The Citizen and the State:
Evidence from Field Experiments

Dissertation

zur Erlangung des wirtschafts- und sozialwissenschaftlichen Doktorgrades
„Doctor rerum politicarum“
der Ruprecht-Karls-Universität Heidelberg

UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

vorgelegt von

Zain Chaudhry

July 2023

The Citizen and the State:
Evidence from Field Experiments

Zain Chaudhry

FIRST SUPERVISOR AND EXAMINER OF THE DISSERTATION

Prof. Dr. Axel Dreher

Professor of Economics

Chair of International and Development Politics

Alfred-Weber-Institute for Economics

Ruprecht-Karls-Universität Heidelberg

SECOND SUPERVISOR AND EXAMINER OF THE DISSERTATION

PD Dr. Alexandra Avdeenko

Chair of International and Development Politics

Alfred-Weber-Institute for Economics

Ruprecht-Karls-Universität Heidelberg

THIRD EXAMINER OF THE DISSERTATION

Prof. Dr. Reimut Zohlnhöfer

Professor of Political Science

Chair of Political Science, Comparative Analysis of Political Systems

Institute of Political Science

Ruprecht-Karls-Universität Heidelberg

For Musab

Contents

List of Figures

List of Tables

Introduction

A systematic lack of state capacity is one of the biggest challenges faced by developing countries. This ineffectiveness has both historical and contemporary causes, and while history may substantially affect current outcomes, it does not exclusively determine them, allowing us to understand and mold current economic and social outcomes. One reason for low state capacity is that citizen-state interaction can be stuck in a bad cycle of low trust and low capacity. Such a stark citizen-state trust deficit can lead to poverty and instability (Migdal 1988; Besley and Persson 2009). But when this positive relationship exists and citizens trust the state, it plays an important positive role in the development and functioning of state institutions (Banfield Edward 1958; Besley and Persson 2009) by easing the cooperation of citizens with each other and with the state (Evans 1995).

This necessitates an understanding of how the state functions and how policymakers can improve state functioning. This would strengthen citizens' trust in the state. All the roles that a modern state performs from the provision of security, education and health matter to citizen-state interaction. They are all implemented by agents whose performance affects citizens' relationship with the state.

To understand what policies improve public-service delivery and improve citizens' interaction with the state, we need careful evaluation of policy instruments to help us understand the relationship between policy and outcome. This contributes to good public governance by helping governments improve public policies that can lead to higher quality of public service provision and thus higher economic development. Such policy evaluation helps citizens hold policymakers accountable and iterate over policies and improve them over time.

In light of this, this thesis aims to illuminate different aspects of state capacity and citizen-state interactions. How can we help improve state functioning? How can the state improve cooperation among citizens and with itself? In all chapters, I collect primary data in Pakistan which allows me to study phenomena that are generally hard to study in developing

countries without field experiments because administrative data are either of low quality or entirely unavailable.

My thesis is structured as follows.

In Chapter 1, I focus on understanding intertemporal preferences of individuals and groups and how individual characteristics drive group behavior. In Chapter 2, I focus on the effect of a political campaign that targets women on women's turnout and the vote share of the campaigning politician. In Chapter 3, I analyze the effect of a state-supported campaign to reduce discrimination and improve economic interactions.

In the next paragraphs, I highlight the contributions and connections of the individual chapters in more detail.

In Chapter 1, together with together with Karrar Hussain, I analyze the role played by intertemporal preferences of individuals and groups. I run a field experiment with university students that can eventually help us understand the behavior of public sector workers. Many organizations, in public and private sectors, require team work. In developing country public sectors, police patrolling teams, vaccination teams, and many other types of teams undertake group work. In order to understand how to design incentives at the group level, we need to understand what time preferences groups exhibit. In this chapter, my goal is to understand whether groups are present-biased and how individuals within a group affect group-level intertemporal choices. This work would eventually allow us to design commitment devices or tailored contracts to increase the efficiency of public-sector employees. As police officials and vaccinators are front-line public service providers, their improved performance can lead to higher trust in the state.

To this end, we conducted a field experiment with university students in Pakistan. On the first day, we asked them to divide effort between two days, exactly one and two weeks in the future at different "task rates" (essentially changing the cost of delaying work). After one week, when the first part of the task had to be undertaken, we elicited this effort allocation again —for that day and the week after. We elicited these preferences separately from all individuals, then created random groups of two individuals each, and asked each group to make the same allocations. Using this data, we conducted reduced-form and structural analysis and find that groups are more present-biased than individuals. Connecting group behavior with individual members' behavior, we find that the individuals with higher present bias drive group decisions. Further, groups exhibit greater present bias when differences in discount

rates within the group are larger as theorized. Finally, present bias in the group decisions is reduced when one group member has very high bargaining power (measured through a multinomial logit model regression of both individuals' allocations on group allocation). We have also recently finished collecting data using the same design from female public-sector workers, who play an important role in healthcare provision in many developing countries and often do certain tasks in teams.

In Chapter 2 of the thesis, together with Karrar Hussain and Attique-ur-Rehman, I turn to citizens' relationship with the state in terms of political interaction. We studied how representatives of the state could mobilize the female vote —a typically marginalized group. Our goal was to understand how representatives of the state —legislators —could increase women's interaction with the state. We evaluated a political campaign run by an incumbent legislator in which he informed women through a door-to-door campaign of the public service delivery work undertaken during his tenure. With our help, the legislator randomized the campaign at the precinct level, allowing us to use official electoral data on vote shares and gender-disaggregated turnout. In the first treatment arm, the campaign gave information only to female voters, and in the second to both women and men. Our results suggest that in a highly competitive campaign, the vote share of the campaigning incumbent increased by five percentage points. This increase was primarily driven by women who were contacted independently of their male relatives and is statistically insignificant in precincts where both men and women were mobilized.

In Chapter 3, I try to understand whether local governments can utilize their near-monopoly over religious institutions such as mosques to increase contact between worshipers of different Muslim sects in Pakistan because creating a society with high cooperation and lower conflict potential is an important goal for any state. Does this contact inside mosques reduce prejudice and increase economic efficiency? We randomized thirty-two mosques: in the first treatment group, we sent volunteer worshipers of the opposite sect to pray in each other's mosques, while in the second treatment group, we had the prayer lead make a statement about inter-faith harmony. We test whether worshipers update their beliefs about the other sect regarding religiosity due to this contact using surveys as well as an incentivized lab-in-the-field experiment. We find that the combined treatment reduces prejudice in an incentivized, real-world task where we offer discounted plumbing services from plumbers of both sects.

Data Contribution and Methodological Approach

In my dissertation, I collect novel datasets on individual behavior. My research relies on medium to large-scale field experiments in which I collect extensive survey data and conduct behavioral, lab-in-the-field experiments. A common strand of my research is a focus on gender, either solely or primarily, to understand women's beliefs and preferences.

All three papers are field experiments in a developing country. While no method is perfect, field experiments alleviate what I believe are two substantial constraints compared to secondary data analysis. First, field experiments allow us to collect very careful data of a nature that directly answers our questions of interest. This expands the horizons of what questions we can answer because these richer aspects of behavior and information sets than would otherwise be unavailable in administrative data. This is particularly challenging in developing countries where less administrative data are available and when it is available it is of lower quality. Second, field experiments allow for cleaner identification. Even with massive advances in instrumental variable, difference-in-differences, and other techniques for natural experiments - debates about assumptions being unmet often remain, which complicates policy adoption. From a political economy and policy adoption perspective, field experiments are thus much easier to understand and leave less room for political debates and manoeuver.

In Chapter 1, I collect data on time preferences of individuals and groups which has hitherto not been done simultaneously. This allows us to understand how individual behavior drives group behavior. In Chapter 2, I carefully collect data on polling station boundaries and match it to digitized electoral voting records. These data include information on the total number of votes for each candidate but also gender-disaggregated voting totals. In Chapter 3, I collect data on sectarian discrimination in Pakistan including beliefs, preferences and behaviors of each sect's members. This is a novel dataset that is mostly survey-based but includes two important incentivized real-world tasks.

Summary of the Chapters

Chapter 1

Although many important economic decisions are taken by groups considering costs and benefits over time, we have little empirical evidence on groups' intertemporal behavior. We

conduct a field experiment with university students and use reduced-form and structural methods to study how within-group differences drive group-level behavior. We measure time preferences using effort-based tasks in the field, create groups randomly, and collect data on the time preferences of both groups and their constituent individuals. We find that groups are more present-biased than individuals, and that individuals with higher present bias drive decisions within groups. Further, we find that groups exhibit greater present bias when the difference in discount rates within the group is larger. Finally, we document that present bias in group decisions is reduced when bargaining power in the group is less symmetric, making the group act more like an individual.

Our contribution to this literature is threefold. First, we measure time inconsistency based on the intertemporal allocation of effort (negative consumption), which is a better method of eliciting time preferences than monetary methods (Sprenger 2015; Cohen et al. 2020; Augenblick, Niederle, and Sprenger 2015). Second, we study exogenously formed groups because endogenously formed groups, such as couples, may self-select on many characteristics such as income level and personality traits. Third, we measure time preferences for individuals and groups both, which allows us to understand how individual behavior drives group behavior.

The findings can be used to create personalized contracts based on each individual's time preferences, which can be applied to any team, including public-sector teams.

Chapter 2

I provide the first estimate of a door-to-door political campaign by an incumbent politician targeting women on electoral outcomes in a developing country. Women voters are informed of the public service delivery work undertaken by the incumbent in his tenure. The campaign was randomized at the precinct level, allowing us to use official electoral data on vote shares and gender-disaggregated turnout. Our results suggest that in a highly competitive campaign, the vote share of the campaigning incumbent increased by 5 percentage points. This increase was primarily driven by women who were campaigned independently of their male relatives. In precincts where both men and women were mobilized, the effect is not statistically significant. However, women's turnout in the election was unaffected.

With this chapter, I provide the first rigorous evaluation of a women-focused door-to-door campaign on turnout and vote shares. We run the campaign in a developing country, which is relevant because of strong norms against female empowerment and public participation. Our

treatment is a standard partisan political campaign, thus the campaign can easily be held for women as well. Hence, the amount of effort required from politicians would be minimal. If such a campaign shows positive results for the campaigner, it sends a credible signal about the importance of women as a distinct voting bloc that politicians should court. Finally, as we randomize our treatment at the precinct level, we can use official administrative data on female turnout and politician vote shares, thus measuring actual electoral outcomes and avoiding problems associated with self-reported voting behavior (Campbell 2010; Gelman et al. 2016).

The findings can be used to convince the state and its representatives that women's involvement in politics can be easily improved and can yield positive political returns.

