias in belief updating has been consistently identified in short lab experiments that focus only on beliefs about one's ranking on intelligence tests, and even in experiments that used the same intelligence test as this experiment did. Additionally, the typical result from the literature of initial overconfidence about relative test scores was replicated here.

On the other hand, the results may imply that self-serving belief updating and information avoidance do not drive real job search behavior in important ways. Although other recent work has identified biased updating in self-relevant contexts, these experiments remained entirely in the mental frame of reporting beliefs about oneself, and never added an optimization problem with a new action space on top of the belief updating problem. There may be something special about making search decisions that muddles the self-relevance of underlying beliefs. The search decision may take place in a new mental frame, where beliefs and signals directly impact highly salient payoffs from search. The indirect, subjective costs of belief revisions to one's ego may take a back seat in driving behavior once this mental frame has shifted. It is much more painful for people to directly report that they are below average than it is for them to use a search strategy which only indirectly reveals their belief in being below average.

This explanation seems particularly plausible based on the simulation results shown above. Subjects seem eager to report overly optimistic beliefs about their relative intelligence when the decision problem is entirely framed as reporting beliefs about oneself and the subsequent search problem has not yet been revealed. However, once the frame has shifted to a search optimization problem, subjects act much more as if they are searching from modest, accurate prior beliefs than as if they are using the previously reported overconfident prior beliefs.

Indeed, this interpretation provides a unified explanation for results observed throughout the experiment. First, there is no statistical difference in search behavior between self-relevant and self-irrelevant contexts, implying that conservatism in belief updating and information avoidance are not impacting search behavior. This makes sense if subjects shut down self-image-related biases once the context has shifted to search. Second, this interpretation reconciles the direct observations of posterior beliefs with the search duration data. Most subjects appear to update "too much" when starting from their reported prior belief, and especially so when they are receiving negative signals about their intelligence test score. Yet there are not consistent statistically significant differences in search between self-relevant and self-irrelevant treatments, or between High Scorers Win and Low Scorers Win conditions. A sensible explanation is that subjects are in fact not updating "too much" because they are not updating from the reported overconfident prior. Instead, they are updating from a more accurate and modest prior.

Future research may disentangle the contrasting explanations for the results of this experiment. Are the null results due to a mental framing issue in which the search context weakens self-image biases like self-serving belief updating? Or do certain aspects of real job search, such as pride in one's occupation, activate self-image concerns that were not strongly activated in this experiment? Whichever explanation is closer to the truth, it has important implications for both unemployment insurance policy and behavioral science. If people facing a search optimization problem operate in a mental frame that largely negates self-image-related biases, then policymakers need not worry much about biased updating and information avoidance among the unemployed. Behavioral scientists may uncover how and why the framing difference is able to neutralize biases that are commonly found in contexts that focus purely on beliefs. If self-image-related biases do impact search behavior, but only under certain conditions, it will be important to identify the exact conditions that activate those biases. Policymakers will want to know when to address self-image concerns among the unemployed to improve their outcomes, and when they can be safely ignored. Behavioral scientists will be interested in what specific conditions of decision problems neutralize or activate self-image concerns since these findings may be applicable to a much wider range of decision problems than just search decisions.

Unraveling the mystery of whether and how self-image impacts job search will be an important part of the broader research agenda of applying behavioral insights to understand job-seeker behavior. This paper has contributed to this agenda by showing that self-image concerns do not impact search duration in a simple experimental environment. However, this is not the last word on whether self-image matters in real job search. Future research should try to activate self-image in different ways, including in the field, and test whether cognitive loads and different framing impacts those self-image concerns. The results may lead to better targeted unemployment assistance policies that incorporate behavioral insights and ultimately improve the search outcomes and well-being of job seekers.

2.7 Figures and Tables

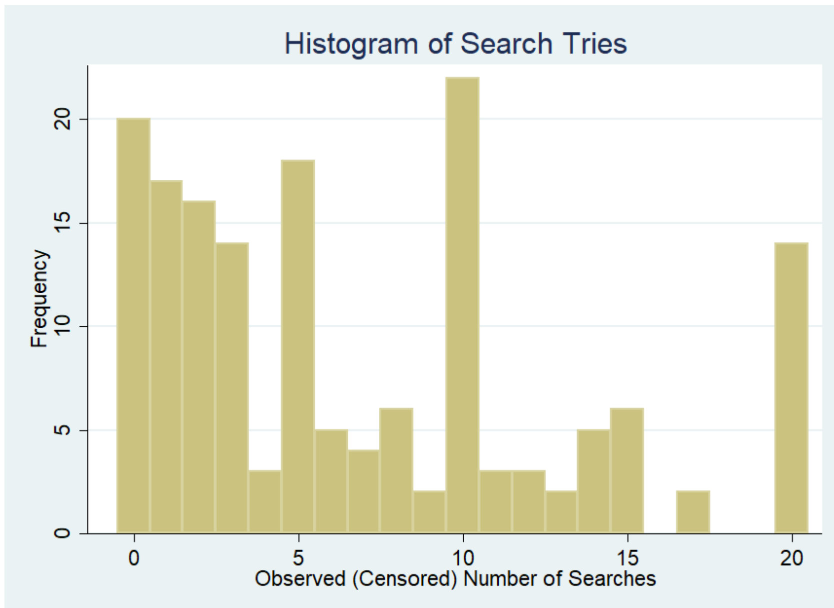


Figure 2.1: Histogram of Search Tries in *Weak Signals*

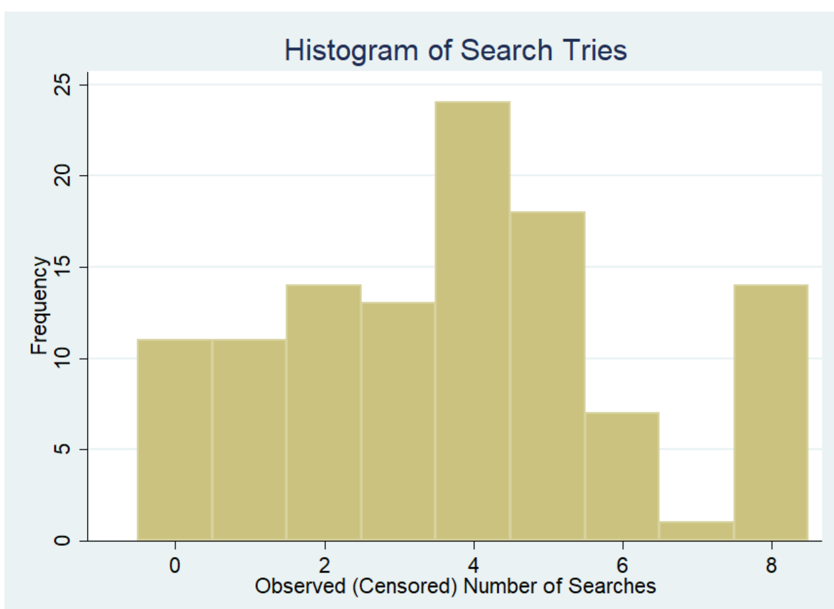


Figure 2.2: Histogram of Search Tries in *Strong Signals*

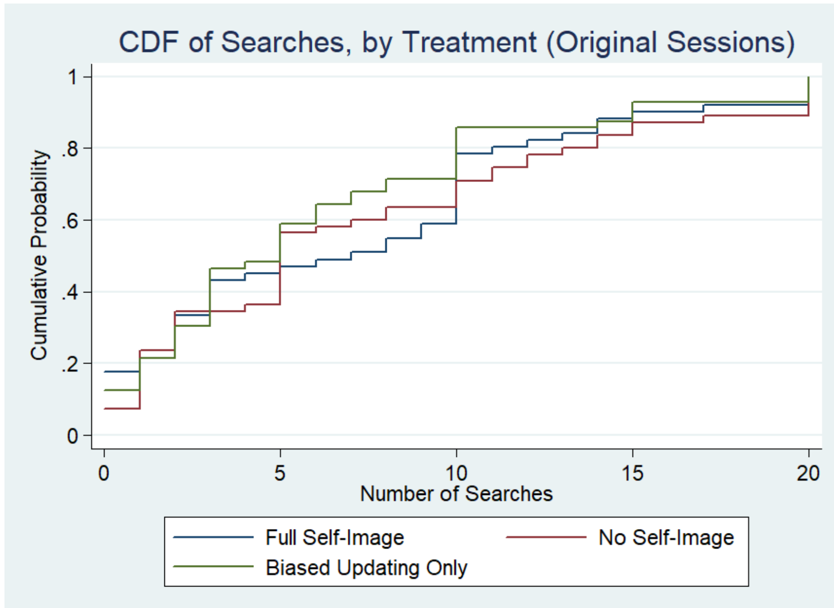


Figure 2.3: CDF of Search Tries by Treatment in *Weak Signals*

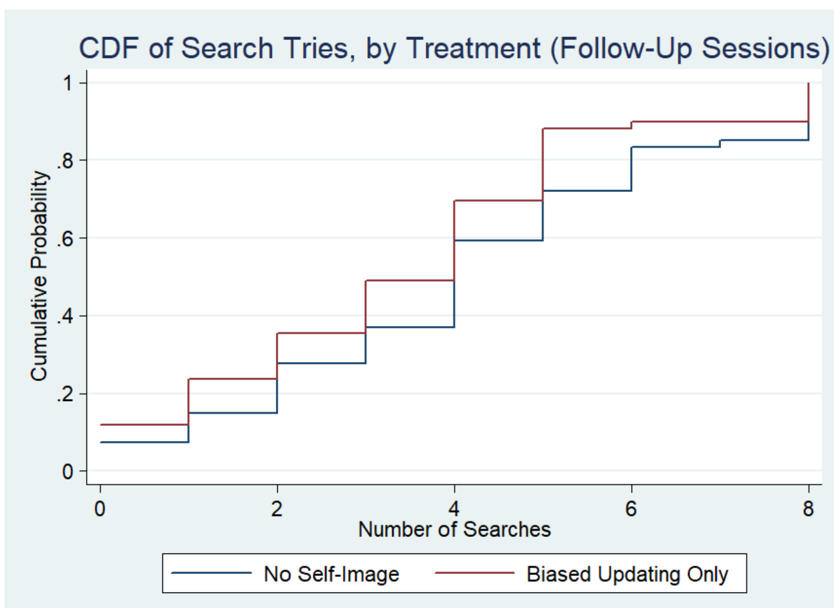


Figure 2.4: CDF of Search Tries by Treatment in *Strong Signals*

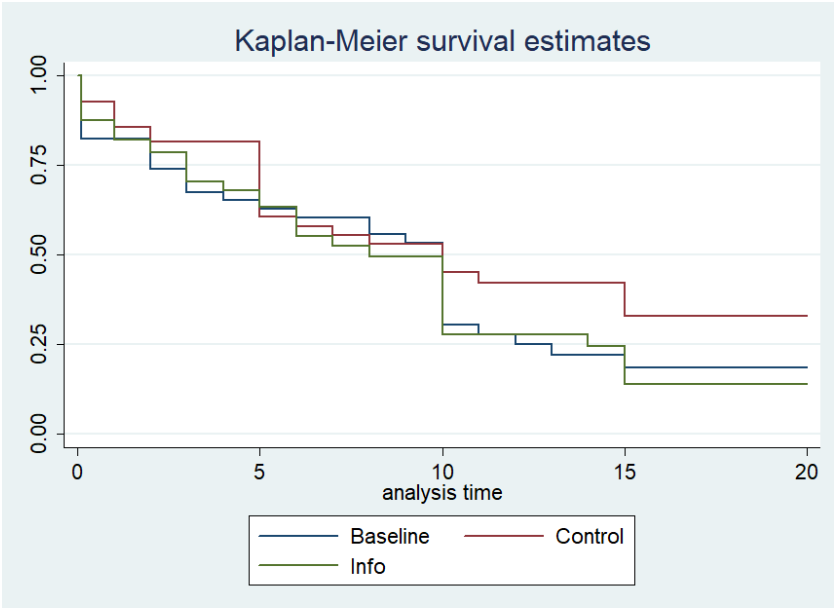


Figure 2.5: Estimated Survival Functions by Treatment in *Weak Signals*

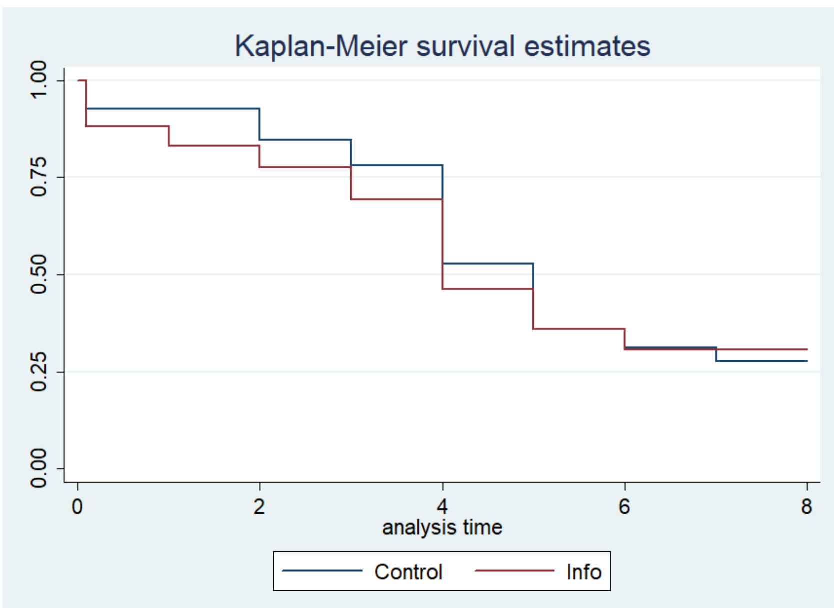


Figure 2.6: Estimated Survival Functions by Treatment in *Strong Signals*

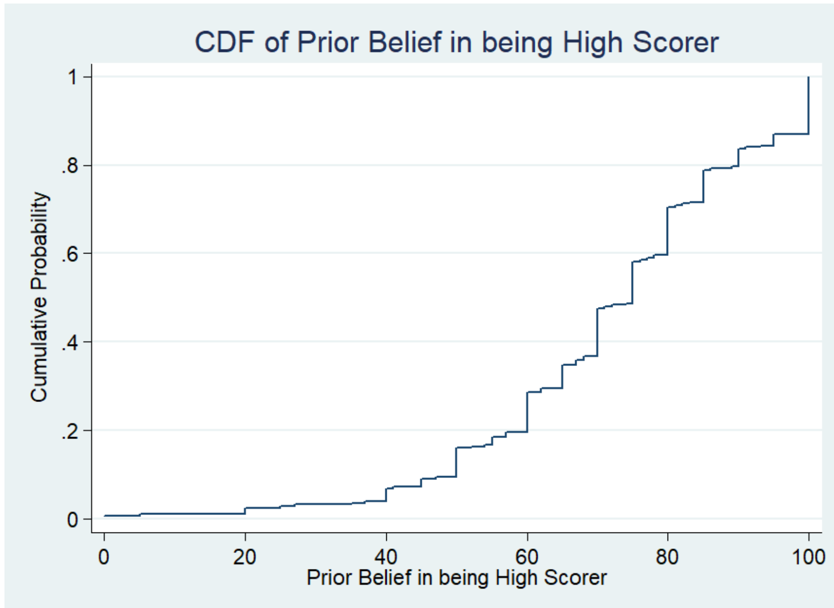


Figure 2.7: CDF of Subject Beliefs in Being High Scorer

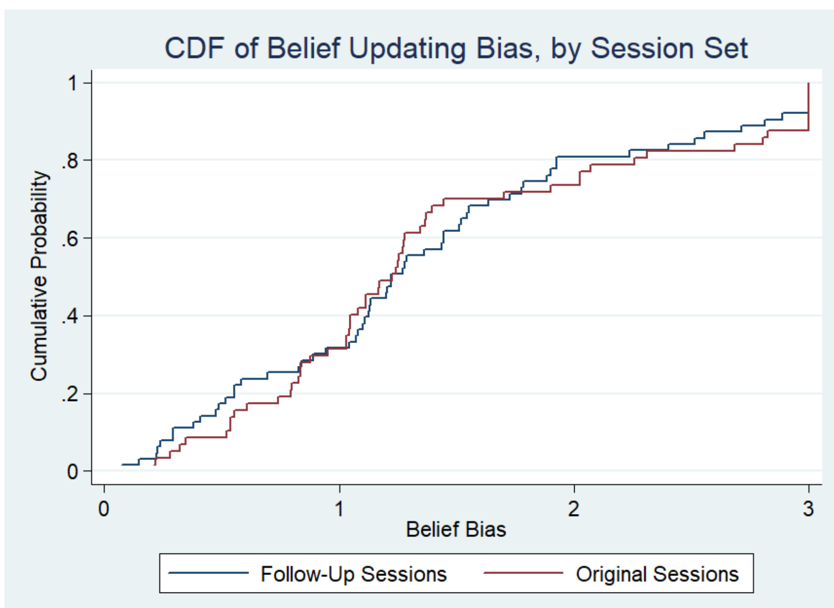


Figure 2.8: CDF of Belief Bias Parameter by Signal Environment

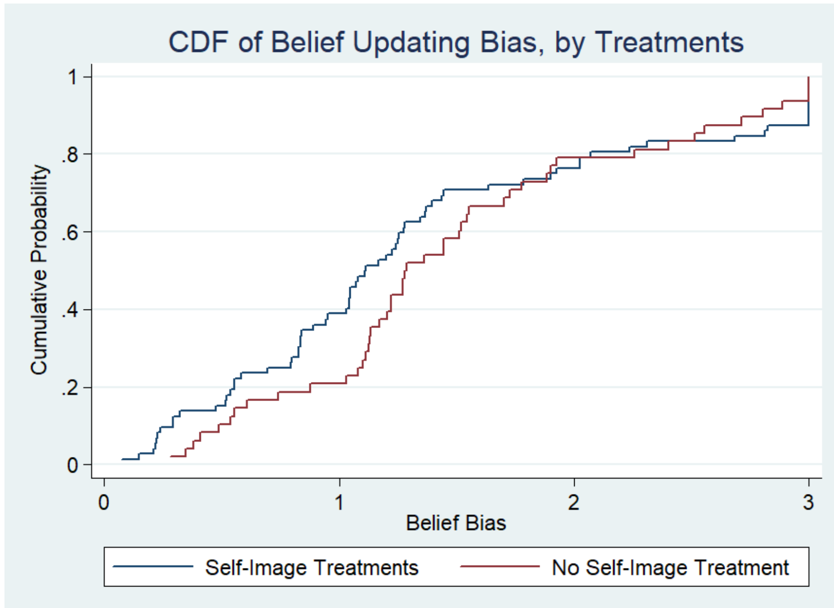


Figure 2.9: CDF of Belief Bias Parameter by Self-Image vs. No Self-Image Treatments

