European University Institute
DEPARTMENT OF ECONOMICS

Essays in Behavioral Economics and Development

Christian Johannes Meyer

Thesis submitted for assessment with a view to obtaining the degree of Doctor of Economics of the European University Institute

Florence, 15 October 2019

European University Institute
Department of Economics

Essays in Behavioral Economics and Development

Christian Johannes Meyer

Thesis submitted for assessment with a view to obtaining the degree of
Doctor of Economics of the European University Institute

European University Institute
Department of Economics

Thesis submitted for assessment with a view to obtaining the degree of
Doctor of Economics of the European University Institute

Essays in Behavioral Economics and Development

Christian Johannes Meyer

First Supervisor:
Professor Michèle Belot

Second Supervisor:
Professor Arthur Schram

August 2019

Abstract

This dissertation presents three independent chapters that build on the tools of behavioral economics to study issues related to labor markets in low-income countries and charitable giving.

In the first chapter, I investigate whether present bias correlates with savings and job search behavior in a population of low-skill workers in Ethiopia. I conduct a field experiment with 460 women who begin employment in the ready-made garment industry. Most are rural-urban migrants without work experience for whom the job represents a stepping stone into the labor market. Almost all workers plan to use their jobs to save money and to look for higher-wage employment, but many fall short of their intentions. I propose self-control problems as a candidate explanation. I elicit a measure of present bias in a tightly-controlled experiment and match results to high-frequency survey data that I collect over a period of three months. Present bias is a significant predictor of job search effort, controlling for liquidity and a broad range of covariates. Present-biased workers spend 57 percent less time on job search per week. As a result of reduced search, present-biased workers generate fewer offers and stay in their jobs significantly longer. In contrast, I find no significant correlation between present bias and savings behavior. I discuss implications for the design of commitment devices in this context.

In the second chapter, co-authored with Egon Tripodi, we study incentivized voluntary contributions to charitable activities. Motivated by the market for blood donations in Germany, we consider a setting where different incentives coexist and agents can choose to donate without receiving monetary compensation. We use a model that interacts image concerns of agents with intrinsic and extrinsic incentives to donate. Laboratory results show that a collection system where compensation can be turned down can improve the efficiency of collection. Image effects and incentive effects do not crowd each other out. A significant share of donors turn down compensation. Heterogeneity in treatment effects suggests gender-specific preferences over signaling.

In the third chapter, also co-authored with Egon Tripodi, we use a field experiment to study how social image concerns affect pledges to engage in a charitable activity. We work with two different blood banks and a municipal government in Germany to offer sign-ups for human whole blood donations. Motivated by a simple signaling framework, we randomly vary the type of organization to donate to and the visibility of the pledge to donate. Our setting also provides natural variation in the group of people that form the “audience” for social image concerns. We find evidence for strong social image concerns when subjects are asked in public whether they would like to pledge a donation with a well-known charity. Almost all subjects renege on their pledge, with no detectable differences between treatments. We discuss avenues for further research and end on a cautionary note for organizations looking to harness pledges to encourage individuals to do good.

To my grandfather,
who has always been a source of inspiration,
and to my parents,
who have always supported me in my endeavors.

Acknowledgments

This chapter of my life would not have been possible without the guidance, patience, support, and kindness of many different people – near and far, in academia and outside.

I am enormously grateful to my committee and to my two supervisors Michèle Belot and Arthur Schram. Thank you for taking me on as a graduate student at a point when I was adrift between various topics and interests, for patiently listening, for being honest, and for always treating me like an intellectual equal. I am extremely lucky to have you as advisors and mentors and I hope that we can continue to work together in the future. Many thanks also to the students and professors at the European University Institute that have provided comments, advice, and guidance over the years.

I would also like to thank Antoinette Schoar, who made it possible for me to spend a significant time of my PhD at MIT Sloan. The MIT community was extremely welcoming and helped me gain a fresh perspective on my PhD – both academically and personally. In Cambridge I also had the pleasure of working with Erez Yoeli, David Rand, and their great colleagues at the MIT Applied Cooperation Team. Thank you for the many thoughtful and fun conversations in the office, over movies, at Grendel's Den, and via Skype.

I am extremely grateful that I could work with and learn from Egon and Anna. I would not have managed to get through th program without the two of you. Thank you for the countless days and nights together at Badyah. Thank you, Egon, for being such a good co-author and partner in crime for various academic and non-academic endeavors. Even when we were yelling at each other over a minor detail of an appendix figure, I was always confident that we would ultimately come up with really good solution. I am also grateful for the many colleagues that became close friends over the years, both in Cambridge and in Florence. Thank you Alessandro, Carmen, Liiri, Matic, and Matthias.

My research would not have been possible without the generous support of various funders that put their trust in me and believed in my projects. I want to thank the UK Department for International Development (DFID), the German Academic Exchange Service (DAAD), the European Commission, the European University Institute, the Institute for the Study of Labor (IZA) and DFID through the GLM-LIC program, New York University Abu Dhabi, the Russel Sage Foundation, and the World Bank. I also cannot begin to express my thanks to Endale, who made my work in Ethiopia possible, who taught me a lot about one of the most fascinating places in the world, and who has become a dear friend.

I am deeply indebted to my former colleagues and friends at the Center for Global Development (CGD) and at the World Bank. My time in DC inspired me to embark on a career in international development and, ultimately, to apply for PhD programs. My two years at CGD still shape the way I think about my research, my ambitions, and my priorities in life. Thank you, Nancy, for continuing to be an inspiration and mentor in my life. Thank you, Alan and Vij, for your unwavering guidance and for countless fun moments. I owe a large part of my current research agenda to the work that we started together at CGD. At the World Bank, I could not have wished for a better colleague and friend than Dean. You showed me what an inspiring and exciting place the Bank can be. Keeping in touch with all of you throughout my PhD made a big difference – thank you.

I want to express my deepest gratitude to my friends near and far that have never let me down throughout this long journey. Sometimes all it took was a short iMessage chat with someone at the other end of the world. I am glad that I managed to see you so many times throughout the years, even if we live in different places. Thank you, Andrew, Angelica, Anna, Jonas, Juan, Julian, Matthias, Mirco, Nabil, Paul, Petter, Sarah, and Stefano.

Finally, I want to thank my parents who have been unwavering in their support and their understanding.

Christian Johannes Meyer
Addis Ababa
August 2019

Contents

List of Figures

List of Tables

Image Concerns in Pledges to Give Blood

List of Acronyms

BT Benabou and Tirole (2006) 92

BL Bole Lemi 11

CAPI computer-assisted personal interviewing 21

CBE Commercial Bank of Ethiopia 24

CRRA constant relative risk aversion 26

CSA Ethiopian Central Statistics Agency 47

CTB convex time budget 7

DRK Deutsches Rotes Kreuz [German Red Cross] 90

EDRI Ethiopian Development Research Institute 47

ESS Ethiopian Socioeconomic Survey 29

GSE General Self-Efficacy 50

ICRC International Committee of the Red Cross 98

MPL multiple price list 7

OLS ordinary least squares 37

RMG ready-made garment 6

WWF World Wildlife Fund 98

SOBC Science of Behavior Change 49

NIH US National Institutes of Health 49

Introduction 1

Research in psychology and economics over the past decades has uncovered a broad range of behavioral quirks and phenomena that shape economic decisions in our lives. This dissertation focuses on two such quirks and studies them using experimental methods: The second chapter of this disseration revisits what might be one of the most commonly-studied behavioral phenomena – present bias – and assesses its link with real-world behaviors in the domains of savings and job search. The third and fourth chapter build on another behavioral phenomenon: While millions of people engage in charitable activities with private cost and public benefit, many are more inclined to do so when observed by others. This effect of social image concerns on charitable giving is well-documented. In joint work with Egon Tripodi, we focus on two different mechanisms through which social image concerns in charitable giving could be leveraged to promote good deeds.

To study the link of experimentally-measured present bias with real-world behaviors, the second chapter combines a lab-in-the-field experiment with high-frequency survey data from 460 women in Ethiopia's ready-made garment industry. Most workers in my sample are recent rural-urban migrants for whom employment in this industry represents the first formal work experience, a stepping stone into the urban labor market of the capital Addis Ababa, and an opportunity to accumulate savings.

My results suggest that self-control problems in the form of experimentally-elicited present bias significantly undermine the ability of workers to use these jobs as such a stepping stone. I show that present bias is a significant predictor of reduced job search effort over a period of three months after starting their job. Present-biased workers search less, as a result generate fewer alternative job offers, and stay at the firm significantly longer. My results offer the first experimental evidence of a theorized link between present bias and job search effort. I do not find evidence for a link between present bias and reduced savings.

An immediate implication of my findings is that individuals looking for work might benefit from policies or devices that commit their future selves to more search. Whether and under what conditions such a commitment device can be welfare-improving depends on the exact welfare criterion, which is particularly difficult to define for on-the-job-search in the context of this study. The conclusion of the second chapter discusses potential interventions and sketches paths forward.

In the third and fourth chapter, we consider different mechanisms through which social image concerns in charitable giving could be leveraged to promote good deeds.

In the third chapter, we focus on image concerns as a channel through which a "dual market", in which agents can choose whether or not they would like to be paid for a charitable activity, can induce sorting into unpaid contributions. Our interest in the role of social image in a "dual market" for charitable giving stems from the market for human whole blood donations in Germany. In most high-income countries, the concern that incentives could backfire is reflected in tight regulation of how blood donations can be collected. Regulations typically do not allow for monetary payments to donors. In many regions of Germany, however, monetary and non-monetary incentives appear to coexist in a "dual market" in which different blood collectors offer different incentives and prospective donors can choose where to donate. Donations at the German Red Cross are always unpaid, while donations at hospitals or commercial blood banks are compensated with 20 to 30 euro.

We use a simple theoretical framework and a laboratory experiment to study this dual market in a stylized environment. Our theoretical framework shows that in a dual market, a positive fraction of donors will choose to be not paid and that this fraction is bigger when actions are taken in public. Offering a compensation and letting agents turn down the compensation lets the collection system leverage the heterogeneity in individual preferences. This enables efficiency gains in the collection similar to those deriving from self-selection in third-degree price discrimination.

Our experiment shows that in a dual market, the availability of extrinsic incentives does not crowd out intrinsic motivations of donors. In fact, giving significantly increases compared to the market design in which donations are not paid. We also find that when given the option to turn down the compensation, a significant fraction of donors indeed chooses to do so, though we find only weak evidence that donors turn down incentives more in public than in private.

In the fourth chapter, we study whether image concerns can encourage pledges to do good in the future. While the effect of social image on giving has been widely demonstrated, prosocial actions themselves can often not be made public. One way for individuals to still benefit from public recognition of their prosocial activities is to demonstrate proof of contribution ex-post, for example by wearing a lapel pin for blood donors or sharing a picture of volunteer activity on social media. Charities often recognize donors on publicly-visible plaques or donor walls. Ex-ante, social pressure can be leveraged by encouraging public pledges to act charitably in the future. Such pledges are used to rally individuals to contribute to future charitable activities, for example through public calls for action in online and offline social networks.

Using the real-world setting and a research design with high ecological validity, we set out to study how social image concerns affect both the uptake and the fulfillment of pledges to donate blood. We conduct a field experiment in a the mid-sized German city where various organizations compete for prospective blood donors.

Our experiment shows that the uptake of pledges is consistent with a theoretical framework in which social image concerns are amplified when subjects care more about being perceived favorably by a reference group of people and when pledging to donate to a more socially-desirable organization. We find evidence for social image concerns when subjects are asked in public whether they would like to pledge a donation with the Red Cross. When subjects are accompanied by friends and family members, public offers significantly increase the likelihood of pledging to come to a donation drive. When subjects are not accompanied by anyone, but just surrounded by other customers waiting in the municipal service center, we do not find significant differences between public and private offers. Similarly, social image concerns do not appear to play a role when subjects are offered to sign up for a remunerated donation with a commercial blood bank.

At the same time, pledges in our particular context do not appear to induce any additional blood donations. Almost all subjects renege on their pledge, with no detectable differences between treatments. We see the lack of fulfillment in our experiment as an important starting point for further academic and policy-oriented work, which we discuss in the conclusion of the fourth chapter.

My choice to embark on a career as an academic economist was always motivated by a desire to improve the lives of people through research that informs and drives better policy. As small as the contributions of this dissertation might be, I hope that they are just the starting point of a wider effort towards this goal.

In Search of a Better Life: Self-Control in the Ethiopian Labor Market 2

2.1 Introduction

Many choices in our lives involve costs and benefits spread out over time. Such choices often suffer from an apparent inconsistency: when we plan for tomorrow, we may decide to save money and search for better jobs. But when tomorrow arrives, we may instead want to spend our money and slack off on our job search. This preference for immediate gratification – present bias – is one of the most robust "anomalies" of intertemporal choice (DellaVigna, 2009; Frederick, Loewenstein, & O'Donoghue, 2002; Loewenstein & Prelec, 1992). This anomaly can be costly. The poor in particular, with less scope to absorb errors, may suffer from not following through on their own plans.

Prominent models that rationalize such self-control problems provide two key predictions (Laibson, 1997; O'Donoghue & Rabin, 1999).[1] First, individuals with self-control

This chapter formed the basis for my job market paper. I want to thank Michèle Belot and Arthur Schram for invaluable advice and support. The chapter has benefited from helpful comments by Liang Bai, Kevin Croke, Anita Glenny, Morgan Hardy, Andrea Ichino, Pamela Jakiela, Gisella Kagy, Philipp Kircher, Peter Kuhn, David K. Levine, Egon Tripodi, Chris Udry, and seminar audiences at the EUI and IZA Summer School 2018. This study would not have been possible without the outstanding work of my field coordinator Endale Geberemedehen, my survey team, and the help of Wendemagegn Zewdu. Tewodros Gebrewolde kindly agreed to serve as local research ethics advisor. Roberto Lescrauwaet and Eyoual Tamrat provided research assistance. Funding from UK Department for International Development (DFID), the European Commission, the European University Institute, the Institute for the Study of Labor (IZA) and DFID through the GLM-LIC program, New York University Abu Dhabi, and the World Bank ieConnect for Impact program is gratefully acknowledged. This study was registered with the AER RCT Registry as #AEARCTR-0002555 and received IRB approval at NYU Abu Dhabi as protocol #020-2018.

[1] Alternative models that can rationalize self-control problems include dual-self models by Thaler and Shefrin (1981) and Fudenberg and Levine (2006) or models that focus on temptation by Banerjee and Mullainathan (2010) and Gul and Pesendorfer (2001). While predictions generated for example by Fudenberg and Levine

problems have characteristic consumption patterns. They consume too little of a good with immediate costs and future rewards (such as saving money or searching for a job) and too much of a good with immediate rewards and future costs (such as spending money on consumption or enjoying leisure time). Second, individuals who are aware of their self-control problems value commitment. They want to improve their welfare by tying their hands. While the literature has established this demand for commitment in many domains (Ashraf, Karlan, & Yin, 2006; DellaVigna & Malmendier, 2006; Duflo, Kremer, & Robinson, 2011; Dupas & Robinson, 2013; Giné, Karlan, & Zinman, 2010; Kaur, Kremer, & Mullainathan, 2015; Thaler & Benartzi, 2004), evidence on the hypothesized link between self-control problems and the consumption patterns described above is relatively scarce (Castillo, Ferraro, Jordan, & Petrie, 2011; Falk et al., 2018; Meier & Sprenger, 2010). This chapter provides an empirical test of the link between self-control problems and behavior. I conduct this test in an environment where failure to follow through can have significant negative consequences.

I use an experiment and high-frequency survey data from 460 women in Ethiopia's ready-made garment (RMG) industry to investigate the correlation between present bias and subsequent choices over savings and job search. I work with an RMG firm in peri-urban Addis Ababa that hired a large number of all-female workers during the study period in the spring of 2018. Workers start homogeneous production jobs (such as sewing t-shirts) without appreciable skill requirements, but with steady hours and the same low wage approximately equal to the local poverty line. Consistent with a narrative of low-skill industrial jobs acting as a safety net (Blattman & Dercon, 2018), workers use the jobs as a stepping stone to a better future in two ways. First, by accumulating assets and then leaving the job, for example to start a small business or engage in off-the-job search. Second, by financing continued on-the-job search for better opportunities. Accordingly, savings and job search are the two intertemporal decisions that I study. I collect data on these and a broad range of other covariates using in-person interviews and phone surveys over three months after workers join the firm. The survey is designed to track workers as they leave their jobs, which many do. I correlate this survey data with structural estimates of present bias, which I obtain from a tightly-controlled experiment that I conduct on the day that workers start their new job.

Three features make this an ideal setting to study the effects of self-control problems. First, intentions to save money and to look for work are pervasive. All but one worker in my sample want to save considerable amounts of money, 47 percent in order to build assets to start their own business. I elicit workers' predictions of their monthly savings on the day they start their job. Workers on average expect that they can realistically save one-third of their wage, but most fall significantly short of their goal. This finding is consistent with workers overestimating either their future self control or their future efficiency in saving money (Acland & Levy, 2015; DellaVigna & Malmendier, 2006). On the day they start their new job, 20 percent of workers are still actively looking for other jobs and another 31 percent would like to search but find it too difficult, costly, or time-consuming. The fractions of those who search and those who would have liked to search in any given week increase over the first three months of employment. While various factors can explain intention-behavior

(2006) are similar to the ones presented, I focus on quasi-hyperbolic models for ease of exposition.

gaps in both domains, the evidence is consistent with self-control problems affecting the ability of workers to follow through on their intertemporal plans.

Second, self-control problems can be consequential. Saving money and searching for other jobs are the only ways workers can meaningfully increase future consumption opportunities while in their jobs. This is because wages at the firm are not only low in absolute terms, but also do not increase significantly with individual performance or tenure at the firm. In addition, workers in my sample are poor in absolute terms and relative to their peers in the same age group in Addis Ababa. This means they have little slack income to absorb the potential costs of self-control problems.

Third, the setting allows for a clean experimental design. Enrolling workers into the study as they start their new job provides a relatively homogeneous sample and a clear starting point to study choices over time in a natural environment.

The analysis in this chapter relies on the measurement of potential self-control problems for each worker in my sample. I use a version of the convex time budget (CTB) task (Andreoni & Sprenger, 2012). Each worker makes 15 allocations of a large experimental budget (20 to 40 percent of the monthly wage) over two points in time. By experimentally varying the timing of the payments and the implied interest rate between both payments I can recover individual-level measures of present bias along with other parameters of each worker's utility function. To implement the task in my setting I closely follow Giné, Goldberg, Silverman, and Yang (2017). Workers make their decisions by dividing a number of beans between two empty dishes that represent the two payoffs. Each of the two dishes is positioned below a small whiteboard that indicates the exact payoff date and the exchange rate at which beans are converted into the local currency.

The CTB method aims to address methodological problems of multiple price list (MPL) approaches that have been widely used in the literature. MPL approaches often assume linear utility, which may lead to biased inference when utility is in fact concave (Andersen, Harrison, Lau, & Rutström, 2008).[2] Previous work that correlates experimental estimates of present bias with actual consumption patterns has relied on the MPL method (Castillo et al., 2011; Meier & Sprenger, 2010), possibly because it is easier to implement in the field. Irrespective of whether time preferences are elicited with MPL or CTB, a number of other confounds may undermine identification of present bias from time-dated payments. Importantly, subjects may exhibit a preference for earlier payments because it is more costly to obtain the later payment or because there is uncertainty over whether the experimenter will deliver the payment as promised. Several recent studies that carefully equalize transaction costs between time-dated payments find little evidence of aggregate present bias (Andersen et al., 2008; Andreoni & Sprenger, 2012; Augenblick, Niederle, & Sprenger, 2015; Giné et al., 2017). I take several steps to address this and other potential confounds commonly found in the literature. One such step is to use Ethiopia's mobile money system for costless and precisely-timed experimental payments (Balakrishnan, Haushofer, & Jakiela, 2017). Mea-

[2] Consider utility from consumption $u(c_t)$ at an initial time period t and after a delay of k periods. The implied discount factor between utility in both periods can be calculated as $\delta_u \approx \left[u\left(c_t\right)/u\left(c_{t+k}\right)\right]^{1/k}$. MPL approaches typically infer discount factors in terms of time-dated consumption, not time-dated utility, so that $\delta_c \approx \left[c_t/c_{t+k}\right]^{1/k}$ and it is explicitly or implicitly assumed that $u(c_t) = c_t$. If utility is concave, as it is in Holt and Laury (2002) and Andersen et al. (2008), we will have $\delta_c < \delta_u$ and the implied discount will be biased upward.

sures of how well workers understand the experiment, individual-level estimates of present bias, and aggregate-level estimates of present bias are in line with other recent implementations of the CTB method. I find that 38 percent of the workers in my sample can be categorized as present-biased when they start their new job.

To guide the empirical analysis, I present a simple model that interprets employment at the industrial firm as akin to the safety net of a welfare system. Involuntary transitions are ruled out and workers are employed at their reservation wage.[3] Search increases the probability of receiving a better wage offer. Workers who expect to leave the firm within a fixed amount of time have an additional precautionary savings motive to smooth consumption.[4] Both search and savings thus represent workers' self-insurance efforts. To formalize self-control problems I assume quasi-hyperbolic (β-δ) preferences developed by Laibson (1997) and O'Donoghue and Rabin (1999), who build on earlier work by Strotz (1955) and Phelps and Pollak (1968). In the standard exponential discounted utility framework (Samuelson, 1937), every future period is discounted by a constant discount factor δ. In the (β-δ) framework, an additional present bias parameter β allows for higher discounting between the current and the next period. In the model I assume that if workers are present-biased, they are not aware of it (*naïve*). Every period the worker thus assumes that her future self will not have self-control problems.

The model shows how present bias undermines self-insurance through job search and savings in intuitive ways. First, an increase in present bias reduces savings and thus the ability of workers to smooth consumption after leaving the firm. Second, an increase in present bias reduces the present value of search (DellaVigna & Paserman, 2005). As a corollary, present-biased workers stay longer at the firm. This illustrates how workers may experience a type of "behavioral job-lock," where voluntary turnover is reduced due to self-control problems.[5]

I provide reduced-form evidence on the predictions of the model. My first set of empirical findings considers the relationship of present bias and savings over the three months after joining the firm. I do not find that baseline present bias is a statistically significant predictor of subsequent savings. My preferred specification, which controls for a broad range of covariates, finds that present-biased workers do save marginally less than workers who are not categorized as present-biased. This difference is, however, not significant at any conventional level. This holds for both savings in absolute terms and savings relative to self-set goals set when joining the firm.

This finding is consistent with the literature to the extent that there is little existing evidence on the relationship between experimentally-elicited measures of present bias and consumption behavior in line with predictions of the quasi-hyperbolic model. An important

[3] In my data, 90 percent of transitions over the first four months of employment are voluntary. The wage at the firm, which is the same for all workers in my sample, is approximately equal to the poverty line. Given that workers can likely not fulfill minimum nutrition requirements below this wage level, it is improbable that the wage at the firm is significantly above the reservation wage of workers.

[4] On the day that they start their new job, 25 percent of workers in my sample report that they plan to leave within a fixed amount of time. The median expected tenure of these workers is 12 months. After three months at the firm, 45 percent of workers report that they plan to leave within a fixed amount of time.

[5] The concept of "job-lock" is typically associated with the finding that employer-provided health insurance plays an important role in job mobility decisions (Madrian, 1994).

exception is the work by Meier and Sprenger (2010), who use the MPL method to show that present-biased individuals are more likely to have credit card debt.[6]

My second set of results considers the relationship of present bias and job search over the three months after joining the firm. I find that baseline present bias is an economically and statistically significant predictor of subsequent job search effort. In my preferred specification that controls for a broad range of covariates, present-biased workers spend on average 57 percent less time on job search (37 minutes per week for present-biased workers compared to 85 minutes for those who are not present-biased). Present-biased workers also place fewer than half as many phone calls in search for a new job (0.3 per week for present-biased workers compared to 0.7 those not present-biased). While my analysis focuses on search intensity, the results also hold on the extensive margin.

As an immediate consequence of less search, present-biased workers stay at the firm significantly longer. Controlling for the same broad set of covariates as before, the hazard of leaving the firm is 52 percent as high for present-biased individuals as it is for individuals who are not present-biased. This effect appears to operate through search effort. Search effort significantly increases the hazard of leaving the firm. When I include both present bias and search effort as predictors in a hazards model, baseline present bias loses its predictive power. Results hold when I restrict my analysis to voluntary departures from the firm and when using data on tenure from firm personnel records.[7] To further confirm the mechanism, I consider data on job search outcomes. Baseline present bias is associated with significantly fewer job offers.

Because the evidence presented is correlational, I consider potential confounders and alternative explanations for my findings. First, I show that individual observable characteristics and environmental factors do not predict experimental responses. Second, I provide evidence that individual financial wealth and liquidity are unlikely to explain my results. Randomized cash drops at baseline do not significantly affect responses in an additional CTB experiment. In addition, I use randomized cash drops to show that – consistent with existing theoretical and empirical findings – more liquidity causes less search. If individual liquidity constraints had caused subjects to both appear present-biased and to search less, we would expect that alleviating these constraints should lead to more search, not less. Third, I use detailed survey data on work experience, cognitive control, and non-cognitive skills to show that human capital is unlikely to be an alternative explanation. Fourth, I demonstrate the limited role of reservation wages in my setting. Fifth, I argue that it is improbable that workers who are categorized as present-biased systematically under-report search effort. While it is still possible that my experimental results reflect variation in unobserved variables that affect job search effort, I argue that self-control problems due to present bias offer the most parsimonious explanation for my results.

[6]Falk et al. (2018) use a hypothetical survey measure to provide global evidence of a link between patience and savings. Ashraf et al. (2006) elicit present bias using hypothetical choices, but do not assess the link with borrowing or savings. Karlan, Ratan, and Zinman (2014) review the literature and conclude that there is a "striking lack of empirical evidence" on correlations between present-bias and under-saving (p. 59).

[7]While the analysis of this chapter focuses on self-reported data, I also collect rich administrative data from firm personnel records. Workers truthfully report tenure, so the significant correlation between baseline present bias and the hazard of leaving the firm holds.

Taken together, this set of results provides the first experimental evidence of the theorized link between present bias and job search effort. DellaVigna and Paserman (2005) formalize how time preferences can affect job search behavior. They test their model in two large panel datasets from the United States and find a negative correlation between proxies for impatience and search effort during unemployment. In absence of experimentally-elicited measures of present bias, they proxy impatience with behavior such as health habits, use of contraceptives, and financial decisions. As they note, these indirect measures may pick up unobserved individual traits and preferences. My data allows for a more direct test. With these results the chapter contributes to a growing literature that has introduced insights from behavioral economics into models of job search (DellaVigna, Lindner, Reizer, & Schmieder, 2017; Paserman, 2008; Spinnewijn, 2015).

An immediate implication of my findings is that individuals looking for work might benefit from policies or devices that commit their future selves to more search. Whether and under what conditions such a commitment device can be welfare-improving depends on the exact welfare criterion, which is not obvious to define when we observe two individual choices that are in conflict with each other.[8] I discuss implications for the design of commitment devices in a conclusion.

More broadly the chapter relates to a literature that studies the effects of low-skill industrial jobs, particularly in the RMG industry, on its workers (Blattman & Dercon, 2018; Heath & Mobarak, 2014). Most workers in my sample are recent rural-urban migrants for whom employment at the study firm represents the first formal work experience. My results suggest that present bias may undermine the ability of workers to use these jobs as a stepping stone into the formal labor market of Addis Ababa. This finding complements work by Atkin (2016), who shows that workers in Mexican *maquiladoras* took present-biased decisions by choosing short-term gains at work over long-term gains through schooling. It appears that present bias not only makes workers take low-skill industrial jobs – it also keeps workers in these jobs.

The chapter proceeds as follows. Section 2.2 reviews the empirical setting and provides descriptive evidence on the job search and savings behavior of workers at the study firm. Section 2.3 uses this data to motivate a simple theoretical job search model to guide the empirical analysis. Section 2.4 discusses the experimental design and the elicitation of present bias. Section 2.5 presents results, and Section 2.6 concludes.

2.2 Setting

Ethiopia is one of the poorest countries in the world. The median person among its population of 107.5 million lives on \$2.75 per day (adjusted for purchasing power parity) and 68 percent of the population works in the agricultural sector.

[8] This is particularly the case in the quasi-hyperbolic model that this chapter builds on. In some dual-self models such as Benhabib and Bisin (2005) or Fudenberg and Levine (2006), the long-run self has the same short-run preferences as the short-run self, so a welfare criterion is more obvious to define. Bryan, Karlan, and Nelson (2010) provide a discussion. Bai, Handel, Miguel, and Rao (2017) provide a field test of theoretically-motivated commitment devices and illustrate how they can be welfare-reducing if individuals are over-optimistic about their effectiveness.

The Ethiopian economy and labor market are however undergoing rapid change. The economy grew at an average rate of 10.7 percent annually from 2003 to 2011. In line with the government's push for structural transformation and related public investment, employment has shifted from low-productivity agriculture to services, construction, and tradable goods. The share of the labor force working in the informal sector more than halved from 50.6 to 22.8 percent between 1999 and 2013 (Seid, Tafesse, & Ali, 2016).

Urbanization is a key part of this transformation, as migrants from rural areas seek employment and education in the cities. From 2000 to 2014, Ethiopia's urban population almost doubled from 9.8 million to 18.4 million (World Bank, 2016). The median age of people who migrated from rural to urban areas in the last five years is 21 years, 56.4 percent are female (Central Statistical Agency, 2014). Most migrants come to the capital Addis Ababa, where overall unemployment is high at 24 percent and unemployment among those aged 20 to 24 is even higher at 33.2 percent (Central Statistical Agency, 2015).

2.2.1 The Ready-Made Garment Industry in Ethiopia and the Study Firm

Ethiopia's government is pursuing an ambitious industrialization strategy that aims to make the country Sub-Saharan Africa's leader in light manufacturing. The cornerstone of this strategy is the construction of industrial parks, which aim to attract foreign direct investment from American, Asian, and European producers of ready-made garments, leather goods, pharmaceuticals, and agricultural products (Oqubay, 2016). These parks produce exclusively for export, not for the domestic market.

For the Ethiopian government, the industrial parks with their labor-intensive industries represent formal employment opportunities for the country's youth. Given that light manufacturing firms often prefer to hire women, the parks also represent an opportunity to increase female labor force participation and empower women by giving them their own stable income. For international investors, the parks represent one of the lowest-cost manufacturing destinations in the world (Gelb, Meyer, Ramachandran, & Wadhwa, 2017). In addition to an abundance of labor, Ethiopia has relatively weak labor laws and currently no minimum wage. Firms pay extremely low wages clustered around the local poverty line.[9] They also offer little to no upward mobility, so that the vast majority of workers will not advance past the level of machine operators. With their stable but extremely low wages and almost no skill requirements, the firms in Ethiopia's industrial parks represent what Blattman and Dercon (2018) call an "industrial safety net."

For this chapter I work with one such firm, located in Bole Lemi (BL) Industrial Park in the outskirts Addis Ababa. Appendix Figure 2.D1 locates the industrial park on a map of Addis Ababa and its surroundings. Inside the park, ten foreign-owned firms produce garments and leather goods in 20 factory buildings with a total capacity of about 20,000

[9] In Bole Lemi Industrial Park, where this study is set, entry level wages are about 1,000 birr (US$ 36.30) per month. The local poverty line is about 958 birr per adult per month. In the Hawassa Industrial Park, the largest industrial park currently in operation, entry-level wages are set at 750 birr ($US 27.23) per month. The local poverty line in Hawassa is about 695 birr per adult per month. Poverty lines are based on the official 2015/16 absolute poverty line of 7,184 birr per adult per year, adjusted to current values using the GDP deflator and adjusted for local prices using the spatial price indices reported in National Planning Commission (2017). Following the methodology of the National Planning Commission, spatial prices for food and non-food are weighted using the food share of the poorest quartile of the population (0.525).

workers. The study firm produces garments for well-known European and North American brands. The firm has about 3,300 workers, about half of which are machine operators, who sew and pack the garments. 95 percent of machine operators are female. Hours and compensation are in line with other firms in this and other industrial parks in Ethiopia.[10] Machine operators earn 1,000 birr (US$ 36.30) in the first month, 1,075 birr (US$ 39.00) in the second month, and 1,150 birr (US$ 41.75) after that. Wage growth and opportunities for promotion are extremely limited and depend on performance evaluations every three months.[11] In addition to their base salary, workers receive limited team-based productivity bonuses, subsidized meals, and free transportation to nearby neighborhoods.

I selected this firm because it was planning to hire a large number of production workers during the study period in the spring of 2018. Job seekers come to the factory gate every day. About 250 to 500 people apply every month. Out of those, 200 to 400 are hired (an average of 15 workers per day). This hiring is partly to expand production and partly to make up for the circa 100 workers who leave every month. Production jobs have no appreciable skill requirements beyond basic motor skills. The most common reasons for not being hired are insufficient Amharic language skills, missing documentation, and failure to meet the minimum education requirements. Due to relatively strict enforcement by the international brands that have their garments manufactured in the industrial park, child labor is not common in this context. If applicants are hired, they return on the next day to begin work.

2.2.2 Workers at the Study Firm

The workers in my sample are exclusively female and tend to be young, low-skill, rural-urban migrants with little to no previous work experience. As such they represent one of the most disadvantaged groups in the urban labor market of Addis Ababa. The median worker has completed primary school while the 75th percentile has completed lower secondary schooling.[12] 73 percent of the workers in my sample were not born in Addis Ababa or its outskirts but moved to the city from rural areas. Out of those, 73 percent report having moved for the purpose of finding work, another 10 percent in order to find education or training. The median worker moved 103 months ago. 41 percent have never been employed. Excluding work as a housemaid, a common job for young female migrants, 75 percent do not have any formal work experience before joining the firm. Most workers live in the outskirts of the city around the industrial park (Appendix Figure 2.D1).

With what goals do workers start their jobs? First, saving money is an important goal for workers joining the firm. While 58 percent of workers report that they did not manage

[10]Workers have 52.5 hour, six-day workweeks. Hours are Monday through Saturday from 7.30am to 5.30pm with 75 minutes of break time.

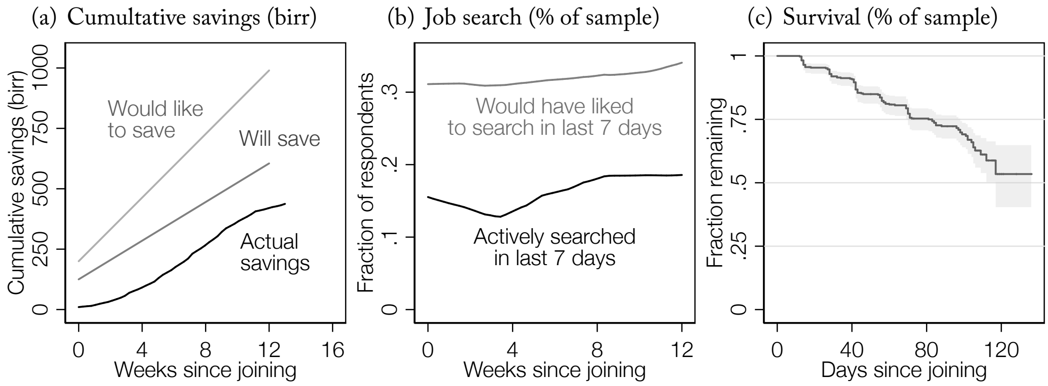
[11]A positive evaluation can increase the salary by 100 birr (US$ 3.63) per month up to a maximum wage of 1,650 birr (US$ 58.05) per month. The firm's human resource department estimates that out of 100 production workers, a maximum of five could ever advance to become team leaders, the next higher level of hierarchy on the factory floor, and even fewer to line supervisors, the group of factory floor managers studied by Macchiavello, Menzel, Rabbani, and Woodruff (2015).

[12]Abebe et al. (2016) illustrate how disadvantaged job seekers with low levels of education are. In their study of urban job seekers in Addis Ababa they show that a worker who has completed secondary schooling is four time less likely to have formal sector job and seven times less likely to have a permanent job than workers with a vocational or university degree.

to save any money in the month before starting employment, all but one respondent report that they are planning to save money during their time at the firm. 47 percent of those state that they are planning to save mainly to build assets, for example to start their own business. Another 36 percent say that they are planning to save money mainly for precautionary reasons.

Most workers indeed build up savings over time, however not as much as they hope. I can compare actual savings over time with a savings goal that workers set when joining the firm. I elicit workers' predictions for savings in two ways: The monthly amount that workers would ideally like to save and the monthly amount that workers think they can realistically save.[13] Panel (a) of Figure 2.1 plots cumulative savings over time compared to the self-set goal and the amount that workers think they can realistically save. Workers clearly fall short of both. One month into the job, the median worker has not accumulated any savings. The mean worker has reached about 53 percent of her ideal monthly savings goal. After three months at the firm, 68 percent of workers feel that they did not save as much as they had hoped when they joined, mostly because they spent more than they were planning (51 percent of those that reported saving less than planned).

Figure 2.1: Job Search, Cumulative Savings, and Survival of Workers During the First Three Months at the Study Firm



Notes: Panels (a) and (b) are local polynomial smoothers at the weekly level. Panel (c) plots the Kaplan-Meier survival estimate for the hazard of leaving the firm at the daily level with a 95 percent confidence interval shaded in gray. Job search intensity in panel (a) is hours of search over the past seven days. Appendix Figure 2.D2 plots all three dimensions of search effort data (hours, number of phone calls, and subjective intensity) and the extensive margin of search. In panel (b), "would like to save" plots the self-set savings goal set on the day that workers join the firm while "will save" refers to the amount that workers think they can realistically save. By construction both measures are linearly increasing in weeks. For comparison with savings goals, the monthly wages are 1,000 birr (month 1), 1,1075 birr (month 2), and 1,150 birr (month 3 and all months after that).

Overall, the intention-behavior gap in savings is consistent with workers overestimating either their future self-control or their future efficiency in saving money. Using adminis-

[13]Workers in my sample have ambitious savings goals. The median reported goal for saving money on the job is 500 birr (US$ 18.15) per month, or approximately half the monthly salary. The median worker thinks that she can realistically save 350 birr (US$ 12.71) per month, or approximately a third of the monthly salary. Appendix 2.B.2 elaborates on the elicitation of subjective probabilities.

trative and experimental data, DellaVigna and Malmendier (2006) and Acland and Levy (2015) show similar overconfidence over future self-control or efficiency for gym attendance.

Second, many workers in my sample see employment at the firm as temporary and use it to finance search for other opportunities. On the day that they join the firm, 25 percent of the workers in my sample reports that they are planning to leave within a fixed amount of time. The median expected tenure of these workers is 12 months. This is reflected in workers' job search efforts at the beginning of employment: 20 percent of the sample are still looking for other opportunities on the day that they join the firm. Of those, 87 percent report that they are looking for higher-wage jobs. Notably, another 31 percent of the sample report on the day that they join that they would like to look for work, but they find it too costly (10 percent of all constrained), time-consuming (31 percent), difficult (20 percent), or face other binding constraints (39 percent). Panel (b) of Figure 2.1 plots the fraction of workers who are searching and the fraction of workers who would have liked to search at a weekly level over the first three month of employment.[14]

High cost of search is consistent with the results of Abebe et al. (2016) and Franklin (2017) in a similar context.[15] On the job search while at the firm mostly works through social networks (61 percent of all workers engaged in search), vacancy boards (21 percent), and job brokers (9 percent). Overall, on the day they join just 49 percent of workers report that are satisfied and do not currently intent to look for other work. The other 51 percent appear to use their job at the firm to queue for better alternatives.

Saving money and searching for work appear to be at least partial substitutes for workers in my sample. As an indication, consider the mean savings goal on the day that workers join the study firm. Workers who are still looking for other employment report that they would like to save 499 birr (95% CI: 447 to 552) over the next month. Workers who are not looking for work report that they would like to save 628 birr (95% CI: 595 to 662) over the next month.[16]

In line with the notion that a significant share of workers sees the job as temporary, turnover from the study firm is high. In my sample, 117 out of 460 workers leave the firm within the first three months. The earliest departure from the firm comes after 12 days, the median exit occurs after 70 days. Panel (c) of Figure 2.1 plots Kaplan-Meier survival estimates. 90 percent of departures from the firm are voluntary. Of all the voluntary departures, 31 percent are due to low pay. The high turnover rate is consistent with previous

[14] Note that this is refering to on-the-job search only, so it excludes search that happens after leaving the firm. I also measure search intensity using the number of hours spent searching, the number of phone calls made in search for work, and a subjective assessment of search intensity. Appendix Figure 2.D2 plots all three measures as well as the extensive margin of search. The overall pattern similar, except for the subjective intensity measure which decreases towards the end of the panel.

[15] Abebe et al. (2016) and Franklin (2017) study search costs in the urban labor market of Addis Ababa and focus on large travel distances to centrally-located vacancy boards that make search costly. Data from my (different) sample suggests that cost in terms of time and effort are the most binding constraints and that social networks, not job boards, are the most common search methods. Appendix Table 2.E1 breaks down reported reasons for not searching while on the job, though the numbers in later survey waves are too small for meaningful analysis.

[16] A similar pattern holds for workers who expect their tenure at the firm to be limited. Workers who do not see the job as temporary report that they would like to save 620 birr (95% CI: 584 to 656) per month, while workers who expect their tenure to be limited report that they would like to save 554 birr (95% CI: 509 to 599) per month.

results from the same context (Abebe, Buehren, & Goldstein, 2018; Blattman & Dercon, 2018).

Overall, the descriptive evidence paints a picture of workers who are disadvantaged in the local labor market and who seek temporary low-wage industrial employment to build assets or finance continued search for better opportunities. There are quasi no entry barriers to employment and few involuntary separations, which suggests that firms in this context represent a safety net, rather than a desirable form of employment.

Importantly, the data above shows that workers are failing to follow through on their goals of saving money and searching for other opportunities. They fall short of their savings goals and they report not looking for work because they find it too time consuming or costly. This indicates that self-control problems may play a role in this context. Notably, both the savings decision and the search decision involve an intertemporal trade-off between large immediate costs and future benefits. A large literature has shown that behavior in each domain can be rationalized using models that allow for time-inconsistent intertemporal choice (DellaVigna, 2009; DellaVigna & Paserman, 2005; Laibson, Repetto, & Tobacman, 1998; Paserman, 2008). In the next section I present a theoretical framework to help build intuition for how such time-inconsistent intertemporal choice can affect both decisions.

2.3 Theoretical Framework

Workers in the study setting can increase their future consumption possibilities in two ways: They can continue to engage in job search and they can use their wage income to save money. Continued search allows workers to find higher-wage employment. Saving allows workers to smooth consumption when leaving the job at the firm, for example when they are fired or if want to become in self-employed.

It is useful to consider both decisions jointly because they can affect each other. If a worker expects to leave the safety net firm after a year (the median expected tenure reported by workers in my experiment) she will want to insure herself against expected income losses in the months leading up to that point. In that case, both search and savings represent self-insurance efforts. Everything else constant, the relative costs of and returns from search and savings will affect how much she searches and how much she saves.[17]

Both the savings decision and the search decision involve an intertemporal trade-off between immediate costs and future benefits. Given the suggestive evidence on self-control problems in the previous section, I allow for workers to be time-inconsistent in this trade-off by using (β-δ) preferences (Laibson, 1997; O'Donoghue & Rabin, 1999).

To guide the interpretation of my reduced-form analysis below, I provide a model that illustrates how both of these choice problems are affected by their relative costs and by time preferences, in particular by present bias. I simplify the environment significantly with the aim of building intuition. My approach rests on the re-interpretation of employment at the

[17]Note that I focus here on a precautionary savings motive of workers holding the amount of initial wealth constant, not the relationship of (initial) wealth on search. Initial wealth could increase for example due to a lump-sum severance pay as in Card, Chetty, and Weber (2007). Lentz and Tranæs (2005) show analytically and with simulations that job search effort is negatively related to initial wealth under the assumption of additively separable utility. This is in line with empirical results, e.g. in Algan, Chéron, Hairault, and Langot (2003) and Bloemen and Stancanelli (2001).

RMG firm as akin to the safety net of a welfare system. This follows directly from the anecdotal evidence presented above, where involuntary separations are rare and many workers continue to search for higher-wage opportunities from the day they start employment at the firm.

2.3.1 Job Search and Savings with Present-Biased Preferences

In this section I provide a discrete-time, partial-equilibrium job search model with endogenous savings that builds on the framework of Lentz and Tranæs (2005) and Card et al. (2007). DellaVigna et al. (2017) present a version of this framework that allows for hyperbolic time preferences (Laibson, 1997; O'Donoghue & Rabin, 1999) to affect both job search behavior (DellaVigna & Paserman, 2005) and endogenous savings.

For tractability reasons, I follow the previous literature in making several key simplifications: First, wages are exogenously fixed. The distribution of wages in the economy is exogenous to the worker and workers are currently employed at their reservation wage. This assumption reflects the fact that the study firm employs all production workers at the same wage near subsistence income levels. Given that workers can likely not fulfill minimum nutrition requirements below this wage level, it is improbable that their wages reflect a reservation wage. The assumption is also in line with empirical evidence, including in my data below, that reservation wages play a limited role in job search (Schmieder, von Wachter, & Bender, 2016) and are not significantly affected by time preferences (DellaVigna & Paserman, 2005; Krueger & Mueller, 2016). Second, once a worker finds a new job, she will stay in this job indefinitely. Third, utility is separable in consumption and search effort. If search costs were monetary and entered a concave utility function, the marginal costs of search would decrease with consumption. To ease interpretation I want to abstract from this case, which is discussed in detail by Lentz and Tranæs (2005). Fourth, if workers are present-biased with discount factor $\beta < 1$, I assume that they are naïve about it. Every period, each worker assumes that her future self will be an exponential discounter with $\beta = 1$. Workers overly optimistic predictions of savings over the course of employment, reported in the previous section, can be interpreted as evidence of such naïve. While there is substantial evidence for naïveté in the literature, including from the widespread lack of commitment (Laibson, 2015), it is more likely that most individuals are neither fully naïve nor fully sophisticated (O'Donoghue & Rabin, 2001). Importantly, unlike individuals who are (partly) sophisticated about their present-bias, naïve individuals will not demand any commitment devices to help them overcome their present-bias.

2.3.1.1 Setup

Consider a worker with finite planning horizon who in each period t chooses assets in the next period A_{t+1} as well as contemporaneous job search effort $s_t \in [0, 1]$, which represents the probability of receiving a job offer at the end of the current period and thus having a new job in $t + 1$. Search costs $k(s_t)$ are twice continuously differentiable and convex with $k'(s_t) > 0$, $k'(s_t) < 0$, $k(0) = 0$, and $k'(0) = 0$. Flow utility in each period is $u(c_t) - k(s_t)$, where c_t is the period consumption and utility from consumption $u(c_t)$ is strictly concave. Income y_t comes from a wage w_t paid at the safety net firm or an outside option $\widetilde{w} > w_t$ in a different job. Once a worker has found and accepted another job at $\widetilde{w}$, the search choice becomes mute. The path of wages at the firm $\{w_t\}_{t=1,\ldots,T}$ is exogenous and reflects the firm's fixed pay scale which depends on tenure. In each period workers can accumulate or run down assets A_t, which earn a return R and are constrained by $A_t \geqslant -L$. The per-period budget constraint is thus $y_t - c_t = \frac{A_{t+1}}{1+R} - A_t$.

2.3.1.2 Value Functions

The formulation of the dynamic programming problem follows DellaVigna et al. (2017), so my exposition here is brief. A worker without present bias who is on her job at the safety net firm chooses s_t and asset level A_{t+1}, which implicitly defines consumption c_t. Her value function is

$$\begin{aligned} V_t^F(A_t) = & \max_{s_t \in [0,1];\, A_{t+1} \geq -L} u(c_t) - k(s_t) \\ & + \delta \left[s_t V_{t+1}^O(A_{t+1}) + (1 - s_t) V_{t+1}^F(A_{t+1}) \right], \end{aligned} \tag{2.1}$$

where δ is the regular per-period discount factor. V^O is the value of an outside job opportunity in period t given by

$$V_t^O(A_t) = \max_{A_{t+1} > 0} u(c_t) + \delta V_{t+1}^O(A_{t+1}). \tag{2.2}$$

The maximization of equations (2.1) and (2.2) is subject to the common budget constraint

$$c_t = A_t + y_t - \frac{A_{t+1}}{1+R} \tag{2.3}$$

and liquidity constraint $A_t \geqslant -L$ for all t. The maximization in V_t^O is a well-behaved sequence problem as the objective is concave, continuous, and the constraint is compact.[18] As noted by Lentz and Tranæs (2005), V_t^F could theoretically be convex, though they show in simulations that nonconcavity never arises. I will follow Card et al. (2007) in simply assuming concavity.