Chapter 3

I conduct a field experiment to analyze the effect of contact and leadership in reducing prejudice and fostering economic cooperation between discordant sectarian groups in Pakistan. We randomly assign 32 mosques with 428 regular worshipers to four groups in which (i) we send volunteer worshipers to mosques of the opposite sect to pray (different sects visibly pray differently), or (ii) have the head of the mosque make a religious announcement in the mosque in support of inter-sectarian harmony, or (iii) a combined treatment with both volunteer visits and announcements, or (iv) a pure control group. We find that the combined treatment reduces prejudice in two incentivized, real-world tasks: treated respondents choose discounted plumbing services from plumbers as well as discounted books of/about the opposite sect. Further, we find that worshipers treated with both treatments have higher trust in the opposite sect and would make higher charitable donations to mosques of the opposite sect.

With this chapter, I am the first to systematically test for separate and combined effects of collaborative contact and supportive of authorities (Paluck, S. Green, and D. Green 2019), showing both the importance of collaborative contact and support from authority figures. We also explore inter-sectarian differences with substantial cross-country geo-political considerations, which have never been explored before. This is an important element not only due to its economic and political importance but also because while intra-religious differences are in many ways similar to issues of race and caste, they nevertheless provide a commonality between the two groups. Finally, a lot of the previous work does not conduct baseline data

collection, which is needed to understand how the contact created by the researchers interacts on top of daily societal interaction.

1 — Fickle Groups: A Field Experiment on Time Preferences

This paper is joint work with Karrar Hussain. It has not yet been published.

Abstract: Although many important economic decisions are taken by groups (e.g. households and committees) considering costs and benefits over time, we have little empirical evidence on groups' intertemporal behavior. We conduct a field experiment with university students and health workers and use reduced-form and structural methods to study how within-group differences drive group-level behavior. We measure time preferences using effort-based tasks in the field, create groups randomly, and collect data on the time preferences of both groups and their constituent individuals. We find that groups are more present-biased than individuals and that individuals with higher present bias drive decisions within groups. Further, we find that groups exhibit greater present bias when the difference in discount rates within the group is larger. Finally, we document that present bias in group decisions is reduced when bargaining power in the group is less symmetric, making the group act more like an individual.

1.1 Introduction

Intertemporal choice - decisions for which costs and benefits are spread out over time - are an important part of day-to-day decision-making for households, policymakers, and managers. These intertemporal choices are relevant to important areas of economic life, such as consumption, saving, and investment. Many important real-life outcomes are strongly associated with individuals' time preferences, including health status (Chabris et al. 2008), educational attainment (Cadena and Keys 2015; Castillo et al. 2011), savings (Beshears et al. 2008; Laibson 1997), physical exercise (DellaVigna and Malmendier 2006) and labor-market earnings (Golsteyn, Grönqvist, and Lindahl 2014). Due to these far-reaching effects, intertemporal choices lead not only to divergent individual-level outcomes but also differences at the macroeconomic level (Rae 1905; Sunde et al. 2022), as consumption and savings are important determinants of economic growth (Mankiw, Romer, and Weil 1992). Therefore, understanding intertemporal choice, using theory and data has been a major focus of economists (Andreoni and Sprenger 2012a; Fudenberg and Levine 2006; Gul and Pesendorfer 2001; Koopmans 1960; Laibson 1997; Noor 2011; O'Donoghue and Rabin 2001; Samuelson 1937).

A rich theoretical and empirical literature has substantially advanced our understanding of intertemporal decision-making since the early work of Samuelson (1937).[1] However, most of this literature focuses on individual decision-making, even though many dimensions of intertemporal choice are better modeled at the group level. For instance, partners within households typically make education, health, and savings decisions together, while finance committees within firms and legislatures allocate budgets over time. These groups often have heterogeneous time preferences, which can lead to tension in important collective decision-making.[2] That preference heterogeneity exists within most groups is evident, for example, from the fact that women and men have different life expectancies and that age

[1] A few seminal examples from the empirical literature, which mostly use structural models to estimate the level and shape of discounting, are Hausman (1979), Lawrance (1991), Warner and Pleeter (2001), Laibson, Repetto, and Tobacman (2007), Harrison, Lau, and Williams (2002), Andersen et al. (2008), Andreoni and Sprenger (2012a), and Andreoni and Sprenger (2012b). For a comprehensive review of the literature, see Ericson and Laibson (2019) and Frederick, Loewenstein, and O'donoghue (2002).

[2] Even in the context of individual choice, one can consider the existence of multiple selves with distinct personalities rather than a single homogeneous decision-making unit. Thaler and Shefrin (1981) contrast the longsighted "planner" within us to the shortsighted "doer," while Metcalfe and Mischel (1999) contrast our "hot" and "cool" systems. More recent work also models multiple selves with competing sets of interests, such as Fudenberg and Levine (2006, 2011), Brocas and Carrillo (2008a,b), and Noor and Takeoka (2022). Such evidence supports the application of collective-choice models to characterize the behavior of individuals.

gaps in partnerships exist, which means that partners have different horizons. Similarly, other decision-making groups, such as committees in firms, exhibit substantial differences in gender, age, and cognitive ability. All these factors have been shown to be determinants of time preferences (Bortolotti et al. 2021; Dohmen et al. 2010; Frederick 2005), which suggests that such groups would have within-group differences in discount rates.

This paper goes beyond the assumption that groups act as a single representative agent and asks: are randomly formed groups present biased, and how do their constituent members' time preferences determine their group time preferences?

To answer this question, we conduct a field experiment with 244 university students in Pakistan to measure individual and group-level time preferences in an effort-allocation task over three weeks. We ask participants to allocate effort to the task of taking photos of a book using an app developed by the research team. The task was conducted by individuals alone or as a group, where both group formation and the determination of individual or group work were random. On the first day of the experiment (Day 1), each individual made the following decisions: allocation of effort for precisely one week (Day 8) and two weeks (Day 15) later. Each decision was made for three different task rates: $R \in \{0.8, 1, 1.2\}$. A task rate of 1:0.8, for example, would mean that every task the participant allocated to the present reduced the number of tasks allocated to the future by 0.8. The participants were asked to make the same choice a week later (Day 8) before they attempted the task for that day. The same decisions were elicited from randomly formed groups as well. Hence, we elicited the same effort-allocation decisions from every individual and group. One of the eighteen decisions made by each participant was chosen to be implemented based on a rule (explained in Section 3.3) about which all participants were informed.

Theoretically, time inconsistency in groups can arise simply from the aggregation of heterogeneous preferences even when individuals alone are time consistent. Such inconsistency may occur because of variations in individual discount rates and innovations in the Pareto weight summarizing the collective decision-making process (Feldstein 1964; Gollier and Zeckhauser 2005; Jackson and Yariv 2015; Marglin 1963). Further, for a uniform distribution of discount rates in an otherwise-homogeneous population, maximizing group utility in a nondictatorial way generates aggregate behavior that corresponds to hyperbolic discounting (Jackson and Yariv 2015). As a result, all else equal, it is optimal to favor inpatient group members in early periods and patient members in later periods.

We analyze our data using reduced-form and structural estimation methods and document three main results.

First, using a two-limit Tobit regression, we find that individuals making immediate, same-day choices allocate around 12% fewer tasks to the earlier task day than those making the same decision a week before the first task day. The corresponding figure for groups is 21%. The results suggest that the degree of time inconsistency is lower for individuals compared to groups, which is in line with previous research.

Second, we estimate a structural model. The structural estimate is close to our reduced-form result, which shows that the theoretical model under consideration is a good fit for the experimental data and corroborates the finding that groups are time inconsistent. Further, we compare individuals' and groups' decisions. While the estimate for the present-bias parameter shows the existence of present bias in effort choice for both individuals and groups, the present-bias estimate is lower for individuals compared to groups. Further, individuals' weekly discount-parameter estimate is less than the groups' estimate. Hence, individuals and groups exhibit different present biases and discount factors.

Third, to better understand the connection between individual and group time preferences, we regress group-decision-estimated time-preference parameters on individual decisions in the group. Because we collect data at the individual and group levels, we can use these within-group estimates of present bias. These results show that the weekly-discount-factor heterogeneity of individuals explains group present bias. We find that group present bias is mostly driven by the individual with greater present bias and that variance in the group members' discount rates and bargaining power explains group present bias. These results are in line with Jackson and Yariv's (2014) theoretical prediction that for a uniform distribution of discount rates in an otherwise-homogeneous population, group utility maximization generates aggregate behavior that corresponds to hyperbolic discounting. If there is some fundamental heterogeneity in temporal preferences in the form of differing discount factors, then the only well-behaved collective utility functions that are both time-consistent and respect unanimity are dictatorial.

Our results are important because we present both non-parametric and parametric characterizations of individual and collective intertemporal choice for the same set of participants

under experimentally controlled environments based upon intertemporal allocations of effort.[3] We begin with an approach free of functional form using experimentally induced exogenous variation, then move to the theory-based parametric analysis of time inconsistency on the individual and group levels. By adopting this approach, our subsequent parametric estimates thus result from restrictive parametric assumptions rather than from a failure of the underlying theoretical framework, which is free of functional form and related to an assessment of the degree of differences between these two kinds of decision environments. In the structural part of our empirical analysis, the preference structure associated with the discounted-utility approach is applied (without modification) to model group behavior. This is in line with the representative-agent modeling structure mostly used in macroeconomics. This unitary approach assumes the collective acts as a single decision-making unit and therefore can be treated as a rational individual.

While a few important papers have empirically tested group-level intertemporal choice (Glätzle-Rützler, Lergetporer, and Sutter 2021; Mazzocco 2007; Schaner 2015), these papers use monetary-choice methods for measurement, study endogenously formed groups (and thus have no exogenous variation in intertemporal preferences at the group level) such as spouses, or elicit intertemporal preferences at the individual or group level but not both.

Our contribution to this literature is threefold. First, we explore intertemporal choice with a better measurement: we measure time inconsistency based on the intertemporal allocation of effort (negative consumption), which is a better method of eliciting time preference (Augenblick, Niederle, and Sprenger 2015; Cohen et al. 2020; Sprenger 2015). Second, we study exogenously formed groups because endogenously formed groups, such as couples, may self-select on time preference, risk preference, income level, and other personality traits. Further, as researchers usually collect data long after couples form, learning effects over time may have caused the individuals' preferences to become more aligned. If couples match assortitatively on the marriage market, there may also be no real differences in time preferences: the data may only show differences because of measurement error correlated with cognitive ability and financial literacy (Schaner 2015). Third, we do not just measure time preferences for individuals or groups but both individuals and groups, which allows us to understand how individual behavior drives group behavior.