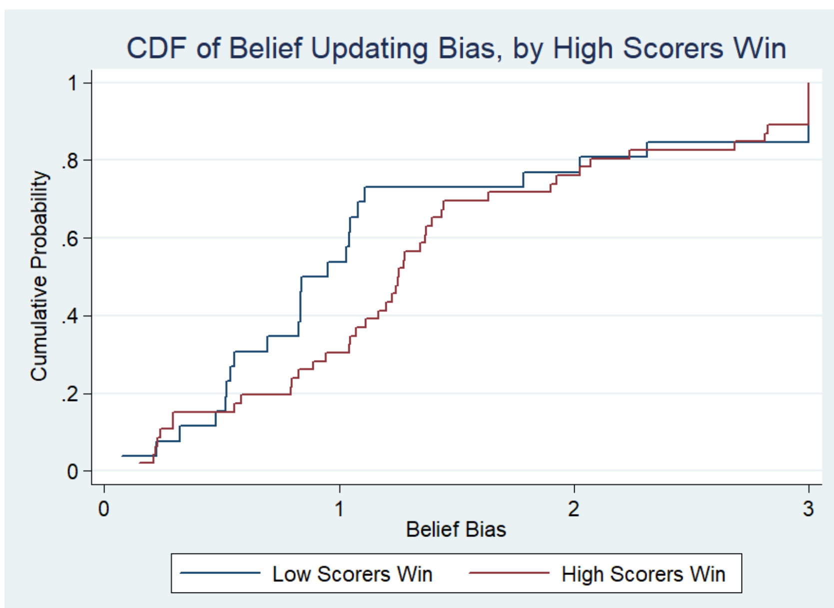


Figure 2.10: CDF of Belief Bias Parameter by Winning Type

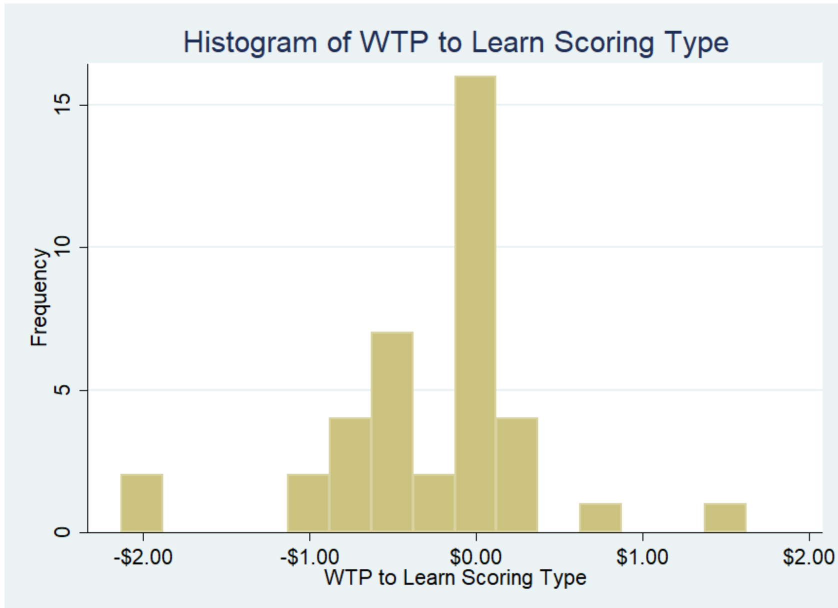


Figure 2.11: Histogram of Willingness to Pay/Accept to Learn Scoring Type

Table 2.1: Covariate Balance between Treatments, *Biased Updating Only* and *Full Self-Image*

	(1) Biased Updating Only		(2) Full Self-Image		(3) t Test	
	mean	sd	mean	sd	p	t
Raven's Matrices Score	14.50	2.97	13.86	3.01	0.21	(-1.26)
High Scorer	0.50	0.50	0.49	0.50	0.95	(-0.06)
Prior Belief	71.98	17.64	72.18	22.53	0.95	(0.06)
Risk Aversion	2.42	1.63	2.82	1.57	0.14	(1.50)
Loss Aversion	3.71	1.53	3.69	1.45	0.92	(-0.11)
Demographic: Age	20.53	2.55	22.61	7.91	0.01*	(2.54)
Demographic: Income	3.68	2.09	3.75	2.22	0.85	(0.19)
Demographic: Female	0.60	0.49	0.59	0.50	0.89	(-0.14)
Demographic: Asian	0.50	0.50	0.25	0.44	0.00**	(-3.06)
Demographic: Hispanic	0.32	0.47	0.29	0.46	0.73	(-0.35)
Demographic: White	0.27	0.45	0.45	0.50	0.02*	(2.33)
Demographic: Black	0.03	0.16	0.02	0.14	0.80	(-0.25)
Observations	115		51		166	

Table 2.2: Covariate Balance between Treatments, *No Self-Image* and *Biased Updating Only*

	(1)		(2)		(3)	
	No Self-Image		Biased Updating Only		t Test	
	mean	sd	mean	sd	p	t
Raven's Matrices Score	14.25	3.05	14.50	2.97	0.54	(-0.62)
High Scorer	0.52	0.50	0.50	0.50	0.68	(0.41)
Prior Belief	71.62	20.21	71.98	17.64	0.89	(-0.14)
Risk Aversion	2.69	1.65	2.42	1.63	0.22	(1.23)
Loss Aversion	3.61	1.38	3.71	1.53	0.58	(-0.55)
Demographic: Age	20.36	1.82	20.53	2.55	0.56	(-0.58)
Demographic: Income	3.60	2.24	3.68	2.09	0.78	(-0.28)
Demographic: Female	0.61	0.49	0.60	0.49	0.93	(0.08)
Demographic: Asian	0.39	0.49	0.50	0.50	0.10	(-1.65)
Demographic: Hispanic	0.32	0.47	0.32	0.47	0.99	(-0.01)
Demographic: White	0.31	0.47	0.27	0.45	0.49	(0.70)
Demographic: Black	0.05	0.21	0.03	0.16	0.43	(0.80)
Observations	109		115		224	

Table 2.3: Covariate Balance between Treatments, *Weak Signals* and *Strong Signals*

	(1)		(2)		(3)	
	Weak Signals		Strong Signals		t Test	
	mean	sd	mean	sd	p	t
Raven's Matrices Score	14.38	2.83	14.14	3.26	0.52	(-0.64)
High Scorer	0.51	0.50	0.50	0.50	0.98	(-0.03)
Prior Belief	72.66	19.45	70.75	19.78	0.43	(-0.79)
Risk Aversion	2.56	1.65	2.65	1.61	0.64	(0.46)
Loss Aversion	3.73	1.47	3.57	1.44	0.35	(-0.94)
Demographic: Age	21.16	4.88	20.40	2.23	0.12	(-1.55)
Demographic: Income	3.78	2.22	3.49	2.08	0.27	(-1.10)
Demographic: Female	0.60	0.49	0.59	0.49	0.84	(-0.20)
Demographic: Asian	0.38	0.49	0.46	0.50	0.20	(1.28)
Demographic: Hispanic	0.28	0.45	0.36	0.48	0.17	(1.38)
Demographic: White	0.36	0.48	0.27	0.44	0.11	(-1.62)
Demographic: Black	0.04	0.19	0.03	0.16	0.63	(-0.48)
Observations	162		113		275	

Table 2.4: Cox Proportional Hazards Model, *Full Self-Image* versus *Biased Updating Only* (Weak Signals)

	(1)	(2)	(3)
Treatment = *Biased Updating*	1.155	1.146	1.246
	(0.655)	(0.671)	(0.572)
High Scorers Win	0.846	1.212	2.065
	(0.876)	(0.867)	(0.574)
Prior Belief in High Scorer	1.001	1.004	1.006
	(0.945)	(0.745)	(0.674)
Treatment = *Biased Updating*	0.794	0.772	0.623
× High Scorers Win	(0.637)	(0.600)	(0.384)
Prior Belief in High Scorer	0.985	0.982	0.976
× High Scorers Win	(0.285)	(0.213)	(0.148)
Intelligence Test Score		0.977	1.009
		(0.597)	(0.850)
Risk Aversion		0.997	1.048
		(0.974)	(0.608)
Loss Aversion		1.114	1.169
		(0.205)	(0.108)
Demographic Controls:	No	No	Yes
N	107	107	107

Exponentiated coefficients; p-values in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.5: Cox Proportional Hazards Model, *Biased Updating Only* versus *No Self-Image* (Weak Signals)

	(1)	(2)	(3)
Treatment = *Biased Updating*	1.123	0.980	1.136
	(0.729)	(0.952)	(0.759)
High Scorers Win	5.444	5.077	9.734
	(0.175)	(0.211)	(0.177)
Prior Belief in High Scorer	1.032**	1.038**	1.044*
	(0.009)	(0.005)	(0.012)
Treatment = *Biased Updating*	1.224	1.333	1.181
× High Scorers Win	(0.697)	(0.587)	(0.790)
Prior Belief in High Scorer	0.955**	0.955**	0.947*
× High Scorers Win	(0.005)	(0.007)	(0.013)
Intelligence Test Score		0.937	0.954
		(0.231)	(0.457)
Risk Aversion		0.884	0.895
		(0.174)	(0.270)
Loss Aversion		1.151	1.181
		(0.090)	(0.097)
Demographic Controls:	No	No	Yes
N	111	111	111

Exponentiated coefficients; p-values in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.6: Cox Proportional Hazards Models, *Biased Updating Only* versus *No Self-Image* (Strong Signals)

	(1)	(2)	(3)
Treatment = *Biased Updating*	1.369	1.052	1.236
	(0.355)	(0.885)	(0.616)
High Scorers Win	38.07***	33.18***	48.18**
	(0.001)	(0.001)	(0.001)
Prior Belief in High Scorer	1.038**	1.046***	1.057***
	(0.002)	(0.000)	(0.000)
Treatment = *Biased Updating*	1.000	1.033	1.244
× High Scorers Win	(1.000)	(0.950)	(0.747)
Prior Belief in High Scorer	0.943***	0.943***	0.934***
× High Scorers Win	(0.000)	(0.000)	(0.000)
Intelligence Test Score		0.991	0.963
		(0.833)	(0.413)
Risk Aversion		0.786**	0.849
		(0.005)	(0.087)
Loss Aversion		1.307**	1.479**
		(0.008)	(0.002)
Demographic Controls:	No	No	Yes
N	113	113	113

Exponentiated coefficients; p-values in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.7: Cox Proportional Hazards Model, High Scorers Win versus Low Scorers Win, within *Full Self-Image* (Weak Signals)

	(1)	(2)	(3)
High Scorers Win	0.893	2.071	1.918
	(0.925)	(0.648)	(0.688)
Prior Belief in High Scorer	0.994	1.001	0.998
	(0.635)	(0.949)	(0.901)
Prior Belief in High Scorer	0.983	0.973	0.972
× High Scorers Win	(0.300)	(0.207)	(0.177)
Intelligence Test Score		0.970	1.087
		(0.665)	(0.385)
Risk Aversion		1.037	1.031
		(0.785)	(0.860)
Loss Aversion		1.170	1.038
		(0.297)	(0.856)
Demographic Controls:	No	No	Yes
N	51	51	51

Exponentiated coefficients; p-values in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.8: Cox Proportional Hazards Model, High Scorers Win versus Low Scorers Win, within *Biased Updating Only* (Weak Signals)

	(1)	(2)	(3)
High Scorers Win	0.397	0.305	0.132
	(0.650)	(0.559)	(0.490)
Prior Belief in High Scorer	1.021	1.026	1.046
	(0.290)	(0.203)	(0.093)
Prior Belief in High Scorer	0.993	0.997	1.006
$\times$ High Scorers Win	(0.792)	(0.909)	(0.866)
Intelligence Test Score		1.001	1.079
		(0.982)	(0.291)
Risk Aversion		0.854	0.801
		(0.224)	(0.159)
Loss Aversion		1.161	1.249
		(0.222)	(0.180)
Demographic Controls:	No	No	Yes
N	56	56	56

Exponentiated coefficients; p-values in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.9: Cox Proportional Hazards Model, High Scorers Win versus Low Scorers Win, within *Biased Updating Only* (Strong Signals)

	(1)	(2)	(3)
High Scorers Win	104.0**	18.37	11.65
	(0.004)	(0.073)	(0.195)
Prior Belief in High Scorer	1.051**	1.042*	1.062*
	(0.007)	(0.020)	(0.013)
Prior Belief in High Scorer	0.931***	0.950*	0.952
× High Scorers Win	(0.001)	(0.022)	(0.072)
Intelligence Test Score		0.978	0.939
		(0.726)	(0.582)
Risk Aversion		0.679**	0.589**
		(0.007)	(0.005)
Loss Aversion		1.391*	1.821**
		(0.014)	(0.002)
Demographic Controls:	No	No	Yes
N	59	59	59

Exponentiated coefficients; p-values in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.10: Discrete Time Logistic Hazards Model, *Full Self-Image* versus *Biased Updating Only* (Weak Signals)

	(1)	(2)	(3)
	y	y	y
Treatment = *Biased Updating*	0.0889	0.0967	0.202
	(0.810)	(0.798)	(0.654)
High Scorers Win	-1.260	-0.955	-1.433
	(0.074)	(0.329)	(0.188)
Prior Belief in High Scorer	-0.0106*	-0.00779	-0.0163
	(0.013)	(0.442)	(0.170)
Treatment = *Biased Updating*	-0.186	-0.216	-0.475
× High Scorers Win	(0.734)	(0.698)	(0.446)
Prior Belief in High Scorer	-0.00643	-0.00985	-0.00408
× High Scorers Win	(0.527)	(0.471)	(0.790)
Intelligence Test Score		-0.0437	-0.00728
		(0.370)	(0.897)
Risk Aversion		-0.00489	0.0182
		(0.959)	(0.855)
Loss Aversion		0.103	0.104
		(0.253)	(0.301)
Period Fixed Effects:	Yes	Yes	Yes
Demographic Controls:	No	No	Yes
N	677	677	677

p-values in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.11: Discrete Time Logistic Hazards Model, *Biased Updating Only* versus *No Self-Image* (Weak Signals)

	(1)	(2)	(3)
	y	y	y
Treatment = *Biased Updating Only*	-0.0744	-0.179	-0.110
	(0.850)	(0.660)	(0.812)
High Scorers Win	-1.532	-0.0586	-0.429
	(0.089)	(0.959)	(0.744)
Prior Belief in High Scorer	-0.00618	0.0218*	0.0182
	(0.176)	(0.045)	(0.143)
Treatment = *Biased Updating* × High Scorers Win	0.451	0.315	0.228
	(0.436)	(0.602)	(0.746)
Prior Belief in High Scorer × High Scorers Win	-0.00982	-0.0299	-0.0270
	(0.435)	(0.055)	(0.135)
Intelligence Test Score		-0.166**	-0.147*
		(0.002)	(0.014)
Risk Aversion		-0.168	-0.103
		(0.102)	(0.371)
Loss Aversion		0.161	0.175
		(0.091)	(0.128)
Period Fixed Effects:	Yes	Yes	Yes
Demographic Controls:	No	No	Yes
N	627	627	619

p-values in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.12: Discrete Time Logistic Hazards Model, *Biased Updating Only* versus *No Self-Image* (Strong Signals)

	(1)	(2)	(3)
	y	y	y
Treatment = *Biased Updating Only*	-0.280	-0.183	-0.371
	(0.427)	(0.640)	(0.402)
High Scorers Win	-0.290	1.263	2.748*
	(0.665)	(0.169)	(0.020)
Prior Belief in High Scorer	-0.0170***	0.0163	0.0314*
	(0.000)	(0.112)	(0.012)
Treatment = *Biased Updating Only* × High Scorers Win	0.674	0.291	0.847
	(0.234)	(0.631)	(0.269)
Prior Belief in High Scorer × High Scorers Win	-0.0101	-0.0348*	-0.0630***
	(0.325)	(0.011)	(0.000)
Intelligence Test Score		-0.110**	-0.123*
		(0.010)	(0.026)
Risk Aversion		-0.433***	-0.412***
		(0.000)	(0.000)
Loss Aversion		0.0479	0.204
		(0.616)	(0.117)
Period Fixed Effects:	Yes	Yes	Yes
Demographic Controls:	No	No	Yes
N	422	422	416

p-values in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.13: Discrete Time Logistical Hazards Model, High Scorers Win versus Low Scorers Win, within *Full Self-Image* (Weak Signals)

	(1)	(2)	(3)
	y	y	y
High Scorers Win	-0.546	-0.202	-0.842
	(0.481)	(0.871)	(0.544)
Prior Belief in High Scorer	-0.0116*	-0.00959	-0.0188
	(0.024)	(0.526)	(0.312)
Prior Belief in High Scorer × High Scorers Win	-0.0168	-0.0210	-0.0190
	(0.159)	(0.245)	(0.330)
Intelligence Test Score		-0.0453	0.0666
		(0.565)	(0.514)
Risk Aversion		-0.000896	-0.0808
		(0.995)	(0.615)
Loss Aversion		0.126	-0.124
		(0.351)	(0.547)
Period Fixed Effects:	Yes	Yes	Yes
Demographic Controls:	No	No	Yes
N	313	313	304

p-values in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.14: Discrete Time Logistic Hazards Model, High Scorers Win versus Low Scorers Win, within *Biased Updating Only* (Weak Signals)

	(1)	(2)	(3)
High Scorers Win	-3.907*	-4.055*	-8.314**
	(0.014)	(0.035)	(0.004)
Prior Belief in High Scorer	-0.00842	-0.00366	0.0150
	(0.093)	(0.806)	(0.523)
Prior Belief in High Scorer × High Scorers Win	0.0262	0.0285	0.0759*
	(0.208)	(0.270)	(0.036)
Intelligence Test Score		-0.0483	0.120
		(0.473)	(0.221)
Risk Aversion		-0.131	-0.285
		(0.380)	(0.131)
Loss Aversion		0.155	0.247
		(0.243)	(0.162)
Period Fixed Effects:	Yes	Yes	Yes
Demographic Controls:	No	No	Yes
N	290	290	282

p-values in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.15: Discrete Time Logistic Hazards Model, High Scorers Win versus Low Scorers Win, within *Biased Updating Only* (Strong Signals)

	(1)	(2)	(3)
	y	y	y
High Scorers Win	0.516	1.415	2.695
	(0.615)	(0.337)	(0.207)
Prior Belief in High Scorer	-0.0137**	0.0210	0.0510*
	(0.008)	(0.127)	(0.022)
Prior Belief in High Scorer $\times$ High Scorers Win	-0.0160	-0.0371	-0.0595
	(0.285)	(0.090)	(0.053)
Intelligence Test Score		-0.143*	-0.310**
		(0.018)	(0.005)
Risk Aversion		-0.666***	-0.910***
		(0.000)	(0.000)
Loss Aversion		0.243	0.684**
		(0.077)	(0.003)
Period Fixed Effects:	Yes	Yes	Yes
Demographic Controls:	No	No	Yes
N	197	197	195

p-values in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.16: Discrete Time Logistic Hazards Model, *Biased Updating Only* versus *No Self-Image* (Pooled Weak and Strong Signals)

	(1)	(2)	(3)
	y	y	y
Treatment = *Biased Updating Only*	-0.0667	-0.0831	-0.0477
	(0.800)	(0.765)	(0.872)
High Scorers Win	-0.304	0.885	1.547
	(0.591)	(0.215)	(0.055)
Prior Belief in High Scorer	-0.00580	0.0239**	0.0299***
	(0.103)	(0.001)	(0.000)
Treatment = *Biasaed Updating Only*	0.460	0.356	0.218
× High Scorers Win	(0.249)	(0.393)	(0.631)
Prior Belief in High Scorer	-0.0182*	-0.0377***	-0.0469***
× High Scorers Win	(0.030)	(0.000)	(0.000)
Strong Signals	-1.080**	-0.824*	-0.724
	(0.003)	(0.031)	(0.067)
Intelligence Test Score		-0.140***	-0.148***
		(0.000)	(0.000)
Risk Aversion		-0.289***	-0.274***
		(0.000)	(0.000)
Loss Aversion		0.109	0.0967
		(0.095)	(0.177)
Period Fixed Effects	Yes	Yes	Yes
Period Fixed Effects × Strong Signals	Yes	Yes	Yes
Demographics	No	No	Yes
N	1049	1049	1049

p-values in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.17: Discrete Time Logistic Hazards Model, High Scorers Win versus Low Scorers Win, within *Biased Updating Only* (Pooled Weak and Strong Signals)

	(1)	(2)	(3)
	y	y	y
High Scorers Win	-0.895	-0.594	-0.417
	(0.304)	(0.603)	(0.761)
Prior Belief in High Scorer	-0.00645	0.0150	0.0202
	(0.141)	(0.119)	(0.108)
Prior Belief in High Scorer	-0.00545	-0.0131	-0.0157
× High Scorers Win	(0.657)	(0.420)	(0.414)
Strong Signals	-0.629	-0.360	-0.349
	(0.165)	(0.469)	(0.495)
Intelligence Test Score		-0.108*	-0.133*
		(0.015)	(0.016)
Risk Aversion		-0.392***	-0.385***
		(0.000)	(0.001)
Loss Aversion		0.186*	0.188
		(0.046)	(0.078)
Period Fixed Effects	Yes	Yes	Yes
Period Fixed Effects × Strong Signals	Yes	Yes	Yes
Demographics	No	No	Yes
N	487	487	487

p-values in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.18: Difference between Reported and Bayesian Posteriors