A worker on her job at the safety net firm chooses s_t to maximize expected utility. Substituting the budget constraint into (2.1), the first-order condition for optimal search intensity of a worker without present bias s^* is

$$c'(s_t^*(A_{t+1})) = \delta \left[V_{t+1}^O(A_{t+1}) - V_{t+1}^F(A_{t+1}) \right]. \tag{2.4}$$

Intuitively, the optimal level of search effort equates the marginal costs of search effort with the marginal gain from search, given by the difference between the value of the outside option and the value of remaining at the safety net firm. The right-hand side of (2.4) is the net value of the outside option.

Compare this to a naïve present-biased worker who is on her job at the safety net firm. She faces the value function

$$\begin{aligned} V_t^{F,n}(A_t) = & \max_{s_t \in [0,1];\, A_{t+1} \geq -L} u(c_t) - k(s_t) \\ & + \beta\delta \left[s_t V_{t+1}^O(A_{t+1}) + (1 - s_t) V_{t+1}^F(A_{t+1}) \right] \end{aligned} \tag{2.5}$$

[18] As is shown, for example, in Adda and Cooper (2003), Chapter 2.3.

where the additional parameter $\beta \leq 1$ allows for the worker to be present-biased between the current period and the future. Recall that I assume naïveté, so that continuation values V_{t+1}^F and V_{t+1}^O above are equivalent to those of the exponential discounters in (2.1) and (2.2), respectively. Intuitively, the naïve present-biased worker assumes in every period that in the next period she will discount the future only by the factor δ. The naïve present-biased worker who found another opportunity faces the value function

$$V_t^{O,n}(A_t) = \max_{A_{t+1}>0} u\left(c_t\right) + \beta\delta V_{t+1}^O(A_{t+1}). \tag{2.6}$$

The first-order condition for optimal search intensity of present-biased workers s^{n*} given budget constraint (2.3) and value function (2.5) is now

$$k'\left(s_t^{n*}\left(A_{t+1}\right)\right) = \beta\delta\left[V_{t+1}^O(A_{t+1}) - V_{t+1}^F(A_{t+1})\right]. \tag{2.7}$$

Due to equivalent continuation values I can directly compare search effort with and without present bias by combining first order equations (2.4) and (2.7):

$$\beta\, k'\left(s_t^*\left(A_{t+1}\right)\right) = k'\left(s_t^{n*}\left(A_{t+1}\right)\right). \tag{2.8}$$

We can now see that search effort is strictly increasing in β due to assumed convexity of the search cost function $k(s_t)$.

It is difficult to fully characterize the model with search and savings analytically.[19] I thus obtain key predictions numerically.

2.3.2 Simulations

To simulate the model from the previous section I make additional functional form assumptions described in Appendix 2.A.1. Parameters are set based on survey data from my experiment, which I present in more detail below.

Before discussing the main results it is worth noting that in the framework presented above, workers at the safety net firm do not have a savings motive if they believe that they will never leave the firm. Recall that involuntary transitions are ruled out, so workers will simply continue to consume their wage indefinitely. While they will continue to invest in search as the value of being at the firm decreases relative to the value of being in another higher-wage job (compare first order condition 2.7), they will not accumulate any assets.

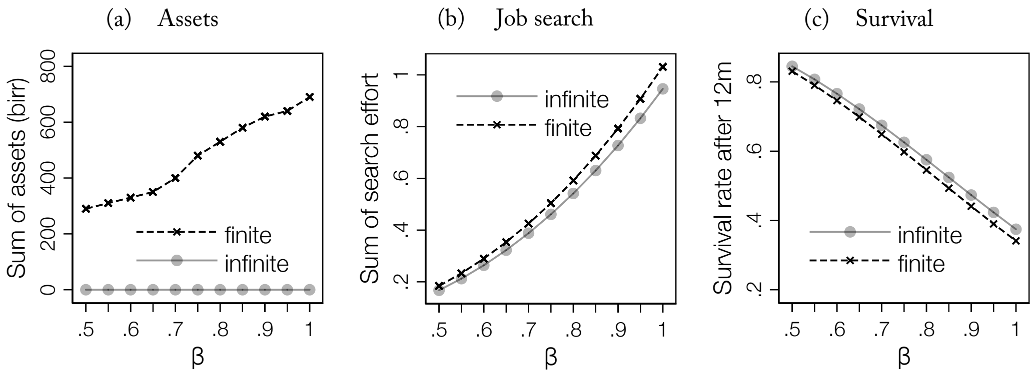
Consider instead the case where I exogenously impose that workers leave after one year, which is equivalent to the median expected tenure that workers in my experiment report on the day that they join the firm. I assume that after leaving income falls to the median pre-employment income reported in my data. In that case, workers have precautionary savings motive that leads them to reduce consumption and accumulate savings towards the end of their expected tenure. Figure 2.A1 in Appendix 2.A.2 plots the optimal paths of search

[19] DellaVigna and Paserman (2005) characterize the relationship of impatience and search effort in a model without savings, where agents choose the reservation wage and search effort. Endogenizing savings leads to a significantly more complicated setup. One simplification that helps to characterize the model analytically is to change the timing so that successful job search in period t leads to a new job in the same period. This is the approach in Lentz and Tranæs (2005) and Card et al. (2007).

effort and asset accumulation and the implied paths of consumption and turnover for both cases.

To see how workers use both search and savings as insurance mechanisms, it is instructive to look at substitution between search and savings. Following Paserman (2008) and DellaVigna et al. (2017) I simulate the model allowing for heterogeneity in search costs with six different worker types who have costs $\{\psi_j\}_{j=1,...,6}$ where $0 < \psi_1 < ... < \psi_6$. I focus on the assumption of finite tenure at the firm, so that the asset choice is not mute. We can see from Figure 2.A3 that types with low search costs search more (plotted on the right y-axis) and save less (plotted on the left y-axis). Intuitively, workers adjust the use of both insurance mechanisms in response to their relative costs.

FIGURE 2.2: How Present Bias Affects Job Search, Asset Accumulation, and Survival (Simulation Results, by Present Bias Parameter)



Notes: Finite refers to simulations in which workers know that they will leave the safety net firm after one year, which is equivalent to the median expected tenure in my survey data. Infinite refers to simulations in which workers assume that they never need to leave the safety net firm, so they do not have a precautionary savings motive. Panels (a) and (b) represent the sum of search effort and assets over all simulated periods. Parameter values as given in text. Figure 2.A1 in Appendix 2.A.2 illustrates the optimal paths of job search and assets and implied consumption and survival for the case of no present bias ($\beta = 1$).

I focus on comparative statics with respect to present bias β. In the framework presented above, both present bias β and the exponential discount parameter δ affect search effort and savings in qualitatively similar ways because they both reduce the present value of the benefits of search and the benefits of saving.[20] Unlike exponential discounting, present bias leads to suboptimal decisions because workers today will under-invest in search and savings because they erroneously expect to search and save more tomorrow. This undermines self-insurance efforts of workers.

First, I examine the relationship of asset accumulation and present bias (Figure 2.2, panel a). As explained above, workers who plan to stay at the firm indefinitely do not have a savings motive and thus do not accumulate any assets. Under the assumption of finite tenure, workers do accumulate assets. The simulations illustrate that assets are increasing in

[20]Appendix Figure 2.A2 illustrates the relative magnitudes with simulations that also vary the discount factor.

β. Lower asset accumulation due to present bias reduces the ability of workers to smooth consumption after leaving the firm. This suggests that there is scope for a welfare-improving intervention that helps present-biased workers save.

Prediction 1 (Present Bias and Savings): Asset accumulation is increasing in β.

Next, I consider the relationship of job search effort and present bias. Panel (b) of Figure 2.2 plots total job search effort by present bias β. As can be seen from the comparison of first order conditions of workers with and without present bias (equation 2.8), search effort is strictly increasing in β. An increase in present bias reduces the present value of search. This effect is identical to the finding of DellaVigna and Paserman (2005).

Prediction 2 (Present Bias and Search): Search effort is increasing in β.

Finally, I turn to survival rates that are implied by worker choices of optimal job search effort and savings (Figure 2.2, panel b). Prediction 3 follows as a direct corollary from Prediction 2 because the probability of leaving the safety net firm depends on the probability of receiving an alternative wage offer and thus directly on search effort s.

Prediction 3 (Present Bias and Turnover): The survival rate at the safety net firm is decreasing in β.

In the next section, I present data on these three predictions that I collected from a sample of workers who join a firm in Ethiopia's RMG industry. I then present results of an intervention designed to alleviate self-control problems in savings due to present bias.

2.4 Experimental Design

2.4.1 Description of Data Collection

For data collection I cooperate with a garment manufacturing firm described in Section 2.2 above. The study firm is typical of the low-skill, low-wage, export-oriented manufacturing industry that can be found in many parts of the developing world. It is located in an industrial park in a peri-urban area of Addis Ababa, the capital of Ethiopia.

Data on a random sample of 460 workers who start employment at the firm is collected in three steps: First, I conduct an in-person baseline survey and a lab-in-the-field experiment to elicit individual time preferences on the day that workers join the study firm. Second, I track individual behavior with a high-frequency phone survey. Third, I conduct an in-person endline survey after the conclusion of the panel. The survey is set up to track workers if they leave the firm. I combine my own data with firm administrative records on tenure and basic demographic data on 238 *pure control* workers who are randomized out of study participation.[21]

Women looking for employment as production workers come to the factory gate of the study firm every day. At the end of each day, I draw a random sample from all job

[21] I also collect rich productivity data for all workers, which we analyze in a companion paper (Hardy, Kagy, & Meyer, 2018).

seekers who are hired by the firm on that day. No other criteria are used for inclusion in the study. Appendix Appendix 2.B.1.1 provides details on the randomization procedure. On the next day, the hired candidates return to the firm to begin employment. Informed consent, baseline interview, and time preference elicitation are administered through in-person interviews on the morning of the second day immediately before candidates are assigned to a production line.[22] As part of the baseline interview, I also collect detailed data on employment histories, subjective expectations over search and savings behavior, and a battery of tests for cognitive and non-cognitive skills. Appendix Sections 2.B.2 and 2.B.3 summarize. Baseline interviews are conducted every day from March to July 2018. On average the team of enumerators conducts 7 baseline interviews per day and 29 per week. Baseline interviews last 85 minutes on average.

In the three months after the baseline, enumerators call respondents every 14 days to collect data on consumption expenditures and savings, job search behavior, job search outcomes, transitions to other jobs (if any), and measures of psychological well-being. Phone surveys are practical in the local context because all workers have a mobile phone and because they allow me to track subjects even if they leave the firm. The same approach has been used successfully in a similar setting by Abebe et al. (2016) and Franklin (2017). Phone calls last 7.5 minutes on average. After three months, enumerators conduct another in-person endline interview. All interviews are conducted in Amharic and Oromiffa using computer-assisted personal interviewing (CAPI). Appendix Appendix 2.B.1.2 gives a detailed description of survey protocols.

Attrition from the phone survey is relatively low. The second follow up call after one month reaches 437 out of 460 or 95 percent of all respondents. The fourth call after two months reaches 331 out of 460 or 72 percent of all respondents. Enumerators stay reasonably well on schedule: At the first follow-up, 93 percent of calls are within 4 days of the scheduled interview day. At the fourth call after two months, 85.5 percent of calls are within four days of the scheduled interview day. Appendix Figure 2.D3 illustrates.

2.4.2 Experimental Elicitation of Time Preferences

I estimate time preferences over money for each respondent using an adapted version of the CTB task by Andreoni and Sprenger (2012). Each subject is asked to allocate an experimental budget $m > 0$ between an amount c_t available at an earlier time t and another amount c_{t+k} available after a delay $k > 0$, i.e. paid out at point $t + k$. Let $(1 + r)$ be the simple gross interest rate to be paid over period k. This means that subjects maximize utility subject to the experimental budget constraint $(1+r)c_t + c_{t+k} = m$. Let the unit of time be days since the experiment and all monetary amounts be measured in experimental tokens.

The CTB method aims to address methodological problems of MPL approaches, which frequently rely on the assumption of linear utility and may lead to biased estimates of time preferences when utility is in fact concave. Importantly, the CTB method lends itself to structural estimation of aggregate and individual-level (β-δ) time preference parameters. Given functional form assumptions, the discounting parameter δ can be identified from

[22] Given that I work closely with the firm, I take a number of precautions to maximize privacy and confidentiality of the respondents. This includes selecting a random subset of workers to be interviewed at home instead of at the factory. Appendix Appendix 2.B.1.3 elaborates.

variation in the interest rate r and delay k while present bias β can be identified from variation in the timing of the earlier payout t. Appendix 2.C provides details on the estimation technique.

Irrespective of whether time preferences are elicited using MPL or CTB approaches, a number of confounds may undermine identification of preference parameters from time-dated payments. One important concern relates to real or perceived transaction costs of receiving payments in the future versus receiving payments today. Subjects may exhibit a preference for earlier payments because it is more costly to obtain the later payment or because of uncertainty over whether the experimenter will deliver the payment as promised. Several recent studies that carefully equalize transaction costs between time-dated payments find little evidence of aggregate present bias (Andersen et al., 2008; Andreoni & Sprenger, 2012; Augenblick et al., 2015; Giné et al., 2017). I take several steps to carefully equalize transaction that I discuss in more detail below.

A second potential confound relates to the assumption of time-invariant utility of experimental subjects. Even if transaction costs are equalized between two time-dated payments, experimental subjects may simply have have preferences that change over time (Halevy, 2015; Janssens, Kramer, & Swart, 2017). One way to rationalize such time-varying preferences are economic conditions outside of the experiment, in particular liquidity constraints on the part of subjects (Dean & Sautmann, 2016). I argue that liquidity are unlikely to be a driver of my experimental results in 2.5.3. I show that a randomized cash drop between the baseline and the endline of the experiment does not affect estimated present bias at endline. In addition, all results below control for measures of baseline liquidity and access to finance.

A third and related concern is whether choices over time-dated monetary payments can identify time preferences over consumption (Augenblick et al., 2015; Cubitt & Read, 2007). If experimental subjects can borrow (save) at a market interest rate that is lower (higher) than the experimental interest rate between earlier and later payments, subjects could allocate their whole budget to the later (earlier) period and arbitrage between the experiment and the market. Such arbitrage likely requires highly sophisticated subjects.[23] Augenblick et al. (2015) overcome arbitrage concerns in the monetary domain by using the CTB method with a real-effort laboratory task and time-dated effort allocations. They find that aggregate present bias is limited for preferences over time-dated effort, but significant for preferences over time-dated monetary payments. While relatively limited access to capital markets in my context combined with the fact that most time-dated allocations in my data are interior solutions suggests that arbitrage concerns are not first order, I do acknowledge this as a limitation of my study.[24]

[23] In a carefully designed laboratory experiment, Coller and Williams (1999) assess the arbitrage argument by providing information about market interest rates. They show that some subjects attempt to exploit arbitrage opportunities between the laboratory experiment and the market, but that they either do not know outside opportunities or fail to determine the correct market rate.

[24] Similar to Augenblick et al. (2015) I also elicit time preferences over real effort by asking subjects to allocate an amount of work at the factory between an earlier and a later point in time. I use a design that builds on Carvalho, Meier, and Wang (2016), where I vary the length of the shorter work assignment and the time by which it needs to be accomplished while holding the delay between earlier and later assignment constant. Results are omitted here because the study firm did not let me implement an incentivized version of this task and all decisions are purely hypothetical. (No deception was involved because the wording of the question was adapted in due time to reflect the hypothetical nature of the task.)

2.4.2.1 Implementation of CTB Task in the Field

I follow the procedures of Giné et al. (2017) to implement the CTB method in the field. Respondents allocate 20 tokens in the form of beans between two empty dishes that represent the two payoffs. Each of the two dishes is positioned below a small whiteboard that indicates the exact payoff date and an exchange rate at which beans can be converted into birr.[25]

I vary the initial payment date t and delay between earlier and later payment k, so that each subject faces three choice sets $(t, k) \in [(1, 14), (1, 28), (15, 14)]$. Within each choice set, a bean on the earlier dish is always worth 10 birr. A bean on the later dish is worth $10 \times (1 + r)$ birr, where $r \in [0.10, 0.25, 0.50, 0.75, 1.00]$. This implies that experimental stakes are large. Subjects can receive between 200 and 400 birr (US\$7.26 to US\$14.52), or 20 to 40 percent of their monthly starting salary.

2.4.2.2 Experimental Protocol

The CTB task is administered as part of the baseline survey and again as part of the endline survey. In this paper I focus on the baseline results. The CTB task is administered towards the middle of the survey questionnaire. This was done so that the enumerator can build trust with the respondent while at the same time minimizing respondent fatigue. The order in which respondents are presented with the CTB choice sets is randomized while the exchange rate for beans on the later dish always increases in each set. As in Giné et al. (2017), enumerators administer a set of unrelated question from the baseline survey between each choice set to reduce the chance that respondents attempt to be consistent between sets.

At the beginning of the CTB task, the enumerator explains the task to the respondent. Respondents are required to pass three questions that test their understanding. The respondent then practices one allocation that will not be implemented. The enumerator also informs the respondent that there is a 50 percent chance that one allocation (with earlier and later payouts) is implemented, determined by a coin flip at the end of the experiment. At the beginning of each choice set, the enumerator uses one of the whiteboards to provide an overview of the value of beans in each of the five decisions in the set. At the beginning of each allocation decision, the enumerator wipes the whiteboards and writes down the payment date and the exchange value of one bean at the top of each board. The enumerator then asks the respondent to allocate all 20 beans to the two dishes. Care was taken to use neutral language and not lead respondents away from corner solutions. Once the respondent has made her allocation, the enumerator calculates the total monetary value on each dish and writes it on the whiteboard next to it. The respondents is asked if she is satisfied with the allocation and is given the opportunity to revise as many times as she likes. For each revised decision the enumerator re-calculates the total monetary amount and again writes it on the board. Once the respondent is satisfied, the enumerator records the decision in the survey software. The software confirms both the number of beans and the total monetary amounts. The coin flip and draw of the payoff-relevant decision are done at the end of the

[25] Respondents use the same beans and dishes for various other parts of the baseline survey before the CTB task (see Appendix Appendix 2.B.2). Extensive piloting over several weeks before the study confirmed that respondents found this a natural and easily-comprehensible way to allocate a budget.

survey. Appendix Appendix 2.B.1.1 gives details. Appendix Figure 2.C1 shows a picture of the allocation decision.

2.4.2.3 Experimental Payments

To obtain unbiased time preference estimates it is critical to equalize real and perceived transaction costs between the earlier and the later time frame. This is particularly true in the setting of this study because respondents are not subjects from a laboratory pool that have a relationship with the experimenter, but a highly mobile and disadvantaged population who may not have full confidence in experimental payments being delivered on time or delivered at all. I take several steps to equalize transaction costs and reduce uncertainty over whether and how payments are delivered.

First, I rely on mobile money payments to respondent cell phones using *CBEbirr*, a system operated by Commercial Bank of Ethiopia (CBE), the country's largest bank. The system is similar to Kenya's popular M-Pesa system and allows users to receive and transfer money and to purchase goods and services using simple text messages. Transfers are immediate and can be precisely timed. Users can withdraw money from any CBE branch or one of many authorized CBEbirr agents. CBE is a well-known and widely-trusted institution, even in more rural areas. All subjects in the study had phones that supported CBEbirr.[26] This equalizes and minimizes costs of receiving and accessing the payments. Mobile money has previously been used for experimental payments from a laboratory CTB task by Balakrishnan et al. (2017).

Second, I follow Andersen et al. (2008) and Giné et al. (2017) in prioritizing symmetry of payments over the opportunity to pay subjects immediately. Even though mobile money payments allow for truly immediate payments as in Balakrishnan et al. (2017), making the earlier payment while enumerators are still physically present with the respondents could favor the earlier over the later allocations if respondents do not have full confidence in the later payment being delivered as promised. I instead choose to make the earliest payment on the next day before noon. While the front-end delay equalizes perceived or real transaction costs, it comes at the cost of not being able to study present bias with respect to truly immediate consumption. It is likely that the small delay attenuates any present bias.[27] Given that present bias is a key input to my empirical analysis, I prefer this more conservative approach over one that increases my ability to detect present bias at the risk of undermining identification.

Third, I follow previous implementations of the CTB task (Andreoni & Sprenger, 2012) in making sure that both experimental payments always happen on the same day of the week. This avoids differential weekday effects between sooner and later payments. Subjects know that all payments are made before noon.

[26] Even though all respondents could receive payments, 4 percent of respondents who won the coin flip preferred to receive their payments using *hawala*, a widely-used money transfer system that is more comparable to a money order. There are several formal and informal hawala operators in Ethiopia. For the small number of cases in which subjects did not want to receive their payments through CBEbirr, we sent formal hawala payments through CBE. Dropping these observations from the analysis does not affect results.

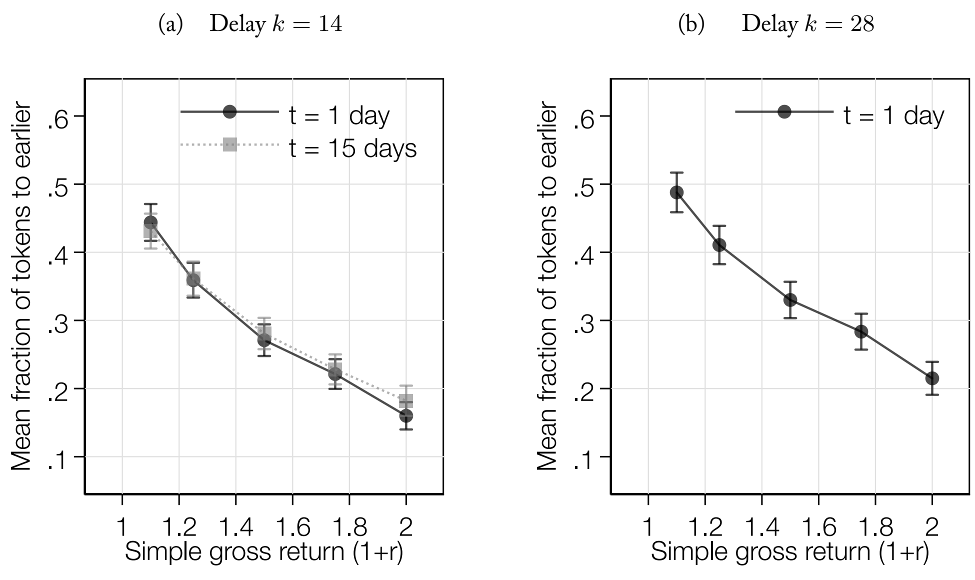
[27] This is indeed one of they key findings of Balakrishnan et al. (2017). Using a CTB task in a laboratory experiment in Nairobi, they find evidence of present bias over money when payments are truly immediate. When payments are slightly delayed until the end of the day, they do not observe aggregate present bias.

Fourth, respondents who win the payout are given a written confirmation or voucher that repeats the payment amounts, payment times and dates, and information on how the payment is delivered. This information serves to reduce the cognitive costs of keeping track of payments.[28] It is is printed on high-quality paper with a watermark of the principal investigator's university, signed by the principal investigator and the enumerator, and includes the private cell phone number of the survey coordinator. Similar to Andreoni and Sprenger (2012), who include a business card of one of the authors and encourage student subjects to reach out to if the payment is not delivered, we encourage subjects to reach out to us should there be any problems with the payment. This is intended to increase trust in the study team and confidence that the payment will be delivered. Appendix Appendix 2.C.3 provides an English translation of the confirmation.

2.4.2.4 Results of Time Preference Elicitation

While aggregate time preferences are not the focus of this study, it is useful to consider aggregate results in order to check if subjects understood the task and to compare results with the existing literature. Figure 2.3 shows the mean fraction of beans allocated to the earlier dish for each experimental interest rate r, separately by front-end delay t and delay between earlier and later allocation k. Error bars indicate the standard error clustered at the individual level.

FIGURE 2.3: Fraction of Experimental Budget Allocated to Earlier, by Front-End Delay t and Delay Between Earlier and Later k



It is clear that in all three choice sets subjects on average respond to the increase in interest rate in line with the law of demand. As the interest rate increases and the price of

[28] Both Andreoni and Sprenger (2012) and Balakrishnan et al. (2017) take such steps to reduce the cognitive transaction costs that may result from keeping track of future payments. Haushofer (2015) explores theoretically how the cognitive cost of keeping track may explain stylized facts of temporal discounting in the literature.

consumption later decreases, subjects monotonically decrease the share of beans allocated to earlier. Appendix Appendix 2.C.4 assesses individual consistency with the law of demand, which compares favorably to previous experiments in the literature. For a given interest rate subjects increase the share of beans allocated to earlier as the delay between earlier and later increases. Appendix Table 2.C1 presents an overview of all subject allocations, including the percentage of corner solutions, for each of the choice sets and each interest rate.

Panel (a) of Figure 2.3 suggests that there is no evidence for aggregate present bias. When comparing the two choice sets with delay $k = 14$ and different front-end delay t, one would expect present-biased subjects to allocate a larger fraction of their budget to earlier when $t = 1$. On aggregate this does not appear to be the case.

For a more rigorous test of aggregate present-bias I follow the parametric assumptions of Andreoni and Sprenger (2012) and Augenblick et al. (2015) to structurally estimate aggregate time preference parameters. I assume constant relative risk aversion (CRRA) utility and, in line with the previous literature, make varying assumptions about background consumption. Appendix Appendix 2.C.5 provides a detailed discussion of identification and estimation via maximum likelihood. Appendix Table 2.C2 presents parameter estimates recovered through non-linear combinations of regressions coefficients from a two-limit Tobit. (β-δ) parameters are precisely estimated and relatively stable across the three assumptions over background consumption, while estimated CRRA curvature varies more strongly with background consumption assumptions. Assuming no background consumption, I estimate present bias $\hat{\beta} = 1.006$ (SE 0.018) and weekly discounting $\hat{\delta} = 0.959$ (SE 0.008). Utility is concave but with relatively limited curvature at $\hat{\alpha} = 0.886$ (SE 0.006). This suggests that money is reasonably fungible between payment dates.

Importantly, in all specifications I fail to reject the null hypothesis of $\beta = 1$. Overall, aggregate parametric results for time preference and utility curvature closely mirror previous estimates by Andreoni and Sprenger (2012), Augenblick et al. (2015), Balakrishnan et al. (2017), and Giné et al. (2017).

Having established that subject choices are consistent with the law of demand and that parametric estimates of aggregate time preferences are in line with the literature, I can now turn to individual-level time preference estimates. I use the same parametric assumptions as for the aggregate estimates to estimate two-limit Tobit regressions for each subject. As before, I make three different assumptions about background consumption. Table 2.1 summarizes the distribution of parameter estimates for individual-level discount parameter $\hat{\delta}_i$, present-bias parameter $\hat{\beta}_i$, and CRRA curvature $\hat{\alpha}_i$. Appendix Figure 2.C4 plots the correlation of present bias and discounting for all three assumptions over background consumption. As in Andreoni and Sprenger (2012) for ease of exposition I focus the rest of my analysis on the assumption of no background consumption (Table 2.1, panel a).

First, the estimation strategy appears to produce reasonable parameter estimates for most subjects, though 21.32 percent of $\hat{\beta}_i$ estimates fall out of the range of $[0.75, 1.5]$. While this share of extreme values is larger than comparable estimates in the literature based on laboratory subjects, it is in line with previous work from the field. Using a different estimation strategy and time preferences over work effort, Andreoni, Callen, Hussain, Khan, and Sprenger (2017) find that 20.3 percent of their sample has $\hat{\beta}_i$ estimates fall out of this range. For the remainder of this chapter, I follow Andreoni et al. (2017) in trimming extreme

Table 2.1: Individual-Level Time Preference Parameter Estimates from Two-Limit Tobit, by Assumptions About Background Consumption

Parameter	N	Mean	5th pctl	50th pctl	95th pctl	Min	Max
			Panel (a) No background consumption $\omega_1 = \omega_2 = 0$				
Present bias $\hat{\beta}_i$	406	1.235	0.581	1.011	1.791	0.018	24.554
Discount factor $\hat{\delta}_i$	406	1.076	0.769	0.972	1.408	0.196	22.569
CRRA curvature $\hat{\alpha}_i$	406	0.645	-0.159	0.825	0.984	-7.677	6.042
			Panel (b) Sample average background consumption $\omega_1 = \omega_2 = \bar{c}$				
Present bias $\hat{\beta}_i$	404	1.389	0.654	1.017	1.618	0.004	108.916
Discount factor $\hat{\delta}_i$	404	1.023	0.791	0.988	1.212	0.194	6.625
CRRA curvature $\hat{\alpha}_i$	404	0.430	-0.372	0.641	0.922	-13.751	1.640
			Panel (c) Individaul background consumption $\omega_1 = \omega_2 = c_i$				
Present bias $\hat{\beta}_i$	404	1.13	0.611	1.017	1.631	0.008	12.99
Discount factor $\hat{\delta}_i$	404	1.05	0.797	0.987	1.247	0.235	20.48
CRRA curvature $\hat{\alpha}_i$	405	0.382	-0.523	0.659	0.944	-34.027	2.108

Notes: Table shows individual-level maximum likelihood estimates of Equation (17) using separate two-limit Tobit models. The three panels reflect different assumptions about background consumption at each point in time (see Equation (13) for details) and correspond to columns (1) to (3) of aggregate estimates presented in Appendix Table 2.C2. Panel (a) assumes no background consumption ($\omega_1 = \omega_2 = 0$). Panel (b) assumes that background consumption is constant and set at the sample average daily consumption expenditure in the seven days before the baseline survey ($\omega_1 = \omega_2 = \bar{c}$). Panel (c) assumes that background consumption is constant and set at the individual daily consumption expenditure in the seven days before the baseline survey ($\omega_1 = \omega_2 = c_i$). Appendix Figure 2.C4 provides a scatter plot of β_i and δ_i for the three different assumptions.

estimates. I trim the top and bottom 5 percent of the sample.[29]

Second, it is worth noting that in line with the aggregate estimates, the individual-level parameter indicate that neither the mean subject nor the median subject are present-biased. CRRA utility curvature for the median subject is different from 1, but relatively close to linear. This differs from previous experimental estimates that do not use the CTB method, which find significantly more curvature Andersen et al. (2008), but is in line with previous estimates based on the CTB method (Andreoni & Sprenger, 2012; Augenblick et al., 2015).

While there is no theoretical reason for using a binary measure of present bias, most of the literature has done so because experimental results did not permit parameter estimation (Ashraf et al., 2006; Meier & Sprenger, 2010). To compare my estimates with existing work and facilitate presentation of results, I will define an indicator for present bias, which takes the value 1 if $\hat{\beta}_i < 1$ and 0 otherwise. I find that 37.8 percent of all subjects are present-biased. This share is in line with proportions reported in the literature.[30]

[29] This is more conservative than Andreoni et al. (2017), who trim the top and bottom 1 percent of their sample.

[30] Augenblick et al. (2015) report that 33 percent of subjects have $\hat{\beta}_i < 0.99$. In my data, 36.5 percent of subjects have $\hat{\beta}_i < 0.99$.

2.5 Empirical Analysis

2.5.1 Sample Description and Comparison to National Household Survey

Table 2.2 reports baseline means and standard errors on the random sample of 460 workers (column 2). It also reports selected demographic data from firm personnel records on the workers that were randomized out of the sample (column 1). On the three observable characteristics available, workers that were randomized out of participation are not significantly different from workers who participated in the study.

As reported in more detail in Section 2.2 above, workers in the sample are exclusively female and tend to be young, low-skill, rural-urban migrants with little to no previous work experience. Table 2.2 presents a comparison of the study sample with household survey data from women in the same age range in Addis Ababa (column 3), the wider population of Addis Ababa (column 4) and Ethiopia (column 5) based on the 2015-16 ESS. The ESS is representative for Addis Ababa and at the national level.[31] Compared with women in the same age range in Addis Ababa, workers in the study sample have less education (completed 6th grade versus completed 10th grade), fewer are married (19 percent versus 30 percent), and slightly fewer identify as Ethiopian Orthodox (73 percent versus 77 percent).

In terms of living standards measured by consumption and asset ownership (Table 2.2, panel b), the study sample is poorer than women in the same age range in Addis Ababa. Workers in the sample spend 129.88 birr (US$ 4.72) per person per week on food consumption, compared to 158.03 birr (US$ 5.74) for the average women in the same age range in Addis Ababa and 149.39 birr (US$ 5.23) for the average person in Addis Ababa. For comparison, the local food poverty line in Addis Ababa is 109.66 birr (US$ 3.98) per person per week.[32] Workers in the study sample also live in households that are notably poorer in assets as measured by a simple additive index. Appendix 2.B.4 provides details on household asset data in my sample.

2.5.2 Correlations of Present Bias with Savings and Job Search

Results are presented in three subsections that follow the predictions derived from the theoretical framework in Section 2.3. Each subsection includes a range of robustness checks. In a fourth subsection I discuss alternative explanations.[33]

[31] I use publicly available microdata from the World Bank's Living Standards Measurement Survey (LSMS) program (Central Statistical Agency and World Bank, 2017). While sampling methods and survey protocols are inherently different from my survey, I aimed to harmonize concepts across survey instruments. The age range of the study population and for column 6 of Table 2.2 is 18 to 31.

[32] For further comparison, total average consumption expenditures at the national level in the 2015-16 ESS was 246.16 birr per person (US$ 8.94). Food consumption from the 2015-16 ESS is adjusted for inflation using the national GDP deflator. Recall period in both survey instruments is seven days. The ESS food aggregate is adjusted by adult equivalent household size. The poverty line is based on the official 2015/16 food poverty line of 3,772 birr per adult equivalent per year, adjusted to current values using the GDP deflator and adjusted for local prices using the spatial food price indices reported in National Planning Commission (2017).

[33] I filed a pre-analysis plan (PAP) with the AER RCT Registry (#AEARCTR-0002555) after piloting the survey instrument and before any data was collected. The current chapter deviates from the PAP in several ways. The PAP was build on the assumption of having daily attendance and productivity data. The firm could unfortunately not provide in time for the analysis. The PAP excluded savings and focused on correlations of present bias with job search effort and labor supply. The current chapter does not consider the administrative

Table 2.2: Baseline Summary Statistics and Comparison to National Household Survey (Means and Standard Errors)

	Pure Control	Study Sample	Wider Population		
			Addis Ababa Young Women	Addis Ababa	Ethiopia
	(1)	(2)	(3)	(4)	(5)
Panel (a) Personal background					
Age	21.44	21.41	24.31	28.20	23.09
	(0.226)	(0.117)	(0.333)	(0.733)	(0.217)
% female	1.00	1.00	1.00	0.56	0.47
	-	-	-	(0.013)	(0.006)
Education (respondent)	8.16	8.17	10.30	6.52	2.33
	(0.135)	(0.243)	(0.496)	(0.298)	(0.078)
% married	0.16	0.19	0.30	0.25	0.26
	(0.028)	(0.018)	(0.033)	(0.012)	(0.003)
% Ethiopian Orthodox faith		0.73	0.77	0.55	0.30
		(0.021)	(0.050)	(0.038)	(0.016)
Panel (b) Living standards					
Food consumption (past 7 days)		129.88	158.03	149.39	103.95
		(6.885)	(14.21)	(11.701)	(2.662)
Non-food consumption (past 7 days)		147.93			
		(9.997)			
Asset index (household)		4.07	8.70	8.45	1.72
		(0.145)	(0.417)	(0.353)	(0.069)
Asset index (self)		1.12			
		(0.077)			
Savings (3 month recall)		384.09			
		(39.582)			
Savings (would like per month)		603.01			
		(14.885)			
Savings (will likely per month)		355.10			
		(8.729)			
N	238	460	183	1,188	27,990
Population represented (using weights)			658,336	4,507,503	117,437,134

Notes: Column 1 presents administrative data from firm personnel records for the group of workers who were randomized out of participation in the study. Columns 3 to 5 present data from the 2015-16 Ethiopian Socioeconomic Survey (ESS) collected as part of the World Bank's Living Standards Measurement Survey program. Column 3 refers to women aged 18 to 31, the age range of the study population. ESS data is weighted and stratified using the provided survey weights. The ESS is representative for Addis Ababa and at the national level. While data in all columns of the table is measured using the same concepts, comparisons between my survey and the ESS should be seen as indicative due to different sampling methods and survey protocols. Food consumption in ESS adjusted for inflation using the national GDP deflator. Recall period for consumption expenditures is seven days both in my survey and in the ESS. The ESS food aggregate is adjusted by adult equivalent household size, while the food aggregate in my survey is per capita. Comparison with non-food data not shown because of differences in survey instrument. Asset index is an additive index of 13 ESS households assets that best predict nominal total household consumption at the national level. Appendix Appendix 2.B.4 for details.

Simple correlations between present bias and the outcome of interest may be biased if present bias is correlated with omitted individual characteristics such as work experience, liquidity, cognitive, and non-cognitive factors that may affect the outcome of interest. Therefore I try to control for a wide range of individual differences. I include controls in four groups, measured in the baseline interview when workers join the firm: Personal characteristics and human capital, liquidity and access to finance, cognitive control, and non-cognitive skills and stress.

Personal characteristics include respondent age, an indicator for respondent marital status, an indicator for whether the respondent has children, an indicator of whether the respondent is a rural-urban migrant, a set of indicators for the respondent's mother tongue and religion, an indicator for whether the respondent has a working spouse, and indicators for the highest education level of the worker. Access to finance includes an indicator for whether the respondent holds any savings at baseline and indicators for how difficult the respondent would find it to take out a small loan to cover an emergency. Cognitive control governs impulse control and affects how well individuals can formulate, maintain, and execute plans and goals. I measure cognitive control with a fully-incentivized spatial Stroop task that respondents complete on a tablet computer. Non-cognitive skills and stress are measured with scores on three psychological scales: Generalized Self-Efficacy, Locus of Control, and the Perceived Stress Scale.

Appendix 2.B provides details on the survey and explains the construction of these variables. Summary statistics on all variables used in the empirical analysis are presented in Appendix Table 2.E2.

2.5.2.1 Present Bias and Savings

Prediction 1 states that asset accumulation is increasing in present bias parameter β. Holding everything else constant, present-biased workers should save less. We can use data from the baseline CTB experiment and follow-up surveys for a reduced-form test of this prediction.

Data on savings per week is a non-negative response variable with a continuous distribution over strictly positive values and most observations at a corner of zero. To take this into account, I model the participation decision and the quantity decision jointly in a latent regression model, estimated using Tobit (Wooldridge, 2010). I pool observations from all follow-up calls and include dummies for each panel wave. To ease interpretation of results, I use a binary measure of present bias, which takes the value 1 if $\hat{\beta}_i < 1$ and 0 otherwise.

The estimation problem can be written as

$$Y_{i,t} = \gamma_0 + \gamma_1 \mathbb{1}_{\{\hat{\beta}_i < 1\}} + \gamma_2 \, \hat{\delta}_i + \sum_{s=0}^{8} \kappa_s \mathbb{1}_{\{t=s\}} + \theta' \mathbf{X}_i + \epsilon_{i,t} \tag{2.9}$$

where for each subject i in period t, $Y_{i,t}$ is the amount of money saved over a seven-day period, $\mathbb{1}_{\{\hat{\beta}_i < 1\}}$ is an indicator for whether subjects are present-biased in the baseline CTB experiment, $\hat{\delta}$ is the discounting parameter from the baseline CTB experiment, $\mathbb{1}_{\{t=0,..,8\}}$

data but instead includes self-reported savings in the analysis. The full administrative data is analyzed in a companion paper (Hardy et al., 2018).

are indicators for each survey wave, $\mathbf{X}_i$ is a vector of respondent observable characteristics and indicators for the enumerator who administered the survey, and ϵ is the error term clustered at the individual level.

Columns 1 and 2 of Table 2.3 present maximum likelihood estimates of Equation (2.9) with and without controlling for the full set of baseline covariates. I report coefficient estimates, which in the Tobit model give the marginal effect of each independent variable on the expected value of the latent variable.

Table 2.3: Savings, Temptation Goods Expenditures, and Present Bias (Regression Coefficient Estimates and Robust Standard Errors)

	Savings		Tempt. goods exp.	
	(1)	(2)	(3)	(4)
	Tobit	*Tobit*	*Tobit*	*Tobit*
Baseline $\hat{\beta}_i < 1$	-86.20	-25.91	13.85	0.670
	(58.089)	(41.836)	(15.158)	(16.063)
Baseline $\hat{\delta}_i$	445.7	84.99	-30.72	-39.16
	(342.470)	(212.468)	(89.965)	(101.663)
Survey wave dummies	Yes	Yes	Yes	Yes
Enumerator dummies	Yes	Yes	Yes	Yes
Personal characteristics	No	Yes	No	Yes
Baseline liquidity	No	Yes	No	Yes
Cognitive control	No	Yes	No	Yes
Non-cognitive ability and stress	No	Yes	No	Yes
N	1901	1832	2102	2028

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses, clustered at individual level. Time preference parameters trimmed at 5th and 95th percentile. Controls for personal characteristics are age, marital status, religion, mother tongue, education, number of children, whether or not respondent moved from a rural area, whether respondent has a working spouse. Controls for baseline liquidity are amount of savings held when starting employment and easy of obtaining credit assessed on a 4 point Likert scale. Control for cognitive control is the final score achieved in the baseline Stroop task. Controls for non-cognitive control and stress are baseline score on the General Self-Efficacy scale, baseline locus of control index, and baseline score on the Perceived Stress Scale (PSS-4). Appendix 2.B provides details on survey protocols and the measurement of control variables. Appendix Figure 2.D4 plots average marginal effects of the present bias indicator in column (2) and (4).

Baseline present-bias appears to reduce savings in line with the theoretical prediction, though the effects are small and not statistically significant at any conventional level. This is true irrespective of whether or not I control for the full set of baseline characteristics. To assess the economic significance of the coefficient one can calculate the average marginal effect of present bias on the expected value of savings. Holding all else constant and using coefficient estimates from column (2) that includes the full set of baseline covariates, present-biased individuals save about 7.7 percent less per week than time-consistent individuals. The predicted savings per week of present-biased individuals is 83.58 birr (US$2.93) compared to 90.57 birr (US$ 3.26) for time-consistent individuals, an insignificant difference of 6.99 birr, $\chi^2(1, N = 1,832) = 0.39$, $p = 0.533$. Panel (a) of Appendix Figure 2.D4 illustrates.

While the theoretical prediction only speaks to the causal link between present bias and savings, it is instructive to also consider consumption expenditure. In particular I consider the amount of consumption expenditures devoted to what Banerjee and Mullainathan (2010) call "temptation goods," that is goods that yield utility in the present as opposed to the future. The surveys collect consumption expenditure data on various categories of goods: food, alcoholic beverages, phone credit, transportation, clothing and shoes, soaps, cosmetics and beauty products, and gifts and donations. I categorize alcoholic beverages, clothing and shoes, and cosmetics and beauty products as temptation goods and use the sum as an alternative dependent variable of Equation (2.9).[34]

Table 2.3 columns 3 and 4 present Tobit coefficient estimates. After controlling for baseline covariates, I find no relationship between baseline present-bias and spending on temptation goods. Using the average marginal effect of present bias on the unconditional expected value of temptation good spending, I find an insignificant difference of 0.4 birr per week, $\chi^2(1, N = 2,028) = 0.00$, $p = 0.967$. Panel (b) of Appendix Figure 2.D4 illustrates.

Results remain qualitatively unchanged under a range of different specifications. First, I replicate the analysis above using $\hat{\beta}_i$ as continuous variable instead of an indicator of present bias. Appendix Table 2.E3 and Table 2.E4 provide coefficient estimates for savings and temptation goods spending, respectively. Second, I consider only the extensive margin, i.e. the decision to accumulate any savings or spend any money on temptation goods in any given week. I estimate Equation (2.9) using a probit model; Appendix Table 2.E5 presents estimated average marginal effects. The direction of effects remains unchanged and estimates remain statistically insignificant at any conventional level.

Present Bias and Search

Prediction 2 states that search effort is increasing in present bias parameter β. Holding everything else constant, present-biased workers should search for work less intensively. For a reduced form test of this prediction I use data from the baseline CTB experiment and my follow-up surveys. Over the whole panel of 3,041 observations, individuals report looking for work in 520 periods (390 on the job search, 130 job search while unemployed). For those 520 observations, I have detailed data on job search effort in three dimensions: Hours spent looking for work, phone calls made in order to look for work, and a subjective assessment of search intensity on a three-point scale ("not very intensively"; "intensively"; "very intensively").[35] Like for all high-frequency data in my panel, the recall period for these three measures is seven days.

Like savings data, data on job search intensity comes in the form of non-negative response variables with continuous distributions over strictly positive values and most observations at a corner of zero. To take this into account, I model the participation decision and the quantity decision jointly in a latent regression model, estimated using Tobit. I pool observations from all follow-up calls and include dummies for each panel wave. To ease interpretation of results, I use a binary measure of present bias, which takes the value 1 if $\hat{\beta}_i < 1$ and 0 otherwise.

[34]This result remains qualitatively unchanged when using the share of total consumption expenditures devoted to this group of goods (not shown).

[35]Subjects found it hard to make the subjective assessment, so I have fewer observations than for the other two measures.

The estimation problem can be written as

$$Y_{i,t} = \gamma_0 + \gamma_1 \mathbb{1}_{\{\hat{\beta}_i<1\}} + \gamma_2\, \hat{\delta}_i + \sum_{s=0}^{8} \kappa_s \mathbb{1}_{\{t=s\}} + \theta' \mathbf{X}_i + \epsilon_{i,t} \tag{2.10}$$

where for each subject i in period t, Y is measured job search effort, $\mathbb{1}_{\{\hat{\beta}_i<1\}}$ is an indicator for whether subjects are present-biased in the baseline CTB experiment, $\hat{\delta}$ is the discounting parameter from the baseline CTB experiment, $\mathbb{1}_{\{t=0,..,8\}}$ are indicators for each survey wave, $\mathbf{X}_i$ is a vector of respondent observable characteristics and indicators for the enumerator who administered the survey, and ϵ is the error term clustered at the individual level. Table 2.4 presents maximum likelihood estimates of Equation (2.10) with and without controlling for the full set of baseline covariates. As before, I report Tobit coefficient estimates.

Table 2.4: Job Search Effort and Present Bias (Regression Coefficient Estimates and Robust Standard Errors)

	(1) Calls *Tobit*	(2) Calls *Tobit*	(3) Hours *Tobit*	(4) Hours *Tobit*	(5) Subjective *Tobit*	(6) Subjective *Tobit*
Baseline $\hat{\beta}_i < 1$	-2.287**	-2.963***	-4.479**	-6.130***	-0.0149	-0.0224
	(1.016)	(0.950)	(1.965)	(1.894)	(0.148)	(0.143)
Baseline $\hat{\delta}_i$	12.22**	12.41***	22.12**	22.28***	0.685	0.401
	(5.067)	(4.216)	(9.621)	(8.593)	(0.768)	(0.675)
Survey wave dummies	Yes	Yes	Yes	Yes	Yes	Yes
Enumerator dummies	Yes	Yes	Yes	Yes	Yes	Yes
Personal characteristics	No	Yes	No	Yes	No	Yes
Baseline liquidity	No	Yes	No	Yes	No	Yes
Cognitive ability	No	Yes	No	Yes	No	Yes
Non-cognitive ability and stress	No	Yes	No	Yes	No	Yes
N	2008	1939	2008	1939	325	320
Log likelihood	-1263	-1166	-1555	-1467	-361	-327

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses, clustered at individual level. Time preference parameters trimmed at 5th and 95th percentile. Controls for personal characteristics are age, marital status, religion, mother tongue, education, number of children, whether or not respondent moved from a rural area, whether respondent has a working spouse. Controls for baseline liquidity are amount of savings held when starting employment and easy of obtaining credit assessed on a 4 point Likert scale. Control for cognitive control is the final score achieved in the baseline Stroop task. Controls for non-cognitive control and stress are baseline score on the General Self-Efficacy scale, baseline locus of control index, and baseline score on the Perceived Stress Scale (PSS-4). Appendix 2.B provides details on survey protocols and the measurement of control variables. Figure 2.D5 plots average marginal effects of the present bias indicator in column (2), (4), and (6).