[3]This occurs in the consumption-choice rather than monetary-choice domain. The monetary methods typically confront several confounding factors in identifying and estimating the shape of time preferences, which we will explain later in this section.

Finally, we use consumption-based measures of intertemporal choice because the assumptions necessary for using time-dated monetary payments to measure intertemporal choice are rarely satisfied (Cohen et al. 2020; Sprenger 2015). For example, in violation of usual assumptions, participants may think of external financial decisions (that is, arbitrage opportunities outside the experiment) (Chabris et al. 2008; Cubitt and Read 2007), they may think of their external consumption choices, or they might not trust the research team enough to neglect future transaction costs and assume payment reliability.[4] Andersen et al. (2008), Andreoni and Sprenger (2012a), and Giné, Goldberg, et al. (2018) document that when closely controlling for transaction costs and payment reliability, dynamic inconsistency in choices over monetary payments is virtually eliminated in the aggregate. All these challenges can create spurious dynamic inconsistencies, as suggested by the fact that this literature has elicited an extremely wide variety of discount rates, ranging from less than 1% (Thaler 1981) to more than 1,000% (Holcomb and Nelson 1992).

The paper proceeds as follows. Section 3.3 presents the details of the experimental design. Section 1.3 provides an overview of the data we collected. Section 1.4 describes the reduced-form regression analysis, including two-limit-Tobit and nonlinear-least-squares estimation, and Section 1.5 presents the structural-estimation results. Section 1.6 explores how individual-level preferences drive group-level preferences. Section 3.5 concludes.

1.2 Experimental Design

To understand dynamic inconsistency in the allocation of effort for individuals and groups, we conducted an experiment with 244 undergraduate students from different majors at Lahore University of Management Sciences (LUMS) over three weeks. LUMS is one of Pakistan's most prestigious universities and attracts the brightest from across the country (and because it boasts large, targeted programs for low-income families, it is not restricted to students from high-income families).

The research team asked participants to allocate effort to one kind of task in the real world with monetary incentives as an individual and as part of a (randomly chosen) group of two. On the first day of the experiment, each individual as well as each group separately

[4]The main idea was originally raised by Thaler (1981), who, when considering the possibility of using incentivized monetary payments in intertemporal choice experiments, noted, "Real money experiments would be interesting but seem to present enormous tactical problems. (Would subjects believe they would get paid in five years?)."

made the following decisions: allocation of effort on Day 8 (exactly one week later) and allocation of effort on Day 15 (exactly two weeks later). The same choices were made one week later, on Day 8, but before the participants were supposed to perform the task. Each decision was made with three different task rates. One of the eighteen allocation decisions was randomly chosen for each participant based on a rule explained below, and they were given a monetary reward of $15 if the work was completed and $0 otherwise.

Timeline: On the first day of the experiment, Day 1, we gave participants detailed instructions on how the experiment would work. We told them that they would make decisions about effort allocation that day (Day 1) to undertake the task exactly one (Day 8) and two weeks (Day 15) later and make the same decisions again exactly one week later (but before the task was set to be performed). They would make these decisions, as individuals and as groups, in sessions with the research team present. Whether a participant first made a decision as an individual or a group was randomized to avoid potential ordering effects.

During all the training sessions, the research team ensured that participants within each group exchanged their email addresses and phone numbers.

Effort Allocations: To further motivate the intertemporal trade-off, an additional factor in the decision-making was the task rate. The decisions were made using the Convex Time Budget methodology proposed by Andreoni and Sprenger (2012a). The allocations were made in a mobile application with slider bars, where every slider bar corresponded to a specific task rate, to make it easy to visualize the decision (see Figure 1.1). We offered three task rates, $R \in \{0.8, 1, 1.2\}$, and a decision had to be made for each one. A task rate of 1:0.8 means that every task the participant allocated to the present (v_1) reduced the number of tasks allocated to the future (v_2) by 0.8. For ease of understanding, the task rates were always represented as $1 : R$, and the participants were fully informed of the value of $R \in \{0.8, 1, 1.2\}$ when making their decisions. Output exceeding v_1 targets on the first day was not transferable to v_2 . The participant's decision can be formulated as allocating tasks $v_1, v_2 = fR, V$ over time, subject to the present-value budget constraint. v_1 and v_2 satisfy the intertemporal budget constraint:

$$v_1 + R \cdot v_2 = V.$$

In Figure 1.1, we show an image of the (translated) main page of the application (the original Urdu version is shown in Appendix Figure 1.6).

To avoid corner solutions in allocation decisions, when an individual or a group decided to allocate all their tasks either to Day 8 or Day 15, the application automatically restricted the minimum number of pictures for the Day 8 task to twelve. With this limit, we observed corner-solution allocation decisions extremely rarely: only in 2.50% of the total 2,196 allocation decisions was $v_1 = 12$. Toward the end of the Day 8 session, all participants were informed which allocation had been (randomly) selected for them out of the eighteen total decisions that they recorded during the experiment. We explained to them at the outset how this decision would be selected. We called the selected decision the "decision that counts." On Day 8, two hours after the allocation-decision session, participants were asked to complete their "decision that counts" allocations in the specified period (21:30 to 22:30 Pakistan Standard Time).

Such intertemporal bonus contracts can be used to investigate intertemporal preferences. The allocations participants make, v_1, v_2, convey information on their discount rates. Additional experimental variation permits us to identify an important behavioral aspect of intertemporal choice: the existence of present-biased preferences.

When participants made decisions on Day 8, we did not remind them of the Day 1 allocations. Importantly, on Day 1, participants were making decisions involving two future work dates (one and two weeks later), whereas, on Day 8, they were making decisions for the same day and the week after. On Day 1, before any decisions were made, participants were told how allocations for the two future dates would function and that only one of the allocation decisions made by them would eventually be chosen for them to implement.

Tasks: The task was to take clear, legible pictures of any book of the participants' choice in one specific hour (21:30 to 22:30 Pakistan Standard Time), using the mobile application provided by the research team, and to upload the pictures to a server (all pictures were automatically geotagged and timestamped). To avoid sample-selection issues, we provided mobile phones and internet packages to all participants and taught them how to use the application and upload the data. The evening time was chosen to ensure that nobody had classes, family obligations, or religious obligations. This equalized outside options that could otherwise have contaminated the purity of the intertemporal choices. A complete practice run was conducted to ensure that everyone understood how the application worked. The target for the task was two hundred pages ($V = 200$) in the individual setting and four hundred

pages ($V = 400$) in the group setting. If the page in a picture were not legible, the picture would not be counted as a completed task.

Recruitment, Selection, and Attrition: Two hundred forty-four students at LUMS took part in the experiment. The participants did not receive a show-up fee but only a completion fee of \$15 if all tasks were accomplished according to their selected allocation.

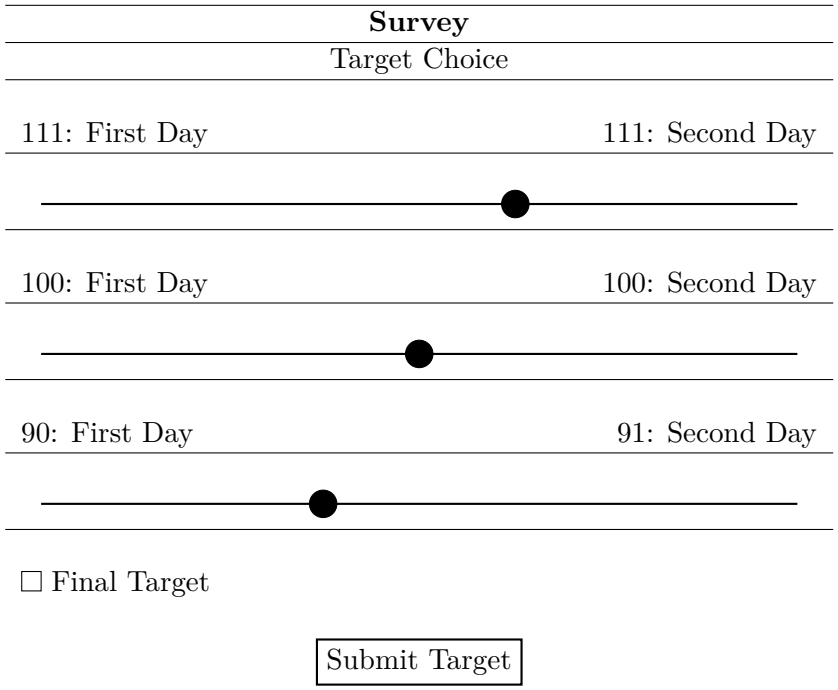


Figure 1.1: Slider Bar for Task Allocations

Notes: The above slider bars, a translation of the original app in Urud, show the individual task allocation decisions over the two weeks, i.e., $V = 200$. The blue letters in translate (literally) to "set target". The next lines (from left to right) translate to "First day: 111; Second day: 111" for slider bar 0.8, "First day: 100; Second day: 100" for slider bar 1, and "First day: 90; Second day: 90" for slider bar 1.2. The text next to the box translates to "finalize target" and the black letters on the bar translate to "set target".

The Allocation That Counts: Participants were informed that three factors determined which allocation out of the eighteen they made would randomly be chosen for them to implement. First, the allocated task could be either from a Day 1 or a Day 8 decision according to 20% and 80% probabilities, respectively. Second, the allocated task could either be an individual or a group task according to 33% and 66% probabilities, respectively. Third, the allocated task could be from one of the three task rates, $R \in \{0.8, 1, 1.2\}$, which were all equally likely. This randomization process, which follows Augenblick, Niederle, and Sprenger (2015), ensured the incentive-compatibility constraint was satisfied for all decisions. The design choice to ensure that participants had a 20% chance of receiving a preference schedule

of the v_1 and v_2 targets from Day 1 was made to allow and intimate to them that they potentially exhibited present-bias.

Monetary Reward: The monetary reward was based on individual or group performance depending on which allocation decision was randomly chosen for each individual. The reward was \$15 for completing the task in line with the (randomly chosen) allocation decision. Any deviation, even of one page, meant the task would be deemed incomplete and carry no reward. In the group task, the reward was a total of \$30, equally split between the two group members irrespective of their contribution to the task. No instructions were given about how to divide the work for the group task.

1.3 Data and Measurement

During the first session (Day 1), we collected data on a variety of participant characteristics. We collected demographic and educational information such as age, ethnicity, major, part-time employment status, family income, whether they had a savings account, used study plans, and knew their paired partner. Following Callen et al. (2015), we collected data on personality using a modified Big Five survey (Barrick and Mount 1991; John, Naumann, and Soto 2008; Van der Linden, Nijenhuis, and Bakker 2010)[5] and survey-based trust in strangers. The Big Five survey consisted of sixty questions and was developed in Urdu and validated for use in Pakistan by the National Institute of Psychology at Quaid-i-Azam University, Islamabad, Pakistan.