	(1) Control	(2) Self-Image	(3) Self-Image: H Wins	(4) Self-Image: L Wins
Mean Signed	-0.0976	-0.0397	-0.0730	0.0206
Difference	(0.191)	(0.206)	(0.218)	(0.172)
Mean Absolute	0.145	0.139	0.166	0.0910
Value Difference	(0.157)	(0.157)	(0.158)	(0.146)
Observations	48	73	47	26

mean coefficients; sd in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.19: Proportion of Subjects Updating Too Much and Too Little Compared to Bayesian Updating

	(1) Control	(2) Self-Image	(3) Self-Image: H Wins	(4) Self-Image: L Wins
Update Too Much	0.778	0.611	0.696	0.462
Update Too Little	0.222	0.389	0.304	0.538
Observations	45	72	46	26

mean coefficients; t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.20: Linear Regression on Search Strategy Method, *Biased Updating Only* versus *No Self-Image* (Strong Signals)

	(1) Search Strategy	(2) Search Strategy	(3) Search Strategy
Treatment = *Biased Updating Only*	-0.836	-0.528	-0.416
	(0.106)	(0.299)	(0.437)
High Scorers Win	-4.075**	-3.537*	-3.488*
	(0.006)	(0.015)	(0.018)
Prior Belief in High Scorer	-0.0466**	-0.0453**	-0.0423**
	(0.001)	(0.002)	(0.006)
Treatment = *Biased Updating Only*	0.517	0.390	-0.0577
× High Scorers Win	(0.492)	(0.598)	(0.944)
Prior Belief in High Scorer	0.0718***	0.0659***	0.0666***
× High Scorers Win	(0.000)	(0.001)	(0.001)
Intelligence Test Score		-0.0439	-0.0224
		(0.497)	(0.748)
Risk Aversion		0.299*	0.299*
		(0.014)	(0.017)
Loss Aversion		-0.171	-0.178
		(0.205)	(0.230)
Demographics	No	No	Yes
N	113	113	113

p-values in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.21: Linear Regression on Search Strategy Method, High Scorers Win vs. Low Scorers Win, within *Biased Updating Only* (Strong Signals)

	(1) Search Strategy	(2) Search Strategy	(3) Search Strategy
High Scorers Win	-4.432*	-2.124	-2.209
	(0.031)	(0.295)	(0.313)
Prior Belief in High Scorer	-0.0581**	-0.0522**	-0.0561**
	(0.002)	(0.004)	(0.005)
Prior Belief in High Scorer × High Scorers Win	0.0846**	0.0511	0.0495
	(0.003)	(0.068)	(0.095)
Intelligence Test Score		-0.106	-0.240
		(0.235)	(0.058)
Risk Aversion		0.383*	0.354
		(0.028)	(0.063)
Loss Aversion		-0.409*	-0.481*
		(0.023)	(0.019)
Demographics	No	No	Yes
N	59	59	59

p-values in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.22: Simulated Search with Correct Beliefs or Reported Beliefs Compared to Observed Search Behavior (Weak Signals)

	Optimal Search w/ Correct Prior		Actual Search		Optimal Search w/ Reported Prior
All Subjects	6.74		6.83	(**)	8.99
Full Self-Image Treatment	5.63		7.06	(*)	9.73
Biased Updating Only Treatment	5.64		6.13	(+)	8.52
No Self-Image Treatment	8.89		7.33		8.80
High Scorers Win (Self-Image Treatments)	5.39	(***)	8.77	(**)	12.25
Low Scorers Win (Self-Image Treatments)	5.94	(*)	3.77		5.07

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.23: Simulated Search with Correct Beliefs or Reported Beliefs Compared to Observed Search Behavior (Strong Signals)

	Optimal Search w/ Correct Prior	Actual Search		Optimal Search w/ Reported Prior
All Subjects	3.57	3.77		3.76
Full Self-Image Treatment	N/A	N/A		N/A
Biased Updating Only Treatment	3.57	3.44	(+)	4.28
No Self-Image Treatment	3.57	4.14	(+)	3.20
High Scorers Win (Self-Image Treatments)	4.30	4.00	(***)	5.91
Low Scorers Win (Self-Image Treatments)	2.81	2.85		2.58

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Chapter 3

The Effect of E-Cigarette Indoor Vaping Restrictions on Adult Prenatal Smoking and Birth Outcomes

3.1 Introduction

Smoking has long been known to be one of the leading causes of poor birth outcomes in the United States (US Department of Heatlh and Human Services, 2014). The rising popularity of e-cigarettes has provided pregnant woman with another smoking cessation product option, similar to the availability of FDA-approved smoking cessation medications and nicotine replacement therapy (NRT) products (Food and Drug Administration, 2017). However, the effectiveness of e-cigarettes as smoking cessation devices and the safety of trying to use e-cigarettes to quit smoking while pregnant is unknown.

E-cigarettes are part of a broader class of devices (including vape pens and tank systems) known as electronic nicotine delivery systems (ENDS). ENDS simulate the smoking experience by "vaping" a mist containing water, propylene glycol, vegetable glycerin, and, in most cases, nicotine and flavorings (Royal College of Physicians in England, 2016). ENDS may also contain trace levels of other toxicants and metals, but in levels estimated to be 9-450 times lower than a conventional cigarette (Goniewicz et al., 2014). Among the general population, there is an emerging consensus that ENDS are substantially safer than combustible tobacco products (Viscusi, 2016). E-cigarettes are unlikely to exceed 5% of the risk associated with combustible tobacco products for non-pregnant women, but the risks may be much higher for pregnant women because nicotine, which is in most ENDS, is a threat to the developing fetus (Royal College of Physicians in England, 2016). The safety of using e-cigarettes for smoking cessation during pregnancy has an inconclusive grade from the United States Preventive Services Task Force (Siu, 2015).

One recent policy change that may impact demand for ENDS is indoor vaping restrictions (IVRs). ENDS IVRs may reduce the attractiveness of ENDS by requiring the user to make an additional investment of time to use the devices either outdoors or in non-regulated indoor areas. Indeed, ENDS IVRs have been found to reduce demand for ENDS in an experimental market (Marti et al., 2016). This added inconvenience of using ENDS may both reduce the use of ENDS and increase cigarette use if fewer people use ENDS for successful smoking cessation. Additionally, ENDS IVRs may provide information to individuals about the relative risks of smoking versus vaping, and cause people to perceive ENDS to be more dangerous, which could also reduce their demand for ENDS.

The objective of this paper is to explore the effect of ENDS IVRs on cigarette use and birth outcomes (e.g. gestation length) among adult pregnant women. A main feature of our paper is that in our preferred specification we take advantage of cigarette use information provided at four points in time for each pregnant woman, which we use to estimate person fixed effect models that free our estimates from unobserved person-specific sources of omitted variable bias.

While several studies have evaluated the effect of ENDS minimum legal purchase age laws on traditional cigarette use (Friedman, 2015; Pesko et al., 2016; Pesko and Currie, 2016; Dave et al., 2017; Abouk and Adams, 2017), no published estimates exist on the effect of ENDS IVRs on cigarette use. In our current study, we estimate the impact of ENDS IVRs on adult pregnant women, rather than teenagers, to study this relationship unhindered by state efforts to restrict youth ENDS access through channels other than ENDS IVRs.

3.2 Background

Pregnant women have a high interest in quitting smoking: 55% of women smoking 3 months before their pregnancy are successfully able to quit smoking during their pregnancy (Centers for Disease Control and Prevention, 2017b). Despite these high rates of successful quitting during pregnancy, few pregnant women use an approved smoking cessation method to help them to quit (Tong et al., 2008). One study found that only 2.6% of pregnant women smokers in a Medicaid managed care program in Maryland used a prescribed pharmacotherapy for smoking cessation (Jarlenski et al., 2015). This low rate may be due to obstetricians and gynecologists (OBGYNs) often not referring pregnant women for behavioral health interventions and often not prescribing NRT because nicotine is a developmental toxicant. Pharmacotherapy interventions for tobacco cessation for pregnant women continue to receive an incomplete grade from the United States Preventive Services Task Force due to uncertain evidence of the overall health benefits (Siu, 2015).

Given the risks in prescribing NRT to pregnant women, it is possible that pregnant women motivated to quit smoking may look elsewhere for help in quitting. Early evidence from randomized controlled trials suggests that ENDS may be effective in eliminating and reducing cigarette consumption (McRobbie et al., 2016). Given low utilization of smoking cessation methods among pregnant women, it is possible that pregnant women may disproportionately look to ENDS to reduce cigarette consumption. According to 2014 National Health Interview Survey data, ever use of e-cigarettes among all child-bearing age adult women (18-45) was 15.2% and past 30-day use was 4.0%. One small randomized controlled trial of pregnant women in Connecticut and Massachusetts that were unable to quit smoking on their own during pregnancy found that 14% vaped during pregnancy, usually to try to quit (Oncken et al., 2017).

States and counties are increasingly using IVRs and taxes to regulate ENDS. New Jersey was the first state to pass an ENDS IVR in 2010, and 10 states and numerous counties/municipalities banned the use of ENDS in at least one type of indoor location by the end of 2016. Minnesota had the first ENDS tax in the nation in 2010, but it wasn't until June, 2015 that a second state (North Carolina) enacted an ENDS tax.

Our study is similar to the literature that has evaluated the effect of ENDS minimum legal purchase age laws. Five studies have used difference-in-differences models to evaluate the effect of enacting these ENDS minimum legal sale age laws on use of tobacco among teens, with four of these studies suggesting that the laws raise teenage cigarette use (Friedman, 2015; Pesko et al., 2016; Pesko and Currie, 2016; Dave et al., 2017) and a fifth study suggesting that it does not and may even lower cigarette use for certain populations (Abouk and Adams, 2017).

Our study is also similar to Markowitz et al. (2011), which used data from the Population Risk Assessment Monitoring System from 2000 to 2005 to find that restricting smoking in the workplace was associated with small increases in gestational length for women 25-34 years of age. A literature review using 11 studies employing interrupted time-series designs also found that smoke-free legislation is associated with small decreases in preterm birth, but with no obvious effects on birth weight (Been et al., 2014).

ENDS IVRs may have different effects depending on smoking status prior to pregnancy. For smokers, ENDS IVRs may have two effects: 1) they may make it less likely that pregnant woman will switch from cigarettes to ENDS; or 2) they may make it more likely that pregnant women will switch from cigarettes to complete abstinence from nicotine. For non-smokers, ENDS IVRs may have two effects: 1) they may make it less likely that pregnant women will use ENDS (which they may have been using earlier); and 2) they may make it more likely that non-smokers will use cigarettes (since access to ENDS has now been restricted).[1] Unfortunately, ENDS use information is not recorded in birth records, but we can test hypotheses examining the effect of ENDS IVRs on cigarette use directly.

Further, we can explore the effects of the laws on birth outcomes, which could deteriorate if more adult pregnant women now smoke and smoking is more dangerous than vaping to the developing fetus, or could improve birth outcomes if the laws increase complete abstinence from nicotine. Given the unique risks of nicotine and the presence of nicotine in cigarettes and most ENDS, it is unclear if pregnant woman who use ENDS instead of smoking traditional cigarettes will experience improved birth outcomes.

Finally, we also explore heterogeneity in the effects of the laws depending on age, education, and health insurance status. According to 2014 National Health Interview Survey data, lifetime and current e-cigarette use is slightly higher among younger adults < 25 (17.3%/4.4%) than older adults ≥ 25 (14.4%/3.9%); is higher among low-educated women (18.3%/5.5%) than high-educated women (13.7%/3.3%); and is higher among uninsured women (18.1%/5.5%) and Medicaid women (19.3%/4.7%) than privately-insured women

[1] However, the magnitude of this latter effect (pregnant women initiating or resuming smoking as a result of ENDS IVRs) is likely minuscule since pregnant women generally attempt to reduce risk during pregnancy.

(13.4%/3.3%). We might expect that populations using e-cigarettes more will be more impacted by ENDS IVRs. However, we find this assumption unconvincing because interest in using ENDS during pregnancy could increase dramatically for women trying to quit smoking. We believe that a more useful prediction for which populations may be disproportionately impacted by ENDS IVRs during pregnancy are those populations with higher prenatal smoking rates.

There are other factors besides level differences in vaping and smoking prior to pregnancy that may also influence responsiveness to ENDS IVRs. For one, higher education has been found to reduce risk beliefs of the impact of e-cigarettes on lung cancer (Viscusi, 2016), so higher educated pregnant women may also be more likely to believe that vaping instead of smoking during pregnancy is healthier as well (and hence these women could be more affected when ENDS IVRs are passed if it disproportionately reduces their attempts to quit smoking). The impact of ENDS IVRs may also be influenced by access to other close substitute smoking cessation products through health insurance programs. In 2010, the Affordable Care Act required coverage of tobacco cessation in all non-grandfathered private plans and for pregnant women on Medicaid (American Lung Association, 2017). Uninsured women, however, may not have access to cost reduction benefits available to insured patients, and so this group could be disproportionately influenced by ENDS IVRs.

Given this background on e-cigarettes, we empirically test the following hypotheses: 1) ENDS IVRs increase prenatal smoking overall, 2) ENDS IVRs increase prenatal smoking rates by pregnant women with higher rates of smoking (e.g. younger, lower educated, and Medicaid recipients), and 3) ENDS IVRs have ambiguous impacts on birth outcomes.

3.3 Data

We combined three data sources to conduct our analysis: birth records data from the National Center for Health Statistics, ENDS IVR data from the American Nonsmokers' Rights Foundation, and population data on counties and municipalities from the U.S. Census Bureau. The birth records provide smoking behavior, demographic information, and infant health outcomes with linked county of residence for the mother. We used county of residence to link the ENDS IVR data and population data to determine mothers' level of exposure to IVRs. This section will describe each of our three data sources and explain the restrictions we imposed on our sample before analysis.

The main data source for our analysis was administrative birth records (the universe of all births in the United States) provided by the National Center for Health Statistics. The Standard Certificate of Live Birth was revised in 2003 and the new form was slowly rolled out in different states over time.[2] The old form asked only about smoking at any time during the pregnancy. The new form collects information about smoking at four points in time: prior to pregnancy and in each trimester. The accuracy of cigarette use during pregnancy is significantly improved in the new form relative to the old form. Maternal smoking in the old form agreed with hospital records 84% of the time in one study (Buescher et al., 1993), but this agreement improved to 94% with the new form (Howland et al., 2015). The introduction of the revised birth record resulted in statistically significant increases in smoking in 21 out of 31 states, suggesting that the old form underreported smoking compared to the new (Curtin and Matthews, 2014). In addition to improving the accuracy of smoking by using the new form, trimester-specific smoking information is also available in the new form which permits us to exploit within-individual variation in cigarette use in response to ENDS IVRs in a panel data analysis.

[2] All states were using the revised birth records in 2015 (Centers for Disease Control and Prevention, 2017b).

Figure 3.1 shows the question capturing cigarette use information as it appears on the revised birth record form. No information about vaping is currently collected in birth records.

We impose the following sample restrictions on our data. The sample includes all mothers whose estimated conception (16 days after pregnancy week 0 or last menstrual period) was between 1/1/2010 and 1/1/2015. We exclude mothers with missing data on weeks of gestation. We exclude mothers under age 18 at the estimated time of conception to avoid confounding from state efforts to regulate youth access to ENDS through other means, such as with minimum legal purchase age laws. We exclude non-singleton births because they often have adverse birth outcomes for reasons unrelated to smoking behavior.

We make several exclusions to derive our final sample of 755 counties from 15 states and Washington D.C. First, we exclude from our sample seventeen states which were not using the revised birth record form by 1/1/2010.[3] Second, three additional states were excluded because they were missing a substantial percentage of data on smoking behavior for entire years.[4] Third, we also exclude fifteen remaining states that did not have comprehensive prohibitions on smoking cigarettes in bars, workplaces, and restaurants (minor exceptions allowed).[5] We obtained this data on indoor smoking restrictions from the CDC State System (Centers for Disease Control and Prevention, 2017a). Governments likely place higher priority on first banning smoking from public places before banning vaping from public places; therefore, we believe that studying the impact of ENDS IVRs in places already comprehensively banning indoor smoking provides more policy-relevant results. We also avoid complex interaction effects between partial smok-

[3]These excluded states are: AK, AL, AR, AZ, CT, HI, LA, MA, ME, MN, MS, NC, NJ, RI, VA, WI, and WV.

[4]These excluded states are: FL, GA, and MI.

[5]These excluded states are: ID, IN, KS, KY, MO, ND, NH, NV, OK, PA, SC, SD, TN, TX, and WY.

ing restrictions and ENDS IVRs. After our various restrictions, we use birth records for approximately 7 million adult pregnant women, representing 37.7% of births nationally during this time period that otherwise meet our exclusion criteria of being singleton births to women conceiving during adulthood and without missing gestational length information on the birth certificate.

IVR data were provided by the American Nonsmokers' Rights Foundation U.S. Tobacco Control Laws Database. These data include all known ENDS IVRs at the state, county, and municipality levels. We focused on the most widespread ENDS IVRs: those applying to bars, workplaces, and restaurants. These data also include IVRs applying to gambling establishments or specific public places such as parks or government buildings, but these policies were rare and did not likely have large impacts on pregnant mothers. Figure 3.2 shows the evolution of the ENDS IVR environment over time (including both partial and comprehensive IVR laws), with states excluded from our sample in gray. From among the 16 states (including D.C.) and 755 counties in our sample, 6 counties had an ENDS IVR in place on 1/1/2011, 38 counties had an ENDS IVR in place on 1/1/2013, and 68 counties had an ENDS IVR in place on 1/1/2015. In Figure 3.3, we show the percent of mothers covered by ENDS IVR laws in these 16 states over time, with approximately 45% of mothers living in these states covered by the beginning of 2015.

Quarterly cigarette excise taxes were used as a control variable and were provided by the CDC State System (Centers for Disease Control and Prevention, 2017a).

The final data source was the U.S. Census which included yearly estimates of county and municipality populations. The population data were merged with the data on bar, workplace, and restaurant ENDS IVRs to calculate the percentage of county population affected by each policy, with each of the three venues receiving equal weight.[6] This policy

[6]Our policy variable is constructed by considering both the percentage of population affected and

variable was merged with the cross-sectional birth records data at the estimated point of conception and at the start of the trimester in our panel data analysis.

Table 3.1 shows descriptive statistics for our sample. The first column shows means for the entire sample, while the remaining columns split the sample into counties in which there was no ENDS IVR adopted prior to 2015 and counties in which there was an ENDS IVR adopted in part or all of the county prior to 2015. Counties that adopted ENDS IVRs have a higher share of minority pregnant women, lower Medicaid enrollment, higher private insurance enrollment, lower prenatal smoking, and somewhat better premature birth outcomes. We model away observable differences between our two groups by including demographics and payment source as control variables in regression analyses. Additionally, our modelling strategy reduces concerns about our outcome variables having different levels since the modelling strategy uses differences in outcome in the pre-adoption period as the baseline level in which to evaluate the policy change in the post-adoption period.

3.4 Methods

Our first approach to estimating the effects of ENDS IVRs on smoking behavior and birth outcomes is to use cross-sectional data and difference-in-difference (DD) models. We analyze two variables that capture smoking behavior. First, smoking participation is an indicator variable that takes a value of one when the mother reports smoking any cigarettes during the pregnancy. Second, we take the average of reported daily cigarettes

by a linear combination of three types of policies. For example, if an entire county population was affected by bar and restaurant ENDS IVRs, but none of the county was affected by a workplace ENDS IVR, the final ENDS IVR variable would take a value of two thirds. As another example, suppose the population of the unincorporated areas of a county was one half of the entire county population. If only these unincorporated areas were affected by comprehensive bar, restaurant, and workplace ENDS IVR, our final policy variable would take a value of one half because only half of the county population was affected.

smoked during each trimester, and multiply it by 30 to calculate average cigarettes smoked per month during pregnancy. We use these two dependent variables in DD models specified as follows:

$$smoking\_any_{ict} = \alpha + \beta_1 IVR_{ct} + \beta_2 cig\_tax_{st} + \delta X_{ict} + \gamma_c + \gamma_t + \epsilon_{ict} \tag{3.1}$$

$$cigs\_per\_month_{ict} = \alpha + \beta_1 IVR_{ct} + \beta_2 cig\_tax_{st} + \delta X_{ict} + \gamma_c + \gamma_t + \epsilon_{ict} \tag{3.2}$$

where i represents mother, c represents county of state s, and t represents year-month combination. We include time fixed effects (both birth-year-month and conception-year-month) and county fixed effects. County fixed effects allows us to control for time-invariant county-level factors that are unobservable and implies that our regression models are identified off within-county variation in ENDS IVRs. Our time fixed effects control for time varying omitted variables that affect the nation as a whole (e.g. national anti-smoking campaigns, national economic conditions). Cigarette taxes are also controlled for since these vary over time.[7] Finally, individual-level controls included in X_{ict} are race/ethnicity, age, mother's delivery payment method,[8] marital status, level of education, and the number of the current birth (i.e. number of previous births plus one).