I find that baseline estimates of present bias (as well as individual discounting) are statistically and economically significant predictors of subsequent job search behavior. In line with prediction 1, present-biased individuals spend significantly fewer hours looking for work and make significantly fewer phone calls. While the sign on the subjective assessment of search intensity is also in line with the prediction, I cannot reject a null effect at any conventional level of significance.

To assess the magnitude of the coefficients I calculate the average marginal effects (average partial effects) of present bias on the expected value of the observed outcome.[36] All else constant, present-biased individuals make on average 0.32 calls per week while time-consistent individuals make 0.68 calls per week, a significant difference of 0.36 calls, $\chi^2(1, N = 1,939) = 9.87$, $p = 0.0017$. Present-biased individuals spend on average 37.2 minutes on search, compared to 84.6 minutes for time-consistent individuals, a significant difference of 47.4 minutes, $\chi^2(1, N = 1,939) = 10.51$, $p = 0.0012$. Finally, present-biased individuals report a subjective intensity of search of 1.62 on a three-point scale from 1 ("not very intensively") to 3 ("very intensively"), while time-consistent individuals report an intensity of 1.6, an insignificant difference of 0.02, $\chi^2(1, N = 320) = 0.02$, $p = 0.8757$.[37] Appendix Figure 2.D5 illustrates the average marginal effects for all three outcome variables.

Results are robust to different specifications. First, I replicate the analysis with $\hat{\beta}_i$ as continuous variable instead of an indicator of present bias. Results remain qualitatively unchanged (Appendix Table 2.E6). $\hat{\beta}_i$ remains a significant predictor of search intensity measured in the number of calls and the time spent per week, though I am marginally less powered to detect effects. The results also hold on the extensive margin. Second, I consider only the extensive margin and abstract from search intensity. I estimate Equation (2.10) using a probit model for the decision to look for week in any given week. Appendix Table 2.E7 presents results. All else constant, present-biased individuals have a 11.86 percent predicted probability of looking for work, while time-consistent individuals a 20.66 percent predicted probability, a significant difference of 8.8 percentage points, $\chi^2(1, N = 1,919) = 8.84$, $p = 0.0029$.

Even though the theoretical framework is silent on the outcomes of job search, it is instructive to assess empirically whether the negative correlation between present bias and search effort also holds for the relationship between present bias and search outcomes. I focus on two outcome measures from my panel data: The number of offers generated by individuals who search and an indicator for voluntary departures from the firm.[38] The estimation problem can be written as

$$\begin{aligned} Y_{i,t} =& \gamma_0 + \gamma_1 \mathbb{1}_{\{\hat{\beta}_i<1\}} + \gamma_2\, \hat{\delta}_i + \gamma_4\, C_{i,t} + \gamma_5\, H_{i,t} + \gamma_6 I_{i,t} \\ & + \gamma_7 II_{i,t} + \sum_{s=0}^{8} \kappa_s \mathbb{1}_{\{t=s\}} + \theta' \mathbf{X}_i + \epsilon_{i,t} \end{aligned} \tag{2.11}$$

[36]For each observation, I calculate the difference in expected values at both values of the present-bias indicator while keeping all other covariates unchanged. The average difference over all observations gives the average marginal effect. Wooldridge (2010) provides an exposition of how to estimate average marginal effects of binary variables in a Tobit model.

[37]Alternatively, one can calculate the average marginal effect of present bias conditional on individuals searching for work: Present-biased individuals who search make 3.04 calls while time-consistent individuals who search make 3.54 calls, a significant difference of 0.5 calls, $\chi^2(1, N = 1,939) = 10.15$, $p = 0.0014$. Similarly, present-biased individuals who search spend 338.4 minutes per week, compared to 402.6 minutes for time-consistent individuals, a significant difference of 64.2 minutes per week, $\chi^2(1, N = 1,939) = 10.89$, $p = 0.001$.

[38]While the survey also asked for wages of job offers generated through search, respondents only reported wages for 65 offers.

where for each subject i in period t, Y is one of two search outcomes, $\mathbb{1}_{\{\hat{\beta}_i<1\}}$ is an indicator for whether subjects are present-biased (with $\hat{\beta} < 1$) in the baseline CTB experiment, $\hat{\delta}$ is the discounting parameter from the baseline CTB experiment, $C_{i,t}$ is the number of calls made in search of a job in the same period, $H_{i,t}$ is the number of hours spent on job search in the same period, $I_{i,t}$ and $II_{i,t}$ are indicators for (very) high job search intensity in the same period based on a subjective assessment, $\mathbb{1}_{\{t=0,..,8\}}$ are indicators for each survey wave, $\mathbf{X}_i$ is a vector of baseline controls, and ϵ is the error term clustered at the individual level. I estimate Equation (2.11) separately for each of two outcome measures. When the outcome variable is the number of offers generated in the same period, I use a Poisson regression. When the outcome variable is an indicator for a voluntary departure in the same period, I use a probit model. Appendix Table 2.E8 presents maximum likelihood estimates with the same set of controls as in Table 2.4. For the probit model I report average marginal effects. For the poisson model I report incidence-rate ratio estimates.

Two results are worth highlighting. First, baseline present bias is associated with fewer job offers generated. This relationship, however, only significant when controlling for job search effort in the same period. Second, present-biased subjects are significantly more likely to depart from the firm voluntarily. This relationship holds when controlling for contemporaneous job search effort.

Present Bias and Turnover

Prediction 3 states that the survival rate at the firm is decreasing in present bias β. In the theoretical framework, this is because the probability of leaving the firm depends on the probability of receiving an alternative wage offer, and thus directly on search effort. Given that present bias undermines search, present-biased individuals should exhibit a higher rate of survival at the firm.

As a first step, I graphically assess exit rates from the firm. I estimate separate Kaplan-Meier survival functions for present-biased workers with $\hat{\beta} < 1$ and time-consistent workers with $\hat{\beta} \geq 1$ (Figure 2.4). Visual inspection of the survival estimates suggests that present-biased workers have a higher rate of survival at the firm in line with prediction 3.

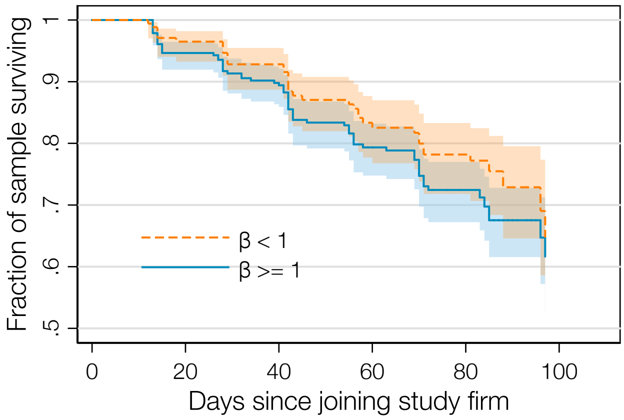
For a more rigorous test that controls for other covariates, I use the Cox (1972) partial likelihood method for the proportional hazard model. Consider the hazard that worker i leaves the firm after t days of work as

$$h_{i,t} = h_0(t)\ \exp\left(\gamma_1 \mathbb{1}_{\{\hat{\beta}_i<1\}} + \gamma_2\, \hat{\delta}_i + \gamma_3\, C_{i,t} + \gamma_4\, H_{i,t} + \theta' \mathbf{X}_i\right) \tag{2.12}$$

where $h_0(t)$ is the baseline hazard, $\mathbb{1}_{\{\hat{\beta}_i<1\}}$ is an indicator for whether subjects are present-biased (with $\hat{\beta} < 1$) in the baseline CTB experiment, $\hat{\delta}$ is the discounting parameter from the baseline CTB experiment, $C_{i,t}$ is the number of calls made in search of a job in the same period, $H_{i,t}$ is the number of hours spent on job search in the same period, and $\mathbf{X}_i$ is a vector of baseline controls. I do not include subjective assessments of job search intensity due to the small number of observations. The main focus of the analysis is parameter γ_1, which measures the correlation of baseline present bias with the hazard of leaving the firm after t days.

Table 2.5 presents maximum likelihood estimates of the exponentiated coefficients (hazard ratios) from Equation (2.12). Columns 1 and 2 only include time preference parameters

Figure 2.4: Kaplan-Meier Survivor Function of Staying at the Firm, by Present Bias Parameter



Notes: Shaded area shows 90 percent pointwise confidence band.

with and without the same set of controls as above, columns 3 and 4 examine the role of savings and job search in isolation. Columns 5 and 6 combine time preferences and job search effort. My preferred specification for a reduced-form test of prediction 3 is presented in column 2.

In line with the theoretical prediction, present-biased individuals have a lower hazard of leaving the firm. The result is statistically significant and economically large. All else equal, the risk of leaving the firm is 52.3 percent as high for present-biased individuals as it is for individuals who are not present-biased. If this effect operates through search effort, as it does in the theoretical framework, we would expect increased search effort to lead to a higher hazard of leaving the firm. We would also expect that the effect of present bias on the hazard of leaving should be less pronounced when controlling for search effort. Both appear to be the case. Search effort measured in hours and in the number of phone calls indeed significantly increases the hazard of leaving the firm (column 4). One more phone call per week increases the hazard of leaving the firm by 20.1 percent. When including both time preferences and search effort, the coefficient on baseline present bias becomes insignificant.

2.5.3 Alternative Explanations

The analysis above already included a number of robustness checks and showed that results are robust to the inclusion of a broad range of observable characteristics. In this section, I discuss alternative explanations.

2.5.3.1 Do Individual Characteristics, Liquidity, and Environmental Factors Explain Experimental Responses?

A focus of this study was to elicit time preferences in a tightly-controlled experiment that eliminates confounders commonly found in the literature (see discussion in Section 2.4.2). Nevertheless, the correlations presented above may be biased if experimental responses are

Table 2.5: Hazard of Leaving the Firm (Cox Hazard Ratios and Robust Standard Errors)

	(1)	(2)	(3)	(4)	(5)	(6)
Baseline $\hat{\beta}_i < 1$	0.554**	0.516***			0.815	0.741
	(0.129)	(0.124)			(0.228)	(0.236)
Baseline $\hat{\delta}_i$	18.162***	9.907**			3.743	2.104
	(18.353)	(11.485)			(4.592)	(3.195)
Search effort, last 7 days (hours)			1.060***	1.078***	1.076***	1.105***
			(0.009)	(0.018)	(0.009)	(0.016)
Search effort, last 7 days (# calls)			1.191***	1.206***	1.183***	1.182***
			(0.025)	(0.041)	(0.026)	(0.043)
Enumerator dummies	Yes	Yes	Yes	Yes	Yes	Yes
Personal characteristics	No	Yes	No	Yes	No	Yes
Baseline liquidity	No	Yes	No	Yes	No	Yes
Cognitive ability	No	Yes	No	Yes	No	Yes
Non-cognitive ability and stress	No	Yes	No	Yes	No	No
N	1830	1767	2267	2195	1792	1732

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Table shows exponentiated coefficients. Robust standard errors in parentheses, clustered at individual level. Time preference parameters trimmed at 5th and 95th percentile. Subjective assessments of job search effort are not included as independent variable due to the small number of observations. Controls for personal characteristics are age, marital status, religion, mother tongue, education, number of children, whether or not respondent moved from a rural area, whether respondent has a working spouse. Controls for baseline liquidity are amount of savings held when starting employment and easy of obtaining credit assessed on a 4 point Likert scale. Control for cognitive control is the final score achieved in the baseline Stroop task. Controls for non-cognitive control and stress are baseline score on the General Self-Efficacy scale, baseline locus of control index, and baseline score on the Perceived Stress Scale (PSS-4). Appendix 2.B provides details on survey protocols and the measurement of control variables.

systematically driven by personal characteristics, liquidity,[39] or environmental factors outside the experiment. To assess if this is the case, I regress an indicator for estimated present bias ($\hat{\beta}_i < 1$) and the estimated present bias parameter $\hat{\beta}_i$ on a range of observable characteristics.

Appendix Table 2.E9 presents regression coefficient estimates from probit and ordinary least squares (OLS) regressions. I find that estimates of present bias are not significantly predicted by most personal observable characteristics. Importantly, neither human capital measured by previous work experience and education, nor access to finance predict estimated present bias.[40]

[39]Cassidy (2018) uses the MPL approach to show that present bias elicited from poor subjects may in fact represent a rational, time-consistent response to liquidity constraints.

[40]Two additional empirical findings are worth highlighting. First, respondent age is negatively correlated with the estimated present bias parameter $\hat{\beta}$, but not a significant predictor of an indicator of present bias. This stands in contrast to Meier and Sprenger (2010), who find that age is a negative predictor of a present bias indicator. Second, higher cognitive control as measured by the spatial Stroop-like task is a significant positive predictor of the estimated present bias parameter $\hat{\beta}$. This is consistent with findings in a recent literature that has investigated the link between cognitive control, or *bandwidth* more generally, and time preferences (Haushofer & Fehr, 2014; Mani, Mullainathan, Shafir, & Zhao, 2013; Schilbach, Schofield, & Mullainathan, 2016).

In addition to the correlations presented here, I can use randomized experimental payouts from the baseline experiment to assess whether liquidity drive time preference estimates from the endline experiment three months later. Recall that in the CTB experiment, individuals have a 50 percent chance of winning their payouts, determined with a coin flip at the end of the experiment. At 20 to 40 percent of the monthly wage, experimental payoffs represent a large cash drop that provides exogenous variation to liquidity. I regress an indicator for estimated endline present bias ($\hat{\beta}_i < 1$) and the estimated endline present bias parameter $\hat{\beta}_i$ on an indicator of whether respondents won their experimental payout at baseline and a range of observable characteristics.

I do not find evidence that experimental responses at endline change as a result of winning the sizable experimental payout at baseline. Appendix Table 2.E10 presents regression coefficient estimates from probit and OLS regressions. While the results indicate that liquidity do not appear to have affected endline present bias, individual liquidity three months after the beginning of the study is likely different than at the beginning of the study.

2.5.3.2 Does Liquidity Explain Search Effort?

In the previous subsection, I provided evidence that individual liquidity does not appear to explain experimental responses. In this subsection, I assess whether financial wealth and liquidity affects search effort.

While my theoretical framework holds initial wealth constant, the effect of liquid savings on search effort is theoretically ambiguous. Lentz and Tranæs (2005) show analytically and with simulations that job search effort is negatively related to initial wealth under the assumption of additively separable utility. This negative relationship is consistent with empirical results in which unemployment spells are positively correlated with initial wealth holdings, e.g. in Algan et al. (2003) and Bloemen and Stancanelli (2001). In my context, however, it is also possible to imagine that liquidity-constrained individuals are not able engage in costly search. If liquidity-constrained subjects were to appear more present-biased, as in Cassidy (2018) but in contrast to the evidence presented above, and if liquidity-constrained subjects also searched less, this would explain the correlational patterns that I find above.

I can assess the link between liquidity and search using randomized experimental payouts from the baseline CTB experiment. I do so in two steps. I first show that subjects who won the experimental payouts and those who did not appear balanced on observable characteristics (Appendix Table 2.E11). There is no significant difference between both groups, except that slightly more participants who are native Oromiffa speakers won the coin flip.

In a second step, I use the same latent regression model as in Equation (2.10) to study the correlation of present bias and search effort, but add an indicator for whether the subject won or lost the coin flip. Given that participants appear balanced on observables, I interpret the coefficient estimate for this indicator as the causal effect of experimental payouts on search. Appendix Table 2.E12 presents maximum likelihood estimates of Equation (2.10) with and without controlling for the full set of baseline covariates. As before, I report Tobit coefficient estimates.

Two findings are worth highlighting. First, winning the experimental payouts causes significantly less search. Second, the significant negative correlation between measured present bias and search effort holds up. The results are consistent with a negative relationship between liquidity and search effort. Subjects with more liquidity search less. Taken together, this suggests that liquidity constraints are unlikely to explain my results above.

2.5.3.3 Does Human Capital Explain Search Effort?

Individuals tend to choose job search effort in response to economic incentives. In particular, job search effort has been found to increase in the expected returns to search (Christensen, Lentz, Mortensen, Neumann, & Werwatz, 2005). As a result, one would expect that workers with higher earnings potential, i.e. larger human capital, are more likely to search or search more intensively. While I already showed that proxies of human capital do not appear to predict measured present bias, it is worth investigating more explicitly if measures of human capital are positively correlated with search effort.

I run a Tobit regression of job search effort (measured in number of calls and time spent on search) on respondent age, an indicator for whether the individual has any formal work experience, a set of indicators for the highest education level completed, and measures of cognitive control and non-cognitive skills.[41] Appendix Table 2.E13 provides regression coefficient estimates. Age, which I control for in all results above, is the only significant predictor of search effort. Without controlling for cognitive control and non-cognitive skills, education is a significant predictor of job search only for those individuals who have completed more than grade 10 schooling.

2.5.3.4 The Role of Reservation Wages

My theoretical and empirical analysis so far has abstracted from workers' reservation wage choices. In the model of DellaVigna and Paserman (2005), individuals with a higher (exponential) discount factor set a higher reservation wage while present bias should essentially be orthogonal to the reservation wage. My data allows for a reduced-form test of this prediction. At each round of the panel, I record self-reported reservation wages using a similar question to the one used by Krueger and Mueller (2016) and in the May 1976 US Current Population Survey.[42]

Appendix Table 2.E14 reports coefficient estimates from an OLS regression of log self-reported reservation wage on baseline time preference parameters and the same set of controls as before. Neither present bias nor the estimated discount parameter are significant predictors of reservation wages. The findings are in line with the empirical results of DellaVigna and Paserman (2005) and Krueger and Mueller (2016), who do not find evidence that time preferences affect the choice of the reservation wage. This indicates that the reservation wage choice is unlikely to play a large role in the context of this study.

[41] Abebe et al. (2016) report that young job seekers in Addis Ababa may find it hard to signal their skills and, as a result, firms often use criteria such as whether workers have any previous work experience.

[42] The question text, translated from Amharic, was: "Suppose someone offered you a job today. What is the lowest monthly pay after taxes that you would accept for the type of work you were looking for?" The question in the 1976 CPS was: "What is the lowest wage or salary you would accept (before deductions) for this type of work?."

2.5.3.5 Reliability of Self-Reported Data on Search Intensity

The analysis above hinges on self-reported data on job search effort. The little existing work that analyzes high-frequency data on job search behavior either uses self-reported survey data similar to mine (Faberman, Mueller, Şahin, & Topa, 2017; Krueger & Mueller, 2016), observational data under highly controlled conditions (Belot, Kircher, & Muller, 2018) or administrative data from online job boards (Faberman & Kudlyak, 2018). In the setting of my study, self-reported data is the only feasible option. It is worth asking if this data is reliable.

First, it is important to note that if questions on job search intensity are affected by experimenter demand or Hawthorne effects, these effects would only confound my results if they are systematically correlated with results from the experimental elicitation of time preferences. Experimenter demand effects occur when respondents systematically alter their answers based on what they believe constitutes desirable or appropriate behavior (Zizzo, 2010). Neither the individuals participating in the experiment nor the team that implements the survey is aware of the research questions. All throughout the experiment, I take care to not provide cues to the respondents. Overall, it is not obvious why individuals categorized as present-biased would report systematically lower job search intensity.

Second, because individuals likely find it difficult to exactly quantify how much they look for work in a given week, I use three different measures of intensity (Cronbach's α = 0.733). I show results separately for each dimension. Results also hold when using an aggregate measure of all three dimensions generated from factor analysis where I retain the first factor.[43] While all three measures likely suffer from measurement error, it is improbable that this measurement error is systematically correlated with results from the experimental elicitation of time preferences.

Third, the results on turnover do not require self-reported data on search. Firm personnel records indicate that workers truthfully report tenure. It is not clear how the results on tenure could be rationalized in the absence of a search channel.

Taken together, it appears unlikely that systematic biases in search intensity data are a significant driver of the results presented above.

2.6 Conclusion

Policymakers in Ethiopia and other low-income countries have promoted labor-intensive light manufacturing as an opportunity to generate a large number of formal employment opportunities. For the low-skill rural-urban migrants in this study, industrial work represents a stepping stone into the formal labor market of Addis Ababa. My results suggest that self-control problems in the form of present bias significantly undermine the ability of workers to use these jobs as such a stepping stone.

I show that present bias is a significant predictor of reduced job search effort over a period of three months after starting a low-skill industrial job in peri-urban Addis Ababa. Present-biased workers search less and – as a result – generate fewer alternative job offers, and stay at the firm significantly longer. My results offer the first experimental evidence of

[43] Not shown, results available upon request.

a theorized link between present bias and job search effort. I do not find evidence for a link between present bias and reduced savings.

An immediate implication of my findings is that individuals looking for work might benefit from policies or devices that commit their future selves to more search. Whether and under what conditions such a commitment device can be welfare-improving depends on the exact welfare criterion, which is not obvious to define when we observe two individual choices that are in conflict with each other. It is particularly difficult in the context of this study, where individuals likely have imperfect information about their own future efficiency in searching for work. It is easy to imagine how individuals could under-estimate how physically and mentally demanding their new job will be and thus over-estimate how easy it will be for them to set aside time to search for alternatives.[44]

What could a potential commitment device for job search look like? For a discussion, it is useful to consider the difference between hard commitment devices, which involve real economic costs, and soft commitment devices, which mainly work through psychological costs (Bryan et al., 2010). Search effort is hard to monitor directly in the study context, so it is difficult to imagine how the market could provide a hard commitment device that directly contracts on workers' search effort or search outcomes. Through focus group interviews with workers in my sample, I identified three feasible alternatives.[45]

A first option would be a soft commitment device in the form of a personal plan for job search. With such a plan, workers could formalize intentions for search and possibly set individual rules for how much to search, when to search, and how to search. This plan could be implemented through a small notebook that workers can carry. This type of soft commitment device builds on a recent literature in economics and psychology, which has found that prompting people to form concrete implementation plans can increase follow-through (Beshears, Milkman, & Schwartzstein, 2016; Milkman, Beshears, Choi, Laibson, & Madrian, 2011).[46] Abel, Burger, Carranza, and Piraino (2017) show that prompting unemployed South African job seekers to make a plan increases the number of applications and diversified search strategies. Because the plan is unenforced, it avoids the challenge of monitoring search effort. Three months after starting their job, 96 percent of individuals in my sample indicated that they would be interested in such a planning device. 74 percent indicated a positive willingness to pay in an unincentivized question.[47]

A second alternative would be a soft and informal commitment device that operates through social pressure in small groups of women. Individuals indicated that beyond social pressure, they would benefit from exchanging knowledge about search in smaller groups. In the context of savings, Kast, Meier, and Pomeranz (2018) test the impact of peer groups with publicly-announced goals and peer monitoring on individual savings. They find large effects on the number of deposits and the savings balance of individuals. 95 percent of

[44] A related design challenge derives from the balance between flexibility and commitment (Amador, Werning, & Angeletos, 2006). For example, an individual taking up a commitment savings plan could demand a provision to withdraw from the contract in case of a medical emergency. The conditions for flexibility are difficult to specify in the context of this study.

[45] Detailed notes are available upon request.

[46] See Bénabou and Tirole (2004) for a theoretical perspective from economics on how unenforced personal plans can serve as a commitment device.

[47] In a companion paper, we implemented a plan that aimed at increasing worker savings, not search. We provide a full evaluation in Hardy et al. (2018).

individuals in my sample indicated their interest in such groups, 69 percent indicated a positive (unincentivized) willingness to pay.

A third option would be a hard and formal commitment device that operates through advance payments to a job broker. 9 percent of workers currently rely on job brokers to assist with search. A formal commitment device could involve an upfront payment to a broker or a fixed fraction of the monthly wage to be paid to job broker. 58 percent of individuals in my sample indicated their interest in such an arrangement.

Overall, my results suggest several avenues for future work. First, future research could use exogenous variation to establish a causal link between present-biased preferences and job search effort. A field experiment that tests potential commitment devices for job search, building on the list above, is a natural starting point. This will require addressing difficult questions around welfare implications. Second, the study of commitment prompts the question of whether workers are aware of their self-control problems or not. The empirical analysis above did not consider to what extent workers are aware of their own self-control problems. Third, the null result on savings behavior merits further analysis.

2.A Appendix: Theoretical Framework

2.A.1 Functional Form Specification

To simulate the model from Section 2.3, I make additional functional form assumptions in line with the previous literature. As in Paserman (2008) and DellaVigna et al. (2017) I use search costs of the form $k(s) = \psi s^{1+\gamma}/(1+\gamma)$ with $\gamma > 0$, where γ determines the convexity of the cost function and ψ is a scaling parameter.[48] ψ is equivalent to the cost of obtaining the outside option job with probability one. I set the curvature parameter of the search cost function $\gamma = 0.4$, taken from Paserman (2008).

I assume that workers exhibit CRRA and derive utility from consumption in the form of $u(c_t) = log(c_t)$.

I use survey data from my experiment to calibrate the remaining parameters of the model. Income at $t = 0$ before joining the safety net firm is set equal to the median cash income of 458 birr (US$16.61).[49] Wages at the safety net firm then follow the wage regime of the firm in my experiment: Workers receive 1,000 birr (US$ 36.30) in the first month, 1,075 birr (US$ 39.02) in the second month, and 1,150 birr (US$ 36.30) in all months after that. The income of workers who leave the firm after 12 months drops back the the pre-employment level of 458 birr. The outside option wage $\widetilde{w}$ is set to equal to 1,835 birr (US$ 66.61), the median wage offer generated by workers after joining the experimental firm.

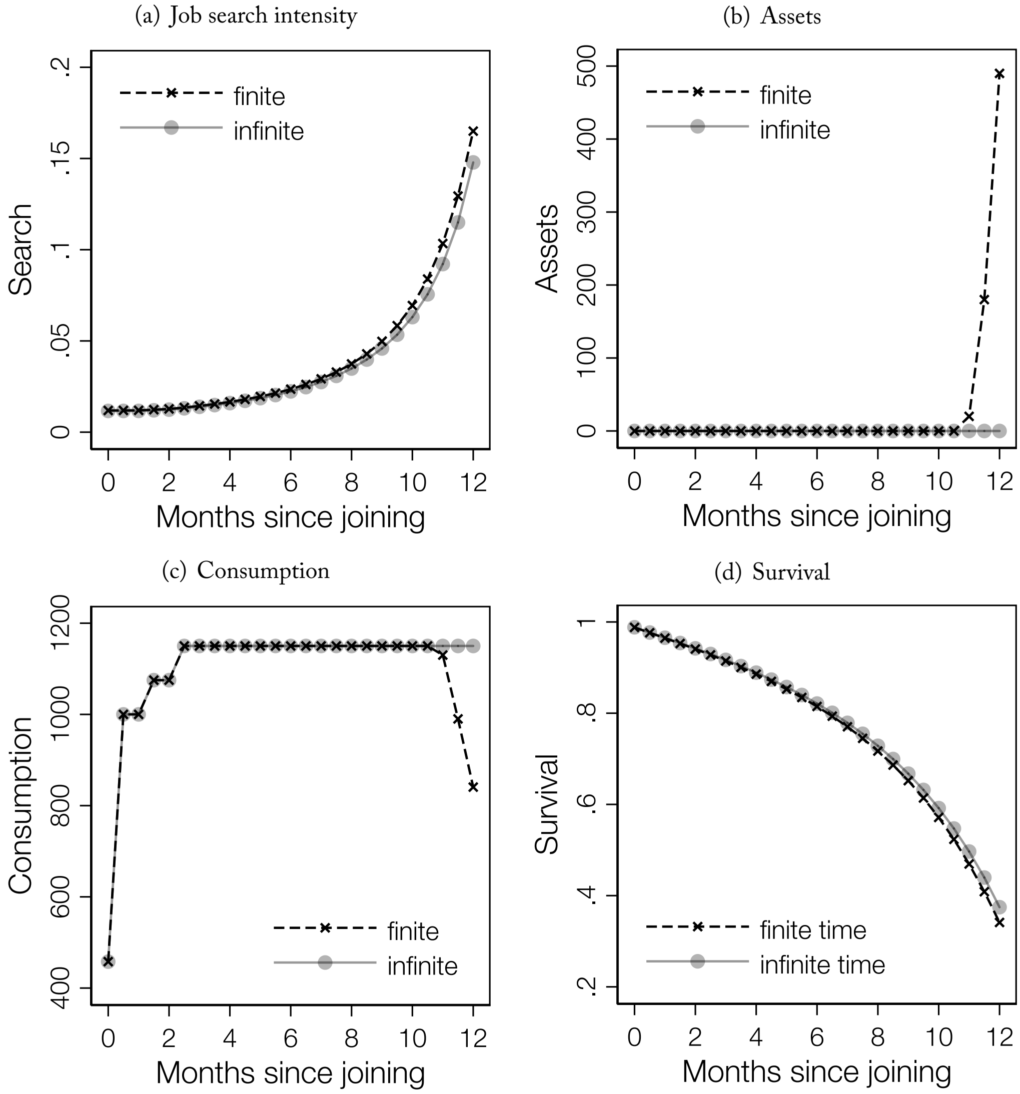
One period in my simulation represents 15 days. The discount factor δ is based on aggregate estimates from a CTB experiment with workers in my sample, conducted on the day before they start their job. My preferred estimate for discounting over 15 days is $\hat{\delta} = 0.92$ (SE 0.008). Assets pay a 15-day return of 0.003 based on the Ethiopian deposit interest rate of 7 percent per year.

[48] As noted by DellaVigna et al. (2017), γ is equal to the elasticity of search effort with respect to the net value of employment. To see this denote $\left[V^O_{t+1}(A_{t+1}) - V^F_{t+1}(A_{t+1})\right] = \Phi$. Using the first order condition with respect to search effort (2.4), rewrite $c'(s^*) = \Phi$. Plugging in the cost function above, we obtain $s^* = (\Phi/\psi)^{1/\gamma}$.

[49] The median wage income over the four weeks before joining the firm is 0 birr. 458 birr is the total cash income including transfers from family members and friends.

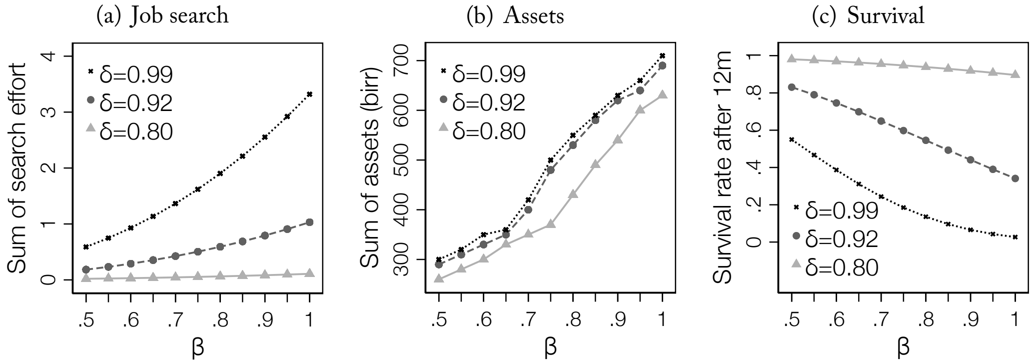
2.A.2 Additional Simulation Results

FIGURE 2.A1: Optimal Path of Search and Assets (Panels a and b) and Implied Consumption and Survival (Panels c and d)



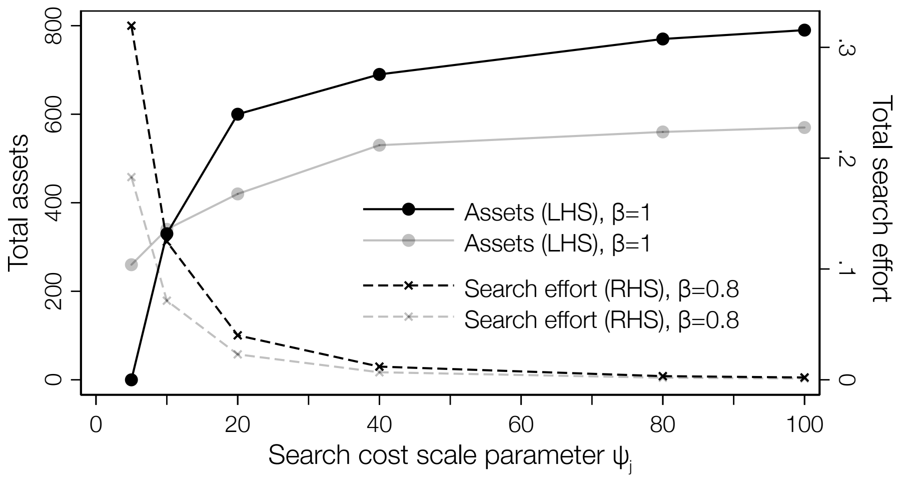
Notes: Assuming exponential discounting ($\beta = 1$). Other parameters are set as described in the text.

Figure 2.A2: How Discounting and Present Bias Affect Job Search, Asset Accumulation, and Survival (Simulation Results, by Present Bias and Discount Parameter, Finite Case)



Notes: Simulation results for the "finite" case in which workers know that they will leave the safety net firm after one year, which is equivalent to the median expected tenure in my survey data. Panels (a) and (b) represent the sum of search effort and assets over all simulated periods. Parameter values as given in text.

Figure 2.A3: How Search Cost Affects Search Effort and Asset Accumulation (Simulation Results, by Search Cost Type and Present Bias)



Notes: Sum of search effort and assets over all simulated periods. Based on simulations allowing for heterogeneity in job search costs and assuming finite tenure at the firm of one year.

2.B Appendix: Procedures and Methods for Survey Data Collection

2.B.1 Survey Procedures

2.B.1.1 Randomization

Randomization is done using a combination of Excel spreadsheets in the survey office, adhoc randomization by the survey software, and a coin flip by the respondent.

Sample selection: Every evening, the field coordinator receives a list with names and contact information of the workers that were hired on that day and that are scheduled to return on the next day to begin employment. This list is entered into an Excel spreadsheet and shuffled in random order using the built-in random number generator. The random order defines whether respondents are approached for an interview at the factory, an interview at home, or are part of a control group that is not interviewed. The randomized list is distributed to all enumerators via email and hard copy. In one case the enumerator did not comply with the survey assignment and the observation was dropped.

Survey and time preference order: The questionnaire modules were administered in the same order to reduce complexity for the enumerator. The three sets of the CTB experiment are presented in random order based on a random number generated by the survey software.

Incentives from time preference elicitation: Subjects have a 50 percent chance of receiving one out of the 15 CTB allocations that they make. Whether or not respondents win is determined using a coin flip at the end of the experiment. The coin is thrown by the respondent under supervision of the enumerator. This was done to maximize transparency for respondents. If respondents win the coin flip, they are asked to make a draw from a bowl with folded cards numbered 1 to 15 that represent the choices of the experiment. Enumerators then enter the drawn number into the survey software.

2.B.1.2 Survey Protocols

Each respondent is randomly assigned to one of 15 enumerators, all but three of which female. Randomization to enumerators is done using the document described in the previous section. All enumerators have several years of work experience in data collection with the Ethiopian Development Research Institute (EDRI) or the Ethiopian Central Statistics Agency (CSA). The survey field coordinator and I conducted six days of study-specific training with all enumerators. In addition, all enumerators conduct training interviews with real respondents for one week. This data is discarded and enumerators are brought back for another day of training and feedback. Care is taken to minimize observer bias and enumerators do not know the specific hypotheses being tested. Training materials are available from the author upon request.

Each respondent is paired with the same enumerator throughout the study to maximize trust and reduce attrition. Informed consent, all interviews, and the time preference elicitation are administered in Amharic or Oromiffa, the most commonly spoken local languages. 67 percent of respondents in the sample report Amharic as their mother tongue, another 20 percent report Oromiffa as their first language. All workers at the firm are required to speak Amharic, but 4.8 percent of the sample still prefer to conduct interviews in Oromiffa. Enumerators are trained in both and are instructed to refuse consent if workers are not comfortable in either of the two languages.

All survey questions were carefully translated from English to Amharic. Where available, the translation was compared to official translations of survey instruments by the CSA. The survey team discussed each question in English and Amharic as a group to make sure that the meaning is correctly translated and understood by all team members. In addition, an Amharic native speaker who was not involved in the study reviewed the translations.

Data for the baseline survey, time preference elicitation, treatments, high-frequency phone survey, and endline survey is collected using CAPI with Android tablets running Open Data Kit / SurveyCTO. The interview GPS location and a randomly-selected 10-second audio segment of each interview are recorded for auditing purposes.

The data presented in this chapter was collected from March 27 to September 28, 2018. In-person interviews are conducted seven days a week, Monday through Saturday at the firm and on Sunday in the homes of respondents. On average, the team of enumerators conducts 7 baseline interviews per day and 29 per week. Phone interviews are conducted every day, mostly in the evening when workers have returned from work. On average, enumerators conduct 108 phone interviews per week.

2.B.1.3 Ensuring Confidentiality

Workers may fear retaliation by the firm for example if they report that they are not planning to stay for long or that they dislike the working conditions. They may also feel the need to respond in a way that is socially or otherwise desirable. In addition to addressing these concerns during the informed consent procedures, I take a number of precautions to alleviate these concerns and maximize respondent privacy.

First, I select a random subset of workers to be interviewed at home instead of at the factory. These interviews follow the exact same protocol as the interviews in the factory, but they happen on Sundays in the privacy of the respondent's home or a safe space in the community. If workers systematically conceal the truth in work-related questions during interviews at the factory, interviews at home should give me a sense of the size and sign of the bias. Similarly, if workers interviewed at home systematically conceal the truth in personal questions (for example about intra-household allocation decisions or gender norms), interviews at the factory should help me assess the bias.

Second, the physical interview location is chosen to maximize privacy. In the factory, baseline interviews are conducted in a cafeteria not visible to other firm staff. For interviews at home, enumerators offer to meet respondents on the property of the local church – a location that is commonly seen as a safe space in the community.

Third, the team of enumerators is instructed to keep their distance from firm management. When working on factory premises, enumerators wear ID badges that identify them as not belonging to the firm.

2.B.2 Elicitation of Subjective Expectations

Throughout the survey when asking subjects to assess various quantities and subjective expectations I use beans as visual aids. Delavande, Giné, and McKenzie (2011) review studies that have used such visual aids and discuss advantages and disadvantages of various methods. In particular, I follow Delavande and Kohler (2009) in explicitly linking beans to probabilities.

My instructions read: *"I want to ask you one question about the chance or likelihood that a certain event is going to happen. There are 10 beans in the bowl [show bowl]. I would like you to choose some beans out of these 10 beans and put them in the empty bowl to express what you think the likelihood or chance is of a specific event happening. One bean represents one chance out of 10. If you do not put any beans in the bowl, it means you are sure that the event will NOT happen. As you add beans, it means that you think the likelihood that the event happens increases. For example, if you put one or two beans, it means you think the event is not likely to happen, but it is still possible. If you put 5 beans, it means that it is just as likely it happens as it does not happen (fifty-fifty). If you pick 6 beans, it means the event is slightly more likely to happen than not to happen. If you put 10 beans in the plate, it means you are sure the event will happen. There is not a right or wrong answer, I just want to know what you think".*

I use their method when asking respondents to assess the probability of reaching their self-set savings goal or their subjective job finding probabilities.

2.B.3 Measures of Self-Regulation, Stress, and Well-being

To measure constructs in the areas of self-regulation and stress, I use measures validated as part of the Science of Behavior Change (SOBC) framework.[50] In particular, I build on Esopo et al. (2018), who have adapted and validated psychological scales that measure self-efficacy and executive control with laboratory subjects in Kenya.

2.B.3.1 Cognitive Control

Cognitive control – sometimes called executive control – is a broad construct in cognitive neuroscience that refers to the processes that organize information for goal-driven decision-making (Mackie, Van Dam, & Fan, 2013). Cognitive control affects how well we can control our impulses and our working memory, and how well we can formulate, maintain, and execute plans and goals. From the viewpoint of economics, Mullainathan and Shafir (2013) see cognitive control as one key component of what they call "bandwidth," that is the mental capacity required to engage in what is sometimes termed "System 2" thinking. In contrast to quick and intuitive System 1 thinking, decision-making under System 2 is "slow, forgetful, deliberate, and costly but typically produces more unbiased and accurate results" (Schilbach et al., 2016, pg. 435). Esopo et al. (2018) provide a summary recent contributions in economics.

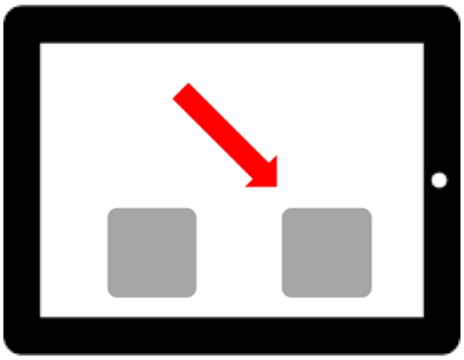
The context of the study, goal-directed behavior of workers that engage in physically demanding and repetitive industrial work with long hours, suggests that bandwidth is an important factor to take into account.

To measure cognitive control I adapt a Stroop-like arrows task for use in the field (Baldo, Shimamura, & Prinzmetal, 1998). Closely following Esopo et al. (2018), subjects are shown red and blue arrows and must press a gray rectangle either in the same direction of the arrow when it is red or the opposite direction direction of the arrow when it is blue. Arrow direction (left or right) and color (red or blue) are randomized. This task is preferable over

[50] SOBC is a large US National Institutes of Health (NIH) program, which aims to improve our understanding of behavior change across a broad range of (mostly health-related) behaviors. SOBC maintains a measures repository at scienceofbehaviorchange.org/measures.

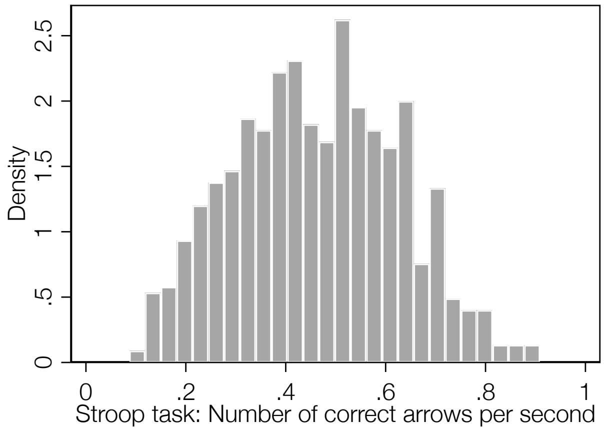
measures that use numbers because it does not require literacy. Subjects complete 20 arrows as quickly as possible on an Android tablet computer that is also used to administer the survey questionnaire. Figure 2.D1 illustrates the tablet screen during the test.

Figure 2.D1: Example Screen of Cognitive Control Measure (Red Arrow: Touch Same Side)



I define correct answers per second as measure of executive control. The task is fully incentivized. Subjects are given 3 birr per correct answer. 1 birr is subtracted per second. The minimum payoff is 0 birr. Figure 2.D2 plots the distribution of scores.

Figure 2.D2: Cognitive Control Task: Histogram of Scores



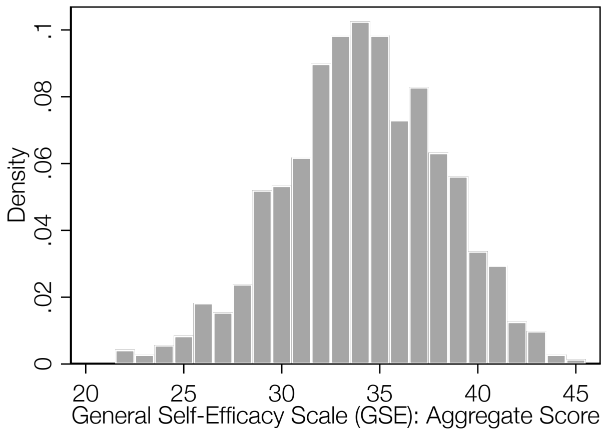
2.B.3.2 Perceived Self-Efficacy

Perceived self-efficacy broadly refers to an individuals belief in his or her own ability to perform well in a specific situation (Bandura, 1997). I hypothesize that self-efficacy could affect the ability of study participants to follow through on their plans. Esopo et al. (2018) review recent evidence that low measures of self-efficacy are correlated with low adherence to exercise regimes and health behaviors.

To measure perceived self-efficacy, I use an adapted version of the General Self-Efficacy (GSE) scale (Schwarzer & Jerusalem, 1995), which is designed to "assess a general sense of perceived self-efficacy with the aim in mind to predict coping with daily hassles as well as adaptation after experiencing all kinds of stressful life events." I use the 12-item scale from the SOBC repository. Responses are anchored on a four-point scale with responses ranging from 1 (strongly disagree) to 4 (strongly agree). The final score is calculated by adding all items, which yields a scale with a range from 12 to 48.

The GSE scale has been successfully used in across different cultural contexts (Luszczynska, Scholz, & Schwarzer, 2005), but to the best of my knowledge not in Ethiopia. I translated and back-translated the English scale to Amharic and piloted it with a focus group before deployment in the baseline survey. Figure 2.D3 plots the distribution of scores.

Figure 2.D3: GSE Measure: Histogram of Scores



2.B.3.3 Locus of Control

Locus of control is a concept from personality psychology. Individuals with strong internal locus of control tend to believe that events in their lives are based on their own decisions, actions, and behaviors. People with an external locus of control believe that events in their life are beyond their control.

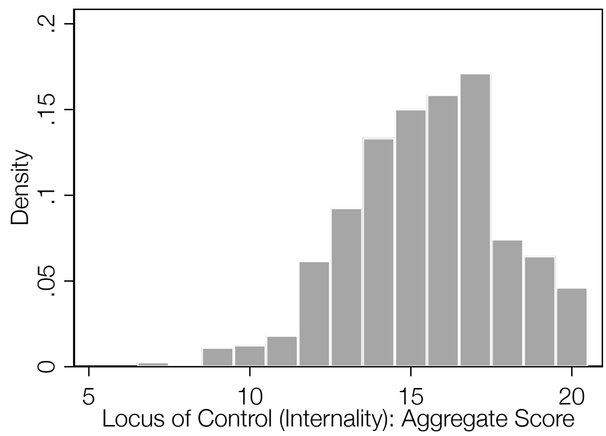
Locus of control has been hypothesized to affect job search behavior. Locus of control could impact an individual's subjective assessment of her own ability to influence job search outcomes (Caliendo, Cobb-Clark, & Uhlendorff, 2015; Falk, Huffman, & Sunde, 2006a, 2006b; McGee & McGee, 2016). To measure locus of control, I use five items from the Internality, Powerful Others, and Chance (IPC) scale developed by Levenson (1981), which is commonly used in applied work. Responses are anchored on a four-point scale ranging from 1 (strongly disagree) to 4 (strongly agree). The final score is calculated by adding all items, which yields a scale with a range from 5 to 20.

As with the GSE scale, I translated and back-translated the English scale to Amharic and piloted it with a focus group before deployment in the baseline survey. Figure 2.D4 plots the distribution of scores.

2.B.3.4 Physical Health

To measure physical health, I rely on self-reported ability to perform "activities of daily living" (ADL). ADL scales are a widely used to measure health in various domains in devel-

Figure 2.D4: Locus of Control Measure: Histogram of Scores



oping and developed countries, originally by clinicians to assess fitness for work, eligibility for disability insurance, or claims for accidents and injuries (McDowell, 2006), and more recently in development program evaluation (Thomas & Strauss, 2007).

ADL scales are preferably over measures that are endogenous to the labor supply decision such as sick days, but come with all the problems of self-reported scales including different interpretation of questions by respondents, endogeneity of self-perceived health to the work experience of respondents, and potential experimenter demand effects.

I create an additive scale of four ADL measures used by Blattman and Dercon (2018) in context of Ethiopian manufacturing workers: walk for 2 kilometers, work outside on your feet for a full day, carry a 20 liter carton of water for 20 meters, and standing at a workbench for 8 hours. Each of the four measures is scored on a four-point scale from 1 (unable) to 4 (easily).

2.B.3.5 Psychological Well-Being

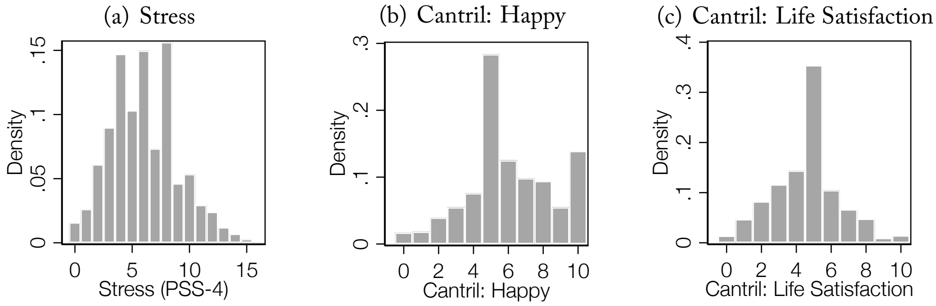
There is evidence that the psychological consequences of poverty can lead to stress and negative affect, which in turn can influence decision-making (Haushofer & Fehr, 2014). To assess if stress and negative affect – feeling unhappy or anxious – are potentially confounding my results, I use two measures.