We describe our sample of participants in Table 1.1. The mean participant age was 20.3 years, and 39% of the sample were women. Sixty-two percent had no access to a formal savings account at the time of the experiment, but 73 percent had access sometime in the past. This is relevant because the behavioral-economics literature has used savings accounts to predict the degree of patience or present bias. As explained above, individuals were randomly paired together. The mean of the group-mate acquaintance index indicates that individual members knew each other at the start of the study (the index ranges from zero to five years of acquaintance). The mean time duration of acquaintance is around thirteen months. In the structural analysis, we explicitly control for the duration of acquaintance. In this experiment,

[5]The Big Five personality traits, according to the Five Factor Model, are five dimensions of human personality that were designed to be descriptive and non-overlapping. These traits are agreeableness, emotional stability, extroversion, conscientiousness, and openness.

Table 1.1: Summary Statistics

	# of Obs	Mean	Standard Deviation	Minimum	Maximum
Demographic					
Age	244	20.32	1.75	17	27
Male	244	0.61	0.48	0	1
No on-campus job	244	0.91	0.28	0	1
Had No savings account	244	0.73	0.44	0	1
Has No savings account	244	0.62	0.48	0	1
Group-mate acquaintance index	244	3.04	1.28	1	5
Acquaintance time duration (months)	244	13.84	21.56	0	60
Big Five Survey					
Openness	212	3.30	0.45	2.17	4.42
Conscientiousness	212	3.43	0.52	1.75	4.92
Extroversion	212	3.25	0.33	2.16	4.44
Agreeableness	212	3.44	0.46	2.33	4.58
Neuroticism	212	2.82	0.57	1.25	4.67

Notes: This table reports summary statistics for our respondent population. We have a full sample of 244 students, though some students were not able to fill out the Big Five survey. We use a 60-question Big Five survey developed in Urdu and validated for use in Pakistan by the National Institute of Psychology at Quaid-i-Azam University, Islamabad, Pakistan. Its variables were recorded on a 1-5 Likert scale.

this variable tries to capture the effect of group dynamics such as coordination externalities at the stage of intertemporal task allocation.

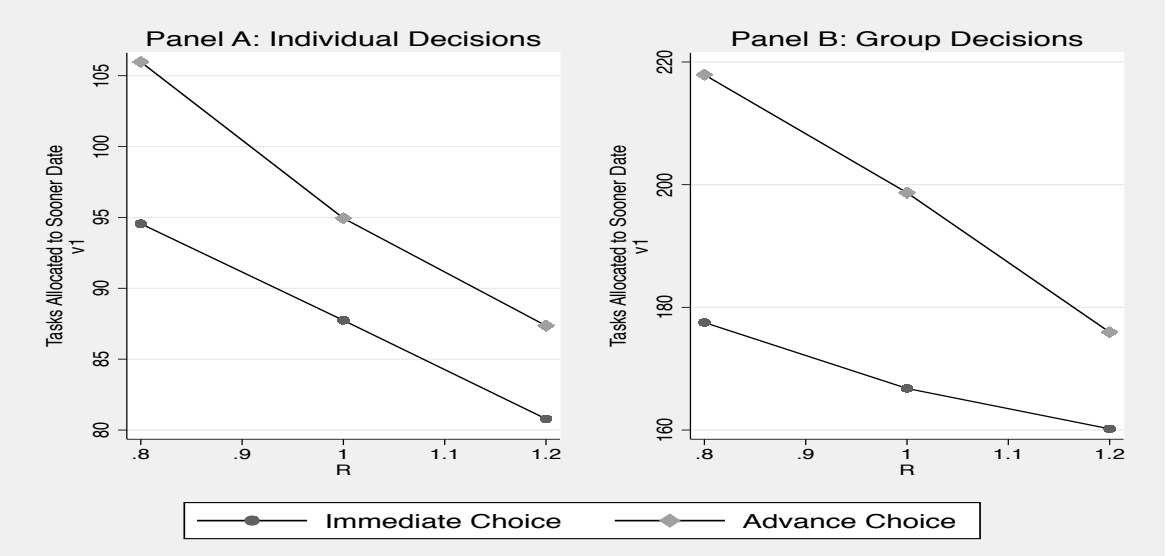


Figure 1.2: Discounting Behavior

Notes: Mean behavior in Individual and Group task allocated to sooner date combined.

1.4 Results: Reduced Form

In this section, we present a reduced-form analysis to test whether individuals and groups are time consistent. We run a two-limit tobit regression to analyze the effect on time preferences

based on all the experimental variations. The advantage of reduced-form analysis is that we do not have to make any assumptions about the functional form of time preferences.

In Table 1.2, our outcome variable is the natural log of v_1 (task allocation for Day 8). In all our regressions, we control for the experimentally induced variations in R by including its natural log. As we set a minimum of twelve tasks to avoid corner solutions in allocation decisions, we use a two-limit tobit regression, which corrects for censoring (Wooldridge 2002).

Table 1.2: Two-Limit Tobit Regression Analysis

Dependent variable:	Log of Tasks Allocated to the 1st Day			
	(1)	(2)	(3)	(4)
β_1 : Log Task Rate	-0.46***	-0.46***	-0.46***	-0.46***
	(0.08)	(0.08)	(0.08)	(0.08)
β_2 : Immediate Decision	-0.15***	-0.21***	-0.14*	-0.20**
	(0.04)	(0.05)	(0.07)	(0.08)
β_3 : Individual Decision		-0.75***		-0.75***
		(0.02)		(0.02)
β_4 : Immediate Individual Decision		0.09**		0.09**
		(0.04)		(0.04)
β_5 : Individual Decision First			-0.04	-0.04
			(0.06)	(0.06)
β_6 : Immediate Individual Decision First			-0.02	-0.02
			(0.08)	(0.08)
β_0 : Constant	4.70***	5.20***	4.72***	5.23***
	(0.03)	(0.03)	(0.04)	(0.04)
# of Obs	2196	2196	2196	2196
# of Groups	122	122	122	122
F-stats	23.68	326.64	13.36	219.78
Hypothesis (*p-values*)				
$H_0 : \beta_2 = 1$	0.00			
$H_0 : \beta_2+\beta_4 = 1$		0.00		
$H_0 : \beta_2+\beta_6 = 1$			0.00	
$H_0 : \beta_2+\beta_4+\beta_6 = 1$				0.00

Notes: $^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. The table presents the effect of Immediate Decision and its interactions with other experimentally induced variations using the Two-Limit Tobit Regression technique. The first presents results with aggregated decisions. The second column captures the estimates of the Immediate Decision for the groups and individuals separately. The third column shows the results of the effect of decision ordering and its interaction with Immediate Decisions. The fourth column includes all previous variables. We cluster standard errors at the group level.

In column (1), we regress v_1 on a dummy variable, *Immediate Decision*, which takes the value of 1 when the allocation decision is immediate, i.e. an allocation made on Day 8 for tasks to be performed that very day. Here we combine both individual and group decisions (also illustrated in Figure 1.2). We show that participants allocate significantly fewer tasks to v_1 when R increases and when the allocation decision is immediate. This means that when the task rate is high (that is, it is more expensive to complete earlier tasks), participants allocate fewer tasks to earlier periods compared to when the task rate is low. This shows that our participants understand the trade-offs involved and make rational choices. Further, we find that on average 15% fewer tasks are chosen on the first day of task completion (Day 8).

In the second column, we add a dummy for decisions made by individuals only and a dummy where we interact *Individual Decision* with the *Immediate Decision* dummy variable. We find that individual participants making immediate choices allocate around 12% fewer tasks to v_1 than those making the same decision a week before the first day of task completion. The corresponding figure for groups is 21%. This shows that time inconsistency is lower for individuals compared to groups as the *Immediate Decision* dummy estimate interacted with the *Individual Decision* dummy increases by 9%. The estimate for *Individual Decision* is negative because of the variation between individuals and their groups in the total number of tasks over the two-week period (the total number of tasks is two hundred and four hundred for individuals and groups respectively). The negative estimate signifies that participants allocate fewer tasks to earlier dates as individuals than as a group.

These findings differ from recent findings (for example, Carbone and Infante (2015) and Denant-Boemont and Loheac (2011)) that groups are less present biased and more time consistent compared to individuals. We attribute this difference to the three factors we explained earlier. First, these studies use monetary methods to elicit time preferences, which, as noted earlier, are imperfect measures of time preference because many variables confound the identification of the shape of time preferences (Augenblick, Niederle, and Sprenger 2015; Cohen et al. 2020; Sprenger 2015). Second, these studies conduct experiments with couples, which are endogenously formed groups, unlike our exogenous groups. Third, group present-bias estimates could be higher than individual estimates (for exactly the same people) because of the underlying bonus structure. Since our tasks are relatively easy to perform and are perfect substitutes, the coordination externality may be an important factor. Later, we explore these issues in more detail by putting some theoretical structure on the estimates.

Finally, in the third column, we analyze the effect of the order in which task-allocation decisions are taken. The order does not have any significant effect on behavior in either the individual or group setting. The estimates of the *Individual Decision First* dummy and its interaction with the *Immediate Decision* dummy are statistically insignificant. This, along with the nonchanging *Immediate Decision* estimate and its interaction with the *Individual Decision* dummy estimates in the last column, indicates the robustness of our finding. In the appendix, we also provide results controlling for demographic variables and show that all our results qualitatively remain the same.

1.5 Results: Structural Analysis

1.5.1 Nonlinear Regression Analysis

Next, we use a structural model to estimate time-preference parameters. We assume quasi-hyperbolic discounting, with subjects allocating effort for a task to an earlier date, v_1, or a later date, v_2. Under the assumptions that the cost-of-effort function is quadratic and that individuals and groups discount the future quasi-hyperbolically (Laibson 1997; O'Donoghue and Rabin 1999), the participants' preferences can be written as follows:

$$b_1 v_1^2 + \beta^{\mathbf{1}_{d=1}} \delta^k \; b_2 v_2^2$$

Normalizing $b_2 = 1$ and therefore dividing the intertemporal effort-cost function by b_2 (to remove scaling effects), the above cost function can be rewritten as follows:

$$\gamma v_1^2 + \beta^{\mathbf{1}_{d=1}} \delta^k \; v_2^2$$

Here, v represents a task performed on a given day (either an earlier or later day), $\gamma > 0$ and $\gamma = 1$ imply that the effort-cost function is stationary over time, and k captures delay length, which in this experiment was fixed at seven days (the gaps between first decision and first task day and between first task day and second task day). The indicator $\mathbf{1}_{d=1}$ captures whether the decision is made immediately or in advance of the first day of task performance. The parameters β and δ encapsulate individual or group discounting, with β capturing the degree of present bias, active for participants who make immediate decisions; that is, $\mathbf{1}_{d=1} = 1$. If $\beta = 1$, the model nests exponential discounting with the discount factor δ, while if $\beta < 1$ the decision-maker exhibits present bias, being less patient in immediate decisions than advance decisions.