In equation 3.1, we estimate the effect of ENDS IVRs on extensive margin smoking using the full sample of births. We estimate equation 3.2 twice, once for just the sample of smokers (to estimate the effect of ENDS IVRs on intensive margin smoking) and again

[7] We do not control for ENDS taxes. Only MN had an ENDS tax for any meaningful period of time in our study, and we already excluded MN for not using revised birth records. North Carolina passed an ENDS tax in June, 2015 and Louisiana passed an ENDS tax in July, 2015. These states were also excluded for not using revised birth records. The District of Colombia passed an ENDS tax in Oct., 2015, but this will not impact our population in any meaningful way since we restricted to women conceiving prior to January 1, 2015.

[8] Payment method also helps control for the mother's economic condition. Besides payment source, birth records do not provide any direct information on the mother's employment status or financial well-being.

using all births (to estimate the effect of ENDS IVRs on extensive and intensive smoking combined).

Our treatment variable IVR_{ct} is continuous in our initial specification, but we explore the robustness of our results to estimating a more traditional DD equation in which we drop counties with partial restrictions so that we derive an indicator, rather than continuous, treatment variable. We lose about 33% of our observations in this specification, but our results are broadly similar with those keeping partially-treated counties.

We hypothesize that our primary coefficient of interest, β_1, will be positive for our extensive margin and combined margin model to indicate an increase in prenatal smoking in response to ENDS IVRs. We suspect that the coefficient will be positive for the intensive margin model as well, although this coefficient will reflect both the change in the number of cigarettes consumed by smokers and any compositional changes in extensive margin smoking as a result of ENDS IVRs.

A necessary assumption for the DD model to recover causal estimates is that the treatment (i.e. counties adopting ENDS IVRs) and the comparison (i.e. counties not adopting ENDS IVRs) would follow the same trend in the post-treatment period, had the treatment states not been treated. However, this assumption is untestable. We instead attempt to provide evidence on this assumption by modifying equations 3.1 and 3.2 to evaluate the parallel trends assumption. We include only mothers who conceived 9 months before ENDS IVR adoption (actual or placebo) and add an interaction between a linear time trend (using date of conception) and an indicator for whether the county ever adopted a law or not.

$$smoking\_any_{ict} = \alpha + \beta_1 ever\_adopt_c * time_t + \beta_2 cig\_tax_{st} + \delta X_{ict} + \gamma_c + \gamma_t + \epsilon_{ict} \tag{3.3}$$

$$cigs\_per\_month_{ict} = \alpha + \beta_1 ever\_adopt_c * time_t + \beta_2 cig\_tax_{st} + \delta X_{ict} + \gamma_c + \gamma_t + \epsilon_{ict} \quad (3.4)$$

In these equations, we do not need to control for $ever\_adopt_c$ and $time_t$ separately because our fixed effects are perfectly nested smaller units. If we cannot reject the null hypothesis that β_1 is zero in these two equations, then this suggests that trends are parallel in the pre-period. We also modify equations 3.3 and 3.4 by replacing $time_t$ with four one-year intervals leading up to ENDS IVR adoption to test for evidence of non-parallel time trends at different points in time in the pre-adoption period.

After providing evidence that our results satisfy the parallel trends assumption, we modify equations 3.1 and 3.2 to perform an event study (Autor, 2003). We continue to exclude partial IVRs and we replace the previous DD variable with a set of mutually exclusive policy lead and lag variables that divides the time period before and after adoption in 3 month intervals (using date of conception): $>$ 18 months before (reference), 18-15 months before, 15-12 months before, 12-9 months before, 9-6 months before (likely 3rd trimester treatment), 6-3 months before (likely 2nd trimester treatment), 3-0 months before (likely 1st trimester treatment), 0-3 months after (likely full treatment), and $>$ 3 months after. In this way, we can evaluate effects of the law among pregnant women fully and partially treated, as well as test for unexpected changes in smoking in places that will pass an ENDS IVR in the future (which could suggest anticipatory behaviors or reverse causality).

A central feature of our paper is our ability to exploit the impact of ENDS IVRs on within-mother smoking by using a panel data analysis. For this analysis, we use smoking information provided in revised birth records for the three months before pregnancy and for each trimester, yielding four observations per mother. Our panel data regressions have the following specifications:

$$smoking\_any_{ict} = \alpha + \beta_1 IVR_{ct} + \beta_2 cig\_tax_{st} + \gamma_i + \gamma_t + \epsilon_{ict} \tag{3.5}$$

$$cigs\_per\_month_{ict} = \alpha + \beta_1 IVR_{ct} + \beta_2 cig\_tax_{st} + \gamma_i + \gamma_t + \epsilon_{ict} \tag{3.6}$$

where i represents mother, c represents county of state s, and t represents trimester. The ENDS IVR variable now represents the policy at the start of the trimester instead of at the time of conception. All specifications include mother and trimester fixed effects. We hypothesize that our primary coefficient of interest, β_1, will be positive to indicate an increase in smoking in that trimester in response to ENDS IVRs. In some specifications we also add to equations 3.5 and 3.6 trimester-year-month fixed effects, absorbing the effect of being in a certain trimester at a specific time. We do not include county fixed effects in these specifications because birth fixed effects are perfectly nested smaller units. Similar to our cross-sectional results, we estimate equation 3.6 twice, once conditional on smoking in that trimester and a second time without this condition.

We estimate stratified models using our panel data design to test for heterogeneity in effects of ENDS IVRs along lines of age (< 25 and ≥ 25), education (high school degree or less, some college or more), and health insurance (Medicaid, private insurance, and self-pay).

Finally, we estimate the effects of ENDS IVRs on birth outcomes in a cross-sectional DD specification by modifying equations 3.1 and 3.2 and replacing the smoking outcome variables with birth outcomes:

$$birth\_outcome_{ict} = \alpha + \beta_1 IVR_{ct} + \beta_2 cig\_tax_{st} + \delta X_{ict} + \gamma_c + \gamma_t + \epsilon_{ict} \tag{3.7}$$

The birth outcomes we examine include premature birth (indicator for less than 37 weeks of gestation), very premature birth (indicator for less than 32 weeks of gestation), small for gestational age (indicator for 25th percentile or lower birth weight for a given gestational length), extra small for gestational age (indicator for 10th percentile or lower), low weight (indicator for less than 2500 grams), and Apgar 5-minute score (continuous measure from 0-10, with 10 being the healthiest score).[9] For the weight and gestation length outcomes, β_1 could be positive (higher probability of a bad birth outcome) if more adult pregnant women now smoke and smoking is more dangerous than vaping to the developing fetus, or could be negative if the laws increase complete abstinence from nicotine.

Standard errors are clustered at the state level in all regression models, which is the highest level at which our policy varies. We estimate all equations using linear models for continuous variables (cigarettes per month, Apgar 5 score) and linear probability models for indicator variables (smoking participation, premature birth, small for gestational length, low birth weight).

3.5 Results

In Table 3.2, Section A, we show results for our standard DD model from equations 3.1 and 3.2 for three smoking outcomes of any smoking during pregnancy, average number of cigarettes smoked per month during pregnancy (conditional on smoking), and average number of cigarettes smoked per month during pregnancy (non-conditional). The IVR in this case is a continuous variable ranging from 0 to 1, with 1 representing a 100% ENDS IVR (i.e. a complete county-wide ban on indoor vaping in bars, workplaces, and restaurants). Switching from no ENDS IVR to a comprehensive ENDS IVR is

[9] We cannot estimate a panel data model for birth outcomes because each outcome appears only once for each birth.

associated with a 0.9 percentage point increase in smoking during a mother's pregnancy (14.1% of the mean), and smokers consume 10.3 more cigarettes per month (4.2% of the mean). Without conditioning on smoking, the average pregnant woman smoked 3.2 extra cigarettes per month (19.8% of the mean).

In Table 3.2, Section B, we remove from the sample partial ENDS IVRs so that we have an indicator variable for no ENDS IVR or a comprehensive ENDS IVR. We lose 2.3 million mothers from 31 counties[10] that were previously treated by ENDS IVRs that were non-comprehensive in nature, and so our sample size is attenuated from 6.8 million observations to 4.6 million observations. Using this reduced sample, comprehensive ENDS IVRs are associated with a 1.3 percentage point increase in smoking (14.5% of the mean) compared to not having any ENDS IVRs, and are associated with 4.1 extra monthly cigarettes smoked (unconditional, 18.7% of the mean). As a percentage change of the mean, effect sizes were similar when keeping or excluding counties with partial ENDS IVRs.

We estimate equations 3.3 and 3.4 to explore if time trends are different between adopting and non-adopting counties (Table 3.3, Section A). We cannot reject the null hypothesis that counties adopting ENDS IVRs and counties not adopting ENDS IVRs followed similar trends in outcomes in the pre-adoption period, as our interaction term coefficients are small and relatively precise. These results suggest that the parallel trends assumption holds and we can interpret our DD results causally.

In Table 3.3, Section B, we use one-year intervals rather than a continuous time variable to assess the parallel trends assumption. We continue to use the same estimation sample as we did in the linear parallel time trends test. Compared to the year closest to policy adoption, smoking outcomes are not statistically significant different

[10] The largest of these counties were Los Angeles County, CA, Cook County, IL, Alameda County, CA, San Diego County, CA, and Santa Clara County, CA.

between adopting and non-adopting counties in the full second and third year before policy adoption. This suggests that trends are parallel in the three years leading up to policy adoption. Statistically significant differences were observed in year four and earlier (compared to the full year closest to policy adoption), which may be driven by heterogeneity in terms of which counties contribute data to this period of time furthest from policy adoption.[11]

In Table 3.4, we estimate an event study to examine the effects of a policy change at various stages of pregnancy. We continue to use our sample excluding counties with less-than-comprehensive IVRs. We further plot these coefficients in the first two graphs of Figure 3.4. The coefficients on the policy leads 18-15 months before the policy, 15-12 months before, and 12-9 months before are all statistically insignificant and small. Wald tests show that they are jointly statistically insignificant. Collectively, these results suggest the absence of any anticipatory behaviors or reverse causality.

The event study results further show that effect sizes begin to increase 3-0 months before conception (for mothers likely treated with the policy during their first trimester). For mothers conceiving 0-3 months after ENDS IVR adoption, we observe a 0.6 percentage point increase in smoking (6.6% of the mean) and 1.7 extra cigarettes per month (unconditional, 7.7% of the mean). For mothers conceiving more than 3 months after the policy, we observe a 1.4 percentage point increase in smoking (16.2% of the mean) and 4.6 extra cigarettes per month (unconditional, 21.0% of the mean). This may suggest delayed enforcement of the ENDS IVRs or county-level heterogeneity in which the effect of early-adopting ENDS IVRs was larger than later-adopting ENDS IVRs.

In Table 3.5, we estimate our models using a panel data analysis as presented in equations 3.5 and 3.6. This within-pregnancy analysis controls for all time invariant

[11] Among the sample of 1,554,889 mothers used in columns 1 and 3 of Table 3.3, 42.4% of the sample is part of the first full year before ENDS IVRs, 31.6% are part of the second year, 18.1% are part of the third year, and only 7.9% are part of the fourth year or earlier.

mother-specific omitted variables that could potentially bias the estimated effect of ENDS IVRs on smoking. We use all ENDS IVRs, including partial laws, in this evaluation. In columns 1 and 3, we find a larger effect of ENDS IVRs than in the cross-sectional analysis, with smoking participation increasing by 2.0 percentage points and extra cigarettes per month among all pregnant women increasing by 12.6 cigarettes per month. Results were not materially affected when we additionally controlled for trimester-year-month fixed effects in columns 4 through 6. Our consistent finding of a 2.0 percentage point increase in prenatal smoking is relatively twice the most directly comparable cross-sectional results of a 0.9 percentage point increase (Table 3.2, Section A). On the intensive margin (columns 2 and 5), the change in cigarettes per month was similar to the change shown in our cross-sectional model (Table 3.2, Section A), suggesting that the overall larger increase in cigarettes per month in the panel data analysis (columns 3 and 6) is driven by the doubling of smoking on the extensive margin.

In Table 3.6, we use our panel data model to explore heterogeneity in effect sizes of ENDS IVRs depending on age, education, and health insurance. ENDS IVRs had a disproportionate effect on smoking participation for adults under 25 years old (3.2 percentage points) compared to those over 25 years old (1.5 percentage points); however, as a percent of mean smoking both groups responded roughly proportionately to ENDS IVRs (roughly a 30% increase). Lower educated individuals were also more likely than higher educated individuals to smoke in response to ENDS IVRs (2.3 percentage points compared to 1.7 percentage points); however, given the lower smoking rate among higher educated individuals, their smoking increased roughly 45% compared to a 22% increase among lower educated individuals. These results are consistent with our hypothesis that smoking may increase by more among higher educated individuals given evidence that they believe ENDS to be safer products (Viscusi, 2016).

We found similar heterogeneity within payment source stratifications, with the Medicaid population increasing their smoking in response to ENDS IVRs by the most with a 2.4 percentage point increase, compared to a 1.5 percentage point increase for the privately insured and a 1.3 percentage point increase for the uninsured. However, as a share of the mean smoking rate, privately insured mothers' smoking increased roughly 52% compared to more modest increases of roughly 25% for Medicaid mothers and uninsured mothers. This result did not confirm our hypothesis that uninsured pregnant women would be most likely to use ENDS due to higher cost barriers for close substitutes of FDA-approved smoking cessation products.

In Table 3.8, we perform a robustness check using only non-smokers in the 3 months prior to pregnancy, hypothesizing that ENDS IVRs should have little to no impact on this population. Our hypothesis was validated, as we observe precisely estimated zeros. Smoking participation decreased by 0.03 percentage points and unconditional cigarettes smoked per month decreased by roughly 1/10th of a cigarette.[12] In Table 3.8, Sections B and C, we stratify these results by younger pregnant women ($<$ 25 years of age) and older pregnant women ($\geq$ 25 years of age), and while ENDS IVRs had roughly twice the impact on reducing smoking participation among the younger sample, both results continued to be economically insignificant. In sum, we can conclude that the 2.0 percentage point increase in smoking found earlier is being driven by individuals smoking in the 3 months prior to pregnancy rather than by non-smokers in the 3 months prior to pregnancy.

Finally, in Table 3.7 we explore the effect of ENDS IVRs on birth outcomes, using equation 3.7. Both our coefficients and standard errors are small in magnitude, and the coefficients are statistically insignificant.[13] Therefore, it appears that the laws have no

[12]We suspect that these results reflect decreased propensities of relapsing back into smoking (having smoked earlier than 3 months before the pregnancy) rather than initiating smoking for the first time.

[13]For example, 95% confidence intervals permit us to detect statistically significant differences in premature birth of 0.6 percentage points (6.9% of the mean) and statistically significant differences in low birth weight of 0.2 percentage points (3.3% of the mean). Therefore, we believe that our analysis is sufficiently powered to detect reasonable effect sizes.

impact on birth outcomes, possibly because substituting one source of nicotine for another results in no differences in birth outcomes (Royal College of Physicians in England, 2016). We identified similar null effects among non-smokers prior to pregnancy (Table 3.9), whom we hypothesized would be most likely to benefit from ENDS IVRs by potentially promoting complete abstinence from nicotine through reducing stand-alone vaping.

Our event study coefficients for birth outcomes (Figure 3.4 and Table 3.10) basically uphold these null results, except for the finding of premature births and very premature births possibly decreasing for mothers likely exposed to the ENDS IVRs in their first trimester of pregnancy (conceiving 3-0 months before an IVR). Our weak evidence of preterm birth improving and no effect on birth weight as a result of ENDS IVRs appears to echo the literature on the effect of cigarette indoor air laws, which has found small decreases in preterm birth, but with no obvious effects on birth weight (Been et al., 2014; Markowitz et al., 2011).

3.6 Discussion

Our panel data model results document a doubling of the effect of ENDS IVRs on smoking compared to cross-sectional results. This suggests the presence of substantial unobservable individual-level heterogeneity that is correlated with both ENDS IVR adoption and smoking. Given the differences in treatment effects that we observe in our study, future studies evaluating the effect of policies on prenatal smoking should consider adopting a panel data evaluation strategy similar to our own to eliminate the possibility of omitted variable bias from unobservable individual-level heterogeneity.

Our estimate that ENDS IVRs increased prenatal smoking by 2.0 percentage points is large, representing approximately a 31.0% increase in the mean smoking rate. We suspect that these large effect sizes are due to not only the added inconvenience of vaping, but also due to changed perceptions of the risks of ENDS. For a non-pregnant woman, the relative risks of ENDS compared to traditional cigarettes has been estimated to be no more than 5% (Royal College of Physicians in England, 2016). Therefore, if ENDS IVRs are causing perceptions of the risks of ENDS to dramatically increase, and people reduce their attempts to quit smoking by using ENDS as a result, this may result in deteriorated health outcomes overall. Governments should be cautious on the messaging they send, through regulations, on the relative risks of different products. Ideally, products should be regulated proportionate to their level of risk (Chaloupka et al., 2015).

One limitation of our study is that that we are unable to look at the health of the mother or later-life health outcomes of the infants. Among women who quit smoking during pregnancy, 60% were still not smoking within 6 months after delivery (Centers for Disease Control and Prevention, 2017b). Therefore, if ENDS IVRs reduce quitting during pregnancy, some of the largest health effects of the law may be on the mother post-partum. Additionally, higher post-partum smoking rates could also endanger the child's immediate and later-life health due to secondhand smoke exposure (Simon, 2016; US Department of Health and Human Services, 2014).

Another limitation of our study is that birth records have no information about vaping, so we are unable to examine this behavior directly. The trimester-specific smoking information provided in revised birth records provides powerful data for researchers to evaluate the effect of a variety of tobacco control policies. However, with the increasing use of ENDS to obtain nicotine, states should consider adding ENDS use information to the birth records as well. Additionally, states may also wish to consider adding questions on other sources of nicotine exposure, such as through use of NRT.

A final limitation is that the generalizability of our findings may be reduced by using only 37.7% of births nationally that occurred in states using revised birth records since

2010 and that had comprehensive cigarette smoking restrictions in place.[14]

Our study focuses on pregnant women, with the benefit of being able to conduct our analysis using panel data that is rare in studies of tobacco control policies. Future work is needed to evaluate the effect of ENDS IVRs on other populations besides pregnant women, including on the use of tobacco products and perceptions of risk. Additionally, as of January, 2017, seven states now tax e-cigarettes, and these policies will similarly be in need of evaluation (Campaign for Tobacco Free Kids, 2017).[15]

[14]Our primary reasons for imposing these constraints on our analysis were to reduce confounding from cigarette indoor air laws, to increase the policy relevance of our results (since states will likely ban indoor smoking before vaping), and to take advantage of panel data on smoking available in revised birth records.

[15]See footnote 7 for more information.

3.7 Figures and Tables

Figure 3.1: Cigarette Question from Revised Birth Record

37. CIGARETTE SMOKING BEFORE AND DURING PREGNANCY

For each time period, enter either the number of cigarettes or the number of packs of cigarettes smoked. IF NONE, ENTER A0".

Average number of cigarettes or packs of cigarettes smoked per day.