To measure stress, I follow Haushofer and Shapiro (2016) and use the Perceived Stress Scale by Cohen, Kamarck, and Mermelstein (1983). While their original scale contains 14 items, I use the same four items as Haushofer and Shapiro (2016). Respondents are asked how often they felt in certain ways. Answers are anchored on a five-point scale with responses ranging from 0 (never) to 4 (very often). The final score is calculated by adding all items, which yields a scale with a range from 0 to 16.

To measure happiness and life satisfaction, I use Cantril's Self-Anchoring Scale (Cantril, 1965), which is commonly used in applied work and global opinion surveys such as Gallup's World Poll (Kahneman & Deaton, 2010). The scale asks respondents to imagine a staircase or ladder with numbered steps, where the top of the ladder represents the best possible or happiest possible life. Respondents are then asked to place themselves on one step of the ladder ("The top of the ladder represents the best possible life for you, and the bottom of the ladder represents the worst possible life for you. On which step of the ladder would you say you personally feel you stand nowadays?") In the survey, enumerators show respondents a picture of a ladder on the tablet to aid visualization.

Figure 2.D5 plots the distribution of scores for all three measures.

FIGURE 2.D5: Psychological Well-Being: Histogram of Scores



2.B.4 Construction of Other Survey Measures

Assets: Asset indices are commonly used in welfare and poverty analyses. They are particularly useful to assess living standards in settings where expenditure and income data is absent or unreliable. This is often the case in urban or peri-urban areas like the study setting, where consumption expenditures may be underreported because consumption happens outside the household. To construct an index of household and individual assets, I take three steps. First, I use data from the 2015-2016 ESS to identify 13 assets that best predict nominal total household consumption expenditures. I do this using dominance analysis, an approach sometimes used in psychology that calculates the marginal contribution of each independent variable in predicting an outcome by testing all possible combinations of independent variables (Budescu, 1993). I identify the following household assets: shelf for storing goods, energy saving stove, kerosene stove, sofa set, refrigerator, electric stove, electric mitad (an appliance to prepare injera bread), radio, television, CD/DVD player, satellite dish, wristwatch, and water pump. Second, I ask each respondent for the number of these assets owned by the household where the respondent sleeps and the number of these assets owned by the respondent. I use the exact question phrasing of the LSMS survey. Third, I calculate two additive indices of all assets: One for assets owned by the household where the respondent sleeps and one for the respondent herself. Additive indices are attractive due to their simplicity. Filmer and Scott (2012) show that under most conditions the method of aggregation does not significantly affect household rankings.

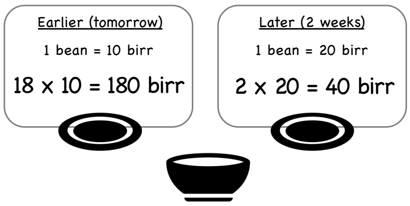
2.C Appendix: Time Preference Elicitation and Estimation

2.C.1 Implementation in the Field

Figure 2.C1: Convex Time Budget Implementation

(a) Picture of Training Interview

(b) Illustration of Whiteboards



Notes: The picture in panel (a) is taken during the training week of the enumerator pictured; data from this respondent was not used. This decision shows choice set $(1, 14)$ with interest rate $r = 1.00$. In this case, the respondent chose to allocate 18 beans to the earlier dish and 2 beans to the later dish.

2.C.2 Selected Experimental Instructions

I provide English translations of key experimental instructions. These are read by the enumerator to the respondent from a tablet screen. The tablet screen shows both the English and the Amharic version, so enumerators can go over the text with respondents. Amharic originals are available upon request.

Introduction

For the next part of our conversation, I would like to ask you to make 15 different decision about how to divide money between two different dates: "earlier" and "later". The money will be represented with beans. I will give you a bowl with 20 beans and ask you to divide the beans between two dishes: one that represents the earlier date and one that represents the later date. The beans that you allocate to "later" will always be worth more than the beans you allocate to "earlier".

It's important that you listen carefully to understand how this exercise works, because you will make decisions about real money. At the end of our conversation, we will flip a coin to decide if you will be paid out one of the decisions. If you win the coin flip, we will have a lottery in which you draw one of the 15 decisions that you made. We will then send you the earlier and later payments for the decision that you draw. You will get each payment exactly on the date specified, not earlier. We will send you the money to your cell phone as CBE Birr payment [a payment system by Commercial Bank of Ethiopia]. After you received the text message you can redeem the money at any CBE branch or any CBE Birr agent. I will now give you an example so that you can better understand the exercise.

Example and Comprehension Check

Let's look at an example together. In this example, the earlier dish represents the amount of money you would like to get tomorrow. The later dish represents the amount you would like to get in 4 weeks. For each dish, I wrote how much one bean is worth. On the earlier dish for money tomorrow, one bean is worth 10 birr. On the later dish for money in 4 weeks, one bean is worth 15 birr.

You can put any number of beans on the earlier dish and on the later dish, but you must use all of your beans. If you decide to put all 20 beans on the earlier dish, this means that you would like to get 200 birr tomorrow and nothing in 4 weeks. If you put 10 beans on the earlier dish and 10 beans on the later dish, this means that you would like to get 100 birr tomorrow (10 beans x 10 birr per bean = 100 birr) and 150 birr in 4 weeks. Notice that beans on the later dish are always worth more than beans on the earlier dish, so putting beans on the later dish means you would get more money in total. If you decide to put all 20 beans on the later dish, this means that you would like to get nothing tomorrow, and wait 4 weeks to get 300 birr (20 x 15 birr per bean = 300 birr).

After you put your beans on the two dishes, I will write down the total amount of money you would get at each time. If you are not happy with the amounts, you can take beans again and change your mind. We can do this as often as you like until you are happy.

Let's try it out. Please go ahead and divide up the 20 beans between the two dishes. Remember you don't need to put all beans on one dish, but you can divide them up between earlier and later as you like.

[ENUMERATOR: Let participant allocate beans. Calculate total in each dish, write on board. Read out to participant.]

Would you be happy with these amounts tomorrow and in 4 weeks?

[ENUMERATOR: Revise if necessary]

Remember this was just an example. We will now go through 15 such decisions between earlier and later. In these 15 decisions, I will change when the earlier dish will be paid out and when the later dish will be paid out. For each combination of earlier and later, I will increase how much one bean in the later dish is worth. At the end of our conversation, we will flip a coin to see if you will receive one of your decisions in the form of two payments.

Do you understand everything so far?

[ENUMERATOR checkpoint. You must ask the following three questions]

Please explain to me when you would get paid the amount on each dish, should you win the coin flip.

Suppose that you put beans on both dishes and win the coin flip, how many payments will you receive?

Please explain to me how much one bean is worth on each dish.

[The respondent needs to answer all three of these questions correctly. If not, please go back and explain again]

Did the respondent correctly answer al questions?

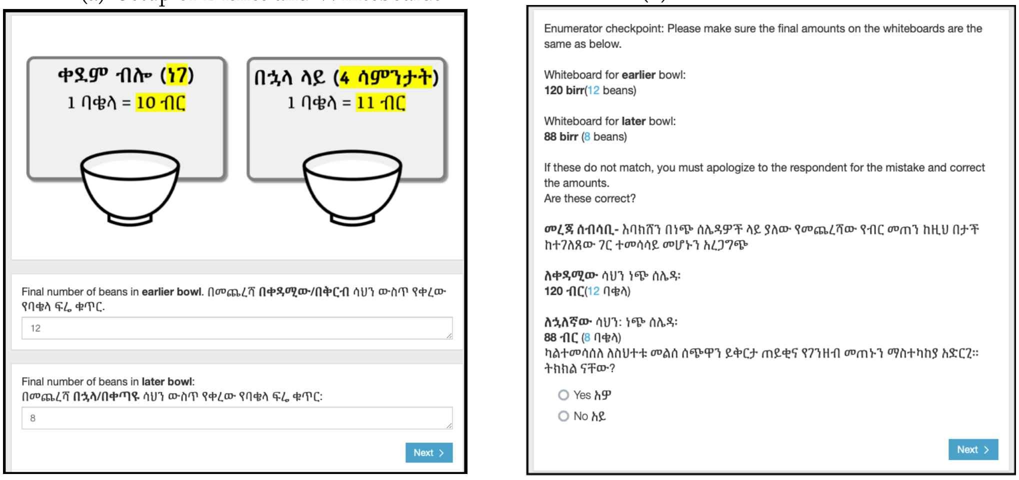
Thank you for listening to my explanations. Let's get started now with your decisions.

Please keep in mind that this is not a test. There are no right or wrong answers. However, it is important to keep in mind that you make decisions over a substantial amount of real money. If you win the coin flip at the end of our conversation, we will pay out one of your decisions. Each of the 15 decision has the same chance of being chosen, so you should think carefully about each of them.

The 15 decisions are done interactively using the beans and the whiteboards. The tablet computer aids the enumerator by visualizing the decision. After illustrating the setup of the white boards, the survey software asks for the final number of beans allocated. Before moving to the next decision, the survey software calculates the total amounts and asks the enumerator to confirm that these were correctly indicated on the white board.

Figure 2.C2: Screenshots of Survey Software During CTB Task

(a) Setup of Dishes and Whiteboards (b) Confirmation of Decision



2.C.3 Payment Confirmation

This shows an English translation of the payment confirmation that subjects receive if they win the coin flip of the CTB experiment. Subjects only receive the original in Amharic. This confirmation is indented to increase confidence in the payment being delivered and reduce the cognitive costs of keeping track. It is printed on high-quality paper with a European University Institute watermark.

Payment Confirmation

As part of our academic study, you made several decisions about whether you would prefer amounts of money earlier or later. Because you won the coin flip at the end of our conversation, you receive the following two amounts:

Earlier payment: *Later Payment:*

[Two large boxes with amount and date for each of the payments]

On the date indicated for each of the payments (but not earlier), you will receive the money to your cell phone as CBEbirr payment. We will send this payment to you before 12 noon on the day indicated. CBEbirr is a payment system by Commercial Bank of Ethiopia. You will receive an SMS from the CBEbirr payment system by the date specified above. The SMS will contain the amount of money that you will receive. The sender will be [survey coordinator name] and the sending phone number will be [survey coordinator phone number]. After you received the text message you can redeem the money at any CBE branch or any CBEbirr agent. When you go to the bank or the CBEbirr agent, please do not forget to bring your Kebele ID and your cell phone with the CBEbirr text message.

If you have any questions, please contact the study coordinator [name] at [coordinator cell phone number].

Recipient name: [subject name]

Date: [date]

Signature Enumerator [signed] *Signature Lead Researcher [signed]*

2.C.4 Consistency and Comparison with Random Choice

Consistency with the law of demand in the CTB experiment can serve as an indicator of whether subjects understood the instructions of the experiment. This is particularly true because the subject pool has relatively lower numeracy and literacy skills than student laboratory subjects.

The average allocations summarized in Table 2.C1 below and visual inspection of Figure 2.3 indicate that on aggregate subject decisions are consistent with the law of demand. As the interest rate increases and the price of consumption later decreases, subjects monotonically decrease the share of beans allocated to earlier.

Table 2.C1: Baseline CTB Allocations to Later (in Ethiopian birr), by Front-End Delay t, Delay k, and Interest Rate r

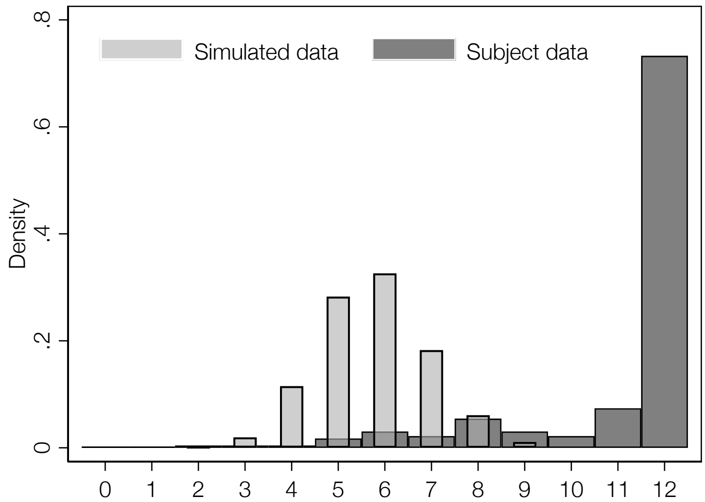
					Percentiles					
t	k	$1+r$	Mean	Std. dev.	10th	25th	50th	75th	90th	Fraction at corner
1	14	1.10	122.24	64.87	0.00	88.00	132.00	165.00	214.50	0.10
1	14	1.25	160.30	69.59	62.50	125.00	175.00	212.50	250.00	0.15
1	14	1.50	218.72	76.00	120.00	180.00	225.00	285.00	300.00	0.23
1	14	1.75	272.64	83.24	175.00	245.00	297.50	350.00	350.00	0.27
1	14	2.00	336.14	87.46	220.00	300.00	380.00	400.00	400.00	0.43
1	28	1.10	112.79	69.76	0.00	55.00	121.00	165.00	209.00	0.10
1	28	1.25	147.48	76.73	0.00	112.50	162.50	200.00	250.00	0.13
1	28	1.50	201.21	87.37	75.00	150.00	225.00	270.00	300.00	0.19
1	28	1.75	251.00	100.61	105.00	210.00	280.00	332.50	350.00	0.23
1	28	2.00	314.19	105.49	180.00	280.00	360.00	400.00	400.00	0.37
15	14	1.10	125.37	61.59	0.00	110.00	132.00	165.00	220.00	0.10
15	14	1.25	159.87	68.37	62.50	125.00	175.00	200.00	250.00	0.15
15	14	1.50	215.95	75.33	120.00	180.00	225.00	270.00	300.00	0.20
15	14	1.75	270.32	84.14	175.00	227.50	280.00	350.00	350.00	0.28
15	14	2.00	327.33	96.83	200.00	300.00	360.00	400.00	400.00	0.42

Notes: Data from all 460 subjects in the baseline survey at $t = 0$.

I follow Giné et al. (2017) in quantifying adherence to the law of demand at the individual level. In each of the three choice sets $(t, k) \in [(1, 14), (1, 28), (15, 14)]$, subjects make five decisions over the same dates but with increasing interest rates. These five decisions can be grouped in four pairs of experimental interest rates between earlier and later where $r' < r''$. In each of these four pairs, subjects should allocate weakly more money to later under r'' than under r'. With 460 subjects in the baseline CTB task, the data has $4 \times 3 \times 460 = 5,520$ such interest rate pairs. Out of those, only 463 (8.4 percent) are not consistent with the law of demand. The median deviation is one bean. This compares favorably to 81 percent of pairs in Giné et al. (2017) and suggests that subjects largely understood the experiment.

Finally, one can compare the number of pairs that are consistent with the law of demand to simulated data in which subjects choose allocations randomly (drawn from a uniform distribution). Figure 2.C3 plots the number of pairs in which subjects allocate weakly more money to later under r'' than under r' for simulated and real baseline data. The comparison suggests that subjects in my experiment indeed made choices that are significantly more consistent than random chance.

Figure 2.C3: Histogram for Consistency with Law of Demand, Subject Decisions Compared to Simulated Random Choice



Notes: In each of three choice sets, subject make five decisions over the same dates but with increasing interest rates. This results four interest rate pairs where $r' < r''$. In total, I can compare $3 \times 4 = 12$ decisions per subjects. The figure plots the number of those decisions that are consistent with the law of demand, such that subjects allocate weakly more money to later under r'' than under r'. Simulated data assumes that for each decision subjects simply allocate their budget based on a random from a uniform distribution.

2.C.5 Theoretical Framework for Parameter Estimation

In this subsection, I outline a simple theoretical framework to estimate time preference parameters based on the CTB experiment. I replicate the approach and parametric assumptions of Andreoni and Sprenger (2012) and Augenblick et al. (2015).

In the lab-in-the-field CTB experiment, each subject chooses to allocate an experimental budget $m > 0$ between an amount c_t available at an earlier time t and a another amount c_{t+k} available after a delay $k > 0$, i.e. paid out at point $t + k$. Let $(1 + r)$ be the simple gross interest rate to be paid over period k. The unit of time are days since the experiment. All monetary amounts are measured in experimental tokens.

Experimental subjects maximize an additively separable utility function with quasi hyperbolic (β-δ) preferences (Laibson, 1997; O'Donoghue & Rabin, 1999) in the form:

$$U(c_t, c_{t+k}) = u(c_t - \omega_t) + \beta^{\mathbb{1}_{t=0}} \delta^k u(c_{t+k} - \omega_{t+k}) \tag{13}$$

where β is the parameter of present bias, $\mathbb{1}_{t=0}$ is an indicator that is 1 when the earlier payoff is realized in period 0, and δ is the long-run discounting parameter. With $\beta = 1$ the framework nests the standard exponential discounting model. ω_t and ω_{t+k} are Stone-Geary (Geary, 1950; Stone, 1954) background consumption or subsistence consumption levels at each point in time. Subjects maximize (13) subject to their experimental budget constraint

$$(1 + r)c_t + c_{t+k} = m. \tag{14}$$

The first order conditions for c_t and c_{t+k} yield the familiar intertemporal Euler equation that must be satisfied by the optimal allocation (c_t^*, c_{t+k}^*)

$$\frac{u'(c_t - \omega_t)}{\beta^{\mathbb{1}_{t=0}} \delta^k u'(c_{t+k} - \omega_{t+k})} = (1 + r). \tag{15}$$

I assume constant CRRA utility in the form $u(c) = 1/\alpha c^{\alpha}$ or, equivalently, $u(c) = c^{1-\theta}/(1 - \theta)$ with θ as the coefficient of relative risk aversion. With that (15) can be written as follows:

$$\frac{c_t - \omega_t}{c_{t+k} - \omega_{t+k}} = \left(\beta^{\mathbb{1}_{t=0}} \delta^k (1 + r)\right)^{1/(\alpha-1)} \tag{16}$$

Assuming ω_t and ω_{t+k} to be fixed, non-estimated values, we can take logs on (16) and obtain

$$\ln\left(\frac{c_t - \omega_t}{c_{t+k} - \omega_{t+k}}\right) = \left(\frac{\ln \beta}{\alpha - 1}\right) \mathbb{1}_{t=0} + \left(\frac{\ln \delta}{\alpha - 1}\right) k + \left(\frac{1}{\alpha - 1}\right) \ln(1 + r) \tag{17}$$

where $(c_t - \omega_t) / (c_{t+k} - \omega_{t+k}) > 0$ by assumption so that the log-transformation is well-defined.

When including an additive error term, Euler equation (17) can be estimated by regression at the level of each subject or in aggregate over all experimental subjects. Because the consumption ratios on the left-hand side are censored by corner solutions, estimation by two-limit Tobit is more appropriate than OLS.

Again following Andreoni and Sprenger (2012) but slightly changing their notation, assume each subject i makes her P budget decisions $j = 1, 2, ..., P$. The estimation problem can be written as follows:

$$\ln\left(\frac{c_t - \omega_t}{c_{t+k} - \omega_{t+k}}\right)_{ij} = \gamma_1 \ \mathbb{1}_{t=0} + \gamma_2 \ k + \gamma_3 \ \ln(1+r) + \epsilon_{ij} \tag{18}$$

where ϵ_{ij} is a mean-zero error. By stacking all P observations for subject i, we obtain

$$\ln\left(\frac{\mathbf{c_t} - \boldsymbol{\omega}_\mathbf{t}}{\mathbf{c_{t+k}} - \boldsymbol{\omega}_\mathbf{t+k}}\right)_\mathbf{i} = \gamma_1 \ \mathbb{1}_{t=0} + \gamma_2 \ \mathbf{k} + \gamma_3 \ln(\mathbf{1} + \mathbf{r}) + \boldsymbol{\epsilon}_i \tag{19}$$

By estimating (19) for each subject i we can obtain the individual-level parameters of interest from non-linear combinations

$$\begin{aligned}\hat{\beta}_i &= \exp(\hat{\gamma}_1/\hat{\gamma}_3)\\ \hat{\delta}_i &= \exp(\hat{\gamma}_2/\hat{\gamma}_3)\\ \hat{\alpha}_i &= (1/\hat{\gamma}_3) + 1\end{aligned}$$

with standard errors from the delta method.

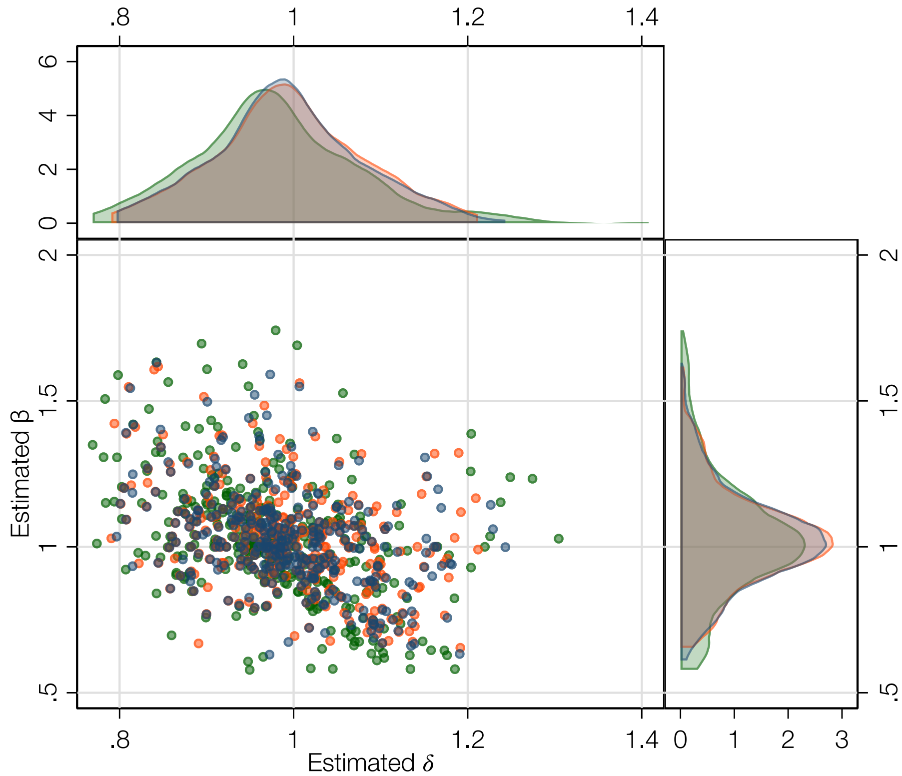
2.C.6 Additional Estimation Results

Table 2.C2: Aggregate Time Preference and CRRA Curvature Estimates (Coefficient Estimates and Standard Errors)

	(1) Tobit	(2) Tobit	(3) Tobit
Present bias $\hat{\beta}$	1.006	1.008	1.002
	(0.018)	(0.013)	(0.014)
Discounting $\hat{\delta}$	0.959	0.977	0.976
	(0.008)	(0.007)	(0.007)
CRRA curvature $\hat{\alpha}$	0.886	0.677	0.734
	(0.006)	(0.011)	(0.012)
$H_0 : \beta = 1$	.119	.338	.0301
p-value	.73	.561	.862
$H_0 : \delta = 1$	23.3	12.4	11.9
p-value	1.39e-06	.000424	.000562
Log likelihood	-17173	-10439	-11902
N	6899	6899	6858
N (uncensored)	4832	4832	4832
Clusters	460	460	459

Notes: Table shows maximum likelihood estimates of Equation (17) using a two-limit Tobit model over the whole sample. Standard errors are clustered at individual level. The three models differ in their assumptions about background consumption at each point in time (see Equation (13) for details). Model (1) assumes no background consumption ($\omega_1 = \omega_2 = 0$). Model (2) assumes that background consumption is constant and set at the sample average daily consumption expenditure in the seven days before the baseline survey ($\omega_1 = \omega_2 = \bar{c}$). Model (3) assumes that background consumption is constant and set at the individual daily consumption expenditure in the seven days before the baseline survey ($\omega_1 = \omega_2 = c_i$).

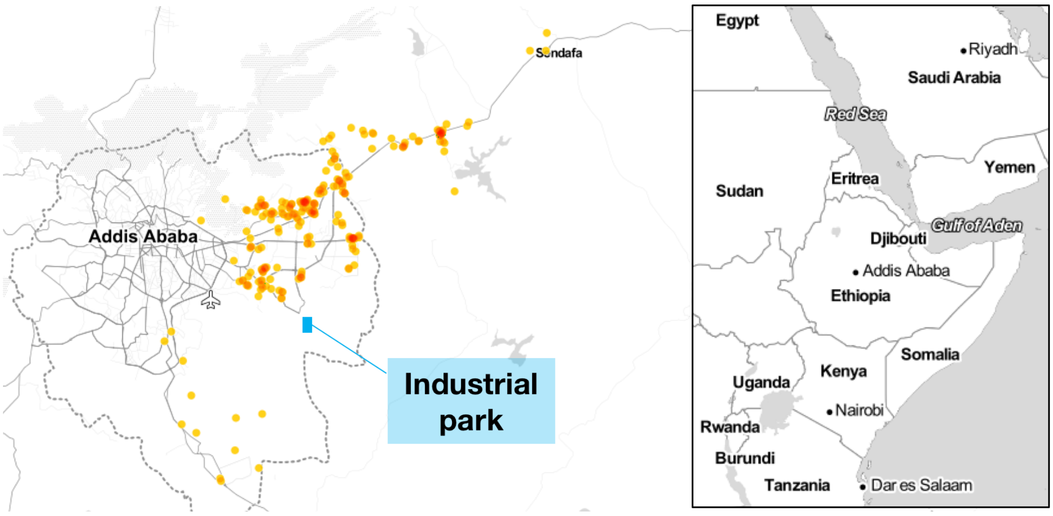
Figure 2.C4: Comparison of Individual-Level Parameter Estimates Using Two-Limit Tobit, by Assumptions About Background Consumption



Notes: Figure shows maximum likelihood estimates of Equation (17) using a two-limit Tobit model at the individual level. Each dot represents one estimate. Different colors represent different assumptions about background consumption: Green markers: assumed zero background consumption; red markers: using the sample average consumption expenditure as background consumption; blue markers: using individually-reported consumption expenditure as background consumption. Sample size differs from the full 460 subjects due to failure of the MLE to converge.

2.D Appendix: Additional Figures

Figure 2.D1: Locations of Survey Firm and Respondent Households



Notes: Study firm is located in Bole Lemi Industrial Park, highlighted in light blue. Travel time from the industrial park to Addis Ababa city center is about 45 minutes to 1 hour, depending on means of transportation and traffic conditions. Yellow dots represent worker households. The more workers live in one locations, the darker the color of the dot.

Sources: Base map data by OpenStreetMap, used under ODbL. Map tiles by Stamen Design, used under Creative Commons CC BY 3.0.

Figure 2.D2: Job Search Intensive and Extensive Margin over Time

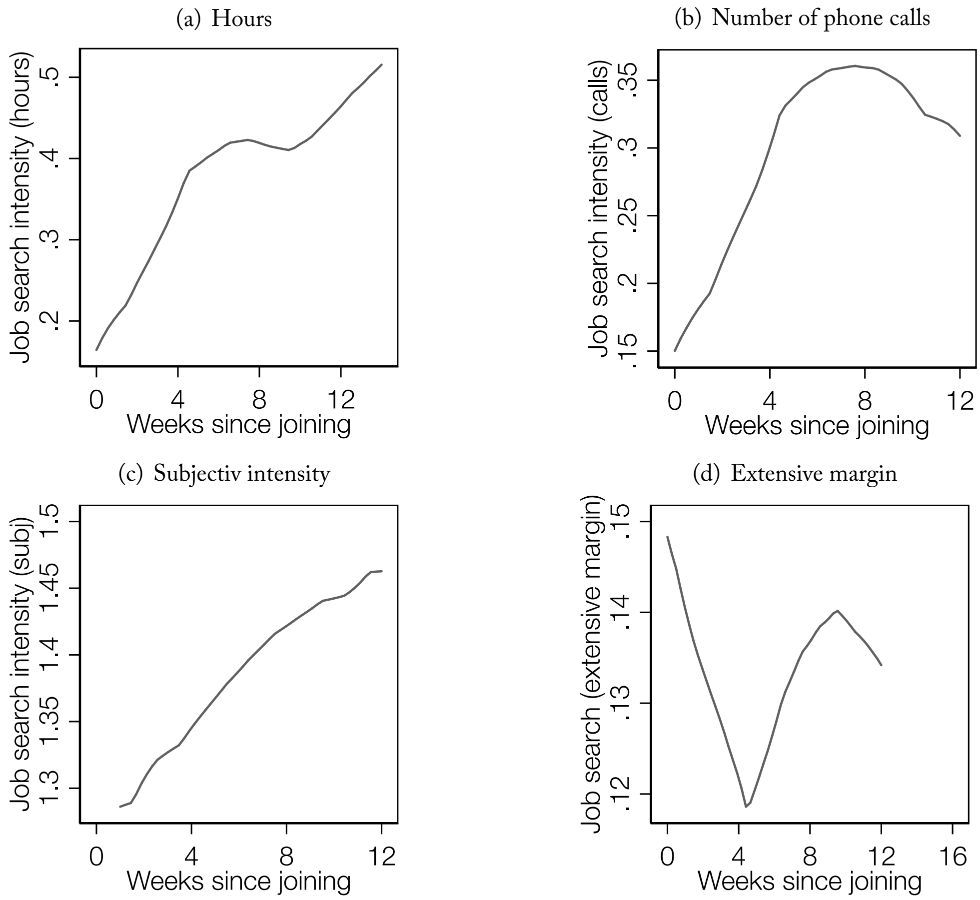
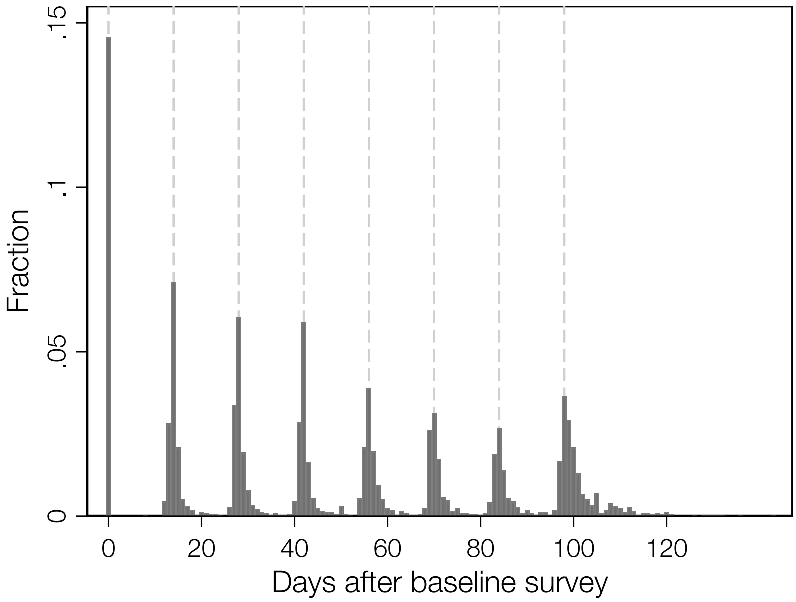
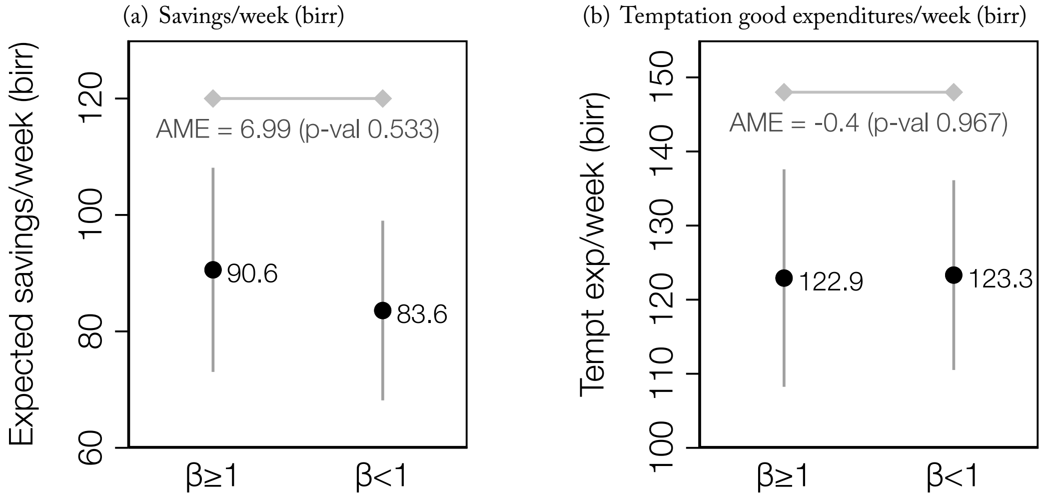


Figure 2.D3: Histogram of Panel Survey Dates After Baseline



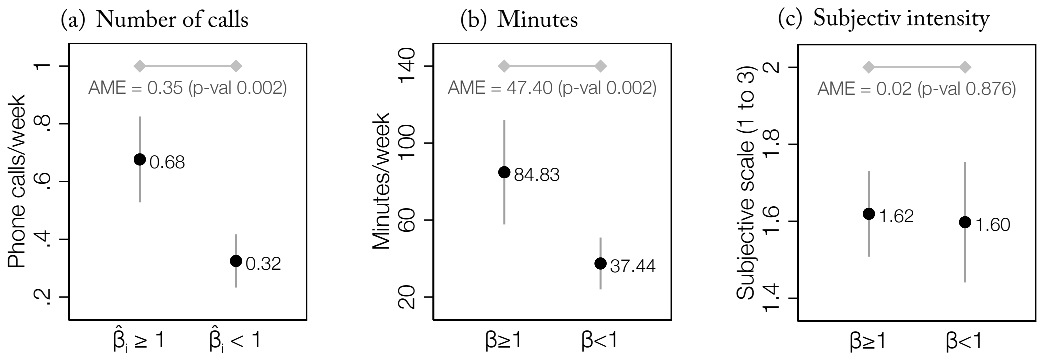
Notes: Dashed lines refer to scheduled survey dates after baseline.

Figure 2.D4: Expected Value of Savings and Temptation Good Expenditure, by Present Bias Indicator



Notes: Plots show the expected value of the observed outcome, based on columns (2) and (4) of Table 2.3. Thin bars indicate 90 percent confidence intervals. AME indicates the average marginal effect for discrete change in present bias indicator, i.e. the difference between the two plotted values.

Figure 2.D5: Expected Value of Search Effort, by Present Bias Indicator



Notes: Plots show the expected value of the observed outcome, based on columns (2), (4), and (6) of Table 2.4. Subjective search intensity is measured on a three-point scale (1 – "not very intensively"; 2 – "intensively"; 3 – "very intensively"). Thin bars indicate 90 percent confidence intervals. AME indicates the average marginal effect for discrete change in present bias indicator, i.e. the difference between the two plotted values.

2.E Appendix: Additional Tables

Table 2.E1: On the Job Search and Reasons for Not Searching (Fractions of Sample)

	Baseline	Week 2	Week 4	Week 6	Week 8	Week 10	Week 12	Week 14	Endline
Searching for work	0.20	0.10	0.12	0.12	0.16	0.14	0.13	0.15	0.14
Would like to search	0.31	0.31	0.33	0.35	0.31	0.31	0.37	0.33	0.33
It takes too much time	0.10	0.13	0.17	0.17	0.14	0.17	0.18	0.15	0.12
It costs too much money	0.03	0.01	0.01	0.02	0.02	0.00	0.01	0.01	0.01
I don't know how/where to look	0.06	0.04	0.02	0.01	0.03	0.03	0.01	0.01	0.04
Other constraints	0.12	0.13	0.13	0.15	0.13	0.11	0.17	0.15	0.16
Not searching	0.49	0.60	0.55	0.53	0.53	0.55	0.49	0.53	0.52
N	460	425	394	321	275	237	209	163	153

Note: The numbers in this table represent workers who are searching while on their job at the study firm, so the decreasing sample size reflects both workers leaving the firm as well as attrition from the panel. Approximately 90 percent of the reasons given as answers under "Other constraints" relate to health problems. To avoid averaging, survey weeks here refer to survey waves of the panel, not the calendar date that the surveys were conducted. Appendix Figure 2.D3 compares scheduled and actual survey dates.

Table 2.E2: Summary Statistics of Baseline Observable Characteristics

Variable	N	Mean	Percentiles					Min	Max
			5th	25th	50th	75th	95th		
Panel (a) Personal characteristics									
Age	460	21.41	18	20	21	23	26	18	31
Married (indicator)	460	0.19	0	0	0	0	1	0	1
Has children (indicator)	460	0.06	0	0	0	0	1	0	1
Has a working spouse (indicator)	460	0.73	0	0	1	1	1	0	1
Rural-urban migrant (indicator)	460	0.67	0	0	1	1	1	0	1
Mother tongue: Amharic (indicator)	460	0.20	0	0	0	0	1	0	1
Mother tongue: Oromiffa (indicator)	460	0.38	0	0	0	1	1	0	1
Religion: Ethiopian Orthodox (indicator)	460	0.73	0	0	1	1	1	0	1
Religion: Muslim (indicator)	460	0.15	0	0	0	0	1	0	1
Education: Has completed 8th grade (indicator)	460	0.17	0	0	0	0	1	0	1
Education: Has completed 10th grade (indicator)	460	0.34	0	0	0	1	1	0	1
Panel (b) Consumption expenditures and savings									
Consumption expenditures: Food	460	141.77	0	0	100	200	500	0	1,200
Consumption expenditures: Non-food	460	147.93	10	36.5	70	155	575	0	1,680
Consumption expenditures: Temptation goods	460	77.96	0	0	20	50	364	0	1,600
Savings	460	469.59	0	0	0	300	2,350	0	30,000
Panel (c) Liquidity and access to finance									
Has bank account (indicator)	460	0.70	0	0	1	1	1	0	1
Ease of obtaining credit (4-point Likert scale)	460	2.24	1	1	2	3	4	1	4
Panel (d) Cognitive control, non-cognitive skills, stress									
Stroop task score	460	0.43	0.18	0.31	0.42	0.54	0.70	0.10	0.89
Generalized Self-Efficacy index	460	34.50	28	32	34	37	41	23	45
Locus of Control index	460	15.59	12	14	16	17	19	7	20
Perceived Stress Scale	460	7.17	3	5	7	9	12	0	16

Notes: Consumption and savings measured with a seven-day recall. Appendix 2.B provides details on survey protocols and the measurement of control variables.

Table 2.E3: Savings and Present Bias as Continuous Variable (Regression Coefficient Estimates and Robust Standard Errors)

	Savings					
	(1)	(2)	(3)	(4)	(5)	(6)
	Tobit	*Tobit*	*Tobit*	*Tobit*	*Tobit*	*Tobit*
Baseline $\hat{\beta}_i$	44.68	75.95		26.97		-42.95
	(101.285)	(170.208)		(91.115)		(111.177)
Baseline $\hat{\delta}_i$		168.5	274.4		35.05	-322.1
		(490.736)	(288.374)		(197.700)	(298.173)
Survey wave dummies	Yes	Yes	Yes	Yes	Yes	Yes
Enumerator dummies	Yes	Yes	Yes	Yes	Yes	Yes
Personal characteristics	No	No	No	Yes	Yes	Yes
Baseline liquidity	No	No	No	Yes	Yes	Yes
Cognitive control	No	No	No	Yes	Yes	Yes
Non-cognitive ability and stress	No	No	No	Yes	Yes	Yes
N	1890	1803	1901	1814	1832	1734

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses, clustered at individual level. Time preference parameters trimmed at 5th and 95th percentile. Controls for personal characteristics are age, marital status, religion, mother tongue, education, number of children, whether or not respondent moved from a rural area, whether respondent has a working spouse. Controls for baseline liquidity are amount of savings held when starting employment and easy of obtaining credit assessed on a 4 point Likert scale. Control for cognitive control is the final score achieved in the baseline Stroop task. Controls for non-cognitive control and stress are baseline score on the General Self-Efficacy scale, baseline locus of control index, and baseline score on the Perceived Stress Scale (PSS-4). Appendix 2.B provides details on survey protocols and the measurement of control variables.

Table 2.E4: Temptation Good Expenditures and Present Bias as Continuous Variable (Regression Coefficient Estimates and Robust Standard Errors)

	Temptation goods expenditure					
	(1)	(2)	(3)	(4)	(5)	(6)
	Tobit	*Tobit*	*Tobit*	*Tobit*	*Tobit*	*Tobit*
Baseline $\hat{\beta}_i$	-4.541		-19.30	20.72		7.973
	(37.355)		(47.697)	(40.736)		(52.269)
Baseline $\hat{\delta}_i$		-1.786	-67.42		-37.85	-90.08
		(86.102)	(123.162)		(96.017)	(134.853)
Survey wave dummies	Yes	Yes	Yes	Yes	Yes	Yes
Enumerator dummies	Yes	Yes	Yes	Yes	Yes	Yes
Personal characteristics	No	No	No	Yes	Yes	Yes
Baseline liquidity	No	No	No	Yes	Yes	Yes
Cognitive control	No	No	No	Yes	Yes	Yes
Non-cognitive ability and stress	No	No	No	Yes	Yes	Yes
N	2090	2102	1995	2009	2028	1921

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses, clustered at individual level. Time preference parameters trimmed at 5th and 95th percentile. Controls for personal characteristics are age, marital status, religion, mother tongue, education, number of children, whether or not respondent moved from a rural area, whether respondent has a working spouse. Controls for baseline liquidity are amount of savings held when starting employment and easy of obtaining credit assessed on a 4 point Likert scale. Control for cognitive control is the final score achieved in the baseline Stroop task. Controls for non-cognitive control and stress are baseline score on the General Self-Efficacy scale, baseline locus of control index, and baseline score on the Perceived Stress Scale (PSS-4). Appendix 2.B provides details on survey protocols and the measurement of control variables.

Table 2.E5: Savings and Temptation Goods at Extensive Margin and Present Bias (Estimated Average Marginal Effects and Robust Standard Errors)

	(1) *Savings Probit*	(2) *Savings Probit*	(3) *Tempt. Probit*	(4) *Tempt. Probit*
Baseline $\hat{\beta}_i < 1$	-0.0351	-0.0192	0.0506*	0.0308
	(0.031)	(0.030)	(0.029)	(0.030)
Baseline $\hat{\delta}_i$	0.108	0.00367	-0.0855	-0.131
	(0.173)	(0.166)	(0.157)	(0.171)
Survey wave dummies	Yes	Yes	Yes	Yes
Enumerator dummies	Yes	Yes	Yes	Yes
Personal characteristics	No	Yes	No	Yes
Baseline liquidity	No	Yes	No	Yes
Cognitive control	No	Yes	No	Yes
Non-cognitive ability and stress	No	Yes	No	Yes
N	2093	2019	2098	1996
Log likelihood				

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses, clustered at individual level. Time preference parameters trimmed at 5th and 95th percentile. Controls for personal characteristics are age, marital status, religion, mother tongue, education, number of children, whether or not respondent moved from a rural area, whether respondent has a working spouse. Controls for baseline liquidity are amount of savings held when starting employment and easy of obtaining credit assessed on a 4 point Likert scale. Control for cognitive control is the final score achieved in the baseline Stroop task. Controls for non-cognitive control and stress are baseline score on the General Self-Efficacy scale, baseline locus of control index, and baseline score on the Perceived Stress Scale (PSS-4). Appendix 2.B provides details on survey protocols and the measurement of control variables.

TABLE 2.E6: Job Search Effort and Present Bias as Continuous Variable (Regression Coefficient Estimates and Robust Standard Errors)

	(1) Calls *Tobit*	(2) Calls *Tobit*	(3) Calls *Tobit*	(4) Hours *Tobit*	(5) Hours *Tobit*	(6) Hours *Tobit*
Baseline $\hat{\beta}_i$	0.817		3.992*	0.566		6.640*
	(1.985)		(2.268)	(3.212)		(3.965)
Baseline $\hat{\delta}_i$		6.599*	12.72**		10.32	24.64**
		(3.988)	(5.290)		(7.914)	(9.950)
Survey wave dummies	Yes	Yes	Yes	Yes	Yes	Yes
Enumerator dummies	Yes	Yes	Yes	Yes	Yes	Yes
Personal characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Baseline liquidity	Yes	Yes	Yes	Yes	Yes	Yes
Cognitive control	Yes	Yes	Yes	Yes	Yes	Yes
Non-cognitive ability and stress	Yes	Yes	Yes	Yes	Yes	Yes
N	1919	1939	1833	1919	1939	1833
Log likelihood	-1194	-1176	-1140	-1510	-1480	-1453

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses, clustered at individual level. Time preference parameters trimmed at 5th and 95th percentile. Controls for personal characteristics are age, marital status, religion, mother tongue, education, number of children, whether or not respondent moved from a rural area, whether respondent has a working spouse. Controls for baseline liquidity are amount of savings held when starting employment and easy of obtaining credit assessed on a 4 point Likert scale. Control for cognitive control is the final score achieved in the baseline Stroop task. Controls for non-cognitive control and stress are baseline score on the General Self-Efficacy scale, baseline locus of control index, and baseline score on the Perceived Stress Scale (PSS-4). Appendix 2.B provides details on survey protocols and the measurement of control variables.

Table 2.E7: Job Search Decision and Present Bias (Estimated Average Marginal Effects and Robust Standard Errors)

	(1) Looking *Probit*	(2) Looking *Probit*
Baseline $\hat{\beta}_i < 1$	-0.0706**	-0.0879***
	(0.031)	(0.030)
Baseline $\hat{\delta}_i$	0.339**	0.325**
	(0.151)	(0.137)
Survey wave dummies	Yes	Yes
Enumerator dummies	Yes	Yes
Personal characteristics	No	Yes
Baseline liquidity	No	Yes
Cognitive control	No	Yes
Non-cognitive ability and stress	No	Yes
N	2004	1919

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses, clustered at individual level. Time preference parameters trimmed at 5th and 95th percentile. Controls for personal characteristics are age, marital status, religion, mother tongue, education, number of children, whether or not respondent moved from a rural area, whether respondent has a working spouse. Controls for baseline liquidity are amount of savings held when starting employment and easy of obtaining credit assessed on a 4 point Likert scale. Control for cognitive control is the final score achieved in the baseline Stroop task. Controls for non-cognitive control and stress are baseline score on the General Self-Efficacy scale, baseline locus of control index, and baseline score on the Perceived Stress Scale (PSS-4). Appendix 2.B provides details on survey protocols and the measurement of control variables.

TABLE 2.E8: Job Search Outcomes, Search Effort, and Present Bias (Incidence-Rate Ratios and Average Marginal Effects with Robust Standard Errors)

	(1) # offers in t *Poisson*	(2) # offers in t *Poisson*	(3) Vol dep in t *Probit*	(4) Vol dep in t *Probit*
Baseline $\hat{\beta}_i < 1$	0.442	0.385*	-0.0257*	-0.189***
	(0.260)	(0.219)	(0.013)	(0.036)
Baseline $\hat{\delta}_i$	178.6**	243.2**	0.0497	0.803***
	(400.895)	(522.120)	(0.065)	(0.274)
Job search effort (hours) in t		0.989		0.00473**
		(0.026)		(0.002)
Job search effort (# calls) in t		1.001		0.0113**
		(0.048)		(0.005)
Job search effort (intensively) in t		1.350		-0.0550
		(0.600)		(0.046)
Job search effort (very intensively) in t		2.129		-0.00128
		(1.148)		(0.065)
Survey wave dummies	Yes	Yes	Yes	Yes
Enumerator dummies	Yes	Yes	Yes	Yes
Personal characteristics	Yes	Yes	Yes	Yes
Baseline liquidity	Yes	Yes	Yes	Yes
Cognitive control	Yes	Yes	Yes	Yes
Non-cognitive ability	Yes	Yes	Yes	Yes
N	320	320	1946	235
Log likelihood	-112	-110		

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses, clustered at individual level. Columns 1 and 2 report incidence-rate ratios from the poisson coefficient estimates while columns 3 and 4 report average marginal effects from probit coefficient estimates. Time preference parameters trimmed at 5th and 95th percentile. Controls for personal characteristics are age, marital status, religion, mother tongue, education, number of children, whether or not respondent moved from a rural area, whether respondent has a working spouse. Controls for baseline liquidity are amount of savings held when starting employment and easy of obtaining credit assessed on a 4 point Likert scale. Control for cognitive control is the final score achieved in the baseline Stroop task. Controls for non-cognitive control and stress are baseline score on the General Self-Efficacy scale, baseline locus of control index, and baseline score on the Perceived Stress Scale (PSS-4). Appendix 2.B provides details on survey protocols and the measurement of control variables.