When modeling group decisions, we assume the group members are characterized by individual preferences and that the group acts as a decision-maker, similar to an individual, whose time-preference parameters can be measured independently of its members' preferences. This modeling technique is in line with the representative-agent setup mostly used in macroeconomic modeling. This unitary approach assumes that the collective acts as a single decision-making unit and therefore can be treated as a rational individual.

Minimizing discounted costs subject to an intertemporal budget constraint yields the following intertemporal Euler equation:

$$\gamma v_1^* R = \beta^{\mathbf{1}_{d=1}} \delta^k v_2^* \tag{1.1}$$

Here v_1^* and v_2^* are the optimal tasks performed on Day 8 and Day 15. This tangency condition implies that when individual or group preferences are dynamically consistent, the optimal $\frac{v_1^*}{v_2^*}$ does not depend on the parameter $\beta^{\mathbf{1}_{d=1}}$ but only depends on the task rate R and the delay length k. Using the Euler equation with the intertemporal budget constraint and rearranging the equations yields the solution function for the optimal v_1^*:

$$v_1^* = \left(\frac{\beta^{\mathbf{1}_{d=1}} \delta^k V}{\gamma R^2 + \beta^{\mathbf{1}_{d=1}} \delta^k}\right)$$

and

$$\boldsymbol{v_1^*} = \begin{cases} \left(\frac{\beta^{\mathbf{1}_{d=1}} \delta^k V}{\gamma R^2 + \beta^{\mathbf{1}_{d=1}} \delta^k}\right) & d = 1 \\ \left(\frac{\delta^k V}{\gamma R^2 + \delta^k}\right) & d = 0. \end{cases}$$

The above equation implies that v_1^* is a nonlinear function of R, $1_{d=1}$, k, and V.[6] If we assume that allocation decisions satisfy the above equation subject to an additive error term, ϵ, we arrive at the nonlinear regression equation:

$$v_{1_{it}}^* = fV, R_{it}, \mathbf{1}_{d=1}, k + \epsilon_{it}. \tag{1.2}$$

The parameters β (present bias), δ (discount factor), and γ (curvature-of-cost function) can be estimated using a nonlinear-least-squares estimation at the individual or group level (see Appendix 1.8 for details). We present these estimates in Table 1.3. Throughout, we cluster standard errors at the group level.

In the first column, we pool all decisions together and show combined results. We show that our main parameter of interest, β, the present bias in effort provision is 0.78 in the aggregate ($s.e. = 0.05$), which means that participants are not time consistent. This estimate is close to the reduced-form result of Table 1.2, which shows that the theoretical model under consideration is a good fit for the experimental data. The weekly discount factor, δ, averages

[6] For this class of effort-cost function, both relative risk aversion and intertemporal elasticity of substitution are functions of v.

around 0.98. Finally, we find that for the cost-parameters ratio, γ, we cannot reject the null hypothesis of a stationary cost-of-effort function; that is, the intertemporal effort-cost function is stationary over time.

Table 1.3: Non-Linear Least Squares Analysis

Dependent Variable:		$v^*_{1_{it}}$			
Combined		*Ind. Vs. Group*		*Decision Order*	
$\beta_{Combined}$	0.78***	β_{Ind}	0.82***	$\beta_{IndFirst}$	0.77***
	(0.05)		(0.06)		(0.05)
$\delta_{Combined}$	0.98***	δ_{Ind}	0.96***	$\delta_{IndFirst}$	0.98***
	(0.04)		(0.03)		(0.06)
$\gamma_{Combined}$	1.07***	γ_{Ind}	1.00***	$\gamma_{IndFirst}$	1.17***
	(0.23)		(0.22)		(0.39)
		β_{Group}	0.71***	$\beta_{IndSecond}$	0.0.79***
			(0.06)		(0.10)
		δ_{Group}	1.01***	$\delta_{IndSecond}$	0.95***
			(0.04)		(0.03)
		γ_{Group}	1.23***	$\gamma_{IndSecond}$	0.85***
			(0.29)		(0.19)
# Observations	2196	# Observations	2196	# Observations	2196
# Groups	122	# Groups	122	# Groups	122
RMSE	0.54	RMSE	0.54	RMSE	0.54
Hypothesis					
$\beta_c = 1$, *p-value*:	0.00	$\beta_I = 1$, *p-value*:	0.00	$\beta_{IF} = 1$, *p-value*:	0.00
$\delta_c = 1$, *p-value*:	0.48	$\beta_G = 1$, *p-value*:	0.00	$\beta_{IS} = 1$, *p-value*:	0.03
$\gamma_c = 1$, *p-value*:	0.76	$\beta_I = \beta_G$, *p-value*:	0.03	$\beta_{IF} = \beta_{IS}$, *p-value*:	0.91

Notes: $^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. The table presents structural estimates of intertemporal hyperbolic discounting model using non-linear least squares estimation. The dependent variable is the amount of optimal tasks allocated for Day 8 of the experiment. The first column presents combined individual and group decision estimates for β, δ, and γ. The second column presents the structural estimates of the model for the individuals and groups separately. The third column presents the structural estimates based on the effect of decision ordering (either individual allocation decision taken first or second). We cluster standard errors at the group level.

In the second column, we compare individuals' decisions and groups' decisions. The estimate for the present-bias parameter shows present bias in effort choice exists for both individuals and groups (which confirms our reduced-form results). However, the present-bias estimate is lower for individuals compared to groups. Further, individuals' weekly discount parameter is less than the groups' parameter estimate. Comparing the individual parameter estimate of present bias with the corresponding group estimate, the null-hypothesis test of $\beta_I = \beta_G$ is rejected, as is the $\delta_I = \delta_G$ hypothesis. Hence, individuals and groups do indeed exhibit different present bias and discount factors. These results are similar to those of our nonparametric-analysis, in which the degree of time inconsistency is, on average, greater in group decisions compared to individual decisions but, at the same time, groups are more patient than individuals. The discount-factor estimates and their individual-versus-group differences are consistent with Milch et al. (2009), who find that participants discount more

as individual decision-makers than they do as members of groups. The finding that groups are more patient also shows that the discount factors for groups are in line with market interest rates, unlike the discount factor for individuals.

Finally, comparing the cost-parameter estimates between individuals and groups, we observe that the estimate of γ is higher for groups compared to individuals, but $\gamma_I - \gamma_G$ has a coefficient of -0.23 with a p-value of 0.13, which indicates that the difference is not statistically significant. The individual decisions' estimated cost-parameters ratio is γ_I=1.00 (s.e.=0.21). The null hypothesis of a stationary cost-of-effort function cannot be rejected, since the $\chi^2 1$ test has a p-value of 0.99. For groups' estimated cost-parameters ratio γ_G=1.23 (s.e.=0.29), under the null hypothesis, groups' stationary cost-of-effort function also cannot be rejected, since the $\chi^2 1$ test has a p-value of 0.43. These results indicate that the underlying cost-of-effort functions for both individuals and groups are stationary over time and are statistically not different from each other.

In the third column, the time-preference parameters are estimated for an ordering effect using the full sample of decisions; that is, we test whether making an individual decision first or second makes any difference. While the coefficients of *Individual Decision First* and *Group Decision First* differ, the difference is statistically insignificant. The same holds true for the discount factor. Hence, we find no support for ordering effects. These results also confirm results from the non-parametric reduced-form estimation shown in Table 1.2.

1.6 Channels: Individuals vs. Groups

The results above lead to two questions. First, what structural channels explain the connection between groups' greater present bias and their constituent members? Second, could any confounding variables help explain groups' present bias? These questions are important for assessing the empirical validity of our theory and understanding what drives group decisions. Below, we present a theoretical model that connects groups' decision-making process to their constituent members. Based on this theoretical model, we primarily test two things. First, we test how absolute differences within a group (within group heterogeneities) in terms of individual members' (i) discount rates, (ii) cost of effort parameters, and (iii) bargaining power within the group affect group present bias. Second, as our groups are randomly created

pairs, we construct $\hat{\beta}^i_{min}$, $\hat{\beta}^i_{max}$, $\hat{\delta}^i_{min}$, and $\hat{\delta}^i_{max}$ for each group and test how the least and most present-biased or patient individuals in a group affect group behavior.

1.6.1 Theory: Collective Decision Functions

We start by introducing a collective cost-of-effort function, which can be thought of as a planner's cost-of-effort function for a group. This aggregation can be done in multiple ways, e.g., utilitarian or Rawlsian approaches. The utilitarian approach involves taking a weighted average of agents' cost-of-effort functions $FCv = {}_i\,\omega_i C_i v$, where C is the cost function and v is the amount of effort provided. The Rawlsian approach involves considering the minimum of agents' cost-of-effort provision $FCv = min_i C_i v$. This means that the greatest consideration is for the least well-off.

In this experiment, we know $C = \delta_1, \beta_1, \gamma_1; ..; \delta_n, \beta_n, \gamma_n$ as we estimated these important parameters using our structural model. This leads us to consider an important class of collective cost-of-effort functions: those that are time separable. This class of functions exhibits a particular type of time inconsistency or intransitivity[7] - namely, present bias - which matches our empirical evidence (Jackson and Yariv 2015).

As shown by Jackson and Yariv (2015), given that all the participants in the experiment are time consistent for any profile $\delta_1, \gamma_1; ..; \delta_n, \gamma_n \in C^n$, a time-separable group cost-of-effort function takes the following form:

$$F\delta_1, \gamma_1, ; ...\delta_n, \gamma_n v = {}_t\,\tilde{\delta}_t C v_t,$$

such that $\tilde{\delta}_t = {}_i\,\omega_i \delta^t_i$. Such time-separable collective functions are standard in the literature. The standard utilitarian aggregation of individual utilities (i.e., ones that put different weights on different individuals) is a special case. According to Jackson and Yariv (2015), for any profile $\delta_1, \gamma_1; ..; \delta_n, \gamma_n \in C^n$ such that for some k, j, $\delta_k \neq \delta_k$, the above collective function, F, is either dictatorial or present biased.[8] In our experiment, before making group decisions, every participant in the experiment was asked to make unanimous decisions in the group since the bonus share was fixed for each group member. Therefore, the collective discount

[7] The intransitivity here is different from Condorcet's (1785) description of the voting paradox and Arrow's impossibility theorem because our collective-decision structure is quite different from the voting settings mentioned in these papers.