	# of cigarettes		# of packs
Three Months Before Pregnancy	________	OR	________
First Three Months of Pregnancy	________	OR	________
Second Three Months of Pregnancy	________	OR	________
Third Trimester of Pregnancy	________	OR	________

Source: CDC, `https://www.cdc.gov/nchs/data/dvs/birth11-03final-ACC.pdf`

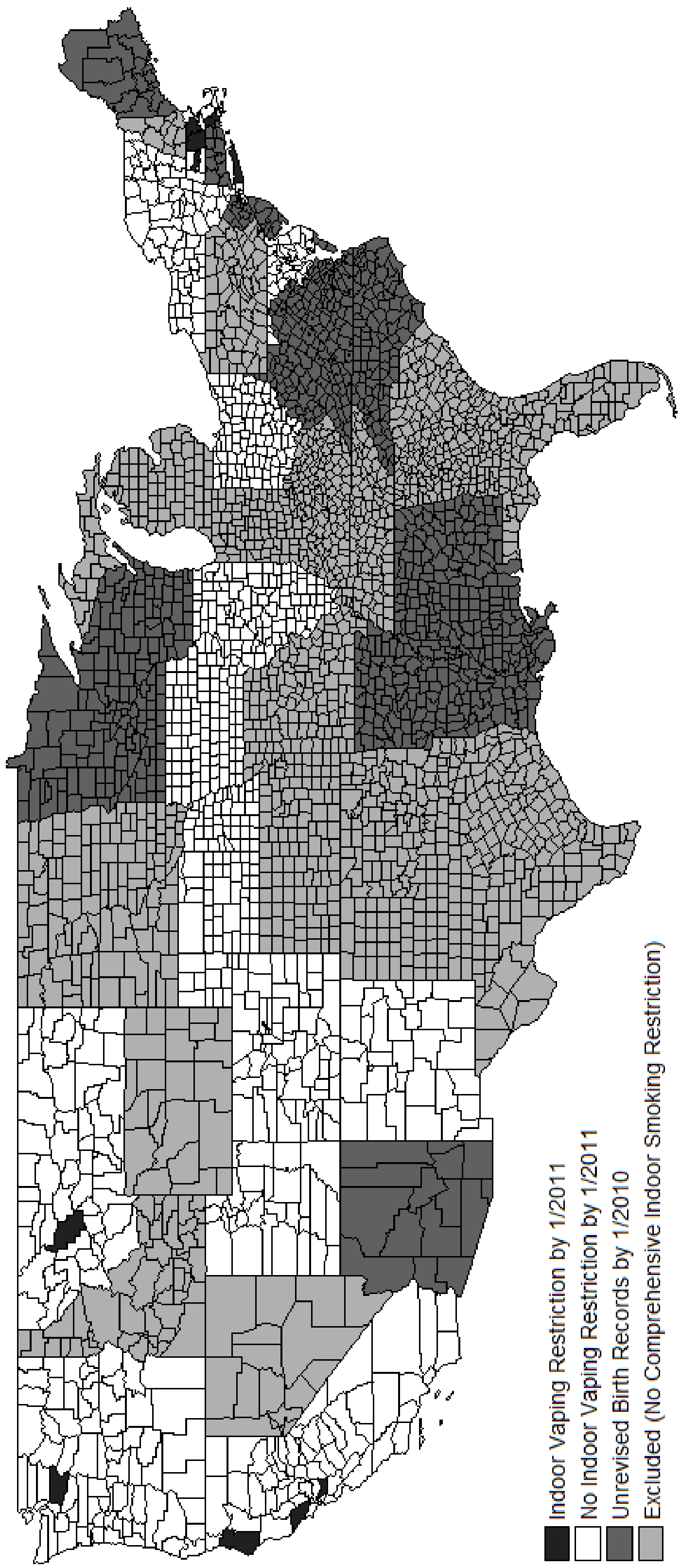


Figure 3.2: Map of ENDS Policy Environment. Section A: January 1, 2011

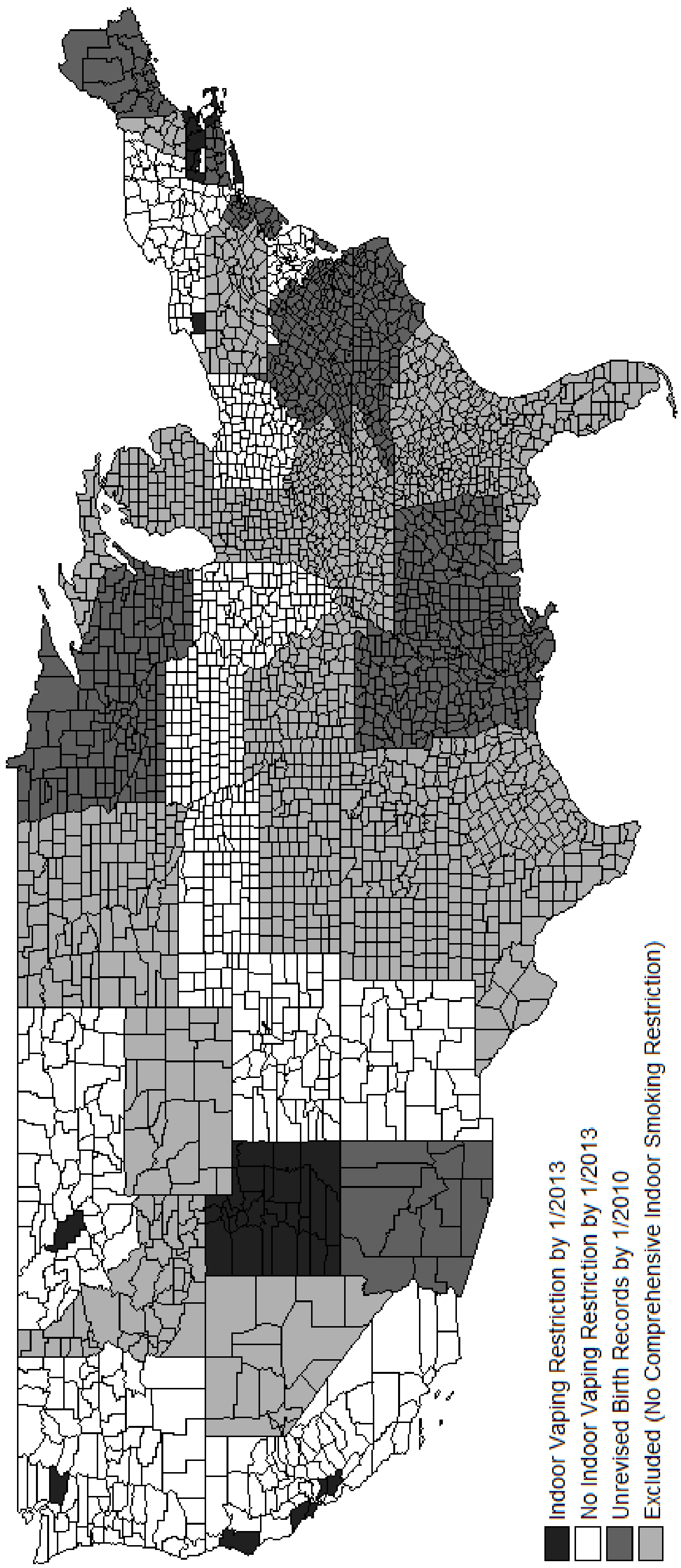


Figure 3.2: Section B: January 1, 2013

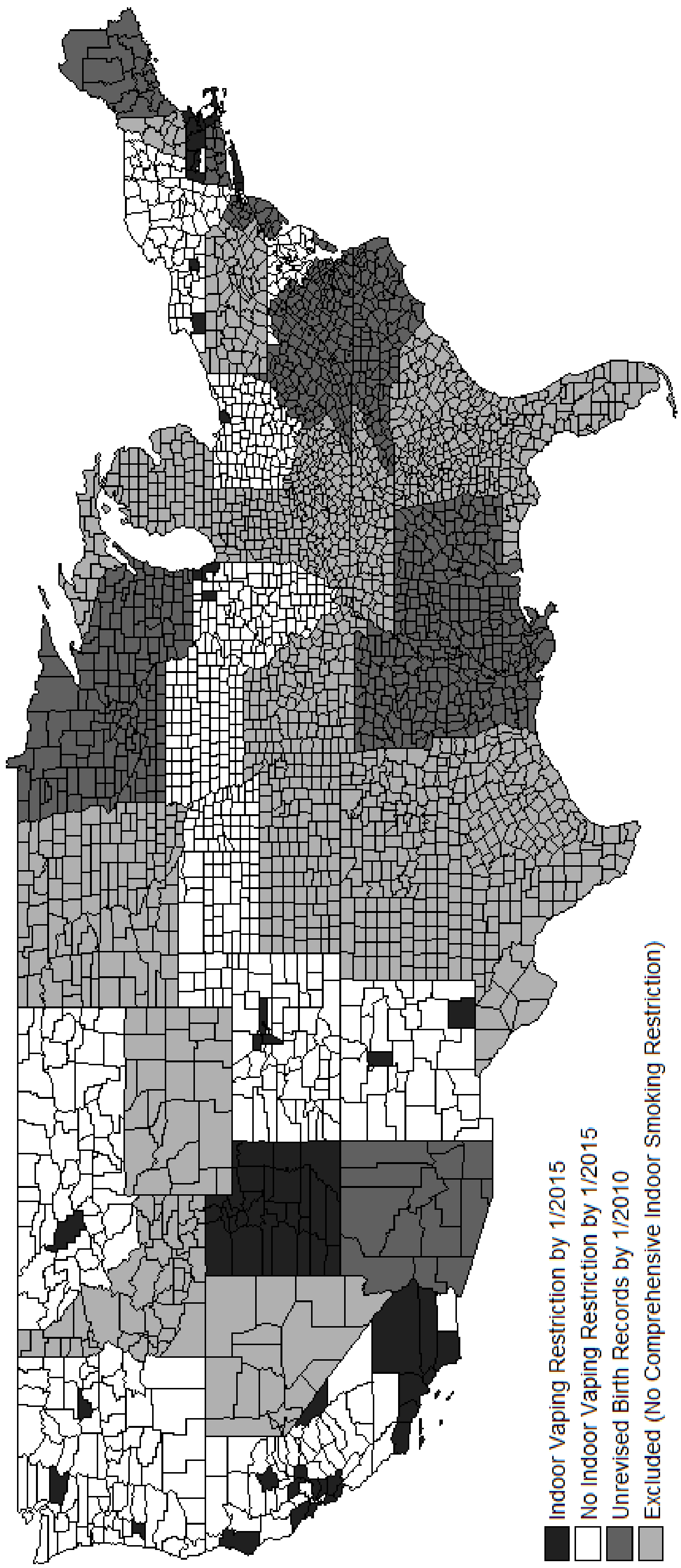
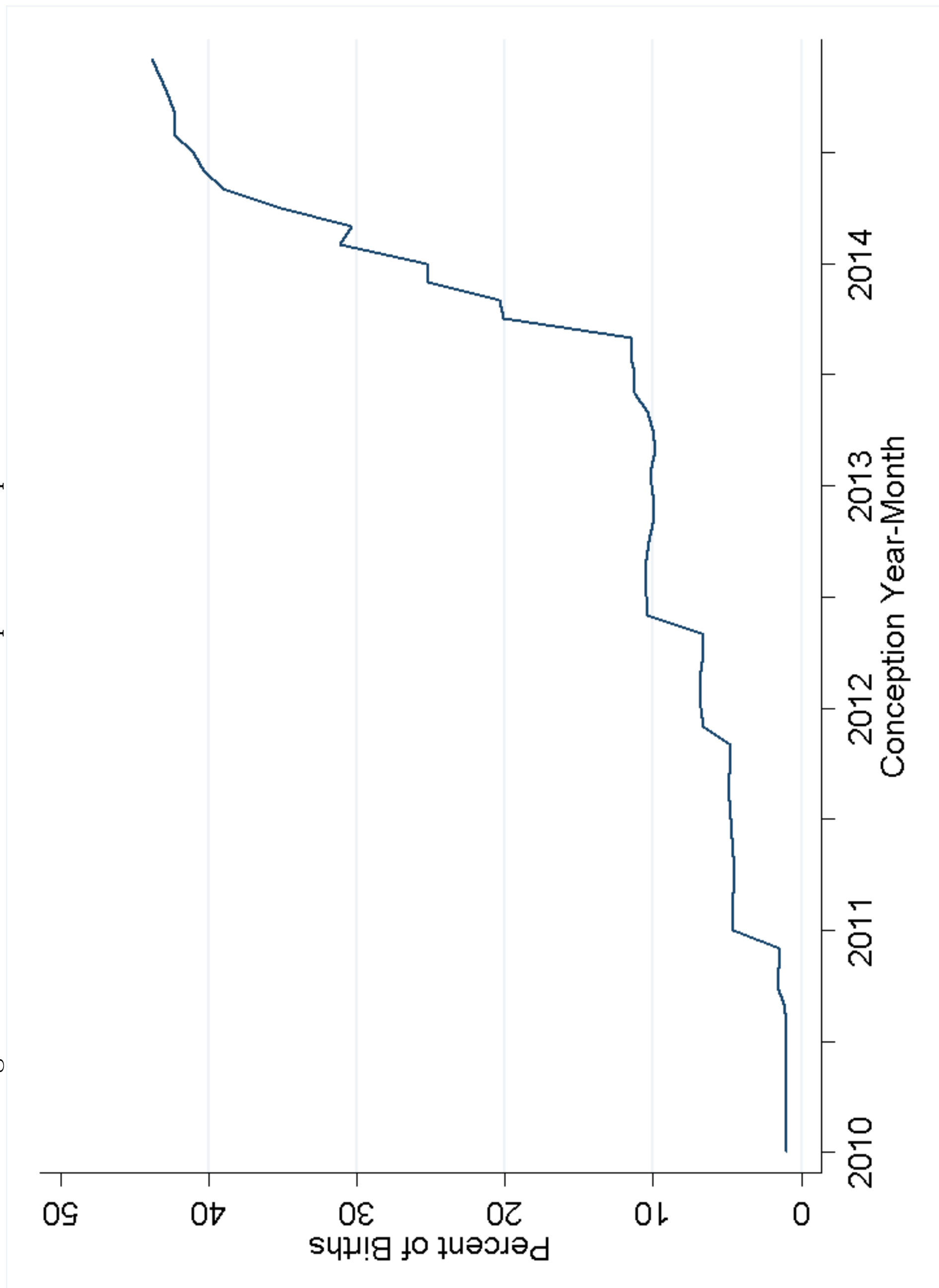


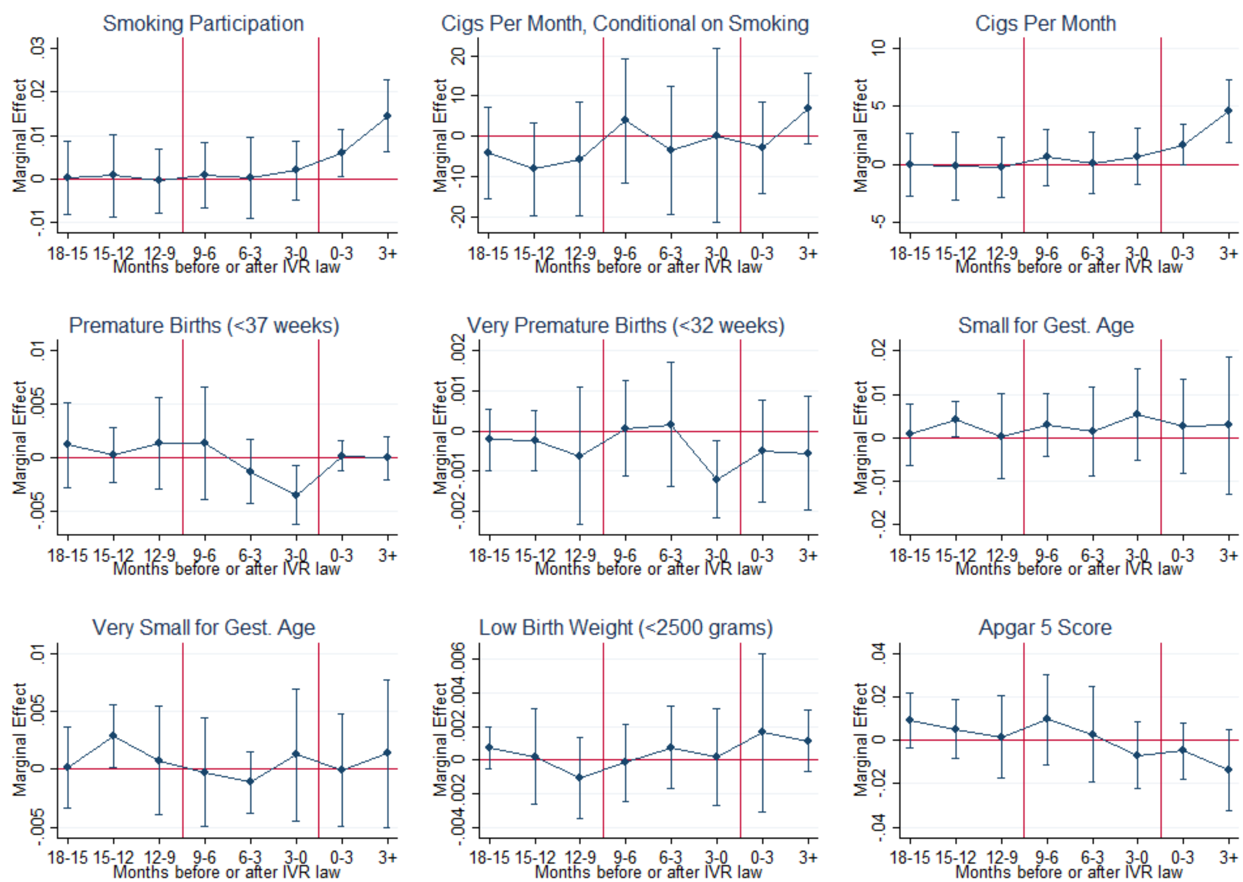
Figure 3.2: Section C: January 1, 2015

Figure 3.3: Percent of Births in Sample with Exposure to ENDS IVRs



Notes: Percent of births in 15 states plus Washington D.C. that are exposed to any IVR at the estimated time of conception.

Figure 3.4: Event Study Marginal Effects



Notes: Marginal effects from event study regressions are plotted. The plotted coefficients and standard errors are taken from the estimates in Table 4 for smoking behavior outcomes and Table A3 for birth outcomes.

Table 3.1: Descriptive Statistics

	(1) All Counties		(2) No IVR before 2015		(3) IVR before 2015	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Demographics:						
Age	28.8514	5.7943	28.2948	5.6783	29.5561	5.8626
Number of Current Birth	2.4971	1.5738	2.5896	1.6331	2.3802	1.4874
Married	0.6254	–	0.6022	–	0.6549	–
Race: White (Non-Hispanic)	0.3303	–	0.3921	–	0.2520	–
Race: Black (Non-Hispanic)	0.1671	–	0.1556	–	0.1818	–
Race: Hispanic	0.4247	–	0.4017	–	0.4539	–
Race: Other Race (Non-Hispanic)	0.0701	–	0.0472	–	0.0992	–
Edu: Less than High School	0.1499	–	0.1511	–	0.1485	–
Edu: High School Diploma	0.2318	–	0.2440	–	0.2162	–
Edu: Some College	0.2848	–	0.3093	–	0.2538	–
Edu: Bachelor's Degree or Higher	0.3152	–	0.2884	–	0.3491	–
Payment Method:						
Medicaid	0.4213	–	0.4375	–	0.4009	–
Private Insurance	0.4948	–	0.4727	–	0.5227	–
Self-Pay	0.0296	–	0.0281	–	0.0315	–
Indian Health	0.0005	–	0.0009	–	0.0001	–
Champus/Tricare	0.0133	–	0.0196	–	0.0053	–
Other Government	0.0194	–	0.0180	–	0.0211	–
Unknown Pay Method	0.0099	–	0.0100	–	0.0098	–
N	7,071,887		3,951,386		3,120,501	

Notes: Table 1 continued on next page. Dataset includes all live singleton births to adult women, with conception occurring between 1/1/2010 and 1/1/2015. Only includes states with comprehensive indoor smoking restrictions on cigarettes during entire sample period. Column 1 includes mothers from all counties. Column 2 includes only mothers from counties that never adopted the policy. Column 3 includes only mothers from counties that eventually adopted the policy within the sample period.