Table 2.E9: Measured Baseline Present Bias and Observable Characteristics (Regression Coefficient Estimates and Robust Standard Errors)

	(1)	(2)	(3)	(4)	(5)	(6)
	$\mathbb{1}_{\{\hat{\beta}_i<1\}}$	$\mathbb{1}_{\{\hat{\beta}_i<1\}}$	$\mathbb{1}_{\{\hat{\beta}_i<1\}}$	$\hat{\beta}_i$	$\hat{\beta}_i$	$\hat{\beta}_i$
	Probit	*Probit*	*Probit*	*OLS*	*OLS*	*OLS*
Panel (a) Personal characterstics and human capital						
Age	0.0263	0.0284	0.0333	-0.00923**	-0.00956**	-0.0104**
	(0.028)	(0.029)	(0.030)	(0.005)	(0.005)	(0.005)
Married (=1)	0.0871	0.0851	0.0839	-0.0439	-0.0415	-0.0409
	(0.185)	(0.187)	(0.193)	(0.031)	(0.032)	(0.033)
Kids (=1)	0.0831	0.0787	0.0824	-0.0340	-0.0342	-0.0342
	(0.285)	(0.287)	(0.287)	(0.050)	(0.050)	(0.050)
Ethiopian Orthodox faith (=1)	0.347	0.363	0.350	-0.0112	-0.0162	-0.0186
	(0.244)	(0.247)	(0.254)	(0.040)	(0.041)	(0.041)
Muslim faith (=1)	0.142	0.179	0.167	-0.000415	-0.0162	-0.0167
	(0.301)	(0.308)	(0.313)	(0.048)	(0.049)	(0.049)
Amharic mother tongue (=1)	-0.394*	-0.391*	-0.415*	-0.0112	-0.0123	-0.0103
	(0.233)	(0.234)	(0.239)	(0.044)	(0.044)	(0.045)
Oromiffa mother tongue (=1)	0.177	0.162	0.139	-0.101**	-0.0941*	-0.0905*
	(0.249)	(0.251)	(0.253)	(0.047)	(0.048)	(0.048)
Rural-urban migrant (=1)	-0.0178	-0.0256	-0.0801	-0.00642	-0.00428	0.00325
	(0.164)	(0.164)	(0.169)	(0.029)	(0.029)	(0.030)
8th grade education completed (=1)	0.210	0.215	0.153	-0.0421	-0.0437	-0.0358
	(0.180)	(0.179)	(0.184)	(0.031)	(0.030)	(0.031)
Previous formal work experience (=1)	-0.0130	-0.00174	-0.0959	-0.000846	-0.00377	0.00719
	(0.150)	(0.152)	(0.157)	(0.025)	(0.026)	(0.027)
Panel (b) Liquidity and access to finance						
Has bank account (=1)		0.0219	0.0468		-0.0192	-0.0225
		(0.159)	(0.159)		(0.027)	(0.027)
Easy to obtain credit (=1)		-0.118	-0.0885		0.0418	0.0426
		(0.156)	(0.166)		(0.027)	(0.029)
Panel (c) Cognitive control, self-regulation, stress						
Cognitive control score			-1.050**			0.150*
			(0.494)			(0.080)
Self-efficacy score			-0.0362*			0.00503
			(0.022)			(0.003)
Locus of control score			-0.00893			-0.00115
			(0.037)			(0.006)
Stress score			0.00813			0.000494
			(0.025)			(0.005)
Panel (d) Other						
Constant	-0.657	-0.668	1.207	1.255***	1.257***	1.025***
	(0.673)	(0.674)	(1.026)	(0.106)	(0.106)	(0.177)
Enumerator dummies	Yes	Yes	Yes	Yes	Yes	Yes
N	366	366	366	366	366	366

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses, clustered at individual level. Dependent variable in columns (1) to (3) is an indicator variable for present bias while the dependent variable in columns (4) to (6) is the untransformed present bias parameter estimate. Appendix 2.B provides details on survey protocols and the measurement of control variables.

Table 2.E10: Measured Endline Present Bias and Cash Drop (Regression Coefficient Estimates and Robust Standard Errors)

	(1)	(2)	(3)	(4)
	$\mathbb{1}_{\{\hat{\beta}_i<1\}}$	$\mathbb{1}_{\{\hat{\beta}_i<1\}}$	$\hat{\beta}_i$	$\hat{\beta}_i$
	Probit	*Probit*	*OLS*	*OLS*
Won coin payoff at baseline (=1)	-0.274	-0.246	0.0113	0.00306
	(0.218)	(0.217)	(0.033)	(0.031)
Enumerator dummies	Yes	Yes	Yes	Yes
Personal characteristics	Yes	Yes	Yes	Yes
Baseline liquidity	No	Yes	No	Yes
Cognitive ability	No	Yes	No	Yes
Non-cognitive ability and stress	No	Yes	No	Yes
N	194	194	198	198

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses, clustered at individual level. Dependent variable in columns (1) and (2) is an indicator variable for present bias estimated in the endline CTB experiment while the dependent variable in columns (3) and (4) is the untransformed present bias parameter estimate at endline. Appendix 2.B provides details on survey protocols and the measurement of control variables.

Table 2.E11: Balance by Experimental Payouts (Means and Standard Errors of the Mean in Parentheses)

Variable	Full sample	Experimental payout		
		No	Yes	Diff p-val
Panel (a) Personal characteristics				
Age	21.41	21.37	21.45	0.735
	(0.117)	(0.186)	(0.149)	
Married (=1)	0.19	0.21	0.17	0.340
	(0.018)	(0.029)	(0.024)	
Kids (=1)	0.06	0.07	0.05	0.369
	(0.011)	(0.019)	(0.014)	
Ethiopian Orthodox faith (=1)	0.73	0.70	0.76	0.149
	(0.021)	(0.033)	(0.027)	
Amharic mother tongue (=1)	0.67	0.69	0.66	0.427
	(0.022)	(0.033)	(0.030)	
Oromiffa mother tongue (=1)	0.20	0.15	0.23	0.039
	(0.019)	(0.026)	(0.026)	
Ethiopian Orthodox faith (=1)	0.73	0.70	0.76	0.149
	(0.021)	(0.033)	(0.027)	
Muslim faith (=1)	0.15	0.17	0.12	0.128
	(0.016)	(0.027)	(0.020)	
Rural-urban migrant (=1)	0.73	0.74	0.72	0.581
	(0.021)	(0.031)	(0.028)	
8th grade education completed (=1)	0.17	0.17	0.17	0.929
	(0.017)	(0.027)	(0.023)	
Previous formal work experience (=1)	0.60	0.59	0.60	0.890
	(0.023)	(0.035)	(0.031)	
Panel (b) Liquidity and access to finance				
Has bank account (=1)	0.70	0.69	0.72	0.463
	(0.021)	(0.033)	(0.028)	
Easy to obtain credit (=1)	0.67	0.68	0.65	0.513
	(0.022)	(0.033)	(0.030)	
Panel (c) Cognitive control, self-regulation, stress				
Cognitive control score	0.43	0.43	0.43	0.943
	(0.007)	(0.011)	(0.010)	
Self-efficacy score	34.50	34.34	34.62	0.446
	(0.182)	(0.260)	(0.252)	
Locus of control score	15.59	15.53	15.64	0.605
	(0.109)	(0.159)	(0.149)	
Stress score	7.17	7.30	7.07	0.405
	(0.138)	(0.213)	(0.182)	
N	460	201	259	

Notes: Experimental payouts are determined by a coin flip at the end of the experiment. "Diff. p-val" refers to the p-value of an t-test for equality of means between subjects who won the coin flip and subjects who did not. Appendix 2.B provides details on survey protocols and the measurement of control variables.

TABLE 2.E12: Job Search Effort and Present Bias Controlling for Experimental Payouts (Regression Coefficient Estimates and Robust Standard Errors)

	(1) Calls *Tobit*	(2) Calls *Tobit*	(3) Calls *Tobit*	(4) Calls *Tobit*	(5) Hours *Tobit*	(6) Hours *Tobit*	(7) Hours *Tobit*	(8) Hours *Tobit*
Baseline $\hat{\beta}_i < 1$	-2.287**	-2.370**	-2.963***	-3.138***	-4.479**	-4.615**	-6.130***	-6.337***
	(1.016)	(1.007)	(0.950)	(0.947)	(1.965)	(1.948)	(1.894)	(1.894)
Baseline $\hat{\delta}_i$	12.22**	10.84**	12.41***	11.46***	22.12**	19.11**	22.28***	19.56**
	(5.067)	(4.901)	(4.216)	(4.102)	(9.621)	(9.057)	(8.593)	(8.118)
Won experimental payout (=1)		-2.049**		-2.128***		-3.502**		-3.763**
		(0.851)		(0.764)		(1.658)		(1.627)
Survey wave dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Enumerator dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Personal characteristics	No	No	Yes	Yes	No	No	Yes	Yes
Baseline liquidity	No	No	Yes	Yes	No	No	Yes	Yes
Cognitive ability	No	No	Yes	Yes	No	No	Yes	Yes
Non-cognitive ability and stress	No	No	Yes	Yes	No	No	Yes	Yes
N	2008	2008	1939	1939	2008	2008	1939	1939
Log likelihood	-1263	-1257	-1166	-1159	-1555	-1549	-1467	-1460

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses, clustered at individual level. Time preference parameters trimmed at 5th and 95th percentile. Controls for personal characteristics are age, marital status, religion, mother tongue, education, number of children, whether or not respondent moved from a rural area, whether respondent has a working spouse. Controls for baseline liquidity are amount of savings held when starting employment and easy of obtaining credit assessed on a 4 point Likert scale. Control for cognitive control is the final score achieved in the baseline Stroop task. Controls for non-cognitive control and stress are baseline score on the General Self-Efficacy scale, baseline locus of control index, and baseline score on the Perceived Stress Scale (PSS-4). Appendix 2.B provides details on survey protocols and the measurement of control variables.

Table 2.E13: Job Search Effort and Human Capital (Regression Coefficient Estimates with Robust Standard Errors)

	(1) Calls *Tobit*	(2) Calls *Tobit*	(3) Calls *Tobit*	(4) Hours *Tobit*	(5) Hours *Tobit*	(6) Hours *Tobit*
Age	0.506***	0.508***	0.493***	0.812***	0.809**	0.771**
	(0.170)	(0.172)	(0.169)	(0.315)	(0.314)	(0.304)
Formal work experience (=1)	-0.415	-0.423	-0.294	-1.992	-1.980	-1.786
	(0.856)	(0.857)	(0.857)	(1.634)	(1.625)	(1.636)
Education level completed (=1)						
(Omitted: Grade 5 or less)						
Grade 6	1.652	1.656	1.476	3.548	3.540	3.343
	(1.672)	(1.672)	(1.684)	(3.219)	(3.216)	(3.189)
Grade 7	-0.605	-0.594	-0.489	-1.320	-1.342	-1.249
	(1.628)	(1.632)	(1.662)	(3.211)	(3.210)	(3.270)
Grade 8	1.476	1.480	1.643	3.498	3.490	3.797
	(1.590)	(1.591)	(1.623)	(3.156)	(3.160)	(3.237)
Grade 9	2.350	2.364	2.471	2.185	2.164	2.145
	(2.186)	(2.191)	(2.249)	(3.397)	(3.395)	(3.428)
Grade 10	0.189	0.205	-0.0109	2.827	2.799	2.502
	(1.467)	(1.473)	(1.495)	(2.874)	(2.862)	(2.865)
More than Grade 10	3.252*	3.280*	3.013	6.571*	6.525*	6.107*
	(1.913)	(1.952)	(1.881)	(3.368)	(3.420)	(3.333)
Cognitive control		-0.376	-0.509		0.568	0.344
		(2.831)	(2.819)		(5.466)	(5.520)
Self-efficacy			0.122			0.136
			(0.126)			(0.255)
Locus of control			0.184			0.488
			(0.223)			(0.410)
Personal characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Enumerator dummies	Yes	Yes	Yes	Yes	Yes	Yes
Survey wave dummies	Yes	Yes	Yes	Yes	Yes	Yes
N	2215	2215	2215	2215	2215	2215
Log likelihood	-1347	-1347	-1343	-1594	-1594	-1590

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses, clustered at individual level. Controls for personal characteristics are age, marital status, religion, mother tongue, number of children, whether or not respondent moved from a rural area, whether respondent has a working spouse. Appendix 2.B provides details on survey protocols and the measurement of control variables.

Table 2.E14: Log Reservation Wage and Time Preference Parameters (OLS Regression Coefficient Estimates with Robust Standard Errors)

	(1)	(2)	(3)	(4)
Baseline $\hat{\beta}_i < 1$	0.969	0.971	0.975	0.977
	(0.028)	(0.028)	(0.027)	(0.028)
Baseline $\hat{\delta}_i$	0.983	0.980	0.984	0.993
	(0.149)	(0.147)	(0.146)	(0.147)
Survey wave dummies	Yes	Yes	Yes	Yes
Enumerator dummies	Yes	Yes	Yes	Yes
Personal characteristics	Yes	Yes	Yes	Yes
Baseline liquidity	No	Yes	Yes	Yes
Cognitive control	No	No	Yes	Yes
Non-cognitive ability	No	No	No	Yes
N	2423	2423	2423	2423
R^2	0.404	0.406	0.407	0.409

Time preference parameters trimmed at 5th and 95th percentile.
Robust standard errors clustered at individual level

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses, clustered at individual level. Appendix 2.B provides details on survey protocols and the measurement of control variables.

2.F References

Abebe, G., Buehren, N., & Goldstein, M. (2018, February). *Welfare Impacts of Factory Jobs: Experimental Evidence from Ethiopia* [Unpublished Manuscript].

Abebe, G., Caria, S., Fafchamps, M., Falco, P., Franklin, S., & Quinn, S. (2016, July). *Curse of Anonymity or Tyranny of Distance? The Impacts of Job-Search Support in Urban Ethiopia* (NBER Working Paper No. 22409). Cambridge, MA: National Bureau of Economic Research. doi: 10.3386/w22409

Abel, M., Burger, R., Carranza, E., & Piraino, P. (2017, September). *Bridging the Intention-Behavior Gap? : The Effect of Plan-Making Prompts on Job Search and Employment* (World Bank Policy Research Paper No. 8181). Washington, DC: The World Bank.

Acland, D., & Levy, M. R. (2015, January). Naiveté, Projection Bias, and Habit Formation in Gym Attendance. *Management Science*, *61*(1), 146-160. doi: 10.1287/mnsc.2014.2091

Adda, J., & Cooper, R. W. (2003). *Dynamic Economics: Quantitative Methods and Applications*. Cambridge, MA: MIT Press.

Algan, Y., Chéron, A., Hairault, J.-O., & Langot, F. (2003, January). Wealth effect on labor market transitions. *Review of Economic Dynamics*, *6*(1), 156-178. doi: 10.1016/S1094-2025(02)00013-3

Amador, M., Werning, I., & Angeletos, G.-M. (2006, March). Commitment vs. Flexibility. *Econometrica*, *74*(2), 365-396. doi: 10.1111/j.1468-0262.2006.00666.x

Andersen, S., Harrison, G. W., Lau, M. I., & Rutström, E. E. (2008, May). Eliciting Risk and Time Preferences. *Econometrica*, *76*(3), 583-618. doi: 10.1111/j.1468-0262.2008.00848.x

Andreoni, J., Callen, M., Hussain, K., Khan, M. Y., & Sprenger, C. (2017). *Using Preference Estimates to Customize Incentives: An Application to Polio Vaccination Drives in Pakistan* [Unpublished Manuscript].

Andreoni, J., & Sprenger, C. (2012, December). Estimating Time Preferences from Convex Budgets. *American Economic Review*, *102*(7), 3333-3356. doi: 10.1257/aer.102.7.3333

Ashraf, N., Karlan, D., & Yin, W. (2006, May). Tying Odysseus to the Mast: Evidence From a Commitment Savings Product in the Philippines. *The Quarterly Journal of Economics*, *121*(2), 635-672. doi: 10.1162/qjec.2006.121.2.635

Atkin, D. (2016, August). Endogenous Skill Acquisition and Export Manufacturing in Mexico. *American Economic Review*, *106*(8), 2046-2085. doi: 10.1257/aer.20120901

Augenblick, N., Niederle, M., & Sprenger, C. (2015, August). Working over Time: Dynamic Inconsistency in Real Effort Tasks. *The Quarterly Journal of Economics*, *130*(3), 1067-1115. doi: 10.1093/qje/qjv020

Bai, L., Handel, B., Miguel, E., & Rao, G. (2017, August). *Self-Control and Demand for Preventive Health: Evidence from Hypertension in India* (Tech. Rep. No. w23727). Cambridge, MA: National Bureau of Economic Research. doi: 10.3386/w23727

Balakrishnan, U., Haushofer, J., & Jakiela, P. (2017, June). *How Soon Is Now? Evidence of Present Bias from Convex Time Budget Experiments* (NBER Working Paper No. 23558). Cambridge, MA: National Bureau of Economic Research. doi: 10.3386/w23558

Baldo, J. V., Shimamura, A. P., & Prinzmetal, W. (1998, April). Mapping symbols to response modalities: Interference effects on Stroop-like tasks. *Perception & Psychophysics*, *60*(3), 427-437.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W.H. Freeman.

Banerjee, A., & Mullainathan, S. (2010, May). *The Shape of Temptation: Implications for the Economic Lives of the Poor* (NBER Working Paper No. 15973). Cambridge, MA: National Bureau of Economic Research. doi: 10.3386/w15973

Belot, M., Kircher, P., & Muller, P. (2018, October). Providing Advice to Job Seekers at Low Cost: An Experimental Study on On-Line Advice. *The Review of Economic Studies*. doi: 10.1093/restud/rdy059

Bénabou, R., & Tirole, J. (2004, August). Willpower and Personal Rules. *Journal of Political Economy*, *112*(4), 848-886. doi: 10.1086/421167

Benhabib, J., & Bisin, A. (2005, August). Modeling internal commitment mechanisms and self-control: A neuroeconomics approach to consumption–saving decisions. *Games and Economic Behavior*, *52*(2), 460-492. doi: 10.1016/j.geb.2004.10.004

Beshears, J., Milkman, K. L., & Schwartzstein, J. (2016, May). Beyond Beta-Delta: The Emerging Economics of Personal Plans. *American Economic Review*, *106*(5), 430-434. doi: 10.1257/aer.p20161100

Blattman, C., & Dercon, S. (2018, July). The Impacts of Industrial and Entrepreneurial Work on Income and Health: Experimental Evidence from Ethiopia. *American Economic Journal: Applied Economics*, *10*(3), 1-38. doi: 10.1257/app.20170173

Bloemen, H. G., & Stancanelli, E. G. F. (2001, April). Individual Wealth, Reservation Wages, and Transitions into Employment. *Journal of Labor Economics*, *19*(2), 400-439. doi: 10.1086/319566

Bryan, G., Karlan, D., & Nelson, S. (2010). Commitment Devices. *Annual Review of Economics*, *2*(1), 671-698. doi: 10.1146/annurev.economics.102308.124324

Budescu, D. V. (1993). Dominance analysis: A new approach to the problem of relative importance of predictors in multiple regression. *Psychological Bulletin*, *114*(3), 542-551. doi: 10.1037/0033-2909.114.3.542

Caliendo, M., Cobb-Clark, D. A., & Uhlendorff, A. (2015, March). Locus of Control and Job Search Strategies. *Review of Economics and Statistics*, *97*(1), 88-103.

Cantril, H. (1965). *The Pattern of Human Concerns*. New Brunswick, NJ: Rutgers University Press.

Card, D., Chetty, R., & Weber, A. (2007, November). Cash-on-Hand and Competing Models of Intertemporal Behavior: New Evidence from the Labor Market. *The Quarterly Journal of Economics*, *122*(4), 1511-1560. doi: 10.1162/qjec.2007.122.4.1511

Carvalho, L. S., Meier, S., & Wang, S. W. (2016, February). Poverty and Economic Decision-Making: Evidence from Changes in Financial Resources at Payday. *American Economic Review*, *106*(2), 260-284. doi: 10.1257/aer.20140481

Cassidy, R. (2018). *Are the poor so present-biased?* (IFS Working Paper No. W18/24). London: Institute for Fiscal Studies.

Castillo, M., Ferraro, P. J., Jordan, J. L., & Petrie, R. (2011, December). The today and tomorrow of kids: Time preferences and educational outcomes of children. *Journal of Public Economics*, *95*(11-12), 1377-1385. doi: 10.1016/j.jpubeco.2011.07.009

Central Statistical Agency. (2014). *Microdata of the 2013 National Labor Force Survey* (Tech. Rep.). Addis Ababa: Central Statistical Agency of the Federal Democratic Republic of Ethiopia.

Central Statistical Agency. (2015). *Statistical Report on the 2015 Urban Employment Unemployment Survey* (Tech. Rep.). Addis Ababa: Central Statistical Agency of the Federal Democratic Republic of Ethiopia.

Central Statistical Agency and World Bank. (2017, February). *LSMS – Integrated Surveys on Agriculture Ethiopia Socioeconomic Survey (ESS) 2015/2016* (Survey Report). Addis Ababa and Washington DC: Central Statistical Agency and Living Standards Measurement Study (LSMS), World Bank.

Christensen, B. J., Lentz, R., Mortensen, D. T., Neumann, G. R., & Werwatz, A. (2005, January). On-the-Job Search and the Wage Distribution. *Journal of Labor Economics*, *23*(1), 31-58. doi: 10.1086/425432

Cohen, S., Kamarck, T., & Mermelstein, R. (1983, December). A Global Measure of Perceived Stress. *Journal of Health and Social Behavior*, *24*(4), 385. doi: 10.2307/2136404

Coller, M., & Williams, M. B. (1999, December). Eliciting individual discount rates. *Experimental Economics*, *2*(2), 107-127. doi: 10.1007/BF01673482

Cox, D. R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, *34*(2), 187-220.

Cubitt, R. P., & Read, D. (2007, December). Can intertemporal choice experiments elicit time preferences for consumption? *Experimental Economics*, *10*(4), 369-389. doi: 10.1007/s10683-006-9140-2

Dean, M., & Sautmann, A. (2016). *Credit Constraints and the Measurement of Time Preferences* [Unpublished Manuscript].

Delavande, A., Giné, X., & McKenzie, D. (2011, March). Measuring subjective expectations in developing countries: A critical review and new evidence. *Journal of Development Economics*, *94*(2), 151-163. doi: 10.1016/j.jdeveco.2010.01.008

Delavande, A., & Kohler, H.-P. (2009, June). Subjective expectations in the context of HIV/AIDS in Malawi. *Demographic Research*, *20*, 817-875. doi: 10.4054/DemRes.2009.20.31

DellaVigna, S. (2009, June). Psychology and Economics: Evidence from the Field. *Journal* *of Economic Literature*, *47*(2), 315-372. doi: 10.1257/jel.47.2.315

DellaVigna, S., Lindner, A., Reizer, B., & Schmieder, J. F. (2017, November). Reference-Dependent Job Search: Evidence from Hungary. *The Quarterly Journal of Economics*, *132*(4), 1969-2018. doi: 10.1093/qje/qjx015

DellaVigna, S., & Malmendier, U. (2006, May). Paying Not to Go to the Gym. *American Economic Review*, *96*(3), 694-719. doi: 10.1257/aer.96.3.694

DellaVigna, S., & Paserman, M. D. (2005, July). Job Search and Impatience. *Journal of Labor Economics*, *23*(3), 527-588. doi: 10.1086/430286

Duflo, E., Kremer, M., & Robinson, J. (2011, October). Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya. *American Economic Review*, *101*(6), 2350-2390. doi: 10.1257/aer.101.6.2350

Dupas, P., & Robinson, J. (2013, June). Why Don't the Poor Save More? Evidence from Health Savings Experiments. *American Economic Review*, *103*(4), 1138-1171. doi: 10.1257/aer.103.4.1138

Esopo, K., Mellow, D., Thomas, C., Uckat, H., Abraham, J., Jain, P., … Haushofer, J. (2018, February). Measuring self-efficacy, executive function, and temporal discounting in Kenya. *Behaviour Research and Therapy*, *101*, 30-45. doi: 10.1016/j.brat.2017.10.002

Faberman, R. J., & Kudlyak, M. (2018). The Intensity of Job Search and Search Duration. *American Economic Journal: Macroeconomics*, *forthcoming*. doi: 10.1257/mac.20170315

Faberman, R. J., Mueller, A., Şahin, A., & Topa, G. (2017, August). *Job Search Behavior among the Employed and Non-Employed* (Tech. Rep. No. w23731). Cambridge, MA: National Bureau of Economic Research. doi: 10.3386/w23731

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., & Sunde, U. (2018). Global Evidence on Economic Preferences. *The Quarterly Journal of Economics*, *forthcoming*. doi: 10.1093/qje/qjy013

Falk, A., Huffman, D., & Sunde, U. (2006a). *Do I Have What It Takes? Equilibrium Search with Type Uncertainty and Non-Participation* (IZA Discussion Paper No. 2531). Bonn: Institute for the Study of Labor (IZA).

Falk, A., Huffman, D., & Sunde, U. (2006b). *Self-Confidence and Search* (IZA Discussion Paper No. 2525). Bonn: Institute for the Study of Labor (IZA).

Filmer, D., & Scott, K. (2012). Assessing Asset Indices. *Demography*, *49*(1), 359-392. doi: 10.1007/s13524-011-0077-5

Franklin, S. (2017, November). Location, Search Costs and Youth Unemployment: Experimental Evidence from Transport Subsidies. *The Economic Journal*. doi: 10.1111/ecoj.12509

Frederick, S., Loewenstein, G., & O'Donoghue, T. (2002, June). Time Discounting and Time Preference: A Critical Review. *Journal of Economic Literature*, *40*(2), 351-401. doi: 10.1257/002205102320161311

Fudenberg, D., & Levine, D. K. (2006, December). A Dual-Self Model of Impulse Control. *American Economic Review*, *96*(5), 1449-1476. doi: 10.1257/aer.96.5.1449

Geary, R. C. (1950). A Note on "A Constant-Utility Index of the Cost of Living". *The Review of Economic Studies*, *18*(1), 65. doi: 10.2307/2296107

Gelb, A. H., Meyer, C. J., Ramachandran, V., & Wadhwa, D. (2017). *Can Africa Be a Manufacturing Destination? Labor Costs in Comparative Perspective* (CGD Working Paper No. 466). Washington, DC: Center for Global Development.

Giné, X., Goldberg, J., Silverman, D., & Yang, D. (2017, March). Revising Commitments: Field Evidence on the Adjustment of Prior Choices. *The Economic Journal*, n/a-n/a. doi: 10.1111/ecoj.12378

Giné, X., Karlan, D., & Zinman, J. (2010, October). Put Your Money Where Your Butt Is: A Commitment Contract for Smoking Cessation. *American Economic Journal: Applied Economics*, *2*(4), 213-235. doi: 10.1257/app.2.4.213

Gul, F., & Pesendorfer, W. (2001, November). Temptation and Self-Control. *Econometrica*, *69*(6), 1403-1435. doi: 10.1111/1468-0262.00252

Halevy, Y. (2015, January). Time Consistency: Stationarity and Time Invariance. *Econometrica*, *83*(1), 335-352. doi: 10.3982/ECTA10872

Hardy, M., Kagy, G., & Meyer, C. J. (2018). *The Devil Is in The Pre-Tales: A Closer Look at Women's First Months in Industrial Work* [Mimeo].

Haushofer, J. (2015). *The Cost of Keeping Track* [Unpublished Manuscript].

Haushofer, J., & Fehr, E. (2014, May). On the psychology of poverty. *Science*, *344*(6186), 862-867. doi: 10.1126/science.1232491

Haushofer, J., & Shapiro, J. (2016, November). The Short-term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya. *The Quarterly Journal of Economics*, *131*(4), 1973-2042. doi: 10.1093/qje/qjw025

Heath, R., & Mobarak, A. M. (2014, August). *Manufacturing Growth and the Lives of Bangladeshi Women* (Working Paper No. 20383). National Bureau of Economic Research. doi: 10.3386/w20383

Holt, C. A., & Laury, S. K. (2002, November). Risk Aversion and Incentive Effects. *American Economic Review*, *92*(5), 1644-1655. doi: 10.1257/000282802762024700

Janssens, W., Kramer, B., & Swart, L. (2017, May). Be patient when measuring hyperbolic discounting: Stationarity, time consistency and time invariance in a field experiment. *Journal of Development Economics*, *126*, 77-90. doi: 10.1016/j.jdeveco.2016.12.011

Kahneman, D., & Deaton, A. (2010, September). High income improves evaluation of life but not emotional well-being. *Proceedings of the National Academy of Sciences*, *107*(38), 16489-16493. doi: 10.1073/pnas.1011492107

Karlan, D., Ratan, A. L., & Zinman, J. (2014, March). Savings by and for the Poor: A Research Review and Agenda. *Review of Income and Wealth*, *60*(1), 36-78. doi: 10.1111/roiw.12101

Kast, F., Meier, S., & Pomeranz, D. (2018, July). Saving more in groups: Field experimental evidence from Chile. *Journal of Development Economics*, *133*, 275-294. doi: 10.1016/j.jdeveco.2018.01.006

Kaur, S., Kremer, M., & Mullainathan, S. (2015, December). Self-Control at Work. *Journal of Political Economy*, *123*(6), 1227-1277. doi: 10.1086/683822

Krueger, A. B., & Mueller, A. I. (2016, February). A Contribution to the Empirics of Reservation Wages. *American Economic Journal: Economic Policy*, *8*(1), 142-179. doi: 10.1257/pol.20140211

Laibson, D. I. (1997, May). Golden Eggs and Hyperbolic Discounting. *The Quarterly Journal of Economics*, *112*(2), 443-478. doi: 10.1162/003355397555253

Laibson, D. I. (2015, May). Why Don't Present-Biased Agents Make Commitments? *American Economic Review*, *105*(5), 267-272. doi: 10.1257/aer.p20151084

Laibson, D. I., Repetto, A., & Tobacman, J. (1998). Self-Control and Saving for Retirement. *Brookings Papers on Economic Activity*, *1998*(1), 91. doi: 10.2307/2534671

Lentz, R., & Tranæs, T. (2005, July). Job Search and Savings: Wealth Effects and Duration Dependence. *Journal of Labor Economics*, *23*(3), 467-489. doi: 10.1086/430284

Levenson, H. (1981). Differentiating among internality, powerful others, and chance. In H. M. Lefcourt (Ed.), *Research with the Locus of Control Construct* (Vol. Volume 1: Assessment Methods, p. 15-63). New York, NY: Academic Press.

Loewenstein, G., & Prelec, D. (1992, May). Anomalies in Intertemporal Choice: Evidence and an Interpretation. *The Quarterly Journal of Economics*, *107*(2), 573-597. doi: 10.2307/2118482

Luszczynska, A., Scholz, U., & Schwarzer, R. (2005, September). The General Self-Efficacy Scale: Multicultural Validation Studies. *The Journal of Psychology*, *139*(5), 439-457. doi: 10.3200/JRLP.139.5.439-457

Macchiavello, R., Menzel, A., Rabbani, A., & Woodruff, C. (2015, December). *Challenges of Change: An Experiment Training Women to Manage in the Bangladeshi Garment Sector* (Centre for Competitive Advantage in the Global Economy Working Paper No. 256). Warwick: The University of Warwick.

Mackie, M.-A., Van Dam, N. T., & Fan, J. (2013, August). Cognitive control and attentional functions. *Brain and Cognition*, *82*(3), 301-312. doi: 10.1016/j.bandc.2013.05.004

Madrian, B. C. (1994, February). Employment-Based Health Insurance and Job Mobility: Is there Evidence of Job-Lock? *The Quarterly Journal of Economics*, *109*(1), 27-54. doi: 10.2307/2118427

Mani, A., Mullainathan, S., Shafir, E., & Zhao, J. (2013, August). Poverty Impedes Cognitive Function. *Science*, *341*(6149), 976-980. doi: 10.1126/science.1238041

McDowell, I. (2006). *Measuring Health: A Guide to Rating Scales and Questionnaires* (3rd ed ed.). Oxford ; New York: Oxford University Press.

McGee, A., & McGee, P. (2016, June). Search, effort, and locus of control. *Journal of*

Economic Behavior & Organization, *126*, 89-101. doi: 10.1016/j.jebo.2016.03.001

Meier, S., & Sprenger, C. (2010, January). Present-Biased Preferences and Credit Card Borrowing. *American Economic Journal: Applied Economics*, *2*(1), 193-210. doi: 10.1257/app.2.1.193

Milkman, K. L., Beshears, J., Choi, J. J., Laibson, D., & Madrian, B. C. (2011, June). Using implementation intentions prompts to enhance influenza vaccination rates. *Proceedings of the National Academy of Sciences*, *108*(26), 10415-10420. doi: 10.1073/pnas.1103170108

Mullainathan, S., & Shafir, E. (2013). *Scarcity: Why having too little means so much*. New York: Henry Holt and Company.

National Planning Commission. (2017). *Ethiopia's Progress Towards Eradicating Poverty: An Interim Report on 2015/16 Poverty Analysis Study* (Tech. Rep.). Addis Ababa: National Planning Commission of the Federal Democratic Republic of Ethiopia.

O'Donoghue, T., & Rabin, M. (1999, March). Doing It Now or Later. *American Economic Review*, *89*(1), 103-124. doi: 10.1257/aer.89.1.103

O'Donoghue, T., & Rabin, M. (2001, February). Choice and Procrastination. *The Quarterly Journal of Economics*, *116*(1), 121-160. doi: 10.1162/003355301556365

Oqubay, A. (2016). *Made in Africa: Industrial Policy in Ethiopia*. Oxford: Oxford University Press.

Paserman, M. D. (2008, August). Job Search and Hyperbolic Discounting: Structural Estimation and Policy Evaluation. *The Economic Journal*, *118*(531), 1418-1452. doi: 10.1111/j.1468-0297.2008.02175.x

Phelps, E. S., & Pollak, R. A. (1968, April). On Second-Best National Saving and Game-Equilibrium Growth. *The Review of Economic Studies*, *35*(2), 185. doi: 10.2307/2296547

Samuelson, P. A. (1937, February). A Note on Measurement of Utility. *The Review of Economic Studies*, *4*(2), 155. doi: 10.2307/2967612

Schilbach, F., Schofield, H., & Mullainathan, S. (2016, May). The Psychological Lives of the Poor. *American Economic Review*, *106*(5), 435-440. doi: 10.1257/aer.p20161101

Schmieder, J. F., von Wachter, T., & Bender, S. (2016, March). The Effect of Unemployment Benefits and Nonemployment Durations on Wages. *American Economic Review*, *106*(3), 739-777. doi: 10.1257/aer.20141566

Schwarzer, R., & Jerusalem, M. (1995). Generalized Self-Efficacy scale. In J. Weinman, S. Wright, & M. Johnston (Eds.), *Measures in health psychology: A user's portfolio.* Windsor: NFER-NELSON.

Seid, Y., Tafesse, A. S., & Ali, S. N. (2016). *Ethiopia – An Agrarian Economy in Transition* (Report). Washington, DC: Brookings Institution.

Spinnewijn, J. (2015, February). Unemployed but Optimistic: Optimal Insurance Design with Biased Beliefs. *Journal of the European Economic Association*, *13*(1), 130-167. doi: 10.1111/jeea.12099

Stone, R. (1954, September). Linear Expenditure Systems and Demand Analysis: An Application to the Pattern of British Demand. *The Economic Journal*, *64*(255), 511. doi: 10.2307/2227743

Strotz, R. H. (1955, January). Myopia and Inconsistency in Dynamic Utility Maximization. *The Review of Economic Studies*, *23*(3), 165-180. doi: 10.2307/2295722

Thaler, R. H., & Benartzi, S. (2004, February). Save More Tomorrow™: Using Behavioral Economics to Increase Employee Saving. *Journal of Political Economy*, *112*(S1), S164-S187. doi: 10.1086/380085

Thaler, R. H., & Shefrin, H. M. (1981, April). An Economic Theory of Self-Control. *Journal of Political Economy*, *89*(2), 392-406. doi: 10.1086/260971

Thomas, D., & Strauss, J. A. (2007). Health over the Life Course. In J. A. Strauss & T. P. Schultz (Eds.), *Handbook of Development Economics* (Vol. 4, p. 3375-3474). Elsevier. doi: 10.1016/S1573-4471(07)04054-5

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data* (2nd ed ed.). Cambridge, MA: MIT Press.

World Bank. (2016). *5th Ethiopia Economic Update: Why so idle? Wages and Employment in a Crowded Labor Market* (Tech. Rep.). Washington, DC: World Bank Group.

Zizzo, D. J. (2010, March). Experimenter demand effects in economic experiments. *Experimental Economics*, *13*(1), 75-98. doi: 10.1007/s10683-009-9230-z

Sorting Into Incentives for Prosocial Behavior

with Egon Tripodi 3

3.1 Introduction

Many public goods rely on voluntary private contributions. Millions of people every year spend their time working as volunteers in their communities, give money to charity, or donate their own blood, organs, and other tissue. For charities seeking volunteers or money and for health care providers seeking blood donations, it is important to understand how to encourage this prosocial behavior.

An often-used way is to provide extrinsic incentives. The economics literature has found mixed evidence on the effects of monetary and non-monetary incentives on giving (Bowles & Polania-Reyes, 2012; Frey & Oberholzer-Gee, 1997). Although a positive effect of extrinsic incentives is in line with standard economic theory, it goes against a considerable literature in psychology and economics, which argues that they can backfire by either crowding out the intrinsic motivation to give (Deci, 1971, 1975; Titmuss, 1971), or ruining the reputation of donors who could be regarded as greedy (Benabou & Tirole, 2006; Exley, 2017). Field experiments have found evidence for extrinsic incentives to have both negative effects on volunteer work (Frey & Goette, 1999) as well as positive effects on organ

We thank Michèle Belot, Leo Bursztyn, Juan Dolado, Klarita Gërxhani, Lorenz Götte, David Huffman, Sebastian Kube, David K. Levine, Moti Michaeli, Daniele Nosenzo, Antoinette Schoar, Arthur Schram, Marta Serra-Garcia, Joël van der Weele, and Erez Yoeli for helpful comments and suggestions. The chapter also benefited from discussions with Danilo Cavapozzi, Anne Hilger, Jiakun Zheng, and audiences at the EUI, the University of Bonn, SIdE-IEA WEEE 2018, the EDP Jamboree 2017, the EBE Summer Meeting 2018, IMEBESS 2018, SMYE 2018, ESA World Meeting 2018, EEA-ESEM 2018, and the Science of Philanthropy Initiative Conference 2018. Marian Weigt provided outstanding research assistance. Funding from the Russel Sage Foundation (Award #98-17-07), the Institute for Applied Microeconomics at the University of Bonn, and the Department of Economics at the European University Institute is gratefully acknowledged. The authors declare no conflicts of interest. Institutional Review Board (IRB) approvals for pilots of our experimental design were obtained from the Ethics Board of the European University Institute and at MIT with IRB Protocol #1612793044.

(Lacetera, Macis, & Stith, 2014) and blood donations (Lacetera, Macis, & Slonim, 2012, 2014).[1]

While the role of incentives has been analyzed in a wide range of domains, they have been mostly studied in isolation and contrasted to the absence of incentives. In this chapter, we study a setting where different incentives coexist. In this environment, agents can turn down an extrinsic incentive to donate. This lets them reveal and signal their individual preferences through their actions.

Our setting is motivated by the market for human whole blood donations in Germany.[2] In most high-income countries, the concern that incentives could backfire is reflected in tight regulation of how blood donations can be collected. Regulations typically do not allow for monetary payments to donors (Council of Europe, 1995; The Lancet, 2005; World Health Organization, 2009). In many regions of Germany, however, monetary and non-monetary incentives appear to coexist in a "dual market" in which different blood collectors offer different incentives and prospective donors can choose where to donate. Donations at the Deutsches Rotes Kreuz [German Red Cross] (DRK) are always unpaid, while donations at hospitals or commercial blood banks are compensated with 20 to 30 euro.

Very little is known about the features of such "dual markets" for the collection of charitable contributions. Does this system of collection increase donations compared to a single market in which either everyone is unpaid or everyone is paid? What are the determinants of the share of unpaid donations in a dual market? In this chapter we focus on two channels that could help explain sorting into unpaid donations in a dual collection system: altruism and social image concerns.

To guide our analysis, we use a model of charitable giving in which prospective donors are motivated to give by intrinsic incentives, extrinsic incentives, and image concerns. We build on the framework by Benabou and Tirole (2006), but introduce two modifications: first, we change the payoff structure so that a potential compensation for the donation is paid out of the value that is generated by the donation. This tension between private and public benefit of the donation introduces a channel through which extrinsic incentives can crowd out intrinsic motivation. Second, we assume that donors do not differ in how much they value extrinsic incentives. This lets us make clear predictions, but comes at the cost of ruling out "reputational crowding out", that is we rule out that donors can have a negative response to the introduction of extrinsic incentives out of concern for appearing greedy. We derive three testable behavioral hypotheses from this model.

[1] Aside from the question of effectiveness, incentives to donate human tissue might be seen as controversial on moral grounds. Only limited incentives appear to be morally acceptable among a sample of people surveyed in the United States (Boulware, Troll, Wang, & Powe, 2006). Becker and Elias (2007) provide a compelling argument in favor of allowing incentives for organ donations. Lacetera (2016) summarizes the debate. In this chapter, we will abstract from the matter of the morality of incentives.

[2] The most common type of human blood donation is a "whole blood" donation, in which approximately one pint of blood is collected over a period of about ten minutes. Men can donate up to six times per year, women up to four times per year. Red blood cells from whole blood donations are typically used for transfusions to other patients and are most commonly seen as motivated by altruistic preferences (Niessen-Ruenzi, Weber, & Becker, 2015). Other types of blood donations include platelet and plasma donations, which take much longer and require donors to be connected to a machine. Donors are commonly compensated in cash for these types of donations.

The first testable hypothesis states that the availability of compensation to donate should increase donations. We call this the "incentive effect". The second hypothesis states that irrespective of whether compensation is available, making actions visible should increase donations. We call this the "social image effect". Our third and novel hypothesis states that in a dual market, where agents can turn down compensation, a positive share of agents will choose to remain unpaid and that this share is larger when actions are taken in public. We call this "sorting", based on the idea that a dual market can bring about efficiency gains in the collection similar to those deriving from self-selection in second-degree price discrimination.

We test these three hypotheses in a laboratory experiment with 329 student subjects. For three rounds, each subject is confronted with the decision to participate in a real effort task. This task generates value for a charity under one of three market designs: donors receive no compensation for the donation (single market *NOT PAID*), donors always receive a compensation for the donation (single market *PAID*), and donors can choose whether they want to receive compensation for the donation (dual market *CHOOSE*). Like for the case of blood collection, any compensation paid out to donors reduces the social value of the donation. This is objectively measured in our controlled setup by the amount of money that goes to charity. We also vary the visibility of actions (*PRIVATE* vs. *PUBLIC*). The combination of market design treatments and visibility treatments in a full 3×2 design produces six distinct treatments, which we run between subjects.

The experimental results mostly support our behavioral hypotheses. We find clear evidence for the incentive effect. In the dual market, the availability of incentives does not crowd out intrinsic motivations of donors, irrespective of whether actions are observable. Moving from a single unpaid market to a dual market significantly increases the number of donations of our experimental subjects.

We also find evidence of strong social image effects. Making actions observable significantly increases donations in all three incentive schemes. Finally, we find support of our sorting hypothesis: when given the option to turn down compensation, a significant share of donors chooses to do so, though we do not find a significant difference between actions taken in private and in public.

Interestingly, and in contrast to similar studies that analyze the effectiveness of conditional and unconditional incentives to act prosocially (Ariely, Bracha, & Meier, 2009; Carpenter & Myers, 2010), we do not find that social image effects attenuate incentive effects. We differ from Ariely et al. (2009) in that subjects decide to donate in the presence of an outside option. Our results suggest that when incentives are small and only partly offset the costs of donating, social image effects and incentive effects need not crowd each other out. In addition, we find heterogeneous effects of social image on contributions that we attribute to gender-specific preferences over signaling. Overall, our findings suggest novel ways to improve mechanisms for the collection of charitable donations by leveraging heterogeneity in individual preferences. Applied to the collection of blood donations, our results may inform the design and regulation of systems that use monetary incentives.

The remainder of the chapter is organized as follows: Appendix 3.2 fixes ideas in a simple theoretical framework and presents testable behavioral hypotheses. Appendix 3.3 details experimental design and procedures. Appendix 3.4 presents the results. Appendix 3.5 concludes with a discussion of the implications of our findings for the market for blood that initially motivated our research.

3.2 Theoretical Framework

In the model by Benabou and Tirole (2006) (BT), being compensated to donate can crowd out donations by spoiling the image of donors. Moreover, any compensation is paid from resources that are exogenous to the economy and is given to donors without affecting the social value of their donation. BT show that whether donors can turn down compensation should not matter, because neither image-indifferent nor image-concerned agents would want to do so. For image-indifferent agents, it would be a dominated strategy to turn down compensation that does not affect the social value of their donation. Image-concerned agents would be worried that their motivation is questioned: turning down incentives could reveal that they are not acting out of altruism, but just to appear as altruistic while in fact (on average) they are not.

For a dual market like in Germany, where prospective donors can choose from a menu of options, the model would thus predict that no one should turn down compensation. Yet we observe that a considerable share of donors chooses to remain unpaid when they have the choice between donating with a 20 to 30 euro compensation or donating without any compensation. Informational frictions and transportation costs may explain part of this outcome, though these do not appear to be empirically significant.

We suggest that a different payoff structure than the one by BT better fits the case of blood donations and many other charitable activities and could explain why prospective donors would choose to turn down incentives. In our version of the model, any potential compensation for the donation is paid out of the value that is generated by the donation. The collector of donations is a charitable organization that transforms collected donations into social value. To increase donations, the collector may find it optimal to pay donors a dividend from their donation as compensation. Increasing private returns from the donation comes at the expense of the value that the donation generates for the rest of the society. This feature of our setup introduces an additional channel through which incentives could potentially crowd out donations: a crowding out of intrinsic motivation. This channel is consistent with an earlier literature stemming from Deci (1971, 1975).

To formulate testable predictions that are directly relevant to our research question, we will substantially simplify the original model by BT. One key simplification is that we assume agents to be homogeneous in their taste for extrinsic incentives. When this is the case, there is no scope for signaling greediness (or a lack thereof). Despite being a common assumption in economics, a potential drawback of making this simplification is that it prevents the reputational crowding out from BT, i.e. a situation where extrinsic incentives reduce the donations of agents who seek to avoid signaling greediness through their actions.

3.2.1 Simple Model

The model economy is characterized by a unit mass of agents indexed by $i = \{1, ..., \infty\}$ and one collector of donations. This economy is analyzed under two different institutional settings. We refer to a *single* market when the collector is bound to pay an exogenously-set compensation $y = \tilde{y} \in \mathcal{R}_+$. We refer to a *dual* market when agents are allowed to choose remuneration $y = \{0, \tilde{y}\}$.

The ***collector*** takes donation d from each agent that decides to contribute and transforms it into social value $B \in \mathcal{R}_+$. For each contribution, the collector pays remuneration $y < B$.

Agents differ along two dimensions: the degree of altruism $a_i \sim F(\cdot)$ with positive bounded support, and the concern for image x_i, which we treat as binary with x_i taking value 1 with probability q (and 0 with probability $1-q$). Both a_i and x_i are independently distributed random variables. Agents make a decision to contribute $d = \{0, 1\}$ in exchange for remuneration y while facing a private cost c. Image concern matters for agents when actions are taken in public ($v = 1$) and is irrelevant when actions are taken in private ($v = 0$).

The utility of agent i can be written as follows:

$$U_i(d, y) = (1 - vx_i)[a_i(B - y) + y - c]d + vx_iE(a|d, y) \tag{3.1}$$

where $E(a|d, y)$ is the image that other agents have of agent i given her actions.