[8] For proofs, see Jackson and Yariv (2014) and Jackson and Yariv (2015).

factor must be a weighted sum of the participants' discount factors and thus must correspond to a weighted utilitarian collective cost-of-effort function.

The above proposition encompasses many different formulations of time-inconsistent preferences. In our structural analysis, we assumed a quasi-hyperbolic formulation, which in this case corresponds to $\tilde{\delta_1} = 1$ and $\tilde{\delta_t} = \beta\delta^{t-1}$ for all $t > 1$. As long as behavior has a separable structure and satisfies unanimity, the above proposition shows that present bias is to be expected.

Using this proposition, a set of testable hypotheses can be generated. We use groups' time-preference estimates (specifically focusing on groups' present bias) and their constituent members' time preference estimates from our structural model, including estimates of the discount rate, present bias and effort-cost parameters, and empirically test the effects of theory-based within-group heterogeneities on group present bias. These within-group heterogeneities include differences in individual members' (i) discount rates, (ii) cost of effort parameters, and (iii) bargaining power within the group.

For this last step, we have to model bargaining power. In a utilitarian formulation, group decisions depend on the choices of, and the relative strengths of the constituent members' weights in the group decisions (captured by Pareto, or bargaining, weights).[9] Following Browning and Chiappori (1998), we express a group's intertemporal effort-cost function as:

$$C_G = \omega_i C_i + \omega_j C_j,$$

where

$$\omega_i + \omega_j = 1 \quad and \quad \omega_i, \omega_j \geqslant 0.$$

These restrictions satisfy the unanimity condition in the group decision-making process:

$$C_l = \ \gamma_l v_{1l}^2 + \beta_l {}^{\mathbf{1}_{d=1}} \delta_l{}^k \ v_{2l}^2 \quad \ and \quad \ l = \{G, i, j\}.$$

Here C_G is the group's intertemporal effort-cost function, C_i and C_j represent the intertemporal effort-cost function of members i and j, respectively, and ω_i and ω_j denote the bargaining power of members i and j, respectively. The latter measures how individual

[9] A bargaining mechanism was introduced into empirical models of the group decision-making process (Manser and Brown 1980; McElroy and Horney 1981). Further, researchers have developed collective models, which assume that groups can achieve efficient decisions (Browning and Chiappori 1998; Chiappori 1992).

preferences are aggregated into groups' joint decisions. In our experimental setting, since we observe both individual and group decisions, we can estimate the extent to which each member influences a group's decisions. Using a multinomial logit model, we estimate ω_i and ω_j by running the following regression equation for each group :

$$v_{1G_p} = \omega_i v_{1i_p} + \omega_j v_{1j_p} + \epsilon_p,$$

$$s.t. \;\; \omega_i + \omega_j = 1 \quad and \quad p = \{1, 2, 3, ..., 122\}.$$

After estimating ω_i, ω_j for each group, we construct $|\Delta\hat{\omega}_{IND}|$, which is the absolute difference between the members' bargaining weights as estimated above. Using these absolute differences, we further construct $|\Delta\hat{\omega}| \approx 1_G = 1$, which is a dummy indicator for groups in which one of the members has virtually all the bargaining power (when $|\Delta\hat{\omega}_{IND}| > 0.98$).

1.6.2 Empirical Results: Absolute Differences Within Groups

Next, we investigate how groups' decisions are determined by the present bias of their members. As mentioned earlier, theoretically, in a group context, inconsistencies can arise simply from the aggregation of heterogeneous preferences because of variations in individual discount rates and cost-function parameters (Jackson and Yariv 2015). Hence, we now test how variations in individual discount rates and effort-cost parameters as well as differences in within-group bargaining weights could generate group behavior that is more present biased than individual behavior. We also test other important individual factors that shape group behavior, such as coordination externalities within groups.

In Table 1.4, we show summary statistics for the individual-level variables that may affect group-level behavior: we show the absolute differences in measures of each group's members, such as weekly discount rate, effort-cost parameter, and bargaining power (we use $\hat{\omega}$ to construct a dummy indicator, $|\Delta\hat{\omega}| \approx 1_G = 1$, which indicates a group in which one of the members has virtually all the bargaining power: $|\Delta\hat{\omega}_{IND}| > 0.98$). We show that the absolute difference between each group's members' weekly discount rate ($|\Delta\hat{\delta}_{IND}|$) has a mean of 6% with a standard deviation of 0.07. The absolute difference between members' cost-of-effort parameters ($|\Delta\hat{\gamma}_{IND}|$) has a mean of 0.36 with a standard deviation of 0.41. The absolute difference between members' bargaining weights ($|\Delta\hat{\omega}_{IND}|$) has a mean of 0.85, indicating that

Table 1.4: Summary Statistics

Variables	Mean	Standard Deviation	Minimum	Maximum
$\lvert\Delta\hat{\delta}_{IND}\rvert$	0.06	0.07	0	0.26
$\lvert\Delta\hat{\gamma}_{IND}\rvert$	0.36	0.41	0	2.15
$\lvert\Delta\hat{\omega}_{IND}\rvert$	0.85	0.29	0	1
$\lvert\Delta\hat{\omega} \approx 1\rvert_G = 1$	0.72	0.44	0	1
$\hat{\beta} \approx 1_{both} = 1$	0.05	0.21	0	1
Acquaintance Time Duration (months)	13.84	21.56	0	60

Notes: No. of observations is 244. $\lvert\Delta\hat{\delta}_{IND}\rvert$ is the absolute difference between the groups' individual members' weekly discount rate. $\lvert\Delta\hat{\gamma}_{IND}\rvert$ is the absolute difference between the groups' individual members' cost of effort parameters. $\lvert\Delta\hat{\omega}_{IND}\rvert$ is the absolute difference between the groups' individual members' bargaining/Pareto weights. $\lvert\Delta\hat{\omega}\rvert \approx 1_G = 1$ is the dummy indicator for the group in which one of the members has approximately all the bargaining power. $\hat{\beta} \approx 1_{both} = 1$ is the dummy indicator, which takes the value of 1 for those groups in which both members are time consistent, approximately.

members have different bargaining power and that for most groups the chances of having a nondictatorial setup are quite high. Finally, we show that the high-bargaining-power variable has a mean of 0.72, which shows that in a majority of groups the probability of having a dictatorial member (borrowing the terminology from Jackson and Yariv (2015)) is high; that is, the groups typically ignore the preferences of all but one agent.

In our design, the possibility that groups in which each participant is time consistent does not pose any additional challenge for our main experimental findings since we observe both group and individual decisions. Focusing on how aggregation relates to time inconsistency, we explicitly controlled for the underlying individual preferences to isolate the effects of aggregation. In cases in which each participant is time consistent, we construct the variable $\hat{\beta} \approx 1_{both} = 1$, a dummy indicator taking the value of 1 for groups in which both members are time consistent (which we define as both members having a $\hat{\beta}$ between 0.95 and 1.05). This variable has a mean of 0.05, indicating that in the overall sample, only 5% of groups have nearly time-consistent members.

Finally, in Table 1.5, we regress individual-level absolute differences for different variables on the group time-preference parameter. In column (1), we regress the absolute value of the difference between each group's members' discount rates, cost functions, bargaining power, and bargaining-power dummy. We test the robustness of this result in column (4), in which we add control variables: the absolute value of the difference in Big Five, age, and gender. These results show that the weekly-discount-factor heterogeneities among individuals explain group present bias. They also show that the presence of a dictator or individual who is dominant (in terms of bargaining power) reduces group present bias. These results are in line with Jackson and Yariv (2014)'s theoretical prediction that for a uniform distribution of discount rates in an otherwise-homogeneous population, group utility maximization generates

Table 1.5: Individual vs. Group Regression Analysis

Dependent variable:	$\hat{\beta}_G$					
	(1)	(2)	(3)	(4)	(5)	(6)
$\lvert\Delta\hat{\delta}_{IND}\rvert$	-3.84**	-3.07**	-3.16**	-3.86**	-3.00**	-3.09**
	(1.25)	(1.19)	(1.28)	(1.26)	(1.23)	(1.32)
$\lvert\Delta\hat{\gamma}_{IND}\rvert$	0.14	0.21	0.18	0.13	0.24	0.19
	(0.26)	(0.24)	(0.25)	(0.28)	(0.24)	(0.25)
$\lvert\Delta\hat{\omega}_{IND}\rvert$	-0.25	-0.17	0.07	-0.21	-0.21	0.06
	(0.23)	(0.27)	(0.31)	(0.27)	(0.26)	(0.28)
$\lvert\Delta\hat{\omega}\rvert \approx 1_G = 1$	0.58**	0.43**	0.30	0.57**	0.43**	0.30
	(0.20)	(0.19)	(0.20)	(0.21)	(0.18)	(0.18)
$\hat{\beta} \approx 1_{both} = 1$	-0.27	-0.36	-0.38	-0.30	-0.30	-0.35
	(0.23)	(0.30)	(0.32)	(0.30)	(0.30)	(0.32)
Acquaintance Duration		0.01*	0.01*		0.01*	0.01*
		(0.01)	(0.01)		(0.01)	(0.01)
$\lvert\Delta Big\ 5\rvert$			-0.08			-0.05
			(0.14)			(0.15)
$\lvert\Delta Age\ in\ years\rvert$				-0.04	0.02	0.06
				(0.06)	(0.05)	(0.07)
$\lvert\Delta Gender\rvert$				-0.09	0.17	0.11
				(0.22)	(0.18)	(0.20)
Constant	1.02***	0.76***	0.72**	1.11***	0.65**	0.56
	(0.14)	(0.19)	(0.25)	(0.24)	(0.24)	(0.34)
# of Groups	122	122	109	122	122	109
Adj R^2	0.06	0.30	0.30	0.07	0.31	0.31
RMSE	1.13	0.98	1.02	1.13	0.97	1.02
$H_0 : Constant = 1$						
p-value	0.87	0.19	0.27	0.65	0.15	0.19

Notes: $^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Standard errors are clustered at the group level. Column (1) presents the estimates of variations in group members' weekly discount factors, cost of effort parameters, bargaining power estimates, presence of dictator, and both time-consistent members dummies. Column (2) presents the estimates of effect of acquaintance duration on group present-bias, controlling for the variables in column (1). Column (3) captures the estimates of effect of within-group differences in Big 5 personality traits, controlling for the variables in column (2). Columns (4), (5), and (6) represent the estimates of theory and non-theory-based factors, controlling for within-group differences in age and gender.

aggregate behavior that corresponds to hyperbolic discounting and that if there is some fundamental heterogeneity in temporal preferences in the form of differing discount factors, then the only well-behaved collective utility functions that are time consistent and respect unanimity are dictatorial. For these two columns, we see that, given that the individuals are exponential discounters and given that there are no variations in individual discount rates and in effort-cost parameters, the groups' allocation decisions would represent time-consistent patterns.