Table 3.1: Descriptive Statistics (Continued)

	(1) All Counties		(2) No IVR before 2015		(3) IVR before 2015	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Smoking Behavior:						
Smoking Participation	0.0654	0.2472	0.0995	0.2994	0.0222	0.1473
Cigs per Month (Conditional on Smoking)	242.6866	213.2711	248.4231	208.6388	210.1141	235.2339
Cigs per Month (Unconditional Mean)	15.8711	81.0827	24.7245	99.3136	4.6626	46.7531
Tobacco Policy:						
Indoor Vaping Restriction	0.0617	0.2236	0.0000	0.0000	0.1399	0.3200
Indoor Smoking Restriction	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000
Cigarette Tax	1.7758	1.2209	1.9949	1.2493	1.4984	1.1243
Birth Outcomes:						
Premature Birth (<37 Weeks)	0.0869	–	0.0919	–	0.0807	–
Very Premature Birth (<32 Weeks)	0.0133	–	0.0145	–	0.0118	–
Small for Gestational Age	0.2376	–	0.2358	–	0.2398	–
Extra Small for Gestational Age	0.0897	–	0.0906	–	0.0885	–
Low Birth Weight (<2500 Grams)	0.0570	–	0.0592	–	0.0543	–
Apgar 5 Score	8.8427	0.7342	8.8182	0.7924	8.8736	0.6518
N	7,071,887		3,951,386		3,120,501	

Notes: Dataset includes all live singleton births to adult women, with conception occurring between 1/1/2010 and 1/1/2015. Only includes states with comprehensive indoor smoking restrictions on cigarettes during entire sample period. Column 1 includes mothers from all counties. Column 2 includes only mothers from counties that never adopted the policy. Column 3 includes only mothers from counties that eventually adopted the policy within the sample period.

Table 3.2: Cross-Sectional, Smoking Outcomes

	(1) Smoking Participation (All Mothers)	(2) Avg. Cigs per Month (Smokers)	(3) Avg. Cigs Per Month (All Mothers)
Section A: All Counties			
Indoor Vaping Restriction	0.0093*	10.2786**	3.1682**
	(0.0035)	(3.0687)	(0.9401)
County FE	Yes	Yes	Yes
Conception Year-Month FE	Yes	Yes	Yes
Birth Year-Month FE	Yes	Yes	Yes
N	6,837,486	451,281	6,837,486
Dep. Var. Mean	0.0660	242.4162	15.9997
Section B: Binary Treated Counties Only			
Indoor Vaping Restriction	0.0128**	8.8345***	4.0867**
	(0.0037)	(1.6321)	(1.1670)
County FE	Yes	Yes	Yes
Conception Year-Month FE	Yes	Yes	Yes
Birth Year-Month FE	Yes	Yes	Yes
N	4,574,436	404,559	4,574,436
Dep. Var. Mean	0.0884	246.3900	21.7905

Notes: Section A includes all counties, while Section B excludes counties that had only part of the county population covered by indoor vaping restrictions at any time in the sample. Only counties with 0% or 100% population covered by all three restrictions (bar, restaurant, workplace) remain. Only mothers age 18 and older included. Controls for mother's age, marital status, race, education, payment source, number of current birth, and cigarette taxes in state at time of conception. Standard errors in parentheses, clustered at state level.
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.3: Test of Parallel Trends Assumption

	(1) Smoking Participation (All Mothers)	(2) Avg. Cigs per Month (Smokers)	(3) Avg. Cigs Per Month (All Mothers)
Section A: Linear Time Trend			
Linear Time Trend * Ever Adopt	0.0001	-0.0653	0.0314
	(0.0001)	(0.1948)	(0.0184)
County FE	Yes	Yes	Yes
Conception Year-Month FE	Yes	Yes	Yes
Birth Year-Month FE	Yes	Yes	Yes
N	1,554,889	141,102	1,554,889
Dep. Var. Mean	0.0907	246.0517	22.3285

Notes: Only mothers that conceived at least 9 months before actual/placebo policy adoption are included to evaluate trends in the period before any partial exposure. Excludes counties that had only part of county population covered by indoor vaping restrictions at any time in sample. Only counties with 0% or 100% population covered by all three restrictions (bar, restaurant, workplace) remain. Only mothers age 18 and older included. Controls for mother's age, marital status, race, education, payment source, number of current birth, and cigarette taxes in state at time of conception. Counties that never adopted an IVR policy are assigned a random placebo policy date, with each day between 1/1/2010 and 12/31/2014 being equally likely. Standard errors in parentheses, clustered at state level.
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.3: Test of Parallel Trends Assumption (Continued)

	(1) Smoking Participation (All Mothers)	(2) Avg. Cigs per Month (Smokers)	(3) Avg. Cigs Per Month (All Mothers)
Section B: Non-Linear Time Trend			
Conception 21-9 Mo. Before	Ref.	Ref.	Ref.
IVR * Ever Adopt			
Conception 33-21 Mo. Before	0.0010	3.4829	-0.0286
IVR * Ever Adopt	(0.0024)	(3.3723)	(0.4500)
Conception 45-33 Mo. Before	-0.0008	3.8169	-0.7388
IVR * Ever Adopt	(0.0020)	(4.3942)	(0.4751)
Conception 45+ Mo. Before	-0.0047*	-3.5644	-1.4011*
IVR * Ever Adopt	(0.0020)	(11.5211)	(0.6343)
County FE	Yes	Yes	Yes
Conception Year-Month FE	Yes	Yes	Yes
Birth Year-Month FE	Yes	Yes	Yes
N	1,554,889	141,102	1,554,889
Dep. Var. Mean	0.0907	246.0517	22.3285

Notes: Only mothers that conceived at least 9 months before actual/placebo policy adoption are included to evaluate trends in the period before any partial exposure. Excludes counties that had only part of county population covered by indoor vaping restrictions at any time in sample. Only counties with 0% or 100% population covered by all three restrictions (bar, restaurant, workplace) remain. Only mothers age 18 and older included. Controls for mother's age, marital status, race, education, payment source, number of current birth, and cigarette taxes in state at time of conception. Counties that never adopted an IVR policy are assigned a random placebo policy date, with each day between 1/1/2010 and 12/31/2014 being equally likely. Standard errors in parentheses, clustered at state level.

$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.4: Cross-Sectional Event Study, Smoking Outcomes

	(1) Smoking Participation (All Mothers)	(2) Avg. Cigs per Month (Smokers)	(3) Avg. Cigs Per Month (All Mothers)
Conception 18-15 Mo. Before IVR	0.0003	-4.3106	-0.0439
	(0.0039)	(5.3555)	(1.2564)
Conception 15-12 Mo. Before IVR	0.0008	-8.2553	-0.1589
	(0.0045)	(5.4538)	(1.3736)
Conception 12-9 Mo. Before IVR	-0.0004	-5.6749	-0.2337
	(0.0034)	(6.6881)	(1.2119)
Conception 9-6 Mo. Before IVR	0.0009	3.7833	0.5932
	(0.0035)	(7.2195)	(1.1582)
Conception 6-3 Mo. Before IVR	0.0002	-3.5442	0.1359
	(0.0044)	(7.5206)	(1.2564)
Conception 3-0 Mo. Before IVR	0.0019	0.1862	0.6947
	(0.0031)	(10.1623)	(1.1541)
Conception 0-3 Mo. After IVR	0.0059^{*}	-2.7730	1.6780^{+}
	(0.0026)	(5.3494)	(0.8235)
Conception 3+ Mo. After IVR	0.0144^{**}	6.8597	4.5990^{**}
	(0.0039)	(4.1224)	(1.2716)
County FE	Yes	Yes	Yes
Conception Year-Month FE	Yes	Yes	Yes
Birth Year-Month FE	Yes	Yes	Yes
N	4,574,436	404,559	4,574,436
Dep. Var. Mean	0.0884	246.3900	21.7905
P-Value for Wald Test of Policy Leads	0.9273	0.3932	0.8989

Notes: Omitted time category is “Conception earlier than 18 mo. before policy.” Excludes counties that had only part of county population covered by indoor vaping restrictions at any time in sample. Only counties with 0% or 100% population covered by all three restrictions (bar, restaurant, workplace) remain. Only mothers age 18 and older included. Wald test of policy leads includes 18-15 months before conception, 15-12 months before conception, and 12-9 months before conception. Controls for mother’s age, marital status, race, education, payment source, number of current birth, and cigarette taxes in state at time of conception. Standard errors in parentheses, clustered at state level.
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.5: Panel Data Analysis

	(1) Smoking Participation (All Mothers)	(2) Avg. Cigs Per Month (Smokers)	(3) Avg. Cigs Per Month (All Mothers)	(4) Smoking Participation (All Mothers)	(5) Avg. Cigs Per Month (Smokers)	(6) Avg. Cigs Per Month (All Mothers)
Indoor Vaping Restriction	0.0198*	11.6961	12.6388*	0.0200*	11.1657	12.5131*
	(0.0069)	(28.0007)	(5.1943)	(0.0071)	(28.0848)	(5.2848)
Mother FE	Yes	Yes	Yes	Yes	Yes	Yes
Trimester FE	Yes	Yes	Yes	Yes	Yes	Yes
Trimester-Year-Month FE	No	No	No	Yes	Yes	Yes
N	28,135,516	1,795,748	28,135,516	28,135,516	1,795,748	28,135,516
Dep. Var. Mean	0.0638	316.3686	20.1922	0.0638	316.3686	20.1922

Notes: Only mothers age 18 and older included. Controls for cigarette taxes in county. Standard errors in parentheses, clustered at state level.
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.6: Panel Data Analysis, Stratified by Age, Education, and Payment Method

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Smoking Participation ($Age < 25$)	Smoking Participation ($Age \geq 25$)	Smoking Participation (Low Edu.)	Smoking Participation (High Edu.)	Smoking Participation (Medicaid)	Smoking Participation (Private Ins.)	Smoking Participation (Self Pay)
Indoor Vaping Restriction	0.0323^{*}	0.0146^{**}	0.0227^{*}	0.0174^{**}	0.0238^{*}	0.0153^{**}	0.0125^{+}
	(0.0128)	(0.0049)	(0.0101)	(0.0056)	(0.0105)	(0.0049)	(0.0068)
Mother FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Trimester FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Trimester-Year-Month FE	No	No	No	No	No	No	No
N	7,194,068	20,941,448	10,716,456	16,915,556	11,829,904	13,955,876	833,496
Dep. Var. Mean	0.1048	0.0497	0.1052	0.0387	0.1037	0.0297	0.0471

Notes: Column 1 includes only mothers between ages 18 and 24 (inclusive). Column 2 includes only mothers age 25 or older. All remaining columns include only mothers age 18 or older. Columns 3 and 4 include only mothers with a high school degree or less, and some college or more, respectively. Columns 5, 6, and 7 include only mothers with Medicaid, private insurance, and self-pay, respectively. Controls for cigarette taxes in county. Standard errors in parentheses, clustered at state level.

$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.7: Cross-Sectional, Birth Outcomes

	(1) Premature (<37 Weeks)	(2) Very Premature (<32 Weeks)	(3) Small for Gest. Age	(4) Extra Small for Gest. Age	(5) Low Weight (<2500 Grams)	(6) Apgar 5 Score
Indoor Vaping Restriction	0.0002	-0.0003	0.0009	0.0006	0.0008	-0.0083
	(0.0014)	(0.0006)	(0.0066)	(0.0028)	(0.0004)	(0.0081)
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Conception Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Birth Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes
N	6,867,005	6,867,005	6,864,621	6,864,621	6,864,622	6,843,308
Dep. Var. Mean	0.0864	0.0131	0.2372	0.0895	0.0565	8.8449

Notes: Only mothers age 18 and older included. Controls for mother's age, marital status, race, education, payment source, number of current birth, and cigarette taxes in state at time of conception. Standard errors in parentheses, clustered at state level.
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.8: Panel Data Analysis, Non-Smokers in 3 Months Prior to Pregnancy Only

	(1) Smoking Participation	(2) Avg. Cigs Per Month	(3) Smoking Participation	(4) Avg. Cigs Per Month
Section A: All Mothers				
Indoor Vaping Restriction	-0.0003*	-0.0723*	-0.0003^{+}	-0.0728^{+}
	(0.0001)	(0.0329)	(0.0001)	(0.0361)
Mother FE	Yes	Yes	Yes	Yes
Trimester FE	Yes	Yes	Yes	Yes
Trimester-Year-Month FE	No	No	Yes	Yes
N	25,710,404	25,710,404	25,710,404	25,710,404
Dep Var. Mean	0.0004	0.1099	0.0004	0.1099
Section B: Mothers Under 25				
Indoor Vaping Restriction	-0.0005^{+}	-0.1098	-0.0005^{+}	-0.1199
	(0.0002)	(0.0819)	(0.0003)	(0.0917)
Mother FE	Yes	Yes	Yes	Yes
Trimester FE	Yes	Yes	Yes	Yes
Trimester-Year-Month FE	No	No	Yes	Yes
N	6,177,412	6,177,412	6,177,412	6,177,412
Dep Var. Mean	0.0007	0.1797	0.0007	0.1797
Section C: Mothers 25 and Older				
Indoor Vaping Restriction	-0.0002*	-0.0597^{+}	-0.0002^{+}	-0.0579^{+}
	(0.0001)	(0.0282)	(0.0001)	(0.0310)
Mother FE	Yes	Yes	Yes	Yes
Trimester FE	Yes	Yes	Yes	Yes
Trimester-Year-Month FE	No	No	Yes	Yes
N	19,532,992	19,532,992	19,532,992	19,532,992
Dep Var. Mean	0.0003	0.0878	0.0003	0.0878

Notes: Only mothers age 18 and older included. Controls for cigarette taxes in county. Standard errors in parentheses, clustered at state level.
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.9: Cross-Sectional, Non-Smokers Prior to Pregnancy Only, Birth Outcomes

	(1) Premature (<37 Weeks)	(2) Very Premature (<32 Weeks)	(3) Small for Gest. Age	(4) Extra Small for Gest. Age	(5) Low Weight (<2500 Grams)	(6) Apgar 5 Score
Indoor Vaping Restriction	0.0002	-0.0002	-0.0006	-0.0004	0.0003	-0.0095
	(0.0015)	(0.0006)	(0.0072)	(0.0031)	(0.0005)	(0.0079)
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Conception Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Birth Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes
N	6,270,487	6,270,487	6,268,429	6,268,429	6,268,430	6,248,741
Dep. Var. Mean	0.0835	0.0123	0.2293	0.0844	0.0529	8.8506

Notes: Only mothers age 18 and older included. Controls for mother's age, marital status, race, education, payment source, number of current birth, and cigarette taxes in state at time of conception. Standard errors in parentheses, clustered at state level.
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.10: Cross-Sectional Event Study, Birth Outcomes

	(1) Premature (<37 Weeks)	(2) Very Premature (<32 Weeks)	(3) Small for Gest. Age	(4) Extra Small for Gest. Age	(5) Low Weight (<2500 Grams)	(6) Apgar 5 Score
Conception 18-15 Mo. Before IVR	0.0012	-0.0002	0.0007	0.0001	0.0007	0.0092
	(0.0019)	(0.0004)	(0.0033)	(0.0016)	(0.0006)	(0.0060)
Conception 15-12 Mo. Before IVR	0.0003	-0.0003	0.0043*	0.0029*	0.0002	0.0052
	(0.0012)	(0.0003)	(0.0019)	(0.0013)	(0.0013)	(0.0063)
Conception 12-9 Mo. Before IVR	0.0014	-0.0006	0.0004	0.0007	-0.0011	0.0014
	(0.0020)	(0.0008)	(0.0046)	(0.0022)	(0.0011)	(0.0089)
Conception 9-6 Mo. Before IVR	0.0014	0.0001	0.0029	-0.0002	-0.0002	0.0095
	(0.0025)	(0.0006)	(0.0034)	(0.0022)	(0.0011)	(0.0098)
Conception 6-3 Mo. Before IVR	-0.0013	0.0002	0.0015	-0.0011	0.0008	0.0028
	(0.0014)	(0.0007)	(0.0048)	(0.0013)	(0.0012)	(0.0102)
Conception 3-0 Mo. Before IVR	-0.0034*	-0.0012*	0.0052	0.0012	0.0002	-0.0069
	(0.0013)	(0.0005)	(0.0049)	(0.0027)	(0.0014)	(0.0072)
Conception 0-3 Mo. After IVR	0.0002	-0.0005	0.0025	-0.0001	0.0017	-0.0049
	(0.0007)	(0.0006)	(0.0051)	(0.0023)	(0.0022)	(0.0060)
Conception 3+ Mo. After IVR	-0.0000	-0.0006	0.0028	0.0014	0.0012	-0.0139
	(0.0009)	(0.0007)	(0.0074)	(0.0030)	(0.0008)	(0.0087)
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Conception Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Birth Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes
N	4,591,515	4,591,515	4,589,490	4,589,490	4,589,491	4,577,551
Dep. Var. Mean	0.0901	0.0138	0.2364	0.0904	0.0580	8.8270
P-Value for Wald Test of Policy Leads	0.8784	0.6002	0.1370	0.1884	0.2105	0.2347

Notes: Excluded time category is "Conception earlier than 18 mo. before policy." Excludes counties that had only part of county population covered by indoor vaping restrictions at any time in sample. Only counties with 0% or 100% population covered by all three restrictions (bar, restaurant, workplace) remain. Only mothers age 18 and older included. Wald test of policy leads includes 18-15 months before conception, 15-12 months before conception, and 12-9 months before conception. Controls for mother's age, marital status, race, education, payment source, number of current birth, and cigarette taxes in state at time of conception. Standard errors in parentheses, clustered at state level.
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Bibliography

[1] Abouk, Rahi, and Scott Adams. "Bans on electronic cigarette sales to minors and smoking among high school students." *Journal of Health Economics* 54 (2017): 17-24.

[2] Acemoglu, Daron, and Joshua D. Angrist. "Consequences of employment protection? The case of the Americans with Disabilities Act." *Journal of Political Economy* 109, no. 5 (2001): 915-957.

[3] Altmann, Steffen, Armin Falk, Simon Jger, and Florian Zimmermann. "Learning about job search: A field experiment with job seekers in Germany." *Journal of Public Economics* 164 (2018): 33-49.

[4] American Lung Association, 2017. "Tobacco–Affordable Care Act Timeline." Retrieved from: `https://www.lung.org/policy-advocacy/healthcare-lung-disease/healthcare-policy/affordable-care-act-timeline`.

[5] Andolfatto, David, Steeve Mongrain, and Gordon M. Myers. "Self-esteem and labour market choices." (2005).

[6] Autor, David H., William R. Kerr, and Adriana D. Kugler. "Does employment protection reduce productivity? Evidence from US states." *The Economic Journal* 117, no. 521 (2007): F189-F217.

[7] Barber, Alison E., Christina L. Daly, Cristina M. Giannantonio, and Jean M. Phillips. "Job search activities: An examination of changes over time." *Personnel psychology* 47, no. 4 (1994): 739-766.

[8] Been, Jasper V., Ulugbek B. Nurmatov, Bianca Cox, Tim S. Nawrot, Constant P. van Schayck, and Aziz Sheikh. "Effect of smoke-free legislation on perinatal and child health: a systematic review and meta-analysis." *The Lancet* 383, no. 9928 (2014): 1549-1560.

[9] Belot, Michele, Philipp Kircher, and Paul Muller. "Providing advice to jobseekers at low cost: an experimental study on online advice." *The Review of Economic Studies* 86, no. 4 (2018): 1411-1447.

[10] Benabou, Roland, and Jean Tirole. "Self-confidence and personal motivation." *The Quarterly Journal of Economics* 117, no. 3 (2002): 871-915.

[11] Berg, Joyce, John Dickhaut, and Kevin McCabe. "Trust, reciprocity, and social history." *Games and Economic Behavior* 10, no. 1 (1995): 122-142.

[12] Brandts, Jordi, and Gary Charness. "The Strategy Versus the Direct-Response Method: A First Survey of Experimental Comparisons." *Experimental Economics* 14, no.3 (2011): 375-398.

[13] Brown, Meta, Christopher J. Flinn, and Andrew Schotter. "Real-time search in the laboratory and the market." *American Economic Review* 101, no. 2 (2011): 948-74.

[14] Buescher, Paul A., Karen P. Taylor, Mary H. Davis, and J. Michael Bowling. "The quality of the new birth certificate data: a validation study in North Carolina." *American Journal of Public Health* 83, no. 8 (1993): 1163-1165.