From this theoretical setting we derive two propositions that underpin our analysis:

Proposition 3.2.1 (Price discrimination) *A dual market for donations increases contributions compared to a single market where no compensation is available. Compared to a single market where compensation cannot be turned down, allowing agents to turn down compensation reduces the cost of collection without affecting the number of donations.*

The proof is presented in Appendix 3.A.

The proposition characterizes the effect of various compensation schemes on donations. It applies when actions are taken in private and in public. Introducing extrinsic incentives to donate increases donations, irrespective of whether these incentives can be turned down. Allowing people to turn down incentives, introduces another margin for people to either express or signal their altruism. Highly altruistic agents donate and choose to turn down the compensation.

As a result, when incentives can be turned down, average cost of collection decreases without compromising supply of donations. These two results illustrate how a dual market, where agents are allowed to choose a remuneration, can bring about efficiency gains in the collection similar to those deriving from self-selection in second-degree price discrimination.

The following proposition is directly linked to the previous and highlights the interaction of image effects with price discrimination.

Proposition 3.2.2 (Image effect) *The visibility of actions (i) increases participation in the single as well as in the dual market, and (ii) lowers the average cost of collection in the dual market.*

The proof of (ii) follows directly from the observation that the objective of image-concerned agents who are sufficiently altruistic to donate in private, but not altruistic enough to turn down compensation $y = \tilde{y}$, changes when acting in public. In order to improve their social image, these agents want to pool with the most altruistic agents, who turn down incentives.[3] Part (i) is due to the fact that image-concerned agents only care about their image when acting in public. As a result, even the least-altruistic of these decide to contribute in public in order to avoid the stigma of looking like the selfish segment of the population.

[3] This signaling game may not have an equilibrium in pure strategy if the share of image-indifferent agents who are altruistic enough to turn down the incentives is positive but small compared to the share of image-concerned agents.

3.2.2 Behavioral Hypotheses

We re-organize the predictions contained in the two propositions above into three testable hypotheses. The *incentive effect* and *social image effect* hypotheses immediately derive from propositions 1 and 2, respectively. The *sorting* hypothesis consolidates predictions from both proposition to summarize the interaction of social image effects and incentive effects in the dual market for charitable giving.

Hypothesis 3.2.3 (Incentive Effect) *Irrespective of whether actions are visible, the availability of incentives increases donations.*

Hypothesis 3.2.4 (Social Image Effect) *Irrespective of whether compensation is available, making actions visible increases donations.*

Hypothesis 3.2.5 (Sorting) *In a dual market, a positive share of agents chooses to be not paid. This share is larger when actions are taken in public.*

The incentive effect is consistent with an empirical literature on incentives for donating blood (Lacetera et al., 2012; Lacetera, Macis, & Slonim, 2013; Mellstrom & Johannesson, 2008; Niessen-Ruenzi et al., 2015). Maybe most closely related to ours is the work by Mellstrom and Johannesson (2008), who conduct an experiment that offers monetary payments to prospective blood donors. Their findings suggest that for women (but not for men), monetary incentives can lead to a net crowding out of donations – thought it is difficult to say whether the results are driven by social signaling or by the fact that incentives lead to a shift in the perception of the incomplete contract, similar to the finding of Gneezy and Rustichini (2000). Moreover, they find that letting women turn down the compensation in favor of a donation to charity fully counteracts this crowding out. Our theoretical setup can partly explain this counteracting effect, in that for the most altruistic donors ($a_i > 1$) introducing incentives for charitable giving causes a net utility loss. Such utility loss can be undone when incentives can be turned down in the dual market. In related work, Chao (2017) suggests that even opt-in gifts could crowd out donations if they shift attention away from the intrinsic motivation. In our framework, we abstract from attention as a potential channel for crowding out. The social image effect is consistent with a growing empirical literature on the effect of social image or social pressure on charitable actions in particular and economic behavior more generally (Ariely et al., 2009; Bursztyn & Jensen, 2017; Carpenter & Myers, 2010; Filiz-Ozbay & Ozbay, 2014; Lacetera & Macis, 2010). Our theoretical setup predicts that, no matter the incentive scheme, making actions visible should increase donations. Consistent with our prediction, Landry, Lange, List, Price, and Rupp (2006) find that both when a charity donation entitles to a lottery ticket and when it does not, social image concerns do increase monetary donations in a door-to-door fundraiser. They also find pronounced gender differences, where men are more likely to contribute to a charity when visited by physically attractive female solicitors. The finding that men are more willing to engage in costly signaling of generosity is consistent with costly signaling theory in evolutionary biology (Gintis, Smith, & Bowles, 2001; Smith & Bird, 2000), which posits that prosocial behavior can be instrumental in signaling good character and attractiveness as a potential match. In particular, there is evidence that women in their mating decision place emphasis on signals indicating resource provision (as opposed to just physical attractiveness), which in turn induces men to strategically signal generosity (Barclay, 2010; Boehm & Regner, 2013; Eagly & Crowley, 1986; Iredale, Van Vugt, & Dunbar, 2008). Van Vugt and Iredale (2013) call men's public good contributions the "human equivalent of a peacock's tail". Although our theoretical setup is silent on gender differences, we are going to investigate these empirically.

Finally, we are not aware of any empirical evidence on the sorting hypothesis as formulated above. It is not obvious whether prospective donors should increase donations when the choice set is augmented in a way to allow signaling of prosocial orientation either through increased donations or by turning down incentives to donate. A large body of evidence on pure and impure altruism suggests that even when donations are completely private, a positive share of prospective donors presented with the possibility to contribute time and effort – with or without compensation – would choose to donate not paid.[4] Signaling motives should increase the latent utility of acting prosocially. Increasing the visibility of actions could strengthen the signaling motive, potentially increasing the share of unpaid donations. The theory of Benabou and Tirole (2006) accommodates sorting as described above, but is hard to test empirically. In our theoretical framework, we chose to make substantial simplifications in order to derive testable hypothesis. We take our experiment as a first step to validate this simplified framework and to test simple hypotheses that could guide the field and inform policy on the properties of dual collection systems for charitable donations.

3.3 Experimental Design and Procedures

3.3.1 General Setup

We test our hypotheses in a laboratory experiment. In our experiment, subjects generate value for a charity by participating in a real-effort task. For the experimental task, we build on the "click for charity" design by Ariely et al. (2009). Different from Ariely et al. (2009), subjects in our framework can choose between participating in the donation task or skipping the task and taking a fixed payoff as outside option.[5] This outside option introduces an homogeneous private cost of donating on top of the individual cost of exerting effort. If subjects choose to participate, they can generate a donation by sequentially entering 400 key sequences on a computer keyboard. One sequence constitutes of four key presses ("*w*", "*e*", "*e*", "*return*"). On their screen, subjects see a bar indicating progress towards the required number of sequences. We chose this task because it is not inherently meaningful or intrinsically rewarding, and allows us to focus on motivation to exert effort for a charity. Other tasks, particularly ones that are more gamified, may be differentially appealing to subjects and thus increase noise and confounds (Charness, Gneezy, & Henderson, 2018). Donations generated with this real-effort task are paid out to a charity chosen by each subject.f

[4]See Ottoni-Wilhelm, Vesterlund, and Xie (2014) for a review of the pure and impure altruism literature.

[5]Without the outside option, the marginal cost of participating in the task could be low enough for lab subjects to be indifferent between exerting effort and waiting while others exert effort. The outside option increases the costs of participating in the donation task, so that subjects that are not altruistic and not concerned about social image should not participate in the task – as predicted by the model.

We employ a full 3×2 between-subject design where we systematically vary the type of incentives offered to engage in the donation task (*PAID, NOT PAID*, *CHOOSE*) and the visibility of actions (*PUBLIC* and *PRIVATE*). Visibility is randomly varied across experimental sessions while the incentives offered are randomly varied across all subjects. Table 3.1 summarizes the design.[6]

Table 3.1: Overview of Treatments

	Not paid $y = 0$	Paid $y = \tilde{y}$	Choose $y \in \{0, \tilde{y}\}$
Private Action $v = 0$	n = 46	n = 48	n = 60
Public Action $v = 1$	n = 47	n = 62	n = 66

Notes: Rows list visibility treatments, columns list incentive treatments.
n refers to number of subjects in each treatment cell (total of 329 subjects).
y refers to the incentive provided, v to the visibility of actions.

After being assigned to one of six treatments, subjects independently engage in the donation task. After the first round, subjects learn that there will be two more rounds of this task. This lets us test our hypotheses both on the extensive and the intensive margin. Irrespective of the treatment, in each of the three rounds can choose between participating in the donation task or skipping. Throughout the experiment, we use tokens as experimental currency. One token is worth 0.04 euro.

3.3.2 Treatments

Along the first dimension of the 3×2 between-subject design we vary the market design, i.e. the availability of incentives to participate in the donation task. In the first two treatments, we either provide monetary incentives to participate in the donation task (single market *PAID* treatment) or no monetary incentives (single market *NOT PAID* treatment). In the third treatment (dual market *CHOOSE* treatment), subjects are presented with both the options of a not paid and a paid donation.

The payoffs are set such that donating generates more value for the charity (100 tokens) than the outside option for the subject (75 tokens). When subjects donate and receive mon-

[6] We conducted a pilot study of our experimental design online on Amazon Mechanical Turk ($N = 408$) to inform the choice between a within-subject and a between-subject design. To address concerns that a crowding-out effect of incentives may arise either only in an environment where incentives are introduced as a policy change (within-subject) or only in a market design where people are unaware of alternative institutional environments, we also considered an experimental design that allowed us to study the transition from a single market *NOT PAID* or single market *PAID* market design to a dual market *CHOOSE* market design. In this alternative design, we introduced the dual market to subjects after a first round in the single market design. We did not find evidence that the single market design has any persistent effects. Between- and within-subject designs led to qualitatively similar results. We conclude that the initial treatment has no impact on the effectiveness of the *CHOOSE* treatment. For the current project, we opt for a between-subject design to minimize potential confounders and demand effects (Charness, Gneezy, & Kuhn, 2012). Appendix 3.E summarizes the pilot.

etary incentives for their donation (50 tokens), those reduce the value to charity (from 100 to 50 tokens). Note that the monetary incentives are always smaller than the outside option. Table 3.2 summarizes the choice set in each of the three treatments and the associated monetary payoffs in tokens.

Table 3.2: Payoffs to Subject and Benefits to Charity, by Treatment and Subject Choice

Treatment	Action space	Payoff to subject	Benefit to charity
NOT PAID	Donate not paid	0	100
	Skip	75	0
PAID	Donate paid	50	50
	Skip	75	0
CHOOSE	Donate not paid	0	100
	Donate paid	50	50
	Skip	75	0

Notes: Experimental Currency: "tokens", 1 token = 0.04 euro

Along the second dimension of the 3×2 between-subject design we vary the visibility of subject actions to make public image salient. In the *PRIVATE* treatment, subjects are informed that their actions will remain anonymous. Subjects are seated at desktop computers separated by divider walls and curtains. To maximize anonymity and to rule out that subjects hear each other type while working on the real-effort task, we play a white noise sound using loudspeakers in the laboratory. We verified that the white noise indeed makes it impossible to hear typing from other workstations. We did not receive any complaints from subjects about this measure. In the *PUBLIC* treatment, before beginning the donation task, we inform subjects that they will be asked to reveal their actions in this task in front of all other subjects in this session. Social image effects thus reflect the full decision environment, including the incentive choice in the dual market *CHOOSE* treatment, that each subject is in. After completing all three rounds we ask subjects to publicly report the number of donations they made.[7] Subjects do so by standing up next to their computer in front of the divider walls. There is no explicit requirement to truthfully report this information.[8] Note, however, that reporting takes place after all decisions have been made.

3.3.3 Procedures

Our theoretical framework asserts that more altruistic individuals are, ceteris paribus, more likely to donate to charity. To check that individual levels of altruism are balanced across treatments, we let all subjects play a simple dictator game before beginning the main ex-

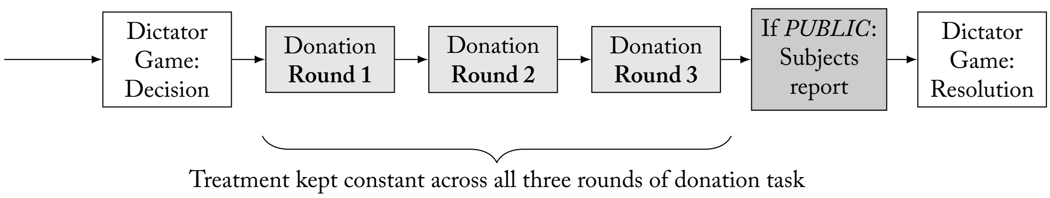
[7] The intention behind not having subjects reveal the incentives under which they donated was to avoid confusion from learning about other treatments.

[8] As an alternative design, we could have publicly announced actual subject choices at the end of the experiment. We decided against this design to stay closer to the theoretical framework of Benabou and Tirole (2006), where the desire to signal altruism has both instrumental and hedonic origins. We allow for both motivations by letting subjects state their own actions. To maintain the ecological validity of revealing a prosocial action, we do not force subjects to say the truth.

perimental task that lets subjects donate to charity.[9] In this dictator game, each subject is randomly and anonymously paired with another subject and chooses to split 20 tokens between herself and the anonymous partner. After testing for subject comprehensions, we let both subjects of the pair play the game as the dictator. At the end of the experiment, the experimental software randomly determines which of the two subjects determines payoffs and the game is resolved.

We then introduce a menu of four charities. Three of those charities are chosen because they are assumed to be well-known among subjects: Doctors Without Borders, the International Committee of the Red Cross (ICRC), and the World Wildlife Fund (WWF). We additionally included the Against Malaria Foundation, which is rated as one of the most effective charities by the independent charity evaluator GiveWell. Subjects are given a short description of each charity. We then let each subject choose the charity that they prefer to donate to throughout the experiment. We do this to reduce potential noise from heterogeneous taste for donations to a specific charity. In order to verify balance across treatments, we ask subjects to rate how they perceive each of the charities and how likely they would be to donate money to each of them. Finally, we let subjects practice the donation task before engaging in it for three rounds. In the *PUBLIC* treatment, subjects publicly report their actions after the third round of the donation task. Figure 3.1 summarizes the sequence of tasks in the experiment.

Figure 3.1: Sequence of the Experiment



At the end of the experiment, we collect demographic data. After each session, we confidentially pay out the show-up fee and any earnings that subjects have generated for themselves in the dictator game and the donation task. We also inform subjects about the amount of money donated to charity on their behalf and provide information on how to obtain a confirmation of the donation on their behalf.

We implement the computerized experiment in oTree with our own modifications written in Python and JavaScript (Chen, Schonger, & Wickens, 2016). A total of 18 experimental sessions were conducted in German at the BonnEconLab in Bonn, Germany, in April 2017 ($n = 329$). Sessions included 20 to 24 subjects and lasted approximately 40 minutes. All subjects are students from various majors at the University of Bonn. They are on average 22 years old, 61 percent are female. Table 3.3 summarizes the sample. On average, participants earned 10.70 euro for themselves and generated 4 euro for charity.[10]

[9] While giving in the dictator game is a well-established measure of generosity vis-à-vis others, it is likely confounded by perceived social norms. As a result, we only rely on our measure of altruism as a balance check, but not to establish key empirical results or to analyze heterogeneous treatment effects.

[10] Subjects from the pool of the BonnEconLab were invited using hroot (Bock, Baetge, & Nicklisch, 2014). Invitations were restricted to students of the University of Bonn, aged 18–25, with no more than one no-show

We can verify that the sample is balanced on observable characteristics, including our measure of altruism measured by the dictator game and preference for the chosen charity. Using a nonparametric one-way ANOVA on ranks (Kruskal-Wallis) test, we fail to reject the null hypothesis that the subject pool exhibits the same characteristics across all treatment groups at the 95 percent level (Table 3.3, column 8).

Table 3.3: Summary Statistics of Observable Characteristics, Full Sample and by Treatment (Means and Standard Errors in Parentheses)

	Full Sample	Private			Public			p-value
		Not paid	Paid	Choose	Not paid	Paid	Choose	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
a) Measured before treatment								
DG: Tokens kept	15.365	14.891	15.271	15.250	15.021	15.677	15.818	0.848
	(0.214)	(0.621)	(0.558)	(0.507)	(0.618)	(0.501)	(0.411)	
Charity rating	4.602	4.783	4.604	4.583	4.660	4.532	4.515	0.131
	(0.043)	(0.087)	(0.129)	(0.072)	(0.102)	(0.123)	(0.100)	
b) Socioeconomic characteristics, measured after treatment								
Age	21.544	21.630	21.708	21.717	21.511	21.210	21.545	0.499
	(0.091)	(0.263)	(0.223)	(0.213)	(0.263)	(0.184)	(0.207)	
Female	0.611	0.630	0.521	0.717	0.574	0.613	0.591	0.429
	(0.027)	(0.072)	(0.073)	(0.059)	(0.073)	(0.062)	(0.061)	
College major	4.398	4.239	4.417	4.400	4.383	4.661	4.258	0.814
	(0.100)	(0.277)	(0.258)	(0.224)	(0.273)	(0.236)	(0.221)	
Observations	329	46	48	60	47	62	66	

Notes: p-value in column (8) is for a one-way ANOVA on ranks (Kruskal-Wallis) test comparing the six treatment groups in columns (2) to (7). DG refers to the dictator game, in which we gave 20 experimental tokens to participants and asked them how many they would like to keep. Charity rating refers to the rating that subjects gave to the charity that they chose to donate to. We asked subjects to agree to the statement "I like the idea of donating money to [chosen charity]" on a 5-point Likert scale where 1 is "strongly disagree" and 5 is "strongly agree". College major is a categorical variable that summarizes the departmental affiliation of our student subjects.

3.4 Results

Recall that in each of the three rounds of the donation task, subjects can decide to participate in or skip the task. In our discussion of results, we consider each participation in the task as one "donation" (all subjects who choose to participate in the donation task complete it). Participation in the first round of the donation task lets us measure the extensive margin of the donation decision. By summing the number of donations across all three rounds, we can additionally analyze an intensive margin of the decision to donate.

Table 3.4 summarizes those measures and gives an overview of donation behavior across treatments. Panel I presents the fraction of subjects who decide to participate in each round while panel II sums the number of rounds that subjects decide to participate in the donation task. For subjects in the dual market *CHOOSE* treatment, columns (4) and (5) report

in prior experiments.

whether subjects choose to be paid. In line with our theoretical predictions, donation behavior in the single market *PAID* and the dual market *CHOOSE* treatments is statistically indistinguishable (column 6), both on the extensive margin and the intensive margin.

Table 3.4: Summary Statistics of Behavior in Donation Task (Fractions and Means, Standard Errors in Parentheses)

	Incentive Treatment			Incentive Choice		p-value
	Not paid (1)	Paid (2)	Choose (3)	Not paid (4)	Paid (5)	H_0: Paid=Choose (6)
I. Fraction of subjects that participated in the task						
a) PRIVATE treatment						
Round 1	0.609	0.604	0.667	0.083	0.583	0.504
	(0.072)	(0.071)	(0.061)	(0.036)	(0.064)	
Round 2	0.174	0.396	0.467	0.083	0.383	0.463
	(0.056)	(0.071)	(0.065)	(0.036)	(0.063)	
Round 3	0.348	0.313	0.383	0.067	0.317	0.446
	(0.070)	(0.067)	(0.063)	(0.032)	(0.061)	
Observations	46	48	60	60	60	
b) PUBLIC treatment						
Round 1	0.766	0.806	0.818	0.136	0.682	0.866
	(0.062)	(0.050)	(0.048)	(0.043)	(0.058)	
Round 2	0.383	0.565	0.591	0.136	0.455	0.763
	(0.071)	(0.063)	(0.061)	(0.043)	(0.062)	
Round 3	0.362	0.484	0.530	0.136	0.394	0.601
	(0.070)	(0.064)	(0.062)	(0.043)	(0.061)	
Observations	47	62	66	66	66	
II. Average total number of rounds participated in the task						
a) PRIVATE treatment						
Sum of all 3 rounds	1.130	1.313	1.517	0.233	1.283	0.290
	(0.129)	(0.142)	(0.135)	(0.072)	(0.132)	
Observations	46	48	60	60	60	
b) PUBLIC treatment						
Sum of all 3 rounds	1.511	1.855	1.939	0.409	1.530	0.545
	(0.124)	(0.121)	(0.127)	(0.105)	(0.136)	
Observations	47	62	66	66	66	
c) Aggregating over both visibility treatments						
Sum of all 3 rounds	1.323	1.618	1.738	0.325	1.413	0.348
	(0.092)	(0.095)	(0.094)	(0.066)	(0.096)	
Observations	93	110	126	126	126	

Notes: Total sample size is 329 subjects. Subjects can always choose between participating in the donation task or skipping. P-value in column (6) is for two-sample Wilcoxon rank-sum (Mann-Whitney) test comparing the outcomes for *PAID* treatment in column (2) and the *CHOOSE* treatment in column (3).

In the rest of this section, we pool together observations from *PAID* and *CHOOSE* treatments to estimate the effects of the availability of incentives on donations behavior. We use this pooled data to provide parametric tests of Hypotheses 1 and 2 on the intensive margin. Appendix 3.C establishes the same results using non-parametric tests. We then use data from the dual market *CHOOSE* treatment to test Hypothesis 3, again on the intensive margin. We test our three hypotheses on the intensive margin due to better statistical power. Results are qualitatively similar on extensive margin based on the first round of the donation task. In addition to tests of our theoretical hypotheses, we discuss the potential interaction between incentive and visibility effects and analyze heterogeneous treatment effects across genders.

3.4.1 Incentive Effects, Social Image Effects, and Sorting

We test our first two hypotheses in a regression framework. Given the count nature of the outcome variable we use maximum likelihood to estimate the following Poisson regression:

$$\begin{aligned} Donations_i = \alpha + \beta_1 PAID\&CHOOSE_i + \beta_2 PUBLIC_i + \\ \beta_3 PAID\&CHOOSE_i \times PUBLIC_i + \boldsymbol{X}_i\gamma + \psi_i \end{aligned} \tag{3.2}$$

where $Donations$ is the total number of donations by subject i over all three rounds of the donation task, *PAID&CHOOSE* is a dummy for the pooled single market *PAID* treatment and the dual market *CHOOSE* treatment, *PUBLIC* is a dummy for the treatment in which subjects have to reveal their actions to other participants, $\boldsymbol{X}$ is a vector of controls, and ψ is a Poisson-distributed error term. Table 3.5 presents average marginal effect estimates while Appendix Table 3.B1 presents the full set of estimated semi-elasticities.

Our results confirm our first behavioral hypothesis, which says that irrespective of whether actions are visible, the availability of incentives increases donations. We find that compared to the single market *NOT PAID* treatment, the availability of incentives does not induce lower participation in the donation task. This is true irrespective of the visibility of actions. The estimated average marginal effect in our specification without any other controls indicates that making incentives available leads to an increase of 0.364 donations over all three rounds (relative to a mean of 1.32 donations in the single market *UNPAID* treatment). The effect size is robust to various sets of controls. Introducing the number of tokens kept in the dictator game as an additional control (Table 3.5, columns 3 to 5) reveals that this measure of altruism is a strong predictor of participation in the donation task.

Result 3.4.1 (Incentive Effect) *Irrespective of whether actions are visible, the availability of incentives increases donations.*

We also find support for our second hypothesis of social image effects. Using the same Poisson regression in Equation (3.2), we find that irrespective of the incentive treatment, making actions visible significantly increases the number of donations over all three rounds. The effect of visibility is of similar magnitude to the incentive effect and is similarly robust to various sets of controls.

Result 3.4.2 (Social Image Effect) *Irrespective of whether compensation is available, making actions visible increases donations.*

We can use our experimental design to assess the potential interaction between incentive and visibility effects. A prominent result in the literature on charitable giving is that incentive effects negatively interact with image effects (Ariely et al., 2009). In our framework, in contrast, we do not find a negative interaction between image effects and incentive effects. In the presence of a salient outside option, small incentives to donate do not appear to spoil the image of donors. Appendix Table 3.B1 presents semi-elasticities estimated from Equation (3.2), including for the interaction-term. We estimate a zero interaction effect that is robust across specifications.

Table 3.5: Poisson Regression for Total Donations: Average Marginal Effects (Coefficient Estimates and Standard Errors in Parentheses)

Dependent variable:	# of donations over the three rounds				
	(1)	(2)	(3)	(4)	(5)
	a) Treatments				
PAID&CHOOSE *(Baseline: NOT PAID)*	0.364***	0.360***	0.432***	0.430***	0.456***
	(0.124)	(0.124)	(0.117)	(0.117)	(0.117)
PUBLIC *(Baseline: PRIVATE)*	0.454***	0.462***	0.498***	0.499***	0.494***
	(0.112)	(0.111)	(0.103)	(0.102)	(0.102)
	b) Controls				
Female		0.238**		0.075	0.030
		(0.116)		(0.110)	(0.110)
DG: Tokens kept			-0.099***	-0.097***	-0.090***
			(0.013)	(0.013)	(0.014)
Other Controls	No	No	No	No	Yes
Observations	329	329	329	329	329

$^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$

Notes: Standard errors are clustered at individual level. *NOT PAID* is the base market design treatment. *PRIVATE* is the base visibility treatment. DG refers to the dictator game, in which we gave 20 experimental tokens to subjects and asked them how many they would like to keep. Other controls are age, chosen charity, and individual rating of chosen charity. Note that due to the presentation of average marginal effects, the interaction (which cannot vary independently) is omitted.

Finally, our third behavioral hypothesis states that in a dual market, a positive fraction of donors chooses to be not paid, and that this fraction is larger when actions are observable. We can test this hypothesis by looking at all subjects in the dual market *CHOOSE* treatment.

In each of the three rounds and in each visibility treatment, the fraction of subjects deciding to not be paid for their donation is significantly larger than zero (Figure 3.2). Aggregating over the three rounds, subjects choose to make 0.23 donations without being paid in *PRIVATE* and 0.41 donations without being paid in *PUBLIC* (Table 3.4, panel II, column 4). This confirms the first part of our third hypothesis.

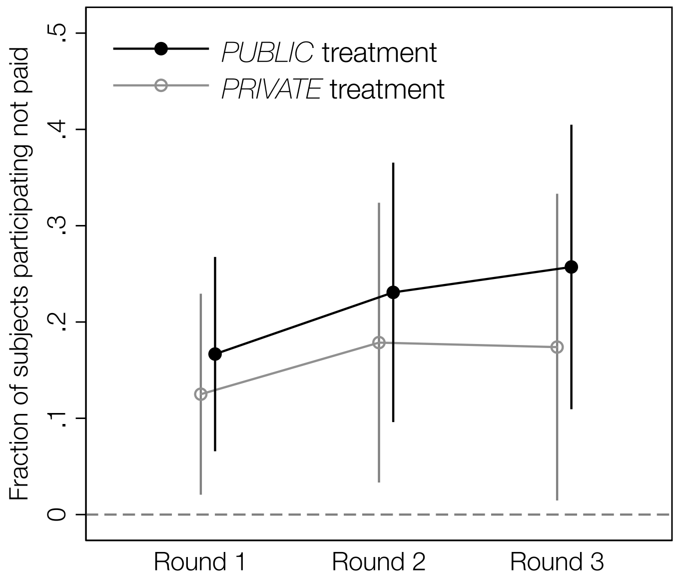
Result 3.4.3 (Sorting Into Unpaid) *In a dual market, a positive share of agents chooses to be not paid.*

In order to analyze sorting into unpaid donations in the dual market *CHOOSE* treatment across visibility conditions, we estimate the following multinomial logit random effect model for the donation decision and the chosen incentive scheme. Each subject i takes decision $d_i \in \{$no participation, unpaid participation, paid participation$\}$:

$$d_{i,t} = \alpha + \beta PUBLIC_i + \boldsymbol{X}_i \boldsymbol{\gamma} + v_{i,t} \tag{3.3}$$

where for each subject i and round t, *PUBLIC* is a dummy for the treatment in which subjects have to reveal their actions, $\boldsymbol{X}$ is a vector of controls, and $v_{i,t} = c_i + u_{i,t}$ is the error term of the random effect model. Treatment assignment is permanent, but exogenous. While time invariance of treatment assignment makes the fixed effect model unidentifiable,

FIGURE 3.2: Fraction of Participating Subjects Turning Down Incentive in Donation Task, by Round



Notes: Bars indicate 95 percent confidence intervals. Standard errors clustered at the individual level.

exogenous treatment assignment meets the random effect assumption and makes this model specification the natural choice.[11]

The multinomial logit random effect model provides estimates for the relative probability of observing not paid rather than paid donations in the *CHOOSE* treatments. In the regression specification without controls, the relative probability increases by 77.3 percent when actions are visible, and the effect size is fairly stable in specifications with controls (see Table 3.B2). While this confirms qualitatively the pattern from Figure 3.2, this increase is not statistically significant. We are not powered to detect a relative risk ratio that is significantly different from unity at any conventional confidence level.

3.4.2 Heterogenous Social Image Effects Across Genders

We find gender-specific effects in the *PUBLIC* treatment that suggest a differential willingness to engage in costly signaling: Making actions visible increases participation in the donation task significantly among men in the *NOT PAID* and *CHOOSE* treatment. For women, we find the inverse in that the increase is only significant in the *PAID* treatment.

Paralleling the analysis above, we use maximum likelihood estimates of a Poisson regres-

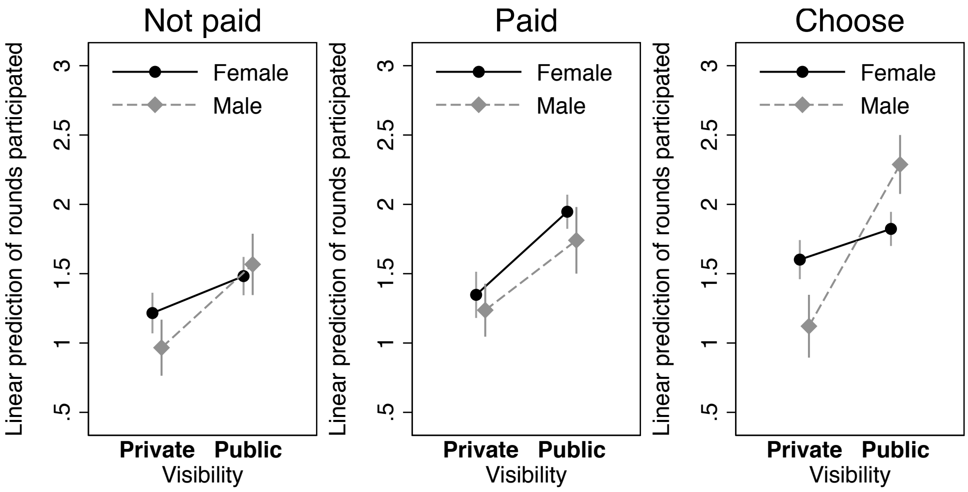
[11] Any specification of the regression equation that includes individual characteristics is prone to bias and would require testing of the random effects assumption.

sion. For each incentive treatment, we separately estimate a model of the form:

$$\begin{aligned} Donations_i = &\alpha + \beta_1 FEMALE + \beta_2 PUBLIC \\ &+ \beta_3 (FEMALE \times PUBLIC) + \beta_4 DG + \psi_i \end{aligned} \tag{3.4}$$

where for each subject i, $Donations_i$ is a count variable for number of individual donations over the three rounds of the donation task, and *DG* is the number of tokens kept in the dictator game. Table 3.B3 presents estimates of the semi-elasticities, which reveal that the social image is significantly different across genders only in the dual market *CHOOSE* treatments. Figure 3.3 provides graphical illustration of the interaction effect by plotting the predicted participation in the donation task for each subsample. The heterogeneous effect of public image is particularly salient in the dual market *CHOOSE* treatment.

Figure 3.3: Gender-Specific Effects of Visibility Treatment, by Incentive Treatment (Linear Prediction of Rounds Participated, Based on Regressions in Table 3.B3)



Notes: Bars indicate 95 percent confidence intervals. Standard errors clustered at the individual level.

We take this as suggestive evidence that men are more willing than women to engage in costly signaling. Recall that in our framework, choosing to participate in the donation task represents a signal that is differentially costly across the three donation treatments. Choosing to participate without being paid (either in the *NOT PAID* or *CHOOSE* treatments) carries the largest reputational gains, since subjects who engage in the real effort task incur the highest opportunity cost by leaving all value to the charity (i.e. they forego the outside option). In the *PAID* treatment, subjects can signal their altruism at a lower opportunity cost (i.e. they forego the outside option minus the individual compensation).

3.5 Discussion and Conclusion

Motivated by the market for blood donations in Germany, where different incentives for altruism coexist and donors can effectively turn down monetary incentives to donate, we set out to study a "dual market" for the collection of charitable donations. While incentives for prosocial behavior have mostly been studied in isolation and contrast to the absence of incentives, we explicitly allow agents to turn down a compensation for their donation.

In the case of blood donations in Germany, different blood collectors offer different incentives and prospective donors can choose where to donate. Donations at the Red Cross are always unpaid, while donations at hospitals or commercial blood banks are compensated with 20 to 30 euro. Everyone who lives in one of the 50 largest communities in Germany can reach an unpaid donation point of the Red Cross within 30 minutes time driving or on public transport. This compares to about 62 percent of the population who can reach a paid donation point within 30 minutes time using the same means of transport (see Table 3.D2 for details and and Figure 3.D2 for the spatial distribution of blood collection centers). In Meyer and Tripodi (2018), Chapter 3 of this dissertation, we survey knowledge of various institutions to donate blood in the city of Bonn and find awareness for paid and unpaid options to be similar (Table 3.D3).[12] While donors appear to be able to choose whether or not they want to be paid, unpaid donations still represent more than 70 percent of all donations in Germany (Paul-Ehrlich-Institut, 2018). Incidentally, the German market also has the highest per capita rate of donations among all 172 countries that report to the WHO and comparatively low wholesale prices for human blood. Germany has the highest number of donations at 57.3 per 1,000 people, compared to 49.2 in Sweden and 43.7 in the United States. The cost of one blood unit on the German wholesale market is among the lowest in the world at about \$110, compared to \$190 in Sweden and Switzerland (Trimborn, 2009) and about \$211 in the United States (Toner et al., 2012).[13] Appendix 3.D provides more details on the German market for whole blood donations.

We study such a dual market in a stylized environment. The results from our laboratory experiment support our three behavioral hypotheses. We confirm our first hypothesis, which predicts that introducing a compensation for a donation should increase giving. In the dual market, the availability of extrinsic incentives does not crowd out intrinsic motivations of donors. In fact, giving significantly increases compared to the market design in which donations are not paid. These findings stand in contrast with the influential work of Titmuss (1971), who argued that paid blood donations could crowd out the intrinsic motivation to donate and lead to a net drop in donations.

For a simple illustration of the effect size, we can use the average marginal effects from the Poisson regression of the number of individual donations over the three rounds on treatment indicators, a gender dummy, and the number of tokens kept in the dictator game (Table 3.B4, column 5). Holding everything else constant, the predicted number of donations in a dual market is 0.473 standard deviations larger than in the single market where donations are not paid. This is equivalent to the estimated effect of moving from the 20th percentile to the 60th percentile in the distribution of "generosity" of subjects as measured by the dictator game, again holding everything else constant.

[12] Meyer and Tripodi (2018) interview about 1,000 randomly sampled customers of the municipal service center in Bonn, a mid-sized city in the west of Germany. Although the data is not representative for Germany, we take awareness of both paid and unpaid collection centers, for a rich set of demographic groups in an urban area, as confirmation that the choice between incentives for donating blood is indeed salient for a non-negligible share of the population.

[13] We calculate per capita donations based on the total number of whole blood donations collected in the years 2011 to 2013 (World Health Organization, 2017). We use the latest year available for all countries that report to the WHO. Population data comes from the World Bank World Development Indicator database.

Offering a compensation and letting agents turn down the compensation lets the collection system leverage the heterogeneity in individual preferences. This enables efficiency gains in the collection similar to those deriving from self-selection in second-degree price discrimination. Our sorting hypothesis states that in a dual market, a positive fraction of donors chooses to be not paid and that this fraction is bigger when actions are taken in public. We find that when given the option to turn down the compensation, a significant fraction of donors choose to do so, though we find only weak evidence that donors turn down incentives more in public than in private.

This result complements the findings of Lacetera, Macis, and Slonim (2014), who conduct a field experiment in which the American Red Cross offers gift cards as incentive to donate blood. They report that after donating, virtually none (2 percent) of the offered cards were turned down. In their setting, the ability to turn down incentives is not salient to prospective donors in their decision to come to the donation drive. Moreover, there is no clear signaling motive for turning down the gift card. In our setting, the two incentive schemes carry different utility in terms of private benefit and signaling value. With this choice between the two different incentives schemes, our dual market should be more effective at leveraging heterogeneity in individual preferences.

Even though we cannot provide strong evidence that sorting operates through social image concerns, we do find robust support of our second hypothesis, which states that visibility of actions increases donations irrespective of the type of available incentives. We can again use the average marginal effects from Poisson regression (Table 3.B4, column 4) to illustrate the effect size of social image. Making actions observable while holding everything else constant increases the predicted number of donations by 0.493 standard deviations. This is slightly larger than the estimated effect of moving from the 20th percentile to the 60th percentile in the distribution of "generosity" of subjects as measured by the dictator game, again holding everything else constant.

The single market *PAID* and *NOT PAID* treatments allow us to compare our findings to the existing literature. In contrast to previous work, we do not find that social image effects attenuate incentive effects (Ariely et al., 2009; Carpenter & Myers, 2010). Individuals in our experiment have an outside option that is larger than the monetary incentives to donate, so that *homo economicus* would never choose to donate. Both our work and Ariely et al. (2009) are based on the theoretical framework of Benabou and Tirole (2006).

Our findings suggest that in this framework, a salient outside option makes incentivized donations more likely to signal altruism and less likely to signal greed. This attenuates the image-spoiling effects of incentives that can bring about a negative interaction between incentive and image effects.

Our findings also suggest a gender-specific willingness to engage in costly signaling that could be interpreted as consistent with gender-specific aversion to standing out (Jones & Linardi, 2014) as well as with costly signaling theory in evolutionary biology (Gintis et al., 2001; Smith & Bird, 2000) and strategic signalling of generosity among men (Barclay, 2010; Boehm & Regner, 2013; Eagly & Crowley, 1986; Iredale et al., 2008).

Our findings have implications for the design of mechanisms for the collection of charitable donations. Applied to the collection of whole blood donations, our results could inform the design and regulation of systems that use monetary incentives. Because voluntary provision of blood donations is often insufficient (Whitaker, Rajbhandary, Kleinman, Harris, & Kamani, 2016), demand for blood is likely increasing in the future (Greinacher, Fendrich, Brzenska, Kiefel, & Hoffmann, 2011), and modern screening technologies appear sufficiently safe to counter adverse selection (Offergeld, Faensen, Ritter, & Hamouda, 2005), several countries are now re-evaluating partial reliance on incentivized or paid donations (Lacetera et al., 2013). Even small efficiency gains in these collection systems can imply economically meaningful savings for public health budgets. In the Unites States alone, about 13.6 million blood units are collected every year at a total value of more than USD 3 billion.[14]

Our results suggest that having different institutions provide distinct incentive schemes can improve the efficiency of the market compared to the case of all institutions offering the same incentives. In such a market, collectors may be able to increase donations by making image concerns more salient. In the case of Germany, the institution that offers unremunerated donations and has most to gain from making donations visible – the Red Cross – in fact largely relies on highly visible mobile drives for its collection.

Our results point to various avenues for future research. First, it would be good to better understand the mechanisms through which sorting into unpaid donations operates both in the German blood market and in general. While our theoretical framework suggests that social image effects should play a key role, our experimental data provides only weak evidence to support this hypothesis. Second, our setting does not appear to suffer from the negative interaction of social image effects and incentive effects that has been found in the previous literature. Empirical studies to determine if and when incentives spoil image utility constitute fruitful avenue for future research. Third, we cannot rule out that specific features of our experimental task undermine the external validity of our findings. While we used a task that is popular in the literature because it is not inherently meaningful and lends itself to a test of subject motivation, there is scope for future work in less stylized settings. Finally, we hope this work stimulates theoretical efforts on the characterization of competitive aspects of dual markets that would allow us to better understand the endogenous formation and social welfare implications of such institutional arrangements–important matters from which we largely abstract in this chapter.

[14]Back-of-the-envelope calculation based on 2007 US data from Toner et al. (2012).

3.A Appendix: Proofs

Proof of Proposition 1

The proposition is composed of two statements. First statement: *"A dual market for donations increases contributions compared to a single market where no incentives are available."*

When actions are private, the utility of any agent i can be re-written as

$$U_i(d, y) = \begin{cases} [a_i(B - y) + y - c]d, & \text{Dual Market: } y \in \{0, \tilde{y}\} \\ [a_i B - c]d, & \text{Single Market - No Incentives } y = 0 \end{cases}$$

Availability of incentives $\tilde{y} > 0$ does not affect donation behavior of highly altruistic agents ($a_i > 1$), who can choose to turn down the incentive, gaining utility

$$a_i B - c > a_i(B - \tilde{y}) + \tilde{y} - c.$$

At the same time, the availability of incentives get agents for whom

$$a_i B - c < 0 < a_i(B - \tilde{y}) + \tilde{y} - c$$

involved in the donation.

When actions take place in public, the same as above applies for image-indifferent agents. Image-concerned agents will now focus instead on taking the action that sends the best possible signal about their degree of altruism. Independence in the distribution of the degree of altruism and image concern implies that image-concerned agents would never refrain from donating, as doing so would send the worst possible signal about their degree of altruism.

Second statement: *"Compared to a single market where conditional incentives are automatic and cannot be turned down, allowing to turn down incentives reduces the cost of collection without affecting the number of donations."*

When actions are private, the utility of any agent i can be re-written as

$$U_i(d, y) = \begin{cases} [a_i(B - y) + y - c]d, & \text{Dual Market: } y \in \{0, \tilde{y}\} \\ [a_i(B - \tilde{y}) + \tilde{y} - c]d, & \text{Single Market - with Incentives} \end{cases}$$

Define the share of highly altruistic agents as $s(a) = Pr(a_i > 1)$. Because $B > c$, a $s(a)$ share of agents would donate irrespective of the availability of incentives, even though their intrinsic motivation to donate is partially crowded out in a single market with incentives. Allowing agents, in a dual market, to sort out of incentives un-does the described crowding out of intrinsic motivation to donate and reduces the average cost of collection.

When actions take place in public, the same as in private applies for image-indifferent agents. For image concerned agents, we need to show that participation is unaffected by the possibility of turning down incentives. Therefore, we need to show that in neither a single incentivized market nor in a dual market image concerned agents want to abstein from donating. The proof goes by contradiction.

In a dual market, suppose there exists a pure strategy equilibrium in which all image concerned agents were to not donate. Any one of these agents could deviate from the equilibrium by donating and turning down the compensation to mimic the most altruistic image indifferent agents. Such deviation would improve the reputation of this agent, hence her utility. A contradiction.

Similarly, in the single incentivized market the profitable deviation is represented by the reputational gain of donating with incentives.

3.B Appendix: Additional Tables

Table 3.B1: Poisson Regression for Total Individual Donations: Semi-Elasticities (Coefficient Estimates and Standard Errors in Parentheses)

Dependent variable:	# of donations over the three rounds				
	(1)	(2)	(3)	(4)	(5)
a) Treatments					
Paid&Choose *(baseline: Not Paid)*	0.232*	0.232*	0.261**	0.260**	0.294**
	(0.133)	(0.132)	(0.121)	(0.121)	(0.120)
Public *(baseline: Private)*	0.290**	0.298**	0.299**	0.301**	0.320**
	(0.141)	(0.140)	(0.132)	(0.131)	(0.129)
Paid&Choose × Public	-0.004	-0.008	0.021	0.020	-0.010
	(0.163)	(0.162)	(0.152)	(0.152)	(0.150)
b) Controls					
Female		0.151**		0.048	0.019
		(0.075)		(0.070)	(0.070)
DG: Dictator kept			-0.062***	-0.061***	-0.057***
			(0.009)	(0.009)	(0.009)
Other controls	No	No	No	No	Yes
Observations	329	329	329	329	329

$^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$

Notes: Standard errors are clustered at individual level. *NOT PAID* is the base market design treatment. *PRIVATE* is the base visibility treatment. Other controls include age, chosen charity, and individual rating of chosen charity.

Table 3.B2: Random Effects Regressions: Relative Risk Ratios (Coefficient Estimates and Standard Errors in Parentheses)

Dependent variable:	Incentive Choice				
	(1)	(2)	(3)	(4)	(5)
		a) Treatment			
Public	1.747	1.652	1.862	1.705	2.229
	(1.278)	(1.155)	(1.439)	(1.257)	(1.585)
		b) Controls			
Female		0.719		0.552	0.601
		(0.531)		(0.420)	(0.456)
DG: Tokens kept			0.928	0.915	0.934
			(0.104)	(0.102)	(0.104)
Other controls	No	No	No	No	Yes
Observations	378	378	378	378	378

$^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$ for relative risk ratios different from unity.
Notes: Standard errors are clustered at the individual level. *PRIVATE* is the base visibility treatment. The incentive choice dependent variable only applies to the 126 subjects in *CHOOSE* treatment over three rounds. Incentive choice takes value ”0” if subject skips, ”1” if participates unpaid, and ”2” if participates paid to the donation task in a given round. The table reports relative risk ratio for outcome ”1” unpaid participation and base outcome ”2” paid participation.

Table 3.B3: Poisson Regression for Total Individual Donations: Semi-Elasticities (Coefficient Estimates and Standard Errors in Parentheses)

	Incentive Treatment Subsamples		
	Not paid (1)	Paid (2)	Choose (3)
	a) Gender dummy × visibility treatment		
Public	0.483*	0.342	0.713***
	(0.253)	(0.210)	(0.268)
Female	0.230	0.086	0.357
	(0.242)	(0.196)	(0.258)
Public × Female	-0.285	0.026	-0.584*
	(0.293)	(0.251)	(0.315)
	b) Controls		
DG: Tokens kept	-0.050***	-0.053***	-0.083***
	(0.015)	(0.015)	(0.019)
Observations	93	110	126

$^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$
Notes: Standard errors are clustered at individual level. *NOT PAID* is the base market design treatment. *PRIVATE* is the base visibility treatment. DG refers to the dictator game, in which we gave 20 experimental tokens to subjects and asked them how many they would like to keep.

TABLE 3.B4: Poisson Regression for Total Individual Donations (Coefficient Estimates and Standard Errors in Parentheses)

Dependent variable:	# of donations over the three rounds					
	Semi-elasticities			Average marginal effects		
	(1)	(2)	(3)	(4)	(5)	(6)
			a) Treatments			
Paid	0.149	0.183	0.205	0.268**	0.322***	0.333***
	(0.157)	(0.142)	(0.140)	(0.129)	(0.121)	(0.118)
Choose	0.294**	0.318**	0.363***	0.409***	0.476***	0.512***
	(0.145)	(0.131)	(0.131)	(0.128)	(0.117)	(0.118)
Public	0.290**	0.301**	0.320**	0.451***	0.496***	0.492***
	(0.141)	(0.131)	(0.129)	(0.107)	(0.098)	(0.097)
Paid × Public	0.056	0.065	0.043			
	(0.189)	(0.175)	(0.172)			
Choose × Public	-0.044	-0.008	-0.043			
	(0.179)	(0.165)	(0.163)			
			b) Controls			
Female		0.040	0.010		0.064	0.016
		(0.069)	(0.070)		(0.109)	(0.111)
DG: Tokens kept		-0.062***	-0.057***		-0.097***	-0.091***
		(0.009)	(0.009)		(0.013)	(0.013)
Other controls	No	No	Yes	No	No	Yes
Observations	329	329	329	329	329	329

$^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$

Notes: Standard errors are clustered at individual level. *NOT PAID* is the base market design treatment. *PRIVATE* is the base visibility treatment. DG refers to the dictator game, in which we gave 20 experimental tokens to subjects and asked them how many they would like to keep. Other controls include age, chosen charity, and individual rating of chosen charity.

3.C Appendix: Non-Parametric Tests of Hypotheses 1 and 2

Given that treatment assignment is random and experimental subjects are balanced on observables across treatments, we can use non-parametric tests as an additional robustness check to our main results.

Non-parametric analysis confirms the first behavioral hypothesis, which says that irrespective of whether actions are visible, the availability of incentives increases donations. When actions are taken in *PRIVATE*, the average number of donations increases from 1.13 in *NOT PAID* to 1.43 in *PAID&CHOOSE*. When actions are taken in *PUBLIC*, the average number of donations increases from 1.51 in *NOT PAID* to 1.9 in *PAID&CHOOSE*.