In columns (2) and (5), we include acquaintance duration as a control in the previous two specifications. The results show the effect of the coordination externality on the groups' present bias estimates. It is natural to think that as *Acquaintance Duration* increases, individuals' coordination problems lessen. This in turn would decrease group present bias. In both columns, the control's estimate is significant at 10%. Under $H_0 : Constant = 1$, $F1, 121$

have p-values that show that, given that there are no variations in individual discount rates and in effort-cost parameters, groups' allocation decisions would represent time-consistent patterns even if the individuals were exponential discounters with *Acquaintance Duration* of 0. Last, in columns (3) and (6), we add the absolute value of the difference in Big Five personality traits to our main specifications. These results show that Big Five personality traits cannot explain groups' present bias. The effect of the presence of a within-group dictator is not significant in these columns, which may be attributable to the fluctuation of the overall sample size.

To summarize, the main message of Table 1.5 is that groups whose members have misaligned discount rates are present biased and that the presence of a dictator within a group improves the group's present-bias estimate. Emphasizing that divergent preferences within a group are sufficient to render time inconsistency, we find that the variation in discount factors significantly affects group present bias even after controlling for time-consistent individuals and the presence of a dictator. Similarly, the effect of coordination externalities, which is captured by *Acquaintance Duration*, is also important for understanding group present bias.

1.6.3 Theory: Individual Minimums and Maximums and Group Parameters

The above stylized theoretical setup motivates another important hypothesis. Having constructed $\hat{\beta}^i_{min}$, $\hat{\beta}^i_{max}$, $\hat{\delta}^i_{min}$, and $\hat{\delta}^i_{max}$ for each randomly created group, we investigate how the least and most present-biased individuals in a group affect group time preferences. We employ a model in which a group's time-preference parameter is a linear function of its members' parameters:

$$\hat{\beta}_g = \alpha_1 + \kappa_1 \hat{\beta}_{min} + \kappa_2 \hat{\beta}_{max} + \epsilon_{1_g} \qquad , \qquad \hat{\delta}_g = \alpha_2 + \eta_1 \hat{\delta}_{min} + \eta_2 \hat{\delta}_{max} + \epsilon_{2_g}$$

First, we investigate whether there is a difference between individual and group decisions that are independent of individuals' time preferences. Specifically, we test the hypothesis that $\alpha_i = 0$ for $i = 1, 2$ ($\alpha_i = 0$ would imply no relationship between individuals' and groups' preferences). Second, we investigate the hypothesis that the coefficients of individual decisions sum to one:

$$\sum_{i=1}^{2} \kappa_i = 1 \quad , \quad \sum_{i=1}^{2} \eta_i = 1,$$

Together, these two tests imply that the group decision is a convex combination of individual decisions. Hence, the coefficient of the latter can be interpreted as the weights of different members in shaping the group decisions.

Third, we examine the mean hypothesis, according to which group decisions are simply a function of mean individual decisions:

$$\hat{\beta}_g = \alpha_1 + \frac{\kappa}{2}\hat{\beta}_{min} + \frac{\kappa}{2}\hat{\beta}_{max} \quad , \quad \hat{\delta}_g = \alpha_2 + \frac{\eta}{2}\hat{\delta}_{min} + \frac{\eta}{2}\hat{\delta}_{max}.$$

This hypothesis implies that the mean is a sufficient statistic for the group's decision. If $\kappa = 1$, then the mean present-bias parameter exactly predicts the component of the group present-bias parameter that varies with individual preferences (that is, we can reject the hypothesis that $\kappa_1 = \kappa_2$).

Fourth, the strong mean hypothesis further requires the mean to exactly predict the group present-bias parameter, and it requires us to test whether $\kappa_1 = \kappa_2 = \frac{1}{2}$. These arguments also hold true for the long-run discount parameter $\hat{\delta}_g$.

1.6.3.1 Empirical Results: Individual Minimums and Maximums and Group Parameters

Now, we present the empirical analysis to show how individual decisions are connected to group outcomes. In Table 1.6, we use the estimates of parameters for individuals' and groups' annual discount rate ($\hat{\delta}$), present bias ($\hat{\beta}$) and cost parameter ($\hat{\gamma}$) calculated using equation (2) to show summary statistics for groups and individuals separately. We report the median, fifth, and ninety-fifth percentiles, and minimum and maximum values. While the median estimated weekly discount rate (0.92) and cost parameters (0.69) are the same for individuals and groups, the median individual present-bias estimate (0.97) is higher than the group median (0.90).

For most individuals and groups, the estimation strategy generates reasonable parameter estimates. However, extreme observations do exist. Figure 1.3 presents histograms of time-preference parameter estimates, $\hat{\beta}$, and discounting parameter estimates, $\hat{\delta}$. We can see that a large proportion of subjects have low discount rates and high present bias.

Table 1.6: Summary Statistics: Discounting, Present-Bias, and Effort Cost Parameter Estimates

	N	Median	5th Percentile	95th Percentile	Minimum	Maximum
Group						
$\hat{\delta}$	122	0.92	0.75	0.99	0.55	0.99
$\hat{\beta}$	122	0.90	0.18	2.32	0.02	10.8
$\hat{\gamma}$	122	0.69	0.19	1.33	0.14	2.80
Individual						
$\hat{\delta}$	244	0.92	0.76	0.97	0.66	1.00
$\hat{\beta}$	244	0.97	0.13	2.79	0.02	15.1
$\hat{\gamma}$	244	0.69	0.18	1.51	0.06	2.80

Notes: We use results from the non-linear least squares (NLS) estimator used in Table 1.3. We show summary statistics for discounting, present-bias and effort cost parameter estimates for all groups and all individuals separately.

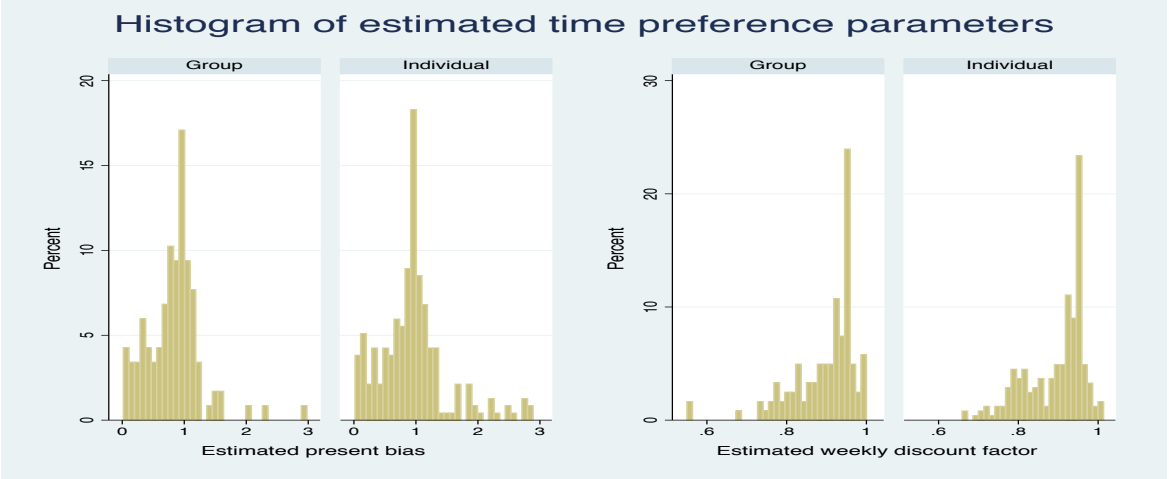


Figure 1.3: Estimates Histogram

In Table 1.7, we use the fact that each group consists of two randomly paired individuals and calculate across-group estimates for the median, fifth, and ninety-fifth percentiles and the minimum and maximum annual discount rate ($\hat{\delta}$), present bias ($\hat{\beta}$) and cost parameter ($\hat{\gamma}$) for individuals. This shows us the spread across the groups and allows us to compare the spread of our estimates for individuals (Table 1.6) with the spread of our estimates for groups (Table 1.7). We see that the group median parameter estimates in Table 1.6 are always greater than the minimum and less than the maximum of the median of the estimates presented in Table 1.7. The same is true for the fifth-percentile and ninety-fifth-percentile estimates of groups compared with their minimums and maxima. This finding is in line with Gollier and Zeckhauser's (2005) model of aggregation of time preferences, in which the rate of

Table 1.7: Summary of Across-Group Min and Max Parameter Estimates

	N	Median	5th Percentile	95th Percentile	Minimum	Maximum
Minimum						
$\hat{\delta}$	122	0.88	0.72	0.95	0.66	0.95
$\hat{\beta}$	122	0.81	0.09	1.23	0.03	11.5
$\hat{\gamma}$	122	0.49	0.15	1.12	0.57	2.80
Maximum						
$\hat{\delta}$	122	0.95	0.79	0.99	0.75	1.00
$\hat{\beta}$	122	1.07	0.39	3.93	0.03	15.1
$\hat{\gamma}$	122	0.79	0.28	1.74	0.14	2.80

Notes: We use results from the non-linear squares (NLS) estimator used in Table 1.3. We calculate summary statistics for individuals across all groups dividing them up into the minimum constituent individual and the maximum constituent individual in each group and present summary statistics for discounting, present-bias, and effort cost parameter estimates.

impatience of the representative agent is a weighted mean of individual rates of impatience, although this may not hold at the extreme points.[10]

In Figure 1.4, we see the distribution of time-preference parameters: the present bias and weekly discount factors for the minimum and maximum in the group along with its corresponding collective estimates. The figure visually highlights two important results. First, groups' time-preference estimates are bounded by their members' estimates. Second, the group present-bias estimates tend to be closer to the within-group minimum estimates. In other words, the group member with a higher present bias dominates the group and drives the group's present bias.

In Table 1.8, we report results from regressing the group-decision estimated time-preference parameters on the individual-decision parameters. Models (1) and (3) show basic linear specifications, while models (2) and (4) test for robustness of the results by including controls for order effects. In both models of present bias, columns (1) and (2), the coefficient, κ_1, is positive and significant. This is an important result signifying that the individual with higher present bias, essentially the more constrained individual, dictates the group present-bias dynamics. The positive constant term indicates complementarity within the group's intertemporal behavior. In both models of the long-run discount factor, columns (3) and (4), we show that although the more patient individual is barely significant in explaining a

[10] Gollier and Zeckhauser (2005) also showed that heterogeneous individual exponential discounting yields collective hyperbolic discounting.