[15] Bull, Clive, Andrew Schotter, and Keith Weigelt. "Tournaments and piece rates: An experimental study." *Journal of Political Economy* 95, no. 1 (1987): 1-33.

[16] Cadsby, C. Bram, Fei Song, and Francis Tapon. "Sorting and incentive effects of pay for performance: An experimental investigation." *Academy of Management Journal* 50, no. 2 (2007): 387-405.

[17] Cahuc, Pierre, and Fabien Postel-Vinay. "Temporary jobs, employment protection and labor market performance." *Labour Economics* 9, no. 1 (2002): 63-91.

[18] Caliendo, Marco, Deborah A. Cobb-Clark, and Arne Uhlendorff. "Locus of control and job search strategies." *Review of Economics and Statistics* 97, no. 1 (2015): 88-103.

[19] Campaign for Tobacco Free Kids, 2017. "State Excise Tax Rates for Non-Cigarette Tobacco Products." Retrieved from: `https://www.tobaccofreekids.org/research/factsheets/pdf/0169.pdf`.

[20] Caplin, Andrew, Daniel Csaba, John Leahy, and Oded Nov. "Rational Inattention, Competitive Supply, and Psychometrics." *The Quarterly Journal of Economics* 135, no. 3 (2020): 1681-1724.

[21] Caplin, Andrew, Mark Dean, and Daniel Martin. "Search and satisficing." *American Economic Review* 101, no. 7 (2011): 2899-2922.

[22] Cappelli, Peter, and Keith Chauvin. "An Interplant Test of the Efficiency Wage Hypothesis." *The Quarterly Journal of Economics* 106, no. 3 (1991): 769-787.

[23] Centers for Disease Control and Prevention, 2017a. "State Tobacco Activities Tracking and Evaluation (STATE) System." Retrieved from: `https://www.cdc.gov/statesystem/`.

[24] Centers for Disease Control and Prevention, 2017b. "Tobacco Use During Pregnancy." Retrieved from: `http://www.cdc.gov/reproductivehealth/maternalinfanthealth/tobaccousepregnancy/l`.

[25] Chaloupka, Frank J., David Sweanor, and Kenneth E. Warner. "Differential Taxes for Differential Risks?Toward Reduced Harm from Nicotine-Yielding Products." *New England Journal of Medicine* 373, no. 7 (2015): 594-597.

[26] Charness, Gary, Uri Gneezy, and Austin Henderson. "Experimental methods: Measuring effort in economics experiments." *Journal of Economic Behavior & Organization* 149 (2018): 74-87.

[27] Chen, Don JQ, and Vivien KG Lim. "Strength in adversity: The influence of psychological capital on job search." *Journal of Organizational Behavior* 33, no. 6 (2012): 811-839.

[28] Chen, Daniel L., Martin Schonger, and Chris Wickens. "oTree – An open-source platform for laboratory, online, and field experiments." *Journal of Behavioral and Experimental Finance* 9 (2016): 88-97.

[29] Corgnet, Brice, Roberto Hernán-González, and Stephen Rassenti. "Firing threats: Incentive effects and impression management." *Games and Economic Behavior* 91 (2015): 97-113.

[30] Corgnet, Brice, Roberto Hernán-González, and Eric Schniter. "Why real leisure really matters: Incentive effects on real effort in the laboratory." *Experimental Economics* 18, no. 2 (2015): 284-301.

[31] Cornsweet, Tom N. "The staircase-method in psychophysics." *The American Journal of Psychology* 75, no. 3 (1962): 485-491.

[32] Cox, James C., and Ronald L. Oaxaca. "Laboratory experiments with a finite-horizon job-search model." *Journal of Risk and Uncertainty* 2, no. 3 (1989): 301-329.

[33] Cromwell, Erich, Sebastian J. Goerg, and Monika Leszczynska. "More than the Money: Payoff-Irrelevant Terms in Relational Contracts." No. 11712. IZA Discussion Papers, 2018.

[34] Curtin, Sally C., and T. J. Matthews. "Smoking prevalence and cessation before and during pregnancy: data from the birth certificate, 2014." *National vital statistics reports: from the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System* 65, no. 1 (2016): 1-14.

[35] Dave, Dhaval, Bo Feng, and Michael Pesko. "The Effects of E-Cigarette Minimum Legal Sale Age Laws on Youth Substance Use." No. w23313. National Bureau of Economic Research, 2017.

[36] DellaVigna, Stefano, Attila Lindner, Balzs Reizer, and Johannes F. Schmieder. "Reference-dependent job search: Evidence from Hungary." *The Quarterly Journal of Economics* 132, no. 4 (2017): 1969-2018.

[37] DellaVigna, Stefano, and M. Daniele Paserman. "Job search and impatience." *Journal of Labor Economics* 23, no. 3 (2005): 527-588.

[38] Dohmen, Thomas, and Armin Falk. "Performance pay and multidimensional sorting: Productivity, preferences, and gender." *American Economic Review* 101, no. 2 (2011): 556-90.

[39] Eil, David, and Justin M. Rao. "The good news-bad news effect: asymmetric processing of objective information about yourself." *American Economic Journal: Microeconomics* 3, no. 2 (2011): 114-38.

[40] Ertac, Seda. "Does self-relevance affect information processing? Experimental evidence on the response to performance and non-performance feedback." *Journal of Economic Behavior & Organization* 80, no. 3 (2011): 532-545.

[41] Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde. "Global evidence on economic preferences." *The Quarterly Journal of Economics* 133, no. 4 (2018): 1645-1692.

[42] Falk, Armin, David Huffman, and Uwe Sunde. "Self-confidence and search." (2006).

[43] Fallucchi, Francesco, Daniele Nosenzo, and Ernesto Reuben. "Measuring preferences for competition with experimentally-validated survey questions." *Journal of Economic Behavior & Organization* 178 (2020): 402-423.

[44] Farber, Henry. "Job Loss and the Decline in Job Security in the United States." In *Labor in the New Economy*, 223-262, University of Chicago Press, 2010.

[45] Fehrenbacher, Dennis D., Steven E. Kaplan, and Burkhard Pedell. "The relation between individual characteristics and compensation contract selection." *Management Accounting Research* 34 (2017): 1-18.

[46] Food and Drug Administration, 2017. "FDA 101: Smoking Cessation Products." Retrieved from: `https://www.fda.gov/ForConsumers/ConsumerUpdates/ucm198176.html`.

[47] Forsythe, Robert, Joel L. Horowitz, Nathan E. Savin, and Martin Sefton. "Fairness in simple bargaining experiments." *Games and Economic Behavior* 6, no. 3 (1994): 347-369.

[48] Freeman, Richard and James Medoff. *What Do Unions Do?* New York: Basic Books.

[49] Friedman, Abigail S. "How does electronic cigarette access affect adolescent smoking?" *Journal of Health Economics* 44 (2015): 300-308.

[50] Gee, Laura K. "The more you know: information effects on job application rates in a large field experiment." *Management Science* 65, no. 5 (2018): 2077-2094.

[51] Golman, Russell, David Hagmann, and George Loewenstein. "Information avoidance." *Journal of Economic Literature* 55, no. 1 (2017): 96-135.

[52] Goniewicz, Maciej Lukasz, Jakub Knysak, Michal Gawron, Leon Kosmider, Andrzej Sobczak, Jolanta Kurek, Adam Prokopowicz et al. "Levels of selected carcinogens and toxicants in vapour from electronic cigarettes." *Tobacco Control* 23, no. 2 (2014): 133-139.

[53] Guiteras, Raymond, and Kelsey Jack. "Productivity in piece-rate labor markets: Evidence from rural Malawi." *Journal of Development Economics* 131 (2018): 42-61.

[54] Harris Poll, 2019. "Layoff Anxiety: A Study in Modern Pressure." June 25-27, 2019. Retrieved from: `http://web.careerarc.com/CS-OP-HarrisPoll-Layoff-Anxiety.html`, last accessed on 25 October 2020.

[55] Hartmann?Boyce, Jamie, Hayden McRobbie, Chris Bullen, Rachna Begh, Lindsay F. Stead, and Peter Hajek. "Electronic cigarettes for smoking cessation." *The Cochrane Database of Systematic Reviews*, no. 9 (2016).

[56] Hossain, Tanjim, and Ryo Okui. "The binarized scoring rule." *Review of Economic Studies* 80, no. 3 (2013): 984-1001.

[57] Howland, Renata E., Candace Mulready-Ward, Ann M. Madsen, Judith Sackoff, Michael Nyland-Funke, Jennifer M. Bombard, and Van T. Tong. "Reliability of reported maternal smoking: comparing the birth certificate to maternal worksheets and prenatal and hospital medical records, New York City and Vermont, 2009." *Maternal and Child Health Journal* 19, no. 9 (2015): 1916-1924.

[58] Jacob, Brian A. "Do principals fire the worst teachers?" *Educational Evaluation and Policy Analysis* 33, no. 4 (2011): 403-434.

[59] Jacob, Brian A. "The effect of employment protection on teacher effort." *Journal of Labor Economics* 31, no. 4 (2013): 727-761.

[60] Jarlenski, Marian P., Margaret S. Chisolm, Sarah Kachur, Donna M. Neale, and Wendy L. Bennett. "Use of pharmacotherapies for smoking cessation: analysis of pregnant and postpartum Medicaid enrollees." *American Journal of Preventive Medicine* 48, no. 5 (2015): 528-534.

[61] Kopányi-Peuker, Anita, Theo Offerman, and Randolph Sloof. "Team Production Benefits from a Permanent Fear of Exclusion." *European Economic Review* 103 (2018): 125-149.

[62] Kwon, Illoong. "Threat of dismissal: incentive or sorting?" *Journal of Labor Economics* 23, no. 4 (2005): 797-838.

[63] Kudlyak, Marianna, Damba Lkhagvasuren, and Roman Sysuyev. "Systematic job search: New evidence from individual job application data." (2013).

[64] Kuhn, Peter. "Wages, effort, and incentive compatibility in life-cycle employment contracts." *Journal of Labor Economics* 4, no. 1 (1986): 28-49.

[65] Lazear, Edward P. "Performance pay and productivity." *The American Economic Review* 90, no. 5 (2000): 1346-1361.

[66] Lazear, Edward P. "Why is there mandatory retirement?" *Journal of Political Economy* 87, no. 6 (1979): 1261-1284.

[67] Lazear, Edward P., Kathryn L. Shaw, and Christopher Stanton. "Making Do with Less: Working Harder during Recessions." *Journal of Labor Economics* 34, no. S1 (2016): S333-S360.

[68] Leuven, Edwin, Hessel Oosterbeek, Joep Sonnemans, and Bas Van Der Klaauw. "Incentives versus sorting in tournaments: Evidence from a field experiment." *Journal of Labor Economics* 29, no. 3 (2011): 637-658.

[69] Markowitz, Sara, E. Kathleen Adams, Patricia M. Dietz, Viji Kannan, and Van Tong. "Smoking policies and birth outcomes: estimates from a new era." No. w17160. National Bureau of Economic Research, 2011.

[70] Marti, Joachim, John Buckell, Johanna Catherine Maclean, and Jody L. Sindelar. "To ?Vape? or Smoke? A discrete choice experiment among US adult smokers." No. w22079. National Bureau of Economic Research, 2016.

[71] McKee-Ryan, Frances, Zhaoli Song, Connie R. Wanberg, and Angelo J. Kinicki. "Psychological and physical well-being during unemployment: a meta-analytic study." *Journal of Applied Psychology* 90, no. 1 (2005): 53.

[72] Miura, Takahiro, Keigo Inukai, and Masaru Sasaki. "The effect of risk attitudes on search behavior: A laboratory search experiment." Graduate School of Economics and Osaka School of International Public Policy (OSIPP) Osaka University Discussion Papers In Economics And Business 17 (2017): 1-16.

[73] Mobius, Markus M., Muriel Niederle, Paul Niehaus, and Tanya S. Rosenblat. "Managing self-confidence: Theory and experimental evidence." No. w17014. National Bureau of Economic Research, 2011.

[74] Niederle, Muriel, and Lise Vesterlund. “Do women shy away from competition? Do men compete too much?” *The Quarterly Journal of Economics* 122, no. 3 (2007): 1067-1101.

[75] Oprea, Ryan. “Survival Versus Profit Maximization in a Dynamic Stochastic Experiment.” *Econometrica* 82, no. 6 (2014): 2225-2255.

[76] Oncken, Cheryl, Karen A. Ricci, Chia-Ling Kuo, Ellen Dornelas, Henry R. Kranzler, and Heather Z. Sankey. “Correlates of electronic cigarettes use before and during pregnancy.” *Nicotine & Tobacco Research* 19, no. 5 (2017): 585-590.

[77] Paserman, M. Daniele. ”Job search and hyperbolic discounting: Structural estimation and policy evaluation.” *The Economic Journal* 118, no. 531 (2008): 1418-1452.

[78] Paul, Karsten I., and Klaus Moser. “Unemployment impairs mental health: Meta-analyses.” *Journal of Vocational behavior* 74, no. 3 (2009): 264-282.

[79] Pesko, Michael F., and Janet M. Currie. “The effect of e-cigarette minimum legal sale age laws on traditional cigarette use and birth outcomes among pregnant teenagers.” No. w22792. National Bureau of Economic Research, 2016.

[80] Pesko, Michael F., Jenna M. Hughes, and Fatima S. Faisal. “The influence of electronic cigarette age purchasing restrictions on adolescent tobacco and marijuana use.” *Preventive Medicine* 87 (2016): 207-212.

[81] Potter, Tristan. “Learning and job search dynamics during the Great Recession.” No. 2017-6. LeBow College of Business, Drexel University, 2018.

[82] Rammstedt, Beatrice, and Oliver P. John. “Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German.” *Journal of Research in Personality* 41, no. 1 (2007): 203-212.

[83] Ross, Stephen A. “The economic theory of agency: The principal’s problem.” *The American Economic Review* 63, no. 2 (1973): 134-139.

[84] Rothschild, Michael. “Searching for the lowest price when the distribution of prices is unknown.” In *Uncertainty in Economics*, pp. 425-454. Academic Press, 1978.

[85] Roy, Andred Donald. “Some Thoughts on the Distribution of Earnings.” *Oxford Economic Papers* 3, no. 2 (1951): 135-146.

[86] Royal College of Physicians in England, 2016. “Nicotine Without Smoke – Tobacco Harm Reduction.” Retrieved from: `https://www.rcplondon.ac.uk/projects/outputs/nicotine-without-smoke-tobacco-harm-reduction-0`.

[87] Schunk, Daniel. “Behavioral heterogeneity in dynamic search situations: Theory and experimental evidence.” *Journal of Economic Dynamics and Control* 33, no. 9 (2009): 1719-1738.

[88] Schunk, Daniel, and Joachim Winter. “The relationship between risk attitudes and heuristics in search tasks: A laboratory experiment.” *Journal of Economic Behavior & Organization* 71, no. 2 (2009): 347-360.

[89] Shapiro, Carl, and Joseph E. Stiglitz. “Equilibrium unemployment as a worker discipline device.” *The American Economic Review* 74, no. 3 (1984): 433-444.

[90] Simon, David. “Does early life exposure to cigarette smoke permanently harm childhood welfare? Evidence from cigarette tax hikes.” *American Economic Journal: Applied Economics* 8, no. 4 (2016): 128-59.

[91] Siu, Albert L. “Behavioral and pharmacotherapy interventions for tobacco smoking cessation in adults, including pregnant women: US Preventive Services Task Force recommendation statement.” *Annals of Internal Medicine* 163, no. 8 (2015): 622-634.

[92] Spinnewijn, Johannes. “Unemployed but optimistic: Optimal insurance design with biased beliefs.” *Journal of the European Economic Association* 13, no. 1 (2015): 130-167.

[93] Stiglitz, Joseph E. “Incentives, risk, and information: notes towards a theory of hierarchy.” *The Bell Journal of Economics* (1975): 552-579.

[94] Tong, Van T., Lucinda J. England, Patricia M. Dietz, and Lisa A. Asare. “Smoking patterns and use of cessation interventions during pregnancy.” *American Journal of Preventive Medicine* 35, no. 4 (2008): 327-333.

[95] US Department of Health and Human Services. “The health consequences of smoking–50 years of progress: a report of the surgeon general.” 2014.

[96] Viscusi, W. Kip. “Risk beliefs and preferences for e-cigarettes.” *American Journal of Health Economics* 2, no. 2 (2016): 213-240.

[97] Wanberg, Connie R. “The individual experience of unemployment.” *Annual Review of Psychology* 63 (2012): 369-396.

[98] Wanberg, Connie R., Jing Zhu, and Edwin AJ Van Hooft. “The job search grind: Perceived progress, self-reactions, and self-regulation of search effort.” *Academy of Management Journal* 53, no. 4 (2010): 788-807.

[99] Wilson, A., and Emanuel Vespa. “Paired-uniform scoring: Implementing a binarized scoring rule with non-mathematical language.” (2018).

Appendix A

Appendix Materials for "Dismissal Threats as an Incentive Device"

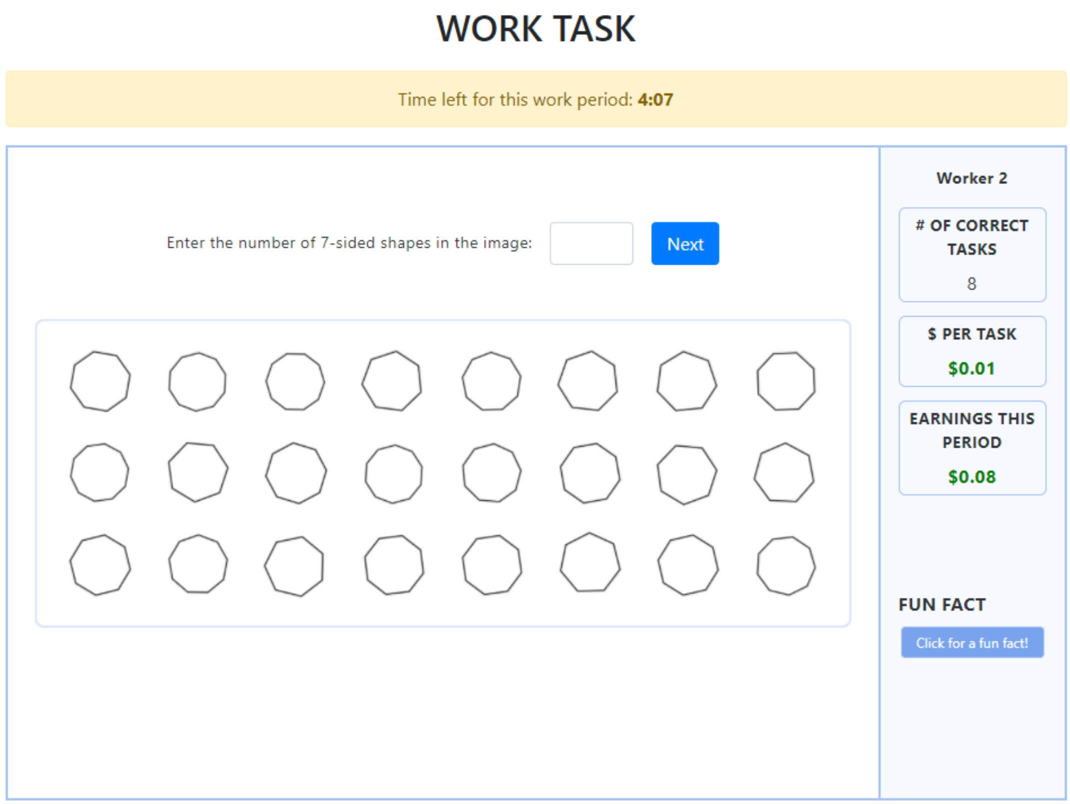
A.1 Real Effort Task

In the real effort task used in this experiment, subjects are shown a randomly generated image with 24 regular polygons (3 rows of 8 shapes each). Each shape may have 7, 8, 9, or 10 sides. In each of the 24 spots for shapes, each of these four polygons is equally likely to appear. Additionally, every shape is rotated a random amount ranging from 0 to 360 degrees. The real effort task for the subjects was to count the number of 7-sided shapes in each image. Although the answer could theoretically range from 0 to 24, in practice most answers were between 4 and 8.

All work periods were timed. To penalize guessing, when a subject submitted an incorrect answer, it did not count as a completed task and a 10-second delay occurred before the next image appeared.