Pairwise two-sided Wilcoxon rank-sum (Mann-Whitney) tests confirm that making incentives available increases the number of donations both in PRIVATE ($z = -1.680$, $p = 0.093$) and in PUBLIC ($z = -2.520$, $p = 0.012$).

We also find support for our second hypothesis of social image effects. Irrespective of the incentive treatment, making actions visible significantly increases the number of donations over all three rounds. In the *NOT PAID* treatment, making actions visible increases the total number of donations from 1.13 in *PRIVATE* to 1.51 in *PUBLIC*. In *PAID&CHOOSE*, making actions visible increases donations from 1.43 to 1.9.

Pairwise two-sided Wilcoxon rank-sum tests (Table 3.C1 columns 3 and 4) reject equal distributions of donations between *PRIVATE* and *PUBLIC* treatments both in *NOT PAID* ($z = -2.247$, $p = 0.025$) and in *PAID&CHOOSE* ($z = -3.512$, $p < 0.001$) market designs. Table 3.C1 below summarizes these test results.

Table 3.C1: Incentive and Social Image Effects: Non-Parametric Analysis

	Not Paid vs. Paid&Choose		Private vs. Public	
	Private (1)	Public (2)	Not Paid (3)	Paid&Choose (4)
Difference in average donations	0.295	0.388	0.380	0.473
z-score	-1.680	-2.520	-2.247	-3.512
p-value	0.093	0.012	0.025	0.000
N	154	175	93	236

Notes: Test statistics are for the total number of rounds participated in the donation task based on a two-sample Wilcoxon rank-sum test.

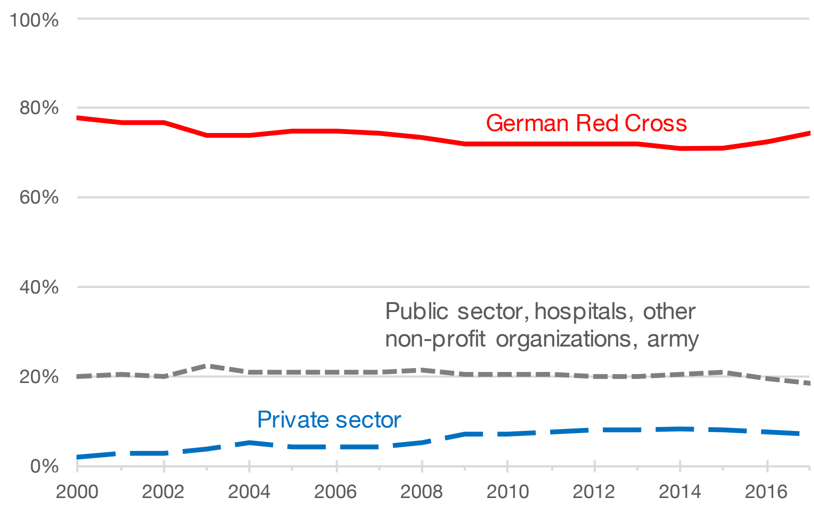
3.D Appendix: Germany's Market for Whole Blood Donations

Our model and experiment is motivated by possible sorting of blood donors in Germany. In this section we provide further details on the German market for whole blood donations. We first provide general institutional backgroound, then summarise a mapping exercise to better understand the possible role of transportation costs in this market, and finally provide survey evidence to shed light on potential informational frictions.

3.D.1 Institutional Background

With a total of 7.2 million donations or 89 donations per 1,000 people, Germany is the fifth-largest blood supplier in the world. Of the 4.4 million whole blood donations collected in 2014, about 71 percent of whole blood donations were collected by the German Red Cross, which generally never pays its donors. The remaining 29 percent represent the military, private donors, and larger hospitals (Paul-Ehrlich-Institut, 2018). Figure 3.D1 plots the share of different whole blood collectors from 2000 to 2017 and illustrates that the share of the German Red Cross has been roughly stable over the past two decades.

Figure 3.D1: Blood Donations in Germany 2000 to 2017, Fraction by Collector



Source: Paul-Ehrlich-Institut (2018). *Notes:* German Red Cross donations are always unpaid.

Hospitals and the private sector commonly pay their whole blood donors, sometimes up to USD 30 per donation.[15] The German legal framework (*Transfusionsgesetz* §10) recommends unpaid donations but provides for an unspecified monetary "compensation" (*Aufwandsentschädigung*). It is interesting to note that the German Red Cross, as quasi-monopolist, has unsuccessfully taken legal action to stop remunerated donations. Most recently in 2012, the Higher Administrative Court of Rhineland-Palatinate (*Oberverwaltungsgericht*

[15] It is difficult to estimate exact numbers because the German government does not publish data on blood donations by type of remuneration, while the relevant WHO database on blood donations is not nationally representative.

Rheinland-Pfalz) dismissed legal action of the German Red Cross against the university hospital in Mainz, who regularly pays its donors. The court found the payment to be lawful. See also Oberverwaltungsgericht Rheinland-Pfalz (2013).

3.D.2 Transportation Costs

To better understand if prospective donors can indeed choose between different options or if the market is simply geographically segmented into different incentive schemes, we map donation points and calculate average travel time to paid and to unpaid donation points for a significant share of the German population. This gives us an idea of how easy or difficult it is to donate at paid and unpaid donation points.

We collect address data for all 35 locations of Germany's largest commercial blood bank (Haema AG), all 36 locations of German university hospitals that have their own blood collection services, and all 30 fixed donation points of the German Red Cross. We also scrape the website of the German Red Cross to obtain locations of all mobile donation drives from November 2016 to early January 2017.

We geocode all 9,306 locations using the Google Maps API (Figure 3.D2). For the 50 largest communities in terms of population in Germany (*politisch selbstständige Gemeinden*), representing about 27 percent of the population, we again use Google Maps API to find all donation sites that are either 30 minutes away from the community midpoint on public transport or 30 minutes away from the midpoint when driving under traffic conditions on October 17, 2016 at 9am. We limit ourselves to the 50 largest communities in order to make the distance calculations and geocoding of addresses more feasible. Similarly, we set an arbitrary 30 minute limit on travel time away from the community midpoint to make computations more feasible.

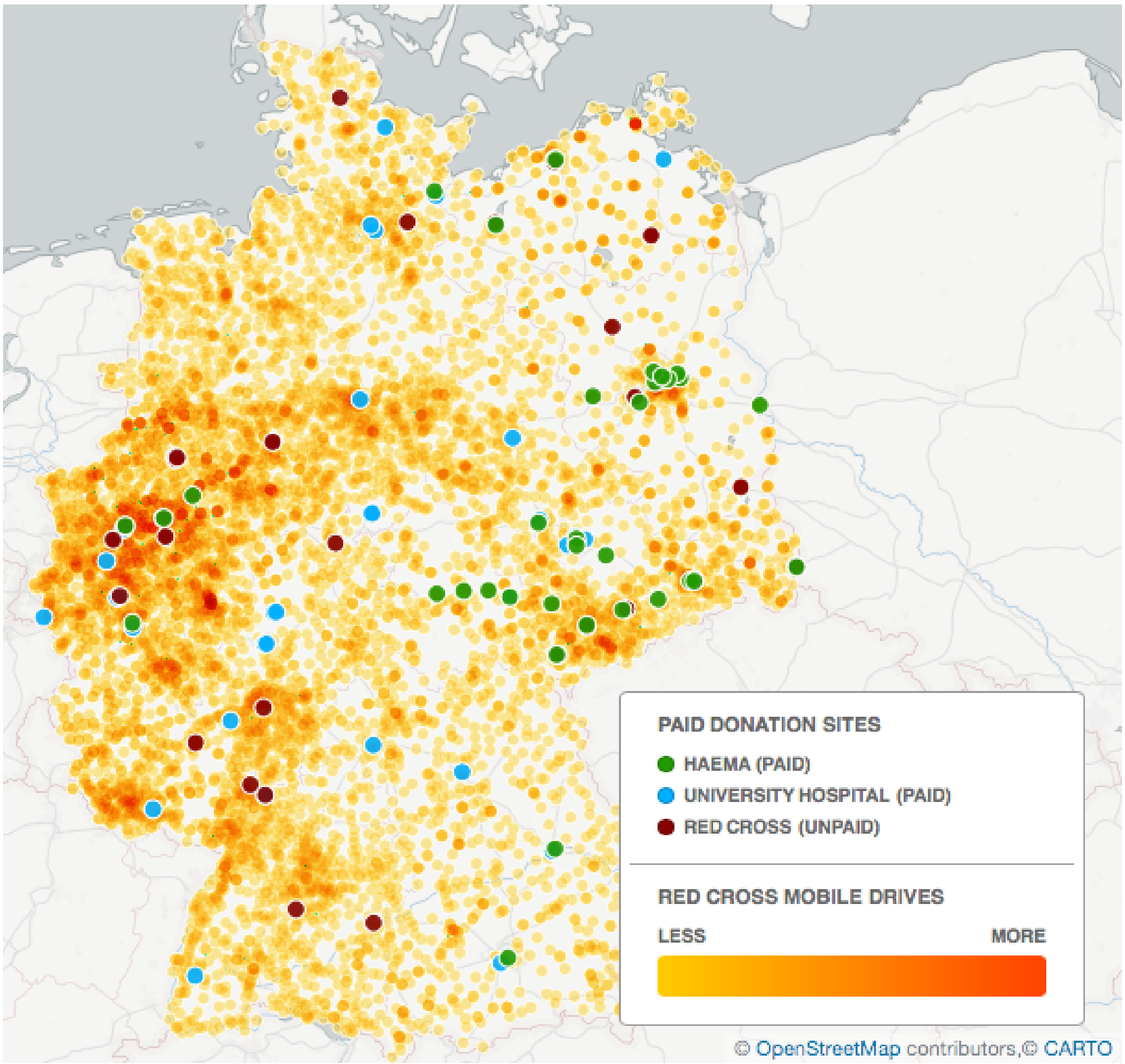
We find that it is not significantly more difficult to reach a paid donation site than it is to reach an unpaid donation site. In the 50 largest communities, it takes about 2.5 minutes longer to reach a paid donation point than it takes to reach an unpaid donation point on public transport or driving. Restricting the analysis to public transport, it takes about 5 minutes longer to reach a paid donation point (Table 3.D1). Put differently, everyone who lives in one of the 50 largest communities in Germany can reach an unpaid (Red Cross) donation point within 30 minutes time driving or on public transport. This compares to about 62 percent of the population who can reach a paid (hospital or Haema AG) donation point within 30 minutes time using the same means of transport (Table 3.D2).

These calculations make numerous simplifying assumptions and should thus be seen as merely indicative. Most importantly, we do not discount the fact that most Red Cross locations in our analysis are temporary (mobile) donation drives that often only collect donations on a specific day. This stands in contrast to the paid donation points that are all fixed and have regular opening hours. The travel times above can thus be seen as a lower bound on how long it takes to reach an unpaid donation site.

3.D.3 Survey Evidence on Awareness of Different Institutions

In Chapter 3 of this dissertation (Meyer & Tripodi, 2018), we conduct a field experiment in Bonn, Germany, to study how social pressure affects pledges to give blood. As part of this experiment, we also assessed individuals' awareness of different institutions to donate blood.

Figure 3.D2: Map of Germany with Fixed Blood Donation Locations and Density of Red Cross Mobile Donation Drives



Source: Own compilation, Google Maps API, CARTO. Map tiles by Stamen Design, used under CC BY 3.0 license. Map data Openstreetmap, used under ODbL license. *Notes:* German Red Cross locations of mobile donation drives from November 2016 to early January 2017 are presented.

Table 3.D1: Travel Time to Nearest Blood Donation Point in 50 Largest German Communities, by Incentive Offered and Mode of Transport (Minutes)

	Driving		Public transport	
Distance from community geographic midpoint	Not paid	Paid	Not paid	Paid
<30 min driving	7.2	11.7	-	-
<30 min public transport or driving	7.8	10.3	8.5	13.4

Sources: Own compilation, Google Maps API, Statistisches Bundesamt (2016).
Notes: Sample consists of the 50 largest communities (*politisch selbständige Gemeinden*) in Germany, dated March 31, 2016. No monetary incentives refers to 9,236 donation centers and mobile donation drives of the German Red Cross. Monetary incentives refers to 35 commercial donation centers of Haema and 36 university hospitals with blood donation units. Travel distances calculated using Google Maps API for traffic conditions on October 17, 2016 at 9am. See text for detailed description of methodology.

Table 3.D2: Share of Population with Access to Blood Donation Points in 50 Largest German Communities, by Incentive Offered

Distance to community geographic midpoint	Access to unpaid donation	Access to paid donation
Less than 30 min driving	1.00	0.69
Less than 30 minutes by public transport or driving	1.00	0.62

Sources: Own compilation, Google Maps API, Statistisches Bundesamt (2016).
Notes: Sample consists of the 50 largest communities (*politisch selbständige Gemeinden*) in Germany, dated March 31, 2016. No monetary incentives refers to 9,236 donation centers and mobile donation drives of the German Red Cross. Monetary incentives refers to 35 commercial donation centers of Haema and 36 university hospitals with blood donation units. Travel distances calculated using Google Maps API for traffic conditions on October 17, 2016 at 9am.

We recruit subjects using an intercept survey among customers of the service centre of the Bonn municipal government. The service centre, centrally located in the city hall, provides a wide range of in-person administrative services such as applications for official documents, driver's licenses, registration of motor vehicles, and payments for city services. Customers arrive at the service centre for appointments that they have previously scheduled online or via telephone. We administer our survey while customers wait for their appointment in a designated waiting area. Participation in the survey is particularly high considering the lack of incentives. About 75 percent of the 1,675 subjects approached agreed to participate and 57 percent completed the survey before being called up for an appointment.

For each blood collecting institution, Table 3.D3 presents the share of interviewed subjects declaring to be aware of the blood collection activity in the city of Bonn. Over the whole sample, 86.5 percent is aware of the German Red Cross, while 72.9 percent are aware of at least one of the paying institutions (among Haema and the Bonn University Hospital).[16] We also break down the share of aware subjects by gender and age group: women

[16]We are aware of other smaller institutions collecting blood in the country. These do not constitute a

seem to be generally more aware than men, and older people slightly more aware than the younger. Over all categories, people seem to be more aware of the unpaid option but not dramatically so. We take this data as suggestive evidence of a dual market for blood in the city of Bonn.

Table 3.D3: Market Awareness in Bonn (Shares and Standard Errors in Parentheses)

	Incentive scheme		Paid		
	Not paid (DRK)	Paid (Haema/Uni)	Haema	Uni	N
Whole sample	0.865	0.729	0.147	0.706	941
	(0.011)	(0.014)	(0.012)	(0.015)	
		by gender			
Female	0.900	0.784	0.184	0.753	490
	(0.014)	(0.019)	(0.018)	(0.020)	
Male	0.827	0.670	0.106	0.654	451
	(0.018)	(0.022)	(0.015)	(0.022)	
		by age group			
18 to 24	0.869	0.777	0.153	0.742	229
	(0.022)	(0.028)	(0.024)	(0.029)	
25 to 34	0.847	0.731	0.197	0.703	320
	(0.020)	(0.025)	(0.022)	(0.026)	
35 to 44	0.850	0.647	0.087	0.642	173
	(0.027)	(0.036)	(0.021)	(0.037)	
45 to 54	0.895	0.737	0.117	0.725	171
	(0.024)	(0.034)	(0.025)	(0.034)	
55 to 64	0.917	0.750	0.104	0.708	48
	(0.040)	(0.063)	(0.045)	(0.066)	

Source: Meyer and Tripodi (2018)
Notes: Data based on a random sample of 941 subjects interviewed in the waiting area of the Bonn city hall.

relevant market share in the city of Bonn and we did not include them in our survey.

3.E Appendix: Online Pilot Study

We conducted a pilot study of our experimental design online on Amazon Mechanical Turk. In the online experiment, we take advantage of the high degree of anonymity to implement only the *PRIVATE* treatment. Instead of the 3×2 between-subject design of our main laboratory experiment, we considered an alternative within-subject design that introduces the dual market treatment after the first donation round. This design lets us study the transition from a single market treatment to a dual market, accounting for potential carryover effects.

3.E.1 Experimental Design and Procedures

Paralleling the laboratory experiment, three treatments determine the incentive scheme under which subjects can perform the real effort task. In addition, subjects can always skip participation and take an outside option of 75 tokens. We provide either monetary incentives to donate (*PAID*; 50 tokens to subject, 50 tokens to charity), or no monetary incentives (*NOT PAID*; 100 tokens to charity), or we let subjects choose among one of the two incentive schemes (*CHOOSE*).

Subjects engage in the real effort task for three rounds. In the first round, we administer the three treatments in a between-subject design. After the first round we introduce the *CHOOSE* treatment for subjects that were PAID in the first round and for a random subsample of subjects that were in the *NOT PAID* treatment in the first round. This results in four distinct treatments.

A total of 408 subjects were recruited for seven session between May and October 2016. Most subjects are from the United States (81.1 percent), have completed college degrees (70.3 percent), and are mostly male (57.1 percent). The average subject is 33 years old. Double participation is ruled out. We pay a show-up fee for completing the experiment of \$0.40. 1 token is worth \$0.04. On average, subjects earned \$1.04 for themselves and generated \$0.37 for charity. Sessions lasted circa 20 minutes.

3.E.2 Results

Table 3.E1 summarises treatment assignment and results for each treatment and round, Figure 3.E1 illustrates the share of subjects participating in the donation task and the share of subjects choosing to not be paid.

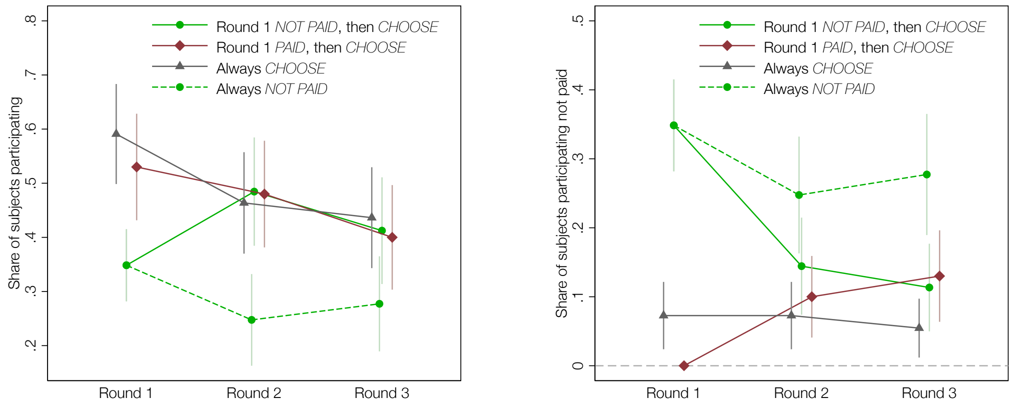
Consistent with findings from the lab, we do not find that introducing monetary incentives crowds out participation in the donation task. In the dual market *CHOOSE* treatment, subjects are significantly more likely to participate in the donation task than in the single market *NOT PAID* treatment. Among subjects in the *NOT PAID* treatment in the first round, those that are randomised into *CHOOSE* in the second round are significantly more likely to participate in the donation. We take this as suggestive evidence that transitioning from a single market design where no one is paid to a dual market design in which donors can choose to be paid to give can increase donations.

Table 3.E1: Distribution of Treatments, Subject Participation, and Subject Incentive Choice in Online Experiment (Number of Subjects)

Treatment	Total	Participation choice		Incentive choice	
		Skip	Participate	Not paid	Paid
		Round 1			
Round 1 *NOT PAID*, then *CHOOSE*	97	65	32	32	-
Round 1 *PAID*, then *CHOOSE*	100	47	53	-	53
Always *CHOOSE*	110	45	65	8	57
Always *NOT PAID*	101	64	37	37	-
		Round 2			
Round 1 *NOT PAID*, then *CHOOSE*	97	50	47	14	33
Round 1 *PAID*, then *CHOOSE*	100	52	48	10	38
Always *CHOOSE*	110	59	51	8	43
Always *NOT PAID*	101	76	25	25	-
		Round 3			
Round 1 *NOT PAID*, then *CHOOSE*	97	57	40	11	29
Round 1 *PAID*, then *CHOOSE*	100	60	40	13	27
Always *CHOOSE*	110	62	48	6	42
Always *NOT PAID*	101	73	28	28	-

Notes: 408 subjects. Last two columns refer to subjects not skipping the donation task.

Figure 3.E1: Subject Participation in Online Donation Task



(a) Overall Participation (b) Participation in *NOT PAID*

Notes: Bars indicate 95 percent confidence intervals. In panel (b), the share of subjects participating not paid is conditional on not skipping the donation task.

3.F References

Ariely, D., Bracha, A., & Meier, S. (2009, February). Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially. *American Economic Review*, *99*(1), 544-555. doi: 10.1257/aer.99.1.544

Barclay, P. (2010, February). Altruism as a courtship display: Some effects of third-party generosity on audience perceptions. *British Journal of Psychology*, *101*(1), 123-135. doi: 10.1348/000712609X435733

Becker, G. S., & Elias, J. J. (2007). Introducing Incentives in the Market for Live and Cadaveric Organ Donations. *The Journal of Economic Perspectives*, *21*(3), 3-24. doi: 10.1257/jep.21.3.3

Benabou, R., & Tirole, J. (2006, December). Incentives and Prosocial Behavior. *American Economic Review*, *96*(5), 1652–1678. doi: 10.1257/aer.96.5.1652

Bock, O., Baetge, I., & Nicklisch, A. (2014, October). hroot: Hamburg Registration and Organization Online Tool. *European Economic Review*, *71*, 117–120. Retrieved 2018-06-09, from http://www.sciencedirect.com/science/article/pii/S0014292114001159 doi: 10.1016/j.euroecorev.2014.07.003

Boehm, R., & Regner, T. (2013, October). Charitable giving among females and males: An empirical test of the competitive altruism hypothesis. *Journal of Bioeconomics*, *15*(3), 251-267. doi: 10.1007/s10818-013-9152-x

Boulware, L. E., Troll, M. U., Wang, N. Y., & Powe, N. R. (2006, November). Public Attitudes Toward Incentives for Organ Donation: A National Study of Different Racial/Ethnic and Income Groups. *American Journal of Transplantation*, *6*(11), 2774-2785. doi: 10.1111/j.1600-6143.2006.01532.x

Bowles, S., & Polania-Reyes, S. (2012, June). Economic Incentives and Social Preferences: Substitutes or Complements? *Journal of Economic Literature*, *50*(2), 368-425. doi: 10.1257/jel.50.2.368

Bursztyn, L., & Jensen, R. (2017, August). Social Image and Economic Behavior in the Field: Identifying, Understanding, and Shaping Social Pressure. *Annual Review of Economics*, *9*, 131-153. doi: 10.1146/annurev-economics-063016-103625

Carpenter, J., & Myers, C. K. (2010, December). Why volunteer? Evidence on the role of altruism, image, and incentives. *Journal of Public Economics*, *94*(11), 911-920. doi: 10.1016/j.jpubeco.2010.07.007

Chao, M. (2017, June). Demotivating incentives and motivation crowding out in charitable giving. *Proceedings of the National Academy of Sciences*, 7301–7306. doi: 10.1073/pnas.1616921114

Charness, G., Gneezy, U., & Henderson, A. (2018, May). Experimental methods: Measuring effort in economics experiments. *Journal of Economic Behavior & Organization*, *149*, 74–87. Retrieved 2018-06-09, from http://www.sciencedirect.com/science/article/pii/S0167268118300556 doi: 10.1016/j.jebo.2018.02.024

Charness, G., Gneezy, U., & Kuhn, M. A. (2012, January). Experimental methods:

Between-subject and within-subject design. *Journal of Economic Behavior & Organization*, *81*(1), 1-8. doi: 10.1016/j.jebo.2011.08.009

Chen, D. L., Schonger, M., & Wickens, C. (2016, March). oTree – An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, *9*, 88-97. doi: 10.1016/j.jbef.2015.12.001

Council of Europe. (1995). *Recommendation No. R (95) 14 of the Committee of Ministers to Member States on the Protection of the Health of Donors and Recipients in the Area of Blood Transfusion. Adopted by the Committee of Ministers on 12 October 1995 at the 545th Meeting of the Ministers' Deputies.* (Tech. Rep.). Strasbourg: Author.

Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology*, *18*(1), 105-115. doi: 10.1037/h0030644

Deci, E. L. (1975). *Intrinsic Motivation*. Boston, MA: Springer US.

Eagly, A. H., & Crowley, M. (1986). Gender and helping behavior: A meta-analytic review of the social psychological literature. *Psychological Bulletin*, *100*(3), 283-308. doi: 10.1037/0033-2909.100.3.283

Exley, C. (2017, March). Incentives for Prosocial Behavior: The Role of Reputations. *Management Science*. doi: 10.1287/mnsc.2016.2685

Filiz-Ozbay, E., & Ozbay, E. Y. (2014, June). Effect of an audience in public goods provision. *Experimental Economics*, *17*(2), 200-214. doi: 10.1007/s10683-013-9363-y

Frey, B. S., & Goette, L. (1999). *Does Pay Motivate Volunteers?* (IEW Working Papers No. 007). Zurich: Institute for Empirical Research in Economics, University of Zurich.

Frey, B. S., & Oberholzer-Gee, F. (1997). The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding-Out. *The American Economic Review*, *87*(4), 746-755.

Gintis, H., Smith, E. A., & Bowles, S. (2001, November). Costly Signaling and Cooperation. *Journal of Theoretical Biology*, *213*(1), 103-119. doi: 10.1006/jtbi.2001.2406

Gneezy, U., & Rustichini, A. (2000). Pay Enough or Don't Pay at All. *The Quarterly Journal of Economics*, *115*(3), 791-810.

Greinacher, A., Fendrich, K., Brzenska, R., Kiefel, V., & Hoffmann, W. (2011, April). Implications of demographics on future blood supply: A population-based cross-sectional study. *Transfusion*, *51*(4), 702-709. doi: 10.1111/j.1537-2995.2010.02882.x

Iredale, W., Van Vugt, M., & Dunbar, R. (2008, July). Showing Off in Humans: Male Generosity as a Mating Signal. *Evolutionary Psychology*, *6*(3), 386–392. doi: 10.1177/147470490800600302

Jones, D., & Linardi, S. (2014, March). Wallflowers: Experimental Evidence of an Aversion to Standing Out. *Management Science*, *60*(7), 1757-1771. doi: 10.1287/mnsc.2013.1837

Lacetera, N. (2016, September). *Incentives and Ethics in the Economics of Body Parts* (Working Paper No. 22673). Cambridge, MA: National Bureau of Economic Research.

Lacetera, N., & Macis, M. (2010, November). Social image concerns and prosocial behavior: Field evidence from a nonlinear incentive scheme. *Journal of Economic Behavior & Organization*, *76*(2), 225-237. doi: 10.1016/j.jebo.2010.08.007

Lacetera, N., Macis, M., & Slonim, R. (2012, February). Will There Be Blood? Incentives and Displacement Effects in Pro-Social Behavior. *American Economic Journal: Economic Policy*, *4*(1), 186-223. doi: 10.1257/pol.4.1.186

Lacetera, N., Macis, M., & Slonim, R. (2013, May). Economic Rewards to Motivate Blood Donations. *Science*, *340*(6135), 927-928. doi: 10.1126/science.1232280

Lacetera, N., Macis, M., & Slonim, R. (2014, May). Rewarding Volunteers: A Field Experiment. *Management Science*, *60*(5), 1107-1129. doi: 10.1287/mnsc.2013.1826

Lacetera, N., Macis, M., & Stith, S. S. (2014, January). Removing financial barriers to organ and bone marrow donation: The effect of leave and tax legislation in the U.S. *Journal of Health Economics*, *33*, 43-56. doi: 10.1016/j.jhealeco.2013.10.006

Landry, C. E., Lange, A., List, J. A., Price, M. K., & Rupp, N. G. (2006). Toward an Understanding of the Economics of Charity: Evidence from a Field Experiment. *The Quarterly Journal of Economics*, *121*(2), 747-782. doi: 10.1162/qjec.2006.121.2.747

Mellstrom, C., & Johannesson, M. (2008, June). Crowding Out in Blood Donation: Was Titmuss Right? *Journal of the European Economic Association*, *6*(4), 845-863. doi: 10.1162/JEEA.2008.6.4.845

Meyer, C. J., & Tripodi, E. (2018). *Image Concerns in Pledges to Give Blood: Evidence from a Field Experiment* (SSRN Working Paper No. 3132289).

Niessen-Ruenzi, A., Weber, M., & Becker, D. M. (2015). *To pay or not to pay – Evidence from whole blood donations in Germany* (Mimeo).

Oberverwaltungsgericht Rheinland-Pfalz. (2013, December). *Urteil DRK-Blutspendedienst West gGmbH v. Land Rheinland-Pfalz* (No. 6 A 10608/13.OVG). Koblenz: Author.

Offergeld, R., Faensen, D., Ritter, S., & Hamouda, O. (2005, February). Human immunodeficiency virus, hepatitis C and hepatitis B infections among blood donors in Germany 2000-2002: Risk of virus transmission and the impact of nucleic acid amplification testing. *Euro Surveillance*, *10*(2), 8-11.

Ottoni-Wilhelm, M., Vesterlund, L., & Xie, H. (2014, September). *Why Do People Give? Testing Pure and Impure Altruism* (Working Paper No. 20497). National Bureau of Economic Research.

Paul-Ehrlich-Institut. (2018). *Preliminary Report for 2018* (Report on Notifications Pursuant to Section 21 TFG (German Transfusion Act)). Langen: Author.

Smith, E. A., & Bird, R. L. B. (2000, July). Turtle hunting and tombstone opening: Public generosity as costly signaling. *Evolution and Human Behavior*, *21*(4), 245-261. doi: 10.1016/S1090-5138(00)00031-3

The Lancet. (2005, June). Blood Supply and Demand. *The Lancet*, *365*(9478), 2151. doi:

10.1016/S0140-6736(05)66749-9

Titmuss, R. M. (1971). *The Gift Relationship: From Human Blood to Social Policy*. New York, NY: Pantheon Books.

Toner, R. W., Pizzi, L., Leas, B., Ballas, S. K., Quigley, A., & Goldfarb, N. I. (2012, August). Costs to hospitals of acquiring and processing blood in the US. *Applied Health Economics and Health Policy*, *9*(1), 29-37. doi: 10.2165/11530740-000000000-00000

Trimborn, M. (2009, May). Rohstoff Blut: 1000 Euro fuer fuenf Liter Blut. *Frankfurter Allgemeine Zeitung*.

Van Vugt, M., & Iredale, W. (2013, February). Men behaving nicely: Public goods as peacock tails. *British Journal of Psychology*, *104*(1), 3-13. doi: 10.1111/j.2044-8295.2011.02093.x

Whitaker, B., Rajbhandary, S., Kleinman, S., Harris, A., & Kamani, N. (2016, June). Trends in United States blood collection and transfusion: Results from the 2013 AABB Blood Collection, Utilization, and Patient Blood Management Survey. *Transfusion*. doi: 10.1111/trf.13676

World Health Organization. (2009). *The Melbourne Declaration on 100% Voluntary Non-Remunerated Donation of Blood and Blood Components*. Author.

World Health Organization. (2017). *Global Status Report on Blood Safety and Availability 2016* (Tech. Rep.). Geneva: Author.

Image Concerns in Pledges to Give Blood: Evidence from a Field Experiment

with Egon Tripodi 4

4.1 Introduction

Eliciting human tissue donations is a constant challenge for healthcare providers and blood banks. Explicitly or implicitly appealing to image concerns of prospective donors is a commonly used tool to elicit such donations. We want to be seen by the people around us as doing good, as being generous and altruistic. Costly prosocial behavior such as donations of time, money, or body tissue, can act as a signal to others that demonstrates such generosity (Benabou & Tirole, 2006). Although people do not always welcome such signaling opportunities (Andreoni, Rao, & Trachtman, 2017; DellaVigna, List, & Malmendier, 2012), social image concerns can be leveraged to induce individuals to behave in socially desirable ways, including giving to charity (Ariely, Bracha, & Meier, 2009; Meyer & Tripodi, 2017), voting (Gerber, Green, & Larimer, 2008), and participating in energy conservation programs (Yoeli, Hoffman, Rand, & Nowak, 2013).

We want to thank Zvonimir Bašić, Michèle Belot, Juan Dolado, Jonas Fooken, Lorenz Götte, Supreet Kaur, David K. Levine, Arthur Schram, and Erez Yoeli for helpful comments. The chapter has also benefited from discussions with seminar audiences at the EUI and the Max Planck Institute for Research on Collective Goods. We are grateful to Thomas Fricke, Nadine Quäsching, and Astrid Schneider at the municipal administration of the City of Bonn, Thomas Herzfeld and Petra Klostermann at DRK-Blutspendedienst West, and Jana Kierdorf and Jan Noack at Haema AG without whom this project would not have been possible. Valerie Stottuth and Marian Weigt provided outstanding research assistance. Funding from the Institute for Applied Microeconomics at the University of Bonn and the Department of Economics at the European University Institute is gratefully acknowledged. The collection, handling, and management of personal information of subjects was reviewed and approved by the data protection officer of the University of Bonn. The findings, interpretations, and conclusions expressed in this chapter are entirely those of the authors and do not necessarily represent the views of our partner organizations.

In many settings, however, prosocial actions themselves cannot be made public. One way for individuals to still benefit from public recognition of their prosocial activities is to demonstrate proof of contribution ex-post, for example by wearing a lapel pin for blood donors or sharing a picture of volunteer activity on social media. Charities often recognize donors on publicly-visible plaques or donor walls. Ex-ante, social pressure can be leveraged by encouraging public pledges to act charitably in the future. Such pledges are used to rally individuals to contribute to future charitable activities, for example through public calls for action in online and offline social networks.[1] These pledges are the focus of our study.

Two steps are involved for a pledge to increase contributions to a charitable activity: First, individuals need to take up the initial commitment. Second, individuals need to follow through and fulfill their pledge. In this chapter, we set out to study how social image concerns affect both of these steps.

In the first step, an observable promise to do good – similar to an observable act of doing good – can be used to signal generosity to others.[2] This signaling process will be the focus of our field experiment. Building on a simple theoretical framework, we study three channels through which social image effects can affect the uptake of pledges: the degree to which an individual cares about being seen by others as "doing good" by taking a pledge, the expectation that an individual has over how socially-desirable others around her perceive her pledge, and the ability of others to update their beliefs about an individual, given her pledge.

In the second step, various mechanisms can explain why individuals would renege or follow through. A pledge can produce an internal commitment that individuals with preferences for moral consistency (Cialdini & Trost, 1998; Cioffi & Garner, 1996) or promise-keeping (Charness & Dufwenberg, 2006) might not want to break.[3] Increasing the psychological costs of reneging could then increase follow up (Andreoni & Serra-Garcia, 2017). Given the challenges of disentangling these mechanisms in our context, we study the second step empirically.

We conduct a field experiment in a the mid-sized German city where various organizations compete for prospective blood donors. In the service center of the municipal government located in the city hall, we approach customers waiting for their appointment with a

[1] A well known example is The Giving Pledge (givingpledge.org), which encourages wealthy people to contribute a majority of their wealth to philanthropic causes. In the domain of human tissue donations, the pharmaceutical company Abbott has been conducting a global media campaign to promote blood donations. The campaign features celebrities in YouTube clips and encourages individuals to "make a pledge and BE THE 1 [to save a life]", with the hope that these individuals fulfill their pledge ("Pledging is only the first step in the BE THE 1 campaign. Now that you've made your pledge, keep your promise by finding your donation center and scheduling an appointment to donate blood and plasma"). It also encourages sharing of this pledge on social media (bethe1donor.com/join).

[2] Andreoni and Serra-Garcia (2017) provide an alternative explanation and illustrate how people may want to substitute immediate donations with pledges to donate to help overcome pressure to give (Andreoni et al., 2017; DellaVigna et al., 2012) and time inconsistency in their preferences for giving (Dreber, Fudenberg, Levine, & Rand, 2016; Saito, 2015).

[3] Vanberg (2008) points out that behavioral accounts that can explain why people stick to a promise include both a social preference of guilt aversion (Ellingsen & Johannesson, 2004; Ostrom, Walker, & Gardner, 1992) and a social norm of promise keeping (Battigalli & Dufwenberg, 2007; Dufwenberg & Gneezy, 2000), and provides evidence for the latter as primary motive for people to stick to pledges.

short questionnaire and an offer to sign up for blood drives scheduled in the city over the following weeks.

The experiment randomly varies treatments over two dimensions: In the first dimension, we vary the organization holding the blood drive. We work with the German Red Cross, which generally never pays its donors, and a private commercial blood bank, which remunerates donors with 20 euro per donation. In the second dimension, we vary whether the sign-up is offered in private on a tablet computer only visible to our subject, or in public by our enumerator. Other customers as well as friends and family members coming along to the appointment serve as a natural "audience" for the public pledge. The sign-up is not binding, but represents a pledge vis-à-vis the blood collector. In the months after the survey, we observe whether our subjects indeed choose to donate by matching their names with the databases of the two collectors.

The share of subjects who sign up to donate blood at the German Red Cross increases by about 30 percent when the pledge is observed by the natural audience in the city hall waiting room, compared to a baseline rate 23.2 percent of pledges made in private. This effect is entirely driven by individuals that visit the municipal service center with friends or family: their pledging rate increases from 23.6 percent in private to 44.8 percent in public. On the other hand, people visiting the service center alone and people for whom we elicit a pledge to donate blood at the commercial blood bank are not affected by the visibility manipulation.

We interpret this evidence as consistent with our theoretical framework in which image returns from prosocial actions are highest when these are unambiguously prosocial and observed by people we care about.[4] Turning to actual blood donations from individuals in our study, we find that less than 1 percent fulfilled their pledges. This strikingly low rate makes it hard to investigate how our treatment manipulations affected donations. We discuss potential explanations for this finding. While our experimental design aimed to be minimally invasive and closely resemble a real-world blood drive, we also discuss potential extensions that could shed more light on the low fulfillment.

This chapter makes at least three contributions to the literature. First, it contributes to a burgeoning literature in psychology and economics concerned with the effect of social image concerns on individual behavior (Bursztyn & Jensen, 2017). We show that image concerns can be leveraged to get people to commit to charitable pledges and that image concerns are likely stronger when the audience of prosocial actions is more closely related to the agent. Second, with the use of charitable pledges, we contribute to a literature on *soft* commitment devices (Bryan, Karlan, & Nelson, 2010).[5]

This literature most commonly studies how such devices can help avoid temptations. Our study uses a soft commitment device that leverages the immediate temptation to give, which alone may not necessarily translate into future giving behavior (Andreoni & Serra-Garcia, 2017). The third contribution is methodological. Our field experiment combines an intercept survey, commonly used in an older marketing literature (Bush & Hair, 1985), with experimental manipulations more commonly used in the modern laboratory. This

[4]Blood donations at the commercial blood bank may not be perceived as unambiguously prosocial because of the monetary incentive associated to the donation of blood.

[5]Bryan et al. (2010) define *hard* commitment devices to impose both material and psychological costs from deviation while *soft* commitment devices only impose psychological costs.

approach increases ecological validity by letting us (i) approach a sample of vastly heterogeneous individuals (ii) from a narrowly-targeted geographical area and by (iii) exploiting natural conditions of the venue of intercept for the identification of behavioral mechanisms.

While our reasoning extends to other forms of costly prosocial behavior, we see our results as particularly relevant in the domain of human tissue donations. In light of aging and thus shrinking donor pools and likely increases in the demand for blood due to medical procedures (Greinacher, Fendrich, Brzenska, Kiefel, & Hoffmann, 2011), medical providers and blood banks are constantly aiming to recruit new donors. Approaching new prospective donors on the street is one such recruitment tool. We closely cooperated with the two blood collectors in this study to maintain high ecological validity and to draw conclusions for donor recruitment. With our focus on human tissue donations, we speak to an emerging literature at the intersection of behavioral economics and health economics (Galizzi & Wiesen, 2018).

4.2 Theoretical Framework

To fix ideas for our empirical analysis of how social image concerns can affect the act to pledge a later donation, we rely on the theoretical framework by Benabou and Tirole (2006), in which the decisions of agents to participate in some prosocial activity carry reputational costs and benefits.[6] We abstract from any direct payoffs from intrinsic and extrinsic motivations that agents might have to participate in the prosocial activity to focus on the implications of visibility. We restrict our attention to social image concerns using the simplified framework provided by Bursztyn and Jensen (2017).

Formally, each agent i in our environment can undertake a binary action, say a pledge to donate, $p_i \in \{0, 1\}$. This action may be visible to a reference group j. Taking the action is informative about the type of agent $\sigma_i \in \{l, h\}$, where to her reference group j type h is seen as more socially-desirable by others than type l. Utility from social image to agent i is then

$$S_i = \lambda_{i,j}\ E_i\left[\omega_j\right]\ \mathrm{Pr}_i(\sigma_i = h|p_i) \tag{4.1}$$

where $\lambda_{i,j}$ is the degree to which the agent cares about being perceived as socially desirable in her reference group j. $E_i\left[\omega_j\right]$ is the expectation that agent i has about how socially-desirable it is to be seen as a high type by other agents in her reference group j, measured by $\omega_j > 0$. Finally, $\mathrm{Pr}_i(\sigma_i = h|p_i)$ is the probability that taking action p_i reveals agent i to be of type h to others in the reference group.

Following this framework, we can empirically detect social image effects in at least three ways. First, and maybe most obviously, social image concerns depend on whether other agents can update their beliefs about the type of agent i, $\mathrm{Pr}_i(\sigma_i = h|p_i)$. When pledges are not observable, agents in the reference group cannot update their beliefs and social image concerns vanish. Conversely, the easier it is for agents in the reference group to observe the actions of agent i, the more salient social image concerns become. Most of the literature has tested for social image concerns by exogenously varying whether actions are observable

[6] This theoretical framework builds on theories of esteem (Bernheim, 1994) and self-signaling (Bodner & Prelec, 2003), and provides a unifying theory to explain prosocial behavior in the presence of incentives.

(Ariely et al., 2009; DellaVigna et al., 2012; DellaVigna, List, Malmendier, & Rao, 2017), we vary the visibility of the pledge to donate.

Second, the degree to which an agent i cares to be perceived in a positive light by others in a reference group j, $\lambda_{i,j}$, amplifies the effect of any social image concerns that might be operative. A test for social image concerns that exogenously varies visibility should thus find a greater effect in when the agent cares more about being perceived in a positive light in group j, i.e. when $\lambda_{i,j}$ is bigger. This can be due to personal preference, the setting in which actions are taken, or the composition of the reference group. We might, for example, care more about how our actions are perceived when reference group consists of close friends and family and more generally to people with whom we prospect future interactions as opposed to complete strangers. Funk (2010) finds evidence that social pressure to vote in Swiss elections is stronger in smaller and more close-kit communities.

Third, the expectation of agent i about social desirability in the reference group j of taking an action and being seen as the high type, $E_i\left[\omega_j\right]$. Similar to concern for being perceived in a positive light by the reference group, the expectation about social desirability amplifies any social image effects that might be at work. Social desirability depends both on the underlying value that the group attaches to being a high type, and the agent's expectations thereof. The former could differ, for example, for different charities. The latter could be affected by social norms. Ariely et al. (2009) manipulate the nature of the cause that subjects of a lab experiment can donate to and show that donations for a "bad cause", in their case the National Rifle Association, are not significantly different in public and in private. We are not aware of any other empirical tests of how the social desirability of taking an action shapes the effect of social image concerns.

4.3 Experimental Design and Procedures

Our experimental setup lets us study the three factors that shape social image concerns in a natural setting. Our experiment uses a 2×2 between-subject design. In the first dimension, we randomly vary the visibility of actions (*PUBLIC* or *PRIVATE*). In the second dimension, we randomly vary the organization that agents can pledge to donate to (*CHARITABLE or COMMERCIAL*).[7] The two organizations that we work with, a well-known charity and a commercial blood bank that pays its donors for giving blood, are likely perceived differently in terms of social desirability in the sample of people in our study. We offer donations to each of these organizations to separate, random subsamples of subjects.

Within this 2×2 design, we have natural variation in the reference groups of prospective donors, which should by construction be orthogonal to our treatments. This variation informs how a change in how much agents care about the opinions of others may shape social image effects.

[7] The initial design had a third treatment in which agents could chose between a charitable pledge and a commercial pledge to donate. The data for this treatment is made available upon request from the authors, but we do not present it in this chapter due to various differences between the collecting institutions that make this treatment uninformative and difficult to interpret.

4.3.1 Local Context and Partner Organizations

We conduct our field experiment in Germany, which stands out among high-income countries in that a sizable share of human whole blood donations are incentivized with cash payments. Germany is the fifth-largest blood supplier in the world (Paul Ehrlich Institut, 2015) and has the highest per capita rate of donations among all countries reporting to the World Health Organization (World Health Organization, 2017). Of the 4.4 million whole blood donations collected in 2014, about 71 percent were collected by the German Red Cross, which never pays its donors. The remaining 29 percent represent the military, private donors, and larger hospitals. The latter two groups commonly pay their donors, sometimes up to about EUR 30 euro per donation.[8] The German legal framework (*Transfusionsgesetz* §10) recommends unpaid donations but allows for an unspecified monetary "compensation" (*Aufwandsentschädigung*).[9]

The fact that paid and unpaid incentive schemes coexist in Germany's market for blood enables us to vary incentives to donate blood in a natural setting. We conduct our field experiment in Bonn, a city of about 310,000 people in the populous Rhine-Ruhr region in the west of Germany. We chose Bonn for its competition among various blood collectors in a well-defined geographic area.[10] In Bonn, prospective donors can donate blood in three different ways: First, in frequent mobile donation drives of the German Red Cross held in public squares in the city center. The German Red Cross never pays its donors. Second, during fixed business hours at a commercial blood bank in the city center, which pays 20 euro per donation. Third, during fixed business hours at Bonn University Hospital located about 6 kilometers outside the city. The hospital pays 25 euro per donation. Figure 4.1 shows the location of all donation points on a map. The Red Cross locations represent mobile donation drives during the period of the field experiment (April to May 2017).

We exclude Bonn University Hospital for three reasons. First, it takes about 30 minutes to reach the hospital using public transport from the city center. In comparison, the commercial blood bank and the Red Cross donation drives are all in walking distance of the city hall.[11] Second, the comparison between the German Red Cross, a well-known charity, and a commercial blood bank presents a starker contrast in terms of social desirability of the donation. This is less clear for the University Hospital, which is in public ownership. Third, our enumerators were clearly identifiable as affiliates of the University of Bonn. Including blood drives at the university-owned hospital could have induced significant experimenter demand effects for those offers.

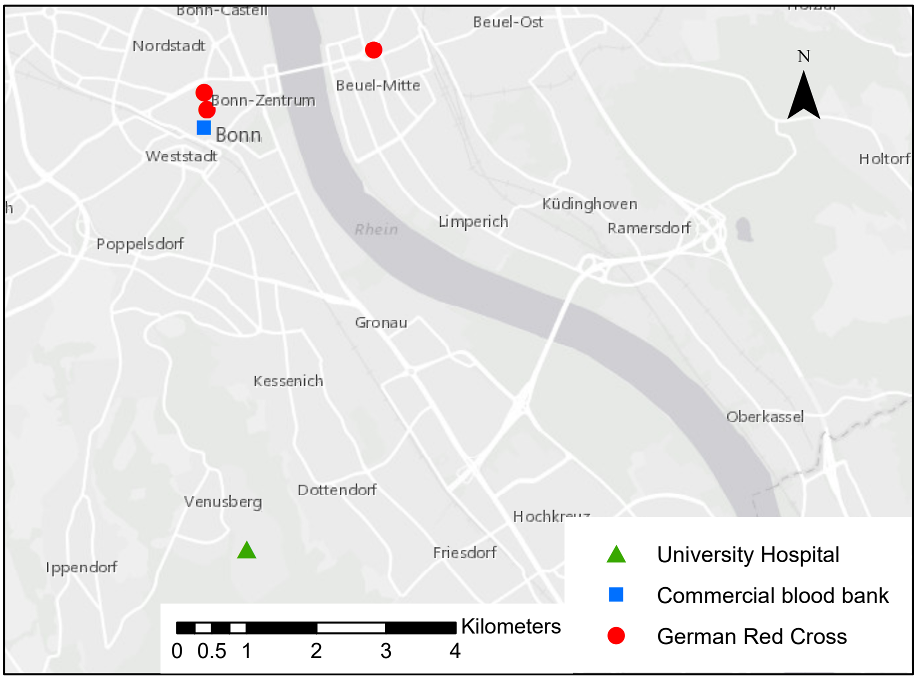
[8] It is difficult to estimate exact numbers because the German government does not publish data on blood donations by type of remuneration, while the relevant WHO database on blood donations is not nationally representative.