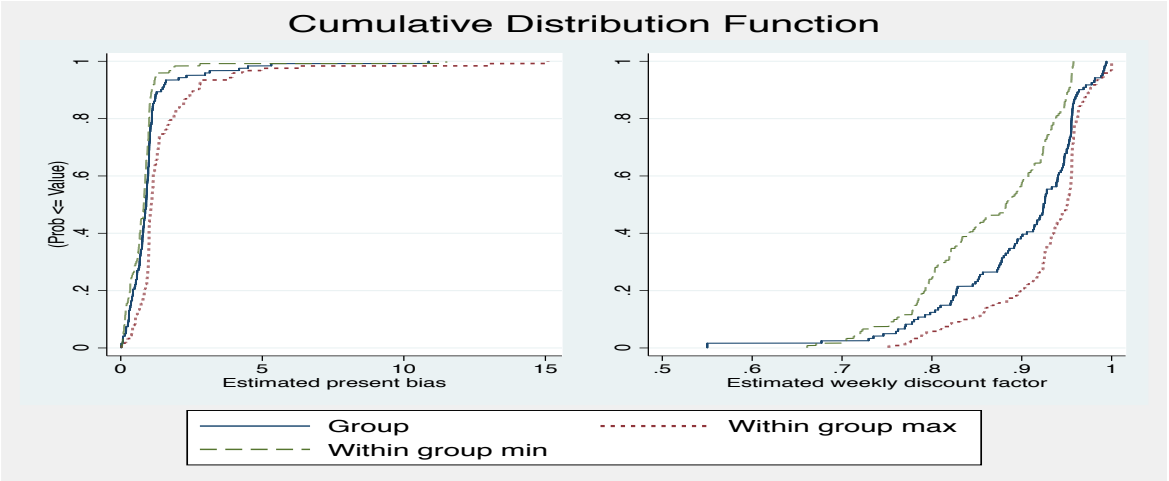


Figure 1.4: Estimated CDF

group's long-run discount rate, the coefficient becomes insignificant when we control for order effects. Hence, the patient individual cannot explain the group's long-run discount factor.

Our results support the hypothesis of a level shift for present bias, while we cannot reject the null hypothesis of no shift in the long-run discount rate - that is, $H_0 : \kappa_2 = 0$. The third part of the table reports the results from our postestimation hypotheses - namely, the weak and strong versions of the mean and convex-combination hypotheses, which we developed for the four OLS models reported in Table 1.8. For $\hat{\beta}_g$, we reject both versions of both hypotheses; that is, individuals' choices are not equally important to explain groups' present bias. On the other hand, for $\hat{\delta}_g$, we are unable to reject either the weak or strong hypotheses, indicating that individuals' mean is a sufficient statistic for the group's decision.

Table 1.8: OLS regressions of group choices on individual choices

Dep. Var:	$\hat{\beta}_g$		$\hat{\delta}_g$		
	(1)	(2)		(3)	(4)
$\kappa_1 : \hat{\beta}_{min}$	0.79***	0.79***	$\eta_1 : \hat{\delta}_{min}$	0.19	0.15
	(0.08)	(0.08)		(0.14)	(0.21)
$\kappa_2 : \hat{\beta}_{max}$	0.11*	0.11*	$\eta_2 : \hat{\delta}_{max}$	0.74*	0.76
	(0.04)	(0.04)		(0.40)	(0.46)
α_1 : Constant	0.20***	0.22**	α_2 : Constant	0.03	0.04
	(0.05)	(0.08)		(0.32)	(0.30)
Order Effect	N	Y	Order Effect	N	Y
# Observations	122	122	# Observations	122	122
R^2	0.71	0.71	R^2	0.11	0.11
RMSE	0.64	0.64	RMSE	0.13	0.14
Hypothesis (*p-values*)					
Weak Mean	0.00	0.00	Weak Mean	0.31	0.35
Strong Mean	0.00	0.00	Strong Mean	0.05	0.13
Convex Combination	0.00	0.00	Convex Combination	0.04	0.01

Notes: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. The table presents OLS regressions of estimated present-bias and discounting for groups on estimated minimum and maximum parameters of individuals. Columns 1-2 present show results for present-bias, and Columns 3-4 for the long-run discount factor. In Columns 2 and 4, we control for decision ordering. We use robust standard errors.

1.7 Conclusion

This paper analyzes individual and collective decisions through the preference elicitation method over unpleasant task consumption. The study uses experimental data to analyze task consumption decisions by groups of individuals who have to reach a consensus regarding the allocation of tasks over time. For this purpose, a joint experimental elicitation of time preferences was performed for the groups as well as for their individual members.

The main results of the paper are as follows: First, on aggregate, a present-bias exists in participants' behavior, i.e., the participants' intertemporal allocation decisions exhibited time inconsistency. Second, the degree of present-bias was more pronounced in a group's task allocation decisions as compared to an individual's task allocation setting. Third, the order in which decisions were made, whether making the individual task allocation first and then the group task allocation or vice versa had no effect on the degree of present-bias. Lastly, using within-group estimates of present-bias and discount factor, the variations in the group's individual members' discount rates do explain group present-bias, as postulated by Jackson and Yariv (2015).

We acknowledge that the results could be partly explained by a selection bias. In our experiment, as in any experiment involving longitudinal measures, subjects were supposed to

commit to three sessions over a time span of three weeks. Here, a specificity of our subjects is probably their ability to commit and schedule (Dohmen et al. 2010; Frederick, Loewenstein, and O'donoghue 2002; Perez-Arce 2017). The estimates of present-bias and discount rates for individual choices we found are no higher than those found in the literature, although the empirical literature on task consumption is very limited. Moreover, we were mainly interested in comparisons. It is plausible that the selection bias impacted all decisions to a similar extent, thus we have no big effect on our comparisons. Finally, our coordinating device allowed groups to quickly converge toward a given decision. In this respect, our results have implications for the way households, boards, and committees can achieve consistent decisions.

1.8 Appendix

A1: Nonlinear Least Squares Method

Let there be N experimental subjects and P Convex Time Budget, (CTBs). Assume that each subject j makes her $v_{1t_{ij}}, i = 1, 2, ..., P$, decisions (both individual and group) according to the non-linear Euler equation mentioned above, but that these decisions are made with some mean-zero, potentially correlated error. That is, let

$$fV, R, \mathbf{1}_{d=1}, k, \gamma, \delta, \beta = \left(\frac{\beta^{\mathbf{1}_{d=1}}\delta^k V}{\gamma R^2 + \beta^{\mathbf{1}_{d=1}}\delta^k}\right),$$

then

$$v^*_{1t_{ij}} = fV, R, \mathbf{1}_{d=1}, k, \gamma, \delta, \beta + e_{t_{ij}}.$$

Stacking the P observations for individual j making her individual and group decisions, we have

$$\boldsymbol{v_{1t}}^*_j = \boldsymbol{f}V, R, \mathbf{1}_{d=1}, k, \gamma, \delta, \beta + \boldsymbol{e}_j.$$

The vector $\boldsymbol{e}_j$ is zero in expectation with covariance matrix $\boldsymbol{V_j}$, a $(P \times P)$ matrix, allowing for arbitrary correlation in the errors e_{ij}. We stack over the N experimental subjects to obtain

$$\boldsymbol{v}^*_{\mathbf{1t}} = \boldsymbol{f}V, R, \mathbf{1}_{d=1}, k, \gamma, \delta, \beta + \boldsymbol{e}.$$

We assume that the terms e_{ij} may be correlated within groups (or individuals within the same group) but that the errors are uncorrelated across groups (or individuals within the same group), $E\boldsymbol{e}_j^{'}\boldsymbol{e}_g = 0$ for $j \neq g$. Therefore, $\boldsymbol{e}$ is zero in expectation with covariance matrix $\boldsymbol{\Omega}$, a block diagonal $(NP \times NP)$ matrix of clusters, with groups, covariance matrices, $\boldsymbol{V_j}$. We define the usual criterion function $SV, R, \mathbf{1}_{d=1}, k, \gamma, \delta, \beta$ as the sum of squared residuals,

$$SV, R, \mathbf{1}_{d=1}, k, \gamma, \delta, \beta = \sum_{j=1}^{N}\sum_{i=1}^{P} v_{1t_{ij}^{*}} - fV, R, \mathbf{1}_{d=1}, k, \gamma, \delta, \beta^2,$$

and minimize S using non-linear least squares with standard errors clustered on the group level to obtain $\hat{\beta}$, $\hat{\delta}$, and γ. NLS procedures permitting the estimation of preference parameters at the aggregate or individual level are implemented in many standard econometrics packages (in our case, *Stata*).

Table 1.9: Two-Limit Tobit Regression Analysis

Dependent variable:	Log of Tasks Allocated to the 1st Day			
	(1)	(2)	(3)	(4)
β_1 : Log Task Rate	-0.49***	-0.50***	-0.49***	-0.50***
	(0.08)	(0.08)	(0.08)	(0.08)
β_2 : Immediate Decision (=1)	-0.12**	-0.12**	-0.10	-0.10
	(0.04)	(0.04)	(0.07)	(0.07)
β_3 : Individual Decision First (=1)			-0.02	-0.04
			(0.06)	(0.06)
β_4 : Immediate Individual Decision First (=1)			-0.03	-0.02
			(0.08)	(0.08)
β_5 : *Age in years*		0.02		0.02*
		(0.01)		(0.01)
β_6 : *Gender Female* = 0		0.03		0.02
		(0.05)		(0.05)
β_7 : *Has a On campus Job Yes* = 0		-0.03		-0.02
		(0.11)		(0.11)
β_8 : *Had a Savings Account Yes* = 0		0.06		0.06
		(0.07)		(0.07)
β_9 : *Has a Savings Account Yes* = 0		-0.02		-0.02
		(0.06)		(0.06)
β_0 : Constant	4.96***	4.42***	4.97***	4.35***
	(0.08)	(0.44)	(0.09)	(0.42)
# of Obs	1464	1464	1464	1464
# of Groups	122	122	122	122
F-stats	21.07	7.26	11.06	5.79
Adj R^2	0.02	0.02	0.02	0.02
Hypothesis (*p-values*)				
$H_0 : \beta_2 = 1$	0.00			
$H_0 : \beta_2 = 1$		0.00		
$H_0 : \beta_2 + \beta_4 = 1$			0.00	
$H_0 : \beta_2 + \beta_4 = 1$				0.00

Notes: $^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Standard errors are clustered at the group level. The table presents the estimates of Immediate Decision (