Importantly, each subject was shown the same randomly generated images in the same order as every other subject within the same work period. That means there cannot be subjects who face particularly difficult or easy images compared to other subjects in the same work period.

A screenshot of the work interface during the low piece rate period is shown below.



A.2 Worker Traits

A.2.1 Measurement Procedures

Part 2 of the experiment involved a series of incentivized decisions to measure worker traits along with a set of unincentivized survey questions. Only one incentivized task or decision was randomly selected for payment from Part 2, and this was explained to subjects once they got to this part of the experiment. Some tasks were single-player decisions, while others were two-player games. Subjects were told that only one task would be randomly selected for payment, but that if a two-player game was selected, that game would also be selected for a counterpart subject in the experiment and the payoffs would be implemented for both players. An algorithm was designed to make it as close to equally likely as possible for each task to be selected for payment, under the constraint that all two-player decisions selected required a counterpart assigned to the same game.

Standard methods were used to elicit the following worker traits in an incentive-compatible way: confidence, performance under high stakes, risk aversion, loss aversion, patience, altruism, trust, and reciprocity. Confidence was measured by asking subjects to predict the rank they earned in a group of 10 randomly selected subjects; if the rank they reported was within 1 rank of the truth, they would be paid \$5. Subjects faced a high-stakes work period in which they completed the same shape-based work task from earlier in the experiment. They would earn \$5 if they completed 4 tasks within 45 seconds. To increase feelings of pressure and stress, the work period timer was significantly enlarged and presented with a bright red background at the top of the screen.

Risk aversion, loss aversion, and patience were all elicited using a variant of the Multiple Price List method called the staircase procedure (see Cornsweet, 1962, for an early definition). In this method, the subject is presented with a series of binary decisions that adjust the options presented in subsequent screens based on past decisions. For example, the initial choice in the risk aversion elicitation is between a 50/50 gamble for \$0 or \$6, or taking a certain amount of \$2.90. If the subject selects the gamble, the certain amount increases for the next offer. If the subject selects the certain amount, the certain amount decreases for the next offer. This staircase method allows precise measurements of subject preferences while requiring much fewer decisions than a standard Multiple Price List would require. For each of these three trait measurements, subjects made 5 decisions but could end up with any of $2^5 = 32$ potential values. The parameters of the staircase choice sets were constructed similarly to those in Falk et al.'s (2018) global survey of economic preferences.

Risk aversion was measured with a series of decisions between a 50/50 gamble for winning \$6 or winning \$0, or taking a variable certain amount. The final value of this certainty equivalent could range from \$0.10 to \$6.30. The loss aversion decision was between a 50/50 gamble for winning \$5 or losing a variable amount of money, or staying at \$0 for this task.[1] The final value of the potential loss could range from \$0.10 to \$6.30. The patience decision was between being paid \$3 today or a variable amount in one week. The final value of the variable amount could range between \$2.90 and \$6.00. Any future dollar amounts selected for payment from this task were actually paid out on Venmo exactly one week after the session ended.

Subjects played two different two-player games, and the strategy method was used to observe the decisions of each subject in both games and both roles (sender and receiver).[2] First, subjects played a Dictator Game in which one player received \$8.00 and could split any amount of it with another player who received \$0 (Forsythe et al., 1994). Second, subjects played a Trust Game in which each player started with \$3 (Berg, Dichaut, and McCabe, 1995). The sender could send a restricted choice set of \$0, \$1, \$2, or \$3 to the receiver, and any amount sent would be tripled. The receiver then input any amount (including decimals) from their new balance to send back to the sender for each potential amount initially sent. The strategy method for the receiver role involved all subjects inputting a dollar amount for each of the four potential amounts initially sent by the

[1] To prevent the possibility of subjects earning below the minimum payment of \$5, subjects were told that \$3 would be added to their earnings only if this task was the one selected for payment. Although this could interfere with the loss framing, it was presented in a less salient manner on a separate screen; the actual gains and losses displayed on the decision page were shown in bold green or red text and did not take into account this extra payment. Based on subject decisions, this appeared to have been successful in invoking the loss framing even when the \$3 payment would have brought them out of the loss domain in this task.

[2] In a survey of experimental methodology, Brandts and Charness (2011) do not find strong evidence of the strategy method changing subject behavior

sender, allowing analysis of the relationship between the amount sent and the amount returned.

A series of unincentivized survey questions were also included in Part 2 to measure other worker traits. Subjects were asked to report the level of stress they felt during the high-stakes work period, along with whether they work well under pressure. They were also asked two questions about preferences for competition: how much they agree with the statement that "Competition brings out the best in me," and whether they agree more with the statement that "Competition is harmful. It brings out the worst in people," or "Competition is good. It stimulates people to work hard and develop new ideas." The phrasing of these questions comes from Fallucchi, Nosenzo, and Reuben (2020) who find that these two survey questions are the most predictive of incentivized measurements of preferences to compete. All of these questions were asked on a scale of one to seven. Finally, a short version of the Big 5 personality trait test was administered (Rammstedt and John, 2007). Subjects answered two questions corresponding to each trait which were combined to create a score from one to five. Finally, subjects were asked to report a few basic demographic variables including age, gender, race, and ethnicity.

At the end of the experiment, total earnings were displayed to subjects including a breakdown of how much they earned from each part of the experiment, and which decision was selected for payment in Part 2. Nothing was explicitly reported to subjects about cumulative earnings up until this point, when all subjects had completed all decisions in the experiment.

A.2.2 Descriptive Statistics

A short description of worker traits and demographics will be provided in this subsection. The mean and standard deviation of each incentivized trait and survey measure are displayed in Tables A1 and A2, respectively. Subjects reported a mean rank in the confidence measurement of 4.93, and they were on average slightly overconfident (overestimating their placement by 0.83 ranks). Note that the confidence measurement is a rank out of 10, with rank 1 being the most productive, so a lower number corresponds to higher confidence. The mean certainty equivalent for a 50/50 gamble between \$0 and \$6 was \$3.04, but this mean is increased by a small number of outliers who appeared highly risk-loving. The median certainty equivalent was \$2.90, and 61.4% of subjects display risk aversion over these small dollar amounts. Subjects show loss aversion on average because the mean acceptable 50/50 gamble involving the loss domain was between winning \$5 or losing \$2.54. In fact, 98% of subjects rejected gambles involving losses with expected values greater than \$0. With regards to patience, many subjects appeared to want to maximize the dollar amount received without regard to timing. Nearly half of subjects (47.5%) were willing to forgo \$3 today to receive any amount greater than or equal to \$3.05 in 1 week. However, no subject accepted \$2.95 in 1 week over \$3 today, suggesting that subjects were attentive and not simply clicking one option repeatedly without thought. Because of this high level of patience among subjects, there was a strong mode in the patience variable and a small standard deviation. However, among the 52.5% of subjects who were not simply selecting the higher dollar amount regardless of timing, there was much more dispersion in patience, ranging from accepted future values of \$3.10 to the maximum of \$6.

Measures of altruism, trust, and reciprocity were calculated from the two-player games. The mean contribution in the Dictator Game was \$2.40, and 24.8% of dictators sent nothing. High levels of trust were observed by senders in the Trust Game: the mean amount sent was \$1.99, and 42.6% of subjects sent the maximum amount of \$3. Reciprocity is not as straightforward to summarize given that it is a function of four different choices by the receiver – one for each potential dollar amount initially sent. To

Table A1: Descriptive Statistics (Incentivized Traits)

	(1) Worker Traits mean/sd
High Piece Rate Tasks	10.61
	(4.91)
Responsiveness to Piece Rate	0.82
	(1.45)
High Stakes Tasks	2.44
	(1.29)
Confidence: Rank Choice	4.93
	(2.13)
Confidence: Actual - Reported Rank	0.83
	(2.50)
Risk Choice	3.04
	(1.31)
Loss Choice	2.54
	(1.31)
Patience Choice	3.40
	(0.66)
Social Prefs: Altruism	2.40
	(1.76)
Social Prefs: Trust	1.99
	(1.02)
Social Prefs: Reciprocity	1.46
	(0.71)
Observations	101

Notes: The table shows the mean value of each trait for subjects in the experiment. Standard deviations are shown in parentheses.

measure reciprocity, a linear relationship was calculated between the amount sent and the amount returned. This was accomplished with an OLS regression run separately for each subject with the amount returned as the outcome variable and the amount initially sent as the explanatory variable (with the intercept forced to zero). The results of this data exercise indicate that for each dollar sent by the Sender (multiplied into \$3 for the Receiver), the Receiver sends back \$1.46 on average (i.e., the Receivers send roughly half of the multiplied money back).

Subject demographics were as follows. The pool of participants consisted of 37.6% male, 59.4% female, and 3.0% other. The race and ethnicity of the participants was 28.7% White, 50.5% Asian, 1% Black, and 19.8% other. The demographic variable on Hispanic identification was collected separately (in line with Census guidelines); 22.8% of subjects identified as Hispanic. The mean age of the subject pool was 20.7 years old, with a minimum age of 18 and a maximum age of 29.

A.2.3 Selection into Dismissal Threats on Worker Traits

This subsection will analyze whether selection into high-paying, strong-dismissal-threat jobs occurs based on worker traits other than productivity. To help answer this question, Tables A3 and A4 present the mean of each worker trait separately for subjects who prefer the Safe Job and subjects who prefer the Risky Job. Standard deviations are displayed below each mean in parentheses. Additionally, the third column presents the p-value resulting from a rank-sum test on the null hypothesis that the two samples of worker traits came from the same distribution. Table A3 shows only incentivized measurements of worker traits, while Table A4 shows survey questions and demographics.

In Table A3, there is strong evidence of sorting on productivity into different jobs, but little evidence of sorting on any other incentivized worker traits. There is a large difference

Table A2: Descriptive Statistics (Survey Traits)

	(1) Worker Traits mean/sd
Personality: Extraversion (scale: 1-5)	2.89 (0.86)
Personality: Agreeableness (scale: 1-5)	3.29 (0.85)
Personality: Conscientiousness (scale: 1-5)	3.57 (0.88)
Personality: Neuroticism (scale: 1-5)	3.16 (0.96)
Personality: Openness (scale: 1-5)	3.58 (0.91)
Survey: High Stakes Stress (scale: 1-7)	5.55 (1.28)
Survey: Well Under Pressure (scale: 1-7)	4.26 (1.39)
Survey: Competition Best In Me (scale: 1-7)	4.64 (1.51)
Survey: Competition Is Good (scale: 1-7)	4.74 (1.53)
Demographic: Female	0.59 (0.49)
Demographic: White	0.29 (0.45)
Demographic: Asian	0.50 (0.50)
Demographic: Hispanic	0.23 (0.42)
Observations	101

Notes: The table shows the mean value of each trait for subjects in the experiment. Standard deviations are shown in parentheses. Survey questions are either on a scale from 1 to 5 or from 1 to 7, as indicated below each trait.

Table A3: Job Preference and Worker Traits (Incentivized)

	(1) Prefer Safe Job mean/sd	(2) Prefer Risky Job mean/sd	(3) Ranksum Test p-value
High Piece Rate Tasks	8.38	12.08	0.0002***
	(3.87)	(4.99)	
Responsiveness to Piece Rate	1.06	0.67	0.8041
	(1.95)	(1.01)	
High Stakes Tasks	2.13	2.64	0.0597+
	(1.34)	(1.23)	
Confidence: Rank Choice	5.67	4.44	0.0028**
	(2.12)	(2.01)	
Confidence: Actual - Reported Rank	1.45	0.43	0.0431*
	(2.07)	(2.69)	
Risk Choice	2.94	3.10	0.5256
	(1.18)	(1.40)	
Loss Choice	2.42	2.61	0.6887
	(1.15)	(1.41)	
Patience Choice	3.42	3.38	0.9912
	(0.71)	(0.63)	
Social Prefs: Altruism	2.40	2.40	0.8606
	(1.87)	(1.70)	
Social Prefs: Trust	1.85	2.08	0.1973
	(0.98)	(1.05)	
Social Prefs: Reciprocity	1.53	1.42	0.5417
	(0.59)	(0.79)	
Observations	40	61	101

Notes: The table shows the mean value of each trait for the subset of workers who preferred the Safe Job in Column 1 and the Risky Job in Column 2. Standard deviations are shown in parentheses. Column 3 shows the p-value from a Wilcoxon rank-sum test comparing the distribution between prefer Safe and prefer Risky workers of the trait in that row.
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A4: Job Preference and Worker Traits (Survey)

	(1) Prefer Safe Job mean/sd	(2) Prefer Risky Job mean/sd	(3) Ranksum Test p-value
Personality: Extraversion	2.70	3.01	0.0965^{+}
	(0.80)	(0.88)	
Personality: Agreeableness	3.33	3.27	0.8241
	(0.80)	(0.88)	
Personality: Conscientiousness	3.34	3.72	0.0214*
	(0.78)	(0.92)	
Personality: Neuroticism	3.35	3.03	0.1387
	(0.83)	(1.02)	
Personality: Openness	3.45	3.66	0.1698
	(0.83)	(0.96)	
Survey: High Stakes Stress	5.72	5.44	0.3057
	(1.20)	(1.32)	
Survey: Well Under Pressure	4.17	4.31	0.6205
	(1.43)	(1.37)	
Survey: Competition Best In Me	4.60	4.67	0.8191
	(1.48)	(1.54)	
Survey: Competition Is Good	4.67	4.79	0.4310
	(1.23)	(1.71)	
Demographic: Female	0.65	0.56	0.3563
	(0.48)	(0.50)	
Demographic: White	0.23	0.33	0.2661
	(0.42)	(0.47)	
Demographic: Asian	0.55	0.48	0.4656
	(0.50)	(0.50)	
Demographic: Hispanic	0.28	0.20	0.3613
	(0.45)	(0.40)	
Observations	40	61	101

Notes: The table shows the mean value of each trait for the subset of workers who preferred the Safe Job in Column 1 and the Risky Job in Column 2. Standard deviations are shown in parentheses. Column 3 shows the p-value from a Wilcoxon rank-sum test comparing the distribution between prefer Safe and prefer Risky workers of the trait in that row.
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

in productivity during the high piece rate period between subjects who preferred the Safe Job versus those who preferred the Risky Job, but this merely echoes the evidence already presented above of selection on worker productivity. Similarly, subjects who prefer the Risky Job tended to do slightly better on during the short high-stakes work period, but this difference was only marginally significant ($p = 0.0597$). Although subjects who prefer the Risky Job were significantly more likely to report a lower rank in the confidence elicitation (i.e., on average they believed they were more productive, since rank 1 is the highest productivity rank), this is likely due to the reported rank's strong correlation with overall productivity in the work task. Turning to overconfidence and underconfidence, although both groups of subjects are overconfident on average, the subjects who prefer the Safe Job were more overconfident than subjects who preferred the Risky Job ($p = 0.0431$). In other words, subjects who prefer the high-paying, strong-dismissal-threat job had more accurate relative evaluations of their productivity. No other incentivized worker traits appear to be significantly different by job preference.

Table A4 presents evidence of sorting on the various survey measures or demographics. Only one personality trait displays a statistically significant difference between subjects who preferred the Safe Job and those who preferred the Risky Job: conscientiousness. This difference implies that subjects who prefer the Risky Job tend to be more organized, careful, and diligent. There is also suggestive evidence of higher levels of extraversion (sociability and talkativeness), although this is only marginally significant ($p < 0.1$)

It is important to consider that many of these worker traits may be correlated with productivity. This is not to say that the univariate comparisons do not matter; firms offering higher pay along with stronger dismissal threats are still likely to observe the differences in traits uncovered in the univariate analysis above. However, it's useful to determine whether workers are attracted to stronger dismissal threats independently based on these traits, or whether it's simply the case that more productive workers happen to also have these traits. An analysis of correlation matrices between these traits (not shown) confirms that most of the traits that show univariate differences by worker preference are indeed correlated with output under the high piece rate.

Probit regressions are run to further establish whether selection on worker traits occurs independent of productivity. Table A5 presents the results. The latent dependent variable is the propensity to select the Risky Job. Table A5 only includes incentivized worker trait measurements as independent variables. Standard errors are clustered by session. Rather than coefficients, marginal effects evaluated at the mean of each trait are displayed. (Another specification was run which additionally included control variables for all survey-based trait measurements, but none of these are significant and their inclusion does not impact the results in any meaningful way.)

It turns out that controlling for worker productivity eliminates any association between job preference and other worker traits. Worker productivity during the piece rate period is highly predictive of preferences over pay and dismissal threats, but none of the other incentivized trait measurements are significant once also controlling for productivity.

Again, this does not mean multidimensional sorting is not occurring – it simply means that the observed sorting on traits is due to their correlation with productivity and not an independent relationship between the trait and job preference. Firms offering steeper dismissal incentives will still observe differences on average in the worker traits mentioned above. This is even more likely to be true when workers have less ability to predict their productivity in a given job, thus muting selection on productivity and empowering selection on other worker traits and preferences.

The experiment design placed all worker trait measurements after the work periods, and therefore a natural question to ask is whether the history observed by subjects may impact these trait measurements and thus change the conclusions on multidimensional

Table A5: Probit Regression of Job Preference on Worker Traits

	(1) Prefer Risky Job
High Piece Rate Tasks	0.0444***
	(0.001)
Responsiveness to Piece Rate	-0.0143
	(0.425)
High Stakes Tasks	-0.0387
	(0.448)
Confidence: Rank Choice	-0.0256
	(0.280)
Risk Choice	0.0355
	(0.448)
Loss Choice	-0.00993
	(0.787)
Patience Choice	0.0510
	(0.543)
Social Prefs: Altruism	0.00730
	(0.840)
Social Prefs: Trust	0.0792
	(0.219)
Social Prefs: Reciprocity	-0.0979
	(0.192)
Observations	98

Notes: The table shows the results of a probit regression. The latent dependent variable is the subjects' propensity to prefer the Risky Job. Each row displays the marginal effect of the variable evaluated at the mean of that variable. Standard errors are clustered at the session level; p-values are shown in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

sorting into stronger dismissal threats. Two examples of this stand out. First, whether a subject was fired along with how many other subjects were announced to have been fired are likely to impact stated confidence. Second, the amount of money earned so far may impact risk and loss aversion. The following analysis will reexamine multidimensional sorting after adjusting for session-specific history. It turns out that even after these adjustments, no qualitative differences are found in sorting patterns from previous analysis.

To determine whether session history impacts statements of confidence, a specification search was conducted to isolate the impact of exogenous events. OLS regressions were run with elicited confidence as the outcome variable and a variety of explanatory variables (standard errors always clustered at the session level). The results consistently indicate that whether a subject was dismissed has an important impact on the subject's statement of confidence, even after controlling for the probability of dismissal. In other words, being exogenously dismissed hurts a subject's confidence regardless of whether the dismissal was warranted. On the other hand, neither the number nor percentage of other subjects who were dismissed in a session seemed to have any impact on statements of confidence in any reasonable specifications.

The final preferred specification with strong explanatory power was used to create an "adjusted confidence" measurement for each subject. The specification controlled for output during the high piece rate period, job preference, job assignment, the probability of being dismissed in each period, and an indicator for whether the subject was actually dismissed in that period. The coefficients on being dismissed in the first and second periods were statistically significant and had values of 1.31 and 0.86, respectively. The interpretation is that controlling for the chance of being dismissed, actually being dismissed results in subjects reporting a worse rank by 1.31 units if dismissed in the first period, and a worse rank by 0.86 units if dismissed in the second period. The adjusted confidence measure was created by simply subtracting these respective values from the reported ranks of subjects who were fired in periods 1 or 2 (in a sense, adjusting them to be more confident if they were randomly fired).

The results of the analysis of multidimensional sorting does not change in any meaningful way when using this new adjusted confidence measure. There is still a highly statistically significant difference in adjusted confidence between subjects who prefer the Risky Job versus the Safe Job, and this difference is not significant in a probit regression that controls for productivity along with all other incentivized worker traits.

Similar methodology was used to determine whether the amount of money earned during the work periods impacted subject risk, loss, and patience preferences. For a series of OLS regressions with each of these three preference as an outcome, the cumulative amount of money earned had no impact on these preferences at the 5% confidence level for any reasonable specification. Therefore, no further adjustments or analysis was done using these traits.