[9] It is interesting to note that the German Red Cross, as quasi-monopolist, has unsuccessfully taken legal action to stop remunerated donations. Most recently in 2012, the Higher Administrative Court of Rhineland-Palatinate (*Oberverwaltungsgericht Rheinland-Pfalz*) dismissed legal action of the German Red Cross against the university hospital in Mainz, who regularly pays its donors. The court found the payment to be lawful. See also Oberverwaltungsgericht Rheinland-Pfalz (2013).

[10] The authors considered various German cities and chose Bonn after studying local competition and after conversations with various actors in the market for blood.

[11] We use the Google Maps Distance Matrix API to calculate travel times from the city hall at 8am on a Tuesday morning.

Figure 4.1: Map of Blood Collection Points in Bonn, Germany



Sources: Esri, DeLorme, HERE, OpenStreetMap contributors.
Notes: German Red Cross locations represent mobile donation drives during the period of the experiment (April and May 2017). The University Hospital (green triangle) is not part of this experiment.

Before the experiment, we reached agreements with the regional chapter of the German Red Cross and the commercial blood bank to cooperate in advertising and data sharing. Subjects are informed that we formally cooperate with both organizations. Accordingly, the survey consent form reflects legal requirements of both organizations. The survey software and the "thank you" notes given to subjects use the official logos of our partners.

We recruit subjects among customers of the service center of the Bonn municipal government. The service center, centrally located in the city hall, provides a wide range of in-person administrative services such as applications for official documents, driver's licenses, registration of motor vehicles, and payments for city services. Customers arrive at the service center for appointments that they have previously scheduled online or via telephone. After signing in with the front desk, they wait for their appointment in a designated waiting area. On an average day, the service center handles about 1,300 appointments during 10 business hours from 8am to 6pm. The average wait time between arrival at the service center and appointment is about 4.5 minutes.

We choose to conduct the experiment in the municipal service center for three reasons: First and most importantly, the population that we can sample from is highly diverse and relevant to study the behavior of potential blood donors. Second, the service center lends itself to an intercept survey because almost all customers have to wait for a few minutes, often with little to do. Third, the physical space of the waiting area with many other people standing and sitting around provides a natural "audience" that we can use to make social image concerns salient. Before the experiment, we agreed with the municipal government on suitable time periods and procedures. Staff of the service center was briefed on our experiment.

4.3.2 Experimental Procedures

Subjects for our experiment are recruited using an intercept survey in the waiting area of the municipal service center. Throughout the operating hours of the service center, our enumerators wait for new customers to arrive in the waiting area. Given the large number of appointments, it was not feasible to interview all customers arriving for appointments. Instead, we opted for a procedure in which our enumerators are instructed to always approach the first new customer to arrive as soon as they have finished with the previous subject. This approach avoids that subjects are influenced by observing other interviews and maximizes the likelihood that our enumerators can complete the interview before subjects are called for their appointment. We restrict our sample to customers that have an appointment, are between 18 and 65 years of age, and are able to speak and read German.

We use CAPI with a tablet computer. Each enumerator uses a 10.1" Android tablet running the Qualtrics Offline Surveys app. Surveys are programmed online in Qualtrics and then downloaded to the tablets for offline use. Tablets are operated in a kiosk mode that does not permit operations other than answering the survey. Responses are stored on the tablet and regularly transmitted to the Qualtrics server using an encrypted connection over a WLAN network in the city hall.

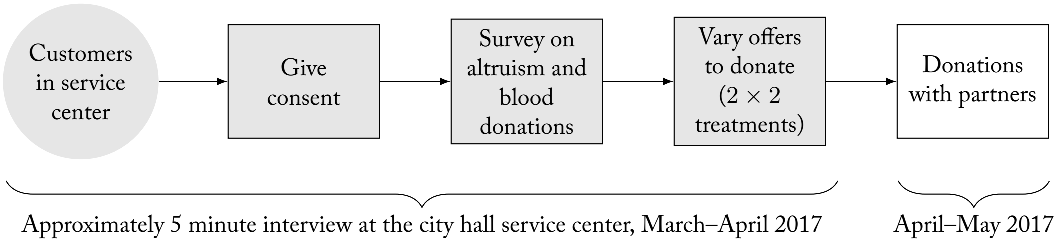
Enumerators can read instructions off the screen of their tablet. When approaching new customers using a rehearsed script, enumerators verbally ask for an initial consent to participate in a short survey. If customers agree, enumerators hand over the tablet to let customers read background information on the survey, a privacy notice, and to let them complete a written informed consent form. Once customers have completed the consent form with their personal details, we treat them as subjects.

Subjects then use the tablet to complete a short self-administered questionnaire about prosocial behavior, perceptions and preferences related to donating blood, awareness of various ways to donate in Bonn, and number of years lived in Bonn. Finally, the survey software administers one of four treatments (see detailed description below). The order of questions is not randomized and the offer to sign up for a donation is always the last element of the survey. Answer choices for categorical questions are displayed on screen in a random order. When beginning the survey, subjects do not know that they will be offered to sign up for a donation at the end.

Throughout the self-administered survey, enumerators stand by to guide and assist subjects in operating the tablet computer. Every time an enumerator approaches a customer of the service center, irrespective of whether the customer gives consent or completes the survey, the enumerator completes a short questionnaire designed to capture observable characteristics and environmental conditions. These include estimated age range, gender, whether customers came in a group, potential problems to complete the questionnaire, how crowded the waiting area was, and reasons for aborting the survey. This information serves to understand potential selection effects in our sample. Figure 4.2 summarizes the sequence of events in the experiment.

We recruited two student research assistants at the University of Bonn as enumerators. We require both enumerators to wear a visible photo ID badge that clearly identifies them as working for the University of Bonn, a large and well-known public university. We conduct

FIGURE 4.2: Sequence and Timeline of the Experiment



a detailed training of enumerators to make sure that the procedures are implemented as described above. In addition to the tablet computer, "thank you" notes, and randomization protocol, enumerators carry a copy of the survey protocols that they can refer to. Surveys were conducted in the municipal service center for 92 hours, spread out over 18 workdays in four weeks in March and April 2017.

4.3.2.1 Treatments

Our experiment offers to sign subjects up to donate blood with one of our partner organizations. We cannot legally require subjects to follow through on their pledge. Instead we explain that we pass on sign-up confirmations for a donation drive to the relevant partner organization, thus emphasizing that the sign-up represents a pledge vis-à-vis the partner organization.[12]

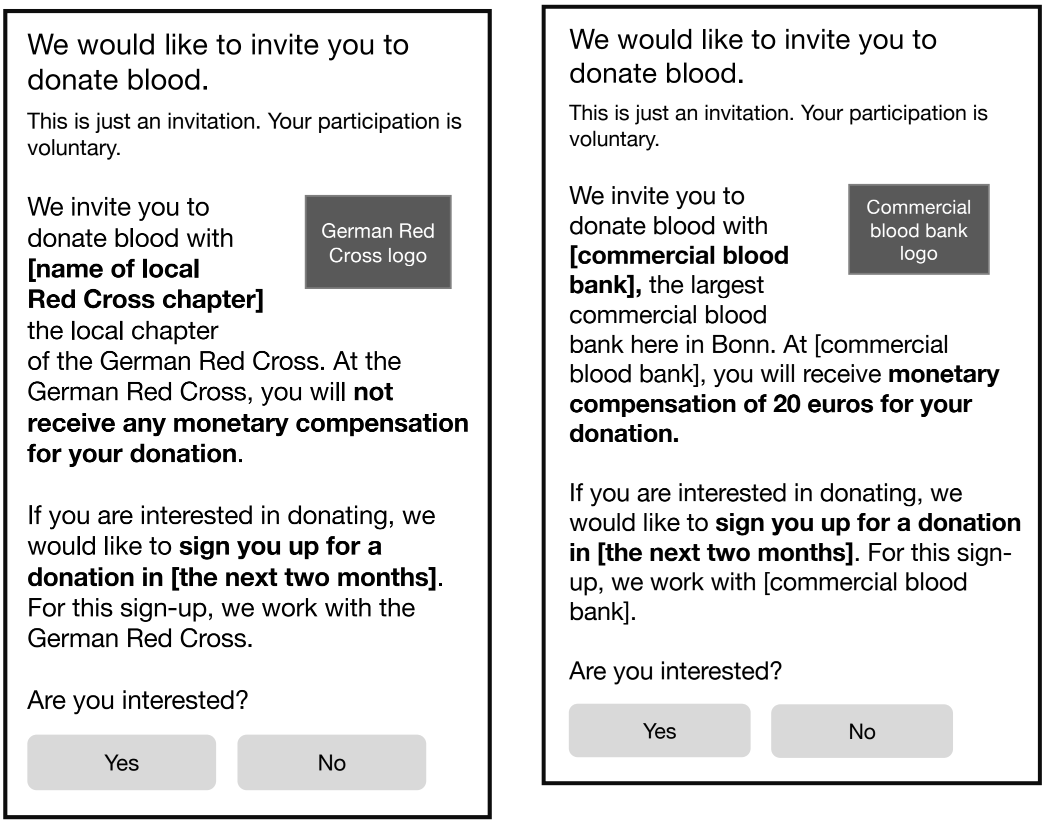
We use a 2×2 between-subject design that varies the sign-up process over two dimensions: Along the first dimension, we vary the organization that subjects can pledge to donate to. This lets us manipulate incentives to donate in a natural way. In the first treatment, subjects can sign up for a donation with the local chapter of the German Red Cross, a well-known charity that does not provide monetary compensation for donations (*CHARITABLE* treatment).

In the second treatment, subjects can sign up for a donation with a commercial blood bank, which provides a monetary compensation of 20 euro per donation (*COMMERCIAL* treatment). Subjects are informed about the compensation (or lack thereof) in their treatment but they do not learn about the other option to donate. Figure 4.3 illustrates the tablet screens for both treatments.

Along the second dimension of the 2×2 between-subject design, we vary the visibility of the offer and sign-up process. In the *PRIVATE* treatment, subjects are presented with the offer to sign up privately on the screen of the tablet. After having completed the survey on altruism and preferences to donate blood, the software presents the offer to sign up for a donation as an additional, last screen. Subjects accept or decline using buttons on the screen. In the *PUBLIC* treatment, the software asks subjects to return the tablet computer

[12]The exact phrasing is as follows: "We invite you to donate blood with [partner organization, depending on treatment, with further explanation]. If you are interested in donating, we would like to sign you up for a donation in [the next two months]. For this sign-up, we work with [partner organization]. Are you interested?".

Figure 4.3: Illustration of Tablet Screens with Donation Offer, by Treatment



(a) *CHARITABLE* treatment (b) *COMMERCIAL* treatment

Notes: This figure is an illustration that approximates the layout of the tablet screens, with instructions translated from German.

to the enumerator after the survey has been completed. The enumerator then advances to a hidden next screen and reads out loud the same offer that in the *PRIVATE* treatment is presented on the screen. Instead of using buttons on the screen, subjects in the *PUBLIC* treatment are required to say out loud whether they would like to sign up for a donation. Other customers waiting in the service center and any friends and family who might be accompanying the subject serve as an "audience" for the public commitment.

All subjects who sign up for a donation receive a "thank you" card for the organization that they signed up with. The cards are printed on high-quality paper and are meant as a token of appreciation to remind subjects of their pledge to donate vis-à-vis the partner organization. They also provide information on where and when they can donate with the relevant partner. Figure 4.4 presents the card design for the German Red Cross (i.e. the card that is given out in the *CHARITABLE* treatment). In the *PRIVATE* treatment, the enumerator learns about the subject choice when the tablet computer is returned to the enumerator. The survey software shows a small graphic at the top of the screen that enables the enumerator to quickly recognize whether the subject chose to sign up.

The type of donation offer was randomized at the hourly level, i.e. over the 92 hours that enumerators were present in the municipal service center. We chose the hourly treatment

Figure 4.4: "Thank You" Card for Red Cross Sign-Up

(a) Front

(b) Back

Notes: Front reads: "Thank you for your participation. We signed you up for a donation with the [local blood donation service of the German Red Cross]. You can find out where and when to donate on the back of this card. We are looking forward to your donation". Back reads "We expect you at one of our donation drives in Bonn over the next few weeks".

assignment because it minimized the chance that subjects would see our enumerators offer donations with a different organization to later subjects (and thus potentially cause inquiries) and because it simplified administration of the survey, in particular the handling of "thank you" cards, for our enumerators. Randomization of donation offers was done before the start of the experiment using Microsoft Excel. The visibility of the offer was randomly allocated between all subjects by our survey software during the experiment.

4.3.3 Donation Drives and Tracking of Subjects

Our study design allows tracking of subjects from the municipal service center to blood drives of our two partner organizations in a period of two months after the initial interviews were conducted. Subject consent and personal information collected during the survey lets us match individual-level data for all subjects, irrespective of treatment, with donation records. For this purpose, the consent form included waivers of medical confidentiality so that both of our partner organizations could report donation behavior (but no other personal or medical information) back to us.

Our two partners pursue different strategies to collect donations. The Red Cross does not have a fixed donation center in Bonn, but offers widely-publicized mobile donation drives in public squares in the city center. The commercial blood bank has regular business hours every day during which it accepts walk-in donors. For the purpose of our experiment, we agreed with our partners on five fixed dates and times that were the same between both partners and that were highlighted on the "thank you" card.[13]

While we specifically invited subjects to come to one of these time slots ("we expect to see you at one of these donation drives"), our data also allows us to track subjects who chose to donate at other times or in other donation drives in the region.

4.4 Empirical Analysis

4.4.1 Sample Characteristics and Balance Across Treatments

Given the random sampling protocol adopted by our enumerators, we would expect our sample to be representative of the population of customers of the municipal service center. Over the four weeks of the study, our enumerators approached a total of 1,072 individuals. From this random sample, 264 refused to participate and 194 dropped out during the survey. Our final sample of completed surveys consists of 614 responses and selects our population of interest by over-representing women, younger individuals, non-immigrants, and individuals who come alone to visit the offices of the city-hall. Compared to the population of the City of Bonn, our final sample has the same gender composition as the city average, but is younger and includes fewer immigrants (Appendix Table 4.A1). Older people appear to have dropped out disproportionately due to difficulties of handling the tablet computer. For immigrants, language difficulties appeared to have been the most common reason for not participating or completing the survey.

[13] The times were 12.30pm to 5pm on April 7, April 20, April 21, May 12, and May 19, 2017. Three of the Red Cross drives were scheduled to start an hour later and to last an hour longer than the time slots at the commercial blood bank.

Among the 194 subjects that abort the survey, just 7 abort after treatment assignment with no differential abortion rates across treatment groups. Our sample is mostly balanced on observables. Table 4.1 presents summary statistics for the final sample of 614 individuals across treatment assignments. We use a one-way ANOVA on ranks (Kruskal-Wallis) to compare observable characteristics, reported by subjects before treatment and measured by the enumerators after the treatment. We document imbalance in pre-treatment measure of awareness of the commercial blood bank in Bonn and in the gender of subjects (as reported by enumerators). In our discussion of results below, we use parametric estimates that include controls for observable characteristics.

4.4.2 Take-up of Pledges in City Hall Experiment

In this section, we study how social image effects shape the take-up of pledges to donate blood. We use the conceptual framework by Bursztyn and Jensen (2017), outlined above, to guide our analysis. Recall that our experimental design randomly varies whether survey respondents were offered to pledge a blood donation publicly in front of a natural audience, or privately on a tablet computer. We also vary randomly the organization that subjects can pledge to donate to. If subjects perceive these organizations to be different in their social desirability, this should change the strength of the social image effect. We then exploit the fact that a considerable share (30 percent in our sample) of customers of the municipal service center are accompanied by one or more friends or family members. We assume that surveyed subjects care to be seen in a positive light by these people. Through this design we are able to first identify social image effects in front of the natural audience, and then test the additional implication of the social signaling model, which predicts a stronger social image effect around people whose opinion matters more to the agent and for actions that are perceived to be more socially-desirable.

Figure 4.5 illustrates the descriptive evidence on the take-up of pledges. We wish to highlight two main patterns: First, we do not observe a strong visibility effect for pledges in the *CHARITABLE* treatment. Second, we observe a strong visibility effect for pledges in the *CHARITABLE* treatment. This effect is largely driven by individuals that visit the city hall with friends and family. Table 4.1 (panel b) lists donation rates. In the remainder of this section, we test the statistical significance of these patterns using parametric analysis.

For each of the two treatments that vary the offer under which the pledge is elicited e (*CHARITABLE*, or *COMMERCIAL*), we estimate a separate logit model with following specification:

$$P_{e,i} = \alpha_e + \beta_e Public_{e,i} + \mathbf{X}'_{e,i}\boldsymbol{\delta}_e + \varepsilon_{e,i} \tag{4.2}$$

where $P_{e,i}$ is a binary variable indicating whether subject i pledged to donate under treatment e, $Public_{e,i}$ denotes the binary variable taking value 0 if i was asked to pledge a blood donation privately on a tablet and 1 if i was asked to pledge a blood donation out loud from the enumerator in front of the natural audience of the waiting area in the municipal service center. $\mathbf{X}'_{e,i}$ is a vector of controls for individual characteristics.

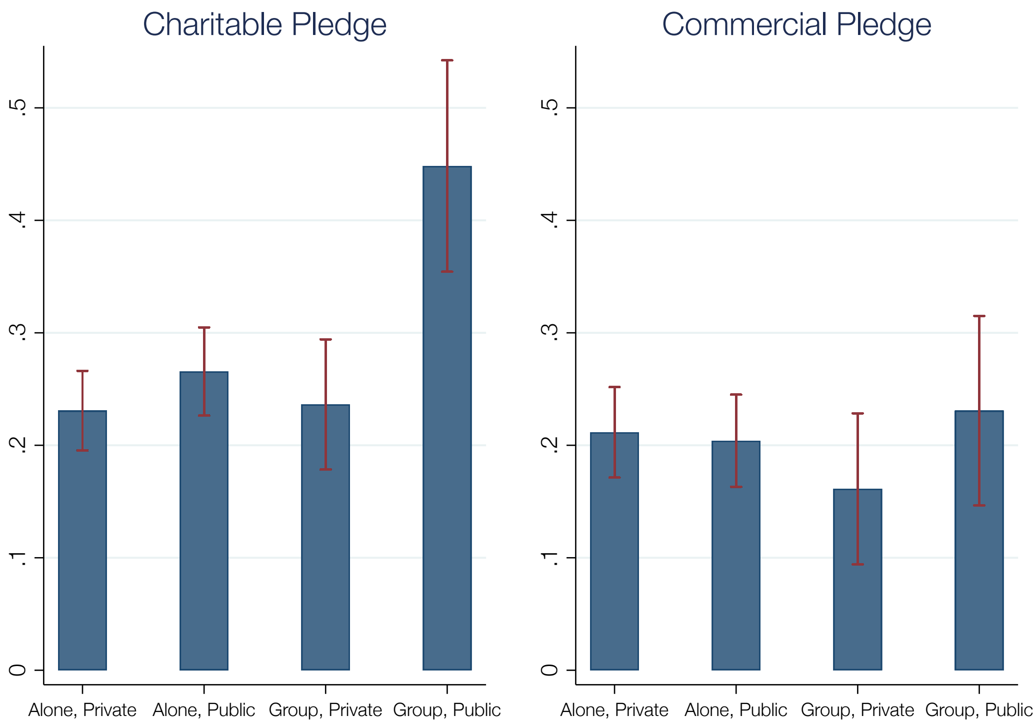
Table 4.2 (columns 1, 2, 5, and 6) presents the average marginal effects from logit estimation of Equation 4.2. We find that varying the visibility of the pledge has no detectable effect on uptake in either of the two subsamples. In our simple theoretical framework, all three factors can explain this null result: (i) it could be that the probability that others can update their assessment of the subject's generosity based on the pledge is very small, i.e. that the signal is not effective, (ii) it could be that subjects do not care to be perceived in

Table 4.1: Summary Statistics for Participating Subjects, by Treatment Assignment (Means and Standard Errors in Parentheses)

	Full sample	Charitable		Commercial		P-value
		Private	Public	Private	Public	
a) Self-reported before treatment						
Frequency of altruistic activity	3.059	3.086	3.019	3.067	3.056	0.922
	(0.037)	(0.064)	(0.078)	(0.076)	(0.082)	
Importance of donating blood	4.007	4.030	3.955	3.881	4.169	0.089
	(0.043)	(0.076)	(0.083)	(0.095)	(0.093)	
Perception of blood donors as altruists	4.153	4.157	4.242	4.074	4.121	0.551
	(0.036)	(0.064)	(0.065)	(0.082)	(0.081)	
Awareness of institutions: DRK	0.855	0.833	0.892	0.815	0.887	0.155
	(0.014)	(0.027)	(0.025)	(0.034)	(0.029)	
Where would you go to donate: DRK	0.412	0.409	0.427	0.422	0.387	0.914
	(0.020)	(0.035)	(0.040)	(0.043)	(0.044)	
Awareness of institutions: Commercial	0.132	0.157	0.089	0.185	0.089	0.031
	(0.014)	(0.026)	(0.023)	(0.034)	(0.026)	
Where would you go to donate: Commercial	0.029	0.030	0.000	0.044	0.048	0.060
	(0.007)	(0.012)	(0.000)	(0.018)	(0.019)	
Awareness of institutions: University	0.705	0.667	0.752	0.741	0.669	0.202
	(0.018)	(0.034)	(0.035)	(0.038)	(0.042)	
Where would you go to donate: University	0.559	0.561	0.533	0.565	0.573	0.918
	(0.020)	(0.035)	(0.040)	(0.043)	(0.045)	
Respondent age	34.415	33.556	34.312	35.807	34.403	0.359
	(0.480)	(0.827)	(0.966)	(1.034)	(1.075)	
Respondent years lived in Bonn	5.666	5.657	5.675	5.689	5.645	0.992
	(0.150)	(0.268)	(0.291)	(0.327)	(0.327)	
b) Uptake of pledges after treatment						
Subject pledged to donate	0.238	0.232	0.299	0.200	0.210	0.179
	(0.017)	(0.030)	(0.037)	(0.035)	(0.037)	
c) Measured by enumerator in post-survey questionnaire						
Male	0.489	0.424	0.459	0.519	0.597	0.018
	(0.020)	(0.035)	(0.040)	(0.043)	(0.044)	
Respondent came in group	0.300	0.364	0.255	0.304	0.250	0.205
	(0.026)	(0.049)	(0.051)	(0.056)	(0.047)	
Respondent immigrant	0.130	0.131	0.127	0.126	0.137	0.993
	(0.014)	(0.024)	(0.027)	(0.029)	(0.031)	
Intensity of social image concern	3.438	3.212	3.618	3.489	3.516	0.004
	(0.045)	(0.081)	(0.085)	(0.097)	(0.101)	
Ability to complete survey	4.203	4.141	4.128	4.348	4.242	0.008
	(0.029)	(0.052)	(0.049)	(0.061)	(0.071)	
Observations	614	198	157	135	124	

Notes: Frequency of altruistic activity asked interviewed subjects how often they engage in altruistic activities, on a 5-point Likert scale where 1 is "never" and 5 is "very often". Importance of donating blood asked interviewed subjects how important they consider donating blood, on a 5-point Likert scale where 1 is "not important" and 5 is "important". Perception of blood donors as altruists asked interviewed subjects to what extent they think is true that a friend or family member is altruistic for donating blood, on a 5-point Likert scale where 1 is "not true" and 5 is "true". Intensity of social image concern asked enumerators to record their perceived intensity of social image, on a 5-point Likert scale where 1 is "very weak" and 5 is "very strong", based on how crowded and how quiet the waiting area was. P-value is for a one-way ANOVA on ranks (Kruskal-Wallis) test comparing the four groups.

FIGURE 4.5: Share of Subjects Pledging a Blood Donation Across Treatments, Split by Whether They Visit the City Hall Alone



Notes: "Public" and "private" are randomly assigned treatments while "alone" and "group" are based on whether or not the subject is accompanied by one or more friends or family members. Error bars indicate standard error of the mean.

a positive light by the group of people in the waiting area, or (iii) it could be that subjects believe that pledging to donate by either of the collectors is not seen as socially desirable by the group of people in the waiting area. We now discuss each of these factors in turn.

Our survey data provides suggestive evidence that (iii) alone is unlikely to explain why social image effects are not operational. When we ask subjects pre-treatment whether they agree or disagree that blood donors are perceived as altruists, we find that 41 percent of subjects strongly agree and another 40 percent agree (overall mean of 4.15 on a 5-point Likert scale).

Turning to (ii), we can explore how the composition of the reference group affects individual propensity to pledge a blood donation by studying heterogeneous treatment effects of the visibility treatment for interviewed subjects who come to the municipal service center alone and those that come in a group. Therefore we extend Equation 4.2 as follows:

$$P_{e,i} = \alpha_e + \beta_e \left(Public_{e,i} \times Group_{e,i} \right) + \mathbf{X}'_{e,i} \boldsymbol{\delta}_e + \varepsilon_{e,i} \tag{4.3}$$

where $Group_{e,i}$ is an indicator variable for whether individual i came to the city hall alone ($Group_{e,i} = 0$) or in a group ($Group_{e,i} = 1$).

Table 4.2 (columns 3, 4, 7, 8) presents average marginal effects. We find that the composition of the reference group indeed shapes social image effects in the uptake of donation

Table 4.2: Logit for Heterogeneous Social Image Effects on Pledge to Donate Blood (Average Marginal Effect Estimates and Standard Errors in Parentheses)

	Charitable				Commercial			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Treatments (baseline: Private × Alone)							
Public	0.070	0.073			0.034	0.035		
	(0.047)	(0.047)			(0.056)	(0.055)		
Private × Group			-0.051	-0.063			-0.055	-0.061
			(0.068)	(0.067)			(0.085)	(0.083)
Public × Alone			0.016	0.014			0.009	0.005
			(0.054)	(0.054)			(0.063)	(0.062)
Public × Group			0.210**	0.217**			0.078	0.090
			(0.103)	(0.102)			(0.111)	(0.112)
	Control variables							
Male	-0.035	-0.042	-0.038	-0.047	0.039	0.031	0.030	0.020
	(0.050)	(0.050)	(0.050)	(0.050)	(0.059)	(0.059)	(0.059)	(0.059)
Group	0.040	0.036			-0.001	0.002		
	(0.057)	(0.057)			(0.069)	(0.069)		
Donor at DRK	0.075	0.081	0.059	0.064	-0.015	-0.017	-0.011	-0.011
	(0.091)	(0.091)	(0.089)	(0.089)	(0.091)	(0.093)	(0.092)	(0.093)
Donor at commercial bank	-0.044	0.014	-0.053	0.010	0.322*	0.214	0.304*	0.191
	(0.188)	(0.220)	(0.181)	(0.216)	(0.176)	(0.195)	(0.177)	(0.193)
Personal characteristics	Y	Y	Y	Y	Y	Y	Y	Y
Awareness blood market	N	Y	N	Y	N	Y	N	Y
Image effect: Group vs Alone Difference (p-value)			0.047	0.029			0.379	0.302
Observations	355	355	355	355	259	259	259	259

$^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$

Notes: Binary variable indicating pledge to donate blood is the dependent variable in all regression models from (1) to (4). Models (1) to (4) estimate regression for pledge to chariable blood donation, (5) to (8) for pledge to commercial blood donation. "Personal characteristics" include age group, migration background, charitable behavior (stated intensity on a 1 to 5 Likert scale), perceived relevance of donating blood (stated intensity on a 1 to 5 Likert scale), and perceived altruism of people donating blood (stated intensity on a 1 to 5 Likert scale). "Awareness blood market" include binary variables indicating awareness of DRK, awareness of the commercial blood bank, and awareness of the Bonn university hospital blood collection center.

pledges. When people are asked to pledge a donation with the German Red Cross, making the request public increases willingness to pledge among people who came in a group ($p < 0.05$) without affecting people coming alone. This effect is consistent with the theoretical mechanism: individuals care more to signal altruism to a socially connected audience.[14]

Interestingly, social image effects for the *COMMERCIAL* treatment are generally much weaker and are not detectable at any conventional level of confidence, both for people coming alone, as well as for people coming in groups. We interpret the stronger effect for the *CHARITABLE* treatment as consistent with the theoretical framework, in that the strength

[14]A related interpretation of theory and evidence is that respondents that visit the city hall with friends or family are systematically more concerned about how others perceive their actions.

of the social image effect should be an increasing function of the perceived social desirability of the action.

Finally, turning to (i), we cannot rule out that our manipulation of visibility failed to increase the probability that others could update their assessment of the subject's generosity based on the pledge.

Pledges are inherently different from actual prosocial behavior because they depend on later fulfillment. It could be that public image concerns are not operational because a reference group of strangers in the municipal service center cannot hold subjects accountable to fulfill the pledge later. This would be consistent with our finding that social image concerns are indeed only operational for subjects who are accompanied by other people.

In addition to these three explanations based on our theoretical framework, we now assess whether our results could be driven by systematic attrition or by competing explanations. We first consider attrition. On the one hand, attrition after treatment assignment is negligible. Of the 194 subjects who dropped out after consenting to participate, only seven did so after the treatment was administered, with no significant difference between treatments.

This suggests that internal validity of our study is not affected by attrition. However, because subjects who drop out before treatment are relatively more likely to have come to the city hall with friends or family, the effect of visibility on pledges for the random draw of people intercepted in groups may be either over- or under-estimated depending on how selection interacts with visibility treatment.

Finally, we consider a competing interpretation of our results. It could be that individuals pledge more frequently when they come in a group because they can learn about the costs and benefits of donating from whoever accompanies them. Under this competing interpretation, we would expect that less-informed subjects drive the visibility effects.

To assess this alternative explanation, we can use the data that we collect prior to treatment. We ask each subject whether they are aware of various organizations collecting blood donations in the area. Subjects who do not know that the DRK is collecting donations in the area are presumably less informed about the costs and benefits of donating blood with this organization. We find no evidence that the visibility effect is driven by these individuals. Appendix Table 4.A2 replicates the analysis of visibility effects but excludes subjects who do not know about the DRK.

4.4.2.1 Fulfillment of Pledges at Blood Drives

Our experiment is designed to investigate the effect of making the choice to commit to a blood donation visible on stated willingness to donate. Because of our partnership with two blood banks, we can also study how social image concerns affect fulfillment rates of pledges, hence actual donations.

We find that the conversion rate of pledges into actual donations is strikingly low. From the initial sample of 614 interviewed subjects we discard 18 observations for which we did not obtain full names to match to our partners' records. Of the 596 remaining observations 141 (23.66 percent) pledged to make a blood donation in April or May 2017. Of those subjects, only one subject donated during the suggested period. Surprisingly, of the 455 subjects who did not pledge to donate, four subjects donated during the same period following our survey (See Table 4.3 for a summary of actual donations across treatments).

These conversion rates are low in comparison to another study that similarly elicits pledges to donate blood among university students (Stutzer, Goette, & Zehnder, 2011).[15]

Among all 596 subjects in our sample who provided a complete name, 65 had previously donated either at the German Red Cross or at the commercial blood bank, with no significant differences between treatment assignment. While the number of actual donations is too small for a statistically meaningful comparison, it appears that subjects who have previously donated blood were slightly more likely to donate following our interview. Among the 65 subjects that had previously donated, 3 subjects (4.6 percent) donated again following the interview. Among the 531 subjects that had not previously donated, 2 subjects (0.38 percent) donated again following the interview.

We can benchmark these numbers to a series of experiments from Goette, Stutzer, Yavuzcan, and Frey (2009), in which a summer blood donation elicitation campaign lead to a conversion rate of approaches over donation of about 0.6 percent for Zurich citizens who had not previously donated and 45.3 percent of registered donors of the Swiss Red Cross in Zurich. Comparatively, our campaign was ineffective at inducing donations particularly among subjects that have previously donated.

We can also benchmark our donation numbers to the national donation rates in Germany. Over the whole year of 2017, the rate of donations in the population was about 4.8 percent (Paul-Ehrlich-Institut, 2018). This rate is higher than the donation rate among first time and previous donors in our study, though we note that subjects in our survey only had a time window of approximately two months after the survey to donate with one of our partner organizations.

Table 4.3: Fulfillment of Pledges at Partner Blood Drives, by Treatment Assignment (Number of Subjects)

	Whole sample	Charitable		Commerical	
		Private	Public	Private	Public
a) Name matching and donor status of study subjects					
All subjects	614	198	157	135	124
of which matched with donor databases	596	193	151	131	121
of which donated with either blood collector	65	18	16	14	17
b) Pledges and donations					
Pledged a donation in study	141	44	45	26	26
of which donated	1	1	0	0	0
Did not pledge a donation in study	455	149	106	105	95
of which donated	4	3	0	0	1

4.5 Discussion and Conclusion

Although pledges to donate are widely used by organizations to encourage contributions, particularly human tissue donations, there is little evidence on their efficacy in changing

[15] Stutzer et al. (2011) document a conversion rate of about 54 percent for pledges over blood donations that take place on the same day.

behavior. In this paper, we aim to shed light on social image concern as an underlying mechanism. Using the real-world setting and a research design with high ecological validity, we set out to study how social image concerns affect both the uptake and the fulfillment of pledges to donate blood.

The results of our field experiment show that the uptake of pledges is consistent with a theoretical framework in which social image concerns are amplified when subjects care more about being perceived favorably by a reference group of people and when pledging to donate to a more socially-desirable organization. We find evidence for social image concerns when subjects are asked in public whether they would like to pledge a donation with the Red Cross.[16] When subjects are accompanied by friends and family members, public offers significantly increase the likelihood of pledging to come to a donation drive. When subjects are not accompanied by anyone, but just surrounded by other customers waiting in the municipal service center, we do not find significant differences between public and private offers. Similarly, social image concerns do not appear to play a role when subjects are offered to sign up for a remunerated donation with a commercial blood bank.

These findings contribute to the growing academic literature on the role played by social image considerations in economic behavior. We document that pledging behavior is consistent with a simple model of social image concerns even when the act of doing good itself is not observable to others.

At the same time, pledges in our particular context do not appear to induce any additional blood donations. Almost all subjects renege on their pledge, with no detectable differences between treatments. This result is in line with Lacetera, Macis, and Mele (2016), who collaborate with a firm that runs fundraising campaigns in an online social network. They provide evidence that individuals may broadcast pledges to donate money in order to signal generosity. While broadcasting appears to be correlated with donations, they show in a separate field experiment that stated support and explicit pledges to donate largely fail translate into actual donations. Our paper can be seen as important complementary evidence to Lacetera et al. (2016) for identifying the effect of an exogenous manipulation of the visibility of pledges, and for providing evidence that pledges are reneged even in the absence of intermediation fees that potentially hamper the effectiveness of donations (Gneezy, Keenan, & Gneezy, 2014) or act as excuses not to give (Exley, 2018). [17]

We see the lack of fulfillment in our experiment as an important starting point for further academic and policy-oriented work:

From an academic perspective, various mechanisms could explain why individuals would renege or follow through on their pledges. While our experiment was not designed to disentangle them, future field experiments could systematically vary the psychological costs of reneging on pledges, for example by varying the time lag between pledge and donation or by varying the framing of the initial pledge. Additional laboratory-based work could help shed light of the relative importance of moral consistency (Cialdini & Trost, 1998; Cioffi &

[16]Consistent with this finding, recent experimental evidence (Karing, 2018) indicates that social image effects on child immunization decisions are stronger for vaccines that are perceived to be more socially desirable.

[17]Similarly, Exley and Naecker (2016) find that hard commitment devices for workshop attendance can be used to engage in social signaling. Although their test is likely underpowered, they also find that providing the commitment device has no effect on workshop attendance.

Garner, 1996) and promise-keeping (Charness & Dufwenberg, 2006) as underlying reasons for not wanting to break a promise to donate.

From a policy perspective, we take our findings as a reminder that simple, behaviorally-informed strategies designed to promote desirable behaviors can have their limits. While such "nudges" can steer people to perform one specific action, they may not have a sustained impact beyond a specific moment, location, or context. Organizations looking to harness pledges should thus consider them in tandem with other strategies to increase conversion rates. One simple strategy can be to reduce the temporal or spatial gap between pledge and donation. When the pledge to donate and the actual donation have to remain separate in time or space, another strategy could involve reminding individuals of their pledge. Andreoni and Serra-Garcia (2017) show that sending 'thank you' cards before the decision to donate can be highly effective in reducing reneging on the pledge.

Compared to simple nudges such as defaults, the efficacy of pledges as a tool to change behavior likely depends on a more complex set of psychological and economic mechanisms. Far more research is needed to understand them.

4.A Appendix: Additional Tables

Table 4.A1: Summary Statistics for City Population and Potential Study Population, by Participation Status (Means and Standard Errors in Parentheses)

	City of Bonn	Study sample	*of which:* participated	*of which:* aborted	*of which:* no consent	(3)=(4)=(5) p-value
	(1)	(2)	(3)	(4)	(5)	(6)
Age	41.8	n/a	34.4	56.0	n/a	
			(0.384)	(0.776)		
Age group	n/a	3.982	3.480	5.804	3.811	0.000
		(0.046)	(0.050)	(0.093)	(0.076)	
Male	0.483	0.519	0.489	0.521	0.587	0.028
		(0.015)	(0.020)	(0.036)	(0.030)	
Immigration background	0.288	0.254	0.130	0.253	0.542	0.003
		(0.013)	(0.014)	(0.031)	(0.031)	
Respondent came in group		0.329	0.300	0.412	0.337	0.000
		(0.019)	(0.026)	(0.043)	(0.039)	
N	327,919	1,072	614	194	264	

Source: Data for Bonn taken from Bonn City Government Statistical Office 2017 population statistics. *Notes:* Respondent age groups: 1 “under 18” 2 “18 to 24” 3 “25 to 34” 4 “35 to 44” 5 “45 to 54” 6 “55 to 64” 7 “64 or older”. We report data for average age separately because could not reconstruct our survey age groups from the publicly available population data. P-value is for a one-way ANOVA on ranks (Kruskal-Wallis) test comparing columns (3), (4), and (5).

Table 4.A2: Logit for Heterogeneous Social Image Effects on Pledge to Donate Blood From Subsample Aware of DRK as Blood Collection Agency (Average Marginal Effect Estimates and Standard Errors in Parentheses)

	Charitable			
	(1)	(2)	(3)	(4)
Treatments (baseline: Private × Alone)				
Public	0.083	0.083		
	(0.051)	(0.051)		
Private × Group			-0.058	-0.068
			(0.075)	(0.074)
Public × Alone			0.029	0.026
			(0.058)	(0.058)
Public × Group			0.232**	0.235**
			(0.111)	(0.111)
Control variables				
Male	-0.071	-0.076	-0.079	-0.085
	(0.054)	(0.053)	(0.053)	(0.053)
Group	0.043	0.040		
	(0.063)	(0.062)		
Donor at DRK	0.105	0.113	0.087	0.094
	(0.096)	(0.096)	(0.095)	(0.095)
Commercial blood bank donor	-0.042	0.017	-0.046	0.019
	(0.190)	(0.223)	(0.186)	(0.221)
Personal characteristics	Y	Y	Y	Y
Awareness rest of the blood market	N	Y	N	Y
Social image effect: Group vs Alone Difference (p-value)			0.055	0.040
Observations	302	302	302	302

$^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$

Notes: Binary variable indicating pledge of a charitable blood donation is the dependent variable in all regression models from (1) to (4). "Personal characteristics" include age group, migration background, charitable behavior (stated intensity on a 1 to 5 Likert scale), perceived relevance of donating blood (stated intensity on a 1 to 5 Likert scale), and perceived altruism of people donating blood (stated intensity on a 1 to 5 Likert scale). "Awareness rest of the blood market" include binary variables indicating awareness of the commercial blood bank and awareness of the Bonn university hospital blood collection center.

4.B References

Andreoni, J., Rao, J. M., & Trachtman, H. (2017, April). Avoiding the Ask: A Field Experiment on Altruism, Empathy, and Charitable Giving. *Journal of Political Economy*, *125*(3), 625–653. doi: 10.1086/691703

Andreoni, J., & Serra-Garcia, M. (2017, September). *Time-Inconsistent Charitable Giving* (Working Paper No. 22824). Cambridge, MA: National Bureau of Economic Research.

Ariely, D., Bracha, A., & Meier, S. (2009, February). Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially. *American Economic Review*, *99*(1), 544–555. doi: 10.1257/aer.99.1.544

Battigalli, P., & Dufwenberg, M. (2007). Guilt in Games. *The American Economic Review*, *97*(2), 170–176. doi: 10.1257/aer.97.2.170

Benabou, R., & Tirole, J. (2006, December). Incentives and Prosocial Behavior. *American Economic Review*, *96*(5), 1652–1678. doi: 10.1257/aer.96.5.1652

Bernheim, B. D. (1994). A Theory of Conformity. *Journal of Political Economy*, *102*(5), 841–877. Retrieved 2017-10-26, from http://www.jstor.org/stable/2138650

Bodner, R., & Prelec, D. (2003). Self-Signaling and Diagnostic Utility in Everyday Decision Making. In I. Brocas & J. D. Carrillo (Eds.), *Rationality and Well-Being* (Vol. 1). Oxford: Oxford University Press.

Bryan, G., Karlan, D., & Nelson, S. (2010). Commitment Devices. *Annual Review of Economics*, *2*(1), 671–698. doi: 10.1146/annurev.economics.102308.124324

Bursztyn, L., & Jensen, R. (2017, August). Social Image and Economic Behavior in the Field: Identifying, Understanding, and Shaping Social Pressure. *Annual Review of Economics*, *9*, 131–153. doi: 10.1146/annurev-economics-063016-103625

Bush, A. J., & Hair, J. F. (1985). An Assessment of the Mall Intercept as a Data Collection Method. *Journal of Marketing Research*, *22*(2), 158–167. doi: 10.2307/3151361

Charness, G., & Dufwenberg, M. (2006, November). Promises and Partnership. *Econometrica*, *74*(6), 1579–1601. doi: 10.1111/j.1468-0262.2006.00719.x

Cialdini, R., & Trost, M. (1998). Social influence: Social norms, conformity and compliance. *The handbook of social psychology*, *D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.)*, 151–192.

Cioffi, D., & Garner, R. (1996, February). On Doing the Decision: Effects of Active versus Passive Choice on Commitment and Self-Perception. *Personality and Social Psychology Bulletin*, *22*(2), 133–147. doi: 10.1177/0146167296222003

DellaVigna, S., List, J. A., & Malmendier, U. (2012, February). Testing for Altruism and Social Pressure in Charitable Giving. *The Quarterly Journal of Economics*, *127*(1), 1–56. doi: 10.1093/qje/qjr050

DellaVigna, S., List, J. A., Malmendier, U., & Rao, G. (2017, January). Voting to Tell Others. *The Review of Economic Studies*, *84*(1), 143–181. doi: 10.1093/restud/rdw056

Dreber, A., Fudenberg, D., Levine, D., & Rand, D. (2016, February). *Self-Control, Social Preferences and the Effect of Delayed Payments* (SSRN Scholarly Paper No. ID 2477454). Rochester, NY: Social Science Research Network.

Dufwenberg, M., & Gneezy, U. (2000, February). Measuring Beliefs in an Experimental Lost Wallet Game. *Games and Economic Behavior*, *30*(2), 163–182. doi: 10.1006/game.1999.0715

Ellingsen, T., & Johannesson, M. (2004, April). Promises, Threats and Fairness*. *The Economic Journal*, *114*(495), 397–420. doi: 10.1111/j.1468-0297.2004.00214.x

Exley, C. L. (2018). *Using charity performance metrics as an excuse not to give* (Tech. Rep.). Working Paper.

Exley, C. L., & Naecker, J. K. (2016, July). Observability Increases the Demand for Commitment Devices. *Management Science*, *63*(10), 3262–3267. doi: 10.1287/mnsc.2016.2501

Funk, P. (2010, September). Social Incentives and Voter Turnout: Evidence from the Swiss Mail Ballot System. *Journal of the European Economic Association*, *8*(5), 1077–1103. doi: 10.1111/j.1542-4774.2010.tb00548.x

Galizzi, M. M., & Wiesen, D. (2018, March). Behavioral Experiments in Health Economics. In *Oxford Research Encyclopedia of Economics and Finance.* Oxford University Press. Retrieved 2019-01-11, from `http://economics.oxfordre.com/view/10.1093/acrefore/9780190625979.001.0001/acrefore-9780190625979-e-244` doi: 10.1093/acrefore/9780190625979.013.244

Gerber, A. S., Green, D. P., & Larimer, C. W. (2008, February). Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment. *American Political Science Review*, *102*(1), 33–48. Retrieved 2018-07-08, from `https://www.cambridge.org/core/journals/american-political-science-review/article/social-pressure-and-voter-turnout-evidence-from-a-largescale-field-experiment/11E84AF4C0B7FBD1D20C855972C2C3EB` doi: 10.1017/S000305540808009X

Gneezy, U., Keenan, E. A., & Gneezy, A. (2014). Avoiding overhead aversion in charity. *Science*, *346*(6209), 632–635.

Goette, L., Stutzer, A., Yavuzcan, G., & Frey, B. M. (2009, March). Free cholesterol testing as a motivation device in blood donations: evidence from field experiments. *Transfusion*, *49*(3), 524–531. doi: 10.1111/j.1537-2995.2008.02007.x

Greinacher, A., Fendrich, K., Brzenska, R., Kiefel, V., & Hoffmann, W. (2011, April). Implications of demographics on future blood supply: a population-based cross-sectional study. *Transfusion*, *51*(4), 702–709. doi: 10.1111/j.1537-2995.2010.02882.x

Karing, A. (2018). Social signaling and childhood immunization: A field experiment in sierra leone. *University of California, Berkeley*.

Lacetera, N., Macis, M., & Mele, A. (2016, April). Viral Altruism? Charitable Giving and Social Contagion in Online Networks. *Sociological Science*, *3*, 202–238. doi: 10.15195/v3.a11

Meyer, C., & Tripodi, E. (2017, October). *Sorting Into Incentives for Prosocial Behavior* (SSRN Working Paper No. 3058195). Rochester, NY: Social Science Research Network. Retrieved 2017-11-14, from https://papers.ssrn.com/abstract=3058195

Oberverwaltungsgericht Rheinland-Pfalz. (2013, December). *Urteil DRK-Blutspendedienst West gGmbH v. Land Rheinland-Pfalz.* Koblenz: Oberverwaltungsgericht Rheinland-Pfalz. Retrieved from http://www3.mjv.rlp.de/rechtspr/DisplayUrteil.asp?rowguid=%7B5F1EE8CA-8A24-476B-A999-864539234DEB%7D

Ostrom, E., Walker, J., & Gardner, R. (1992, June). Covenants with and without a Sword: Self-Governance Is Possible. *American Political Science Review*, *86*(2), 404–417. doi: 10.2307/1964229

Paul-Ehrlich-Institut. (2018). *Tabellarischer Bericht für das Jahr 2017 nach Paragraph 21 Transfusionsgesetz (TFG).* Langen, Germany: Paul-Ehrlich-Institut. Retrieved from https://www.pei.de/DE/infos/meldepflichtige/meldung-blutprodukte-21-transfusionsgesetz/berichte/berichte-21tfg-node.html

Saito, K. (2015, July). Impure altruism and impure selfishness. *Journal of Economic Theory*, *158*(Part A), 336–370. doi: 10.1016/j.jet.2015.05.003

Stutzer, A., Goette, L., & Zehnder, M. (2011, November). Active Decisions and Prosocial Behaviour: a Field Experiment on Blood Donation. *The Economic Journal*, *121*(556), F476–F493. doi: 10.1111/j.1468-0297.2011.02477.x

Vanberg, C. (2008, November). Why Do People Keep Their Promises? An Experimental Test of Two Explanations1. *Econometrica*, *76*(6), 1467–1480. doi: 10.3982/ECTA7673

World Health Organization. (2017). *Global Status Report on Blood Safety and Availability 2016*. Geneva: World Health Organization.

Yoeli, E., Hoffman, M., Rand, D. G., & Nowak, M. A. (2013, June). Powering up with indirect reciprocity in a large-scale field experiment. *Proceedings of the National Academy of Sciences*, *110*(Supplement 2), 10424–10429. Retrieved from http://www.pnas.org/content/110/Supplement_2/10424 doi: 10.1073/pnas.1301210110

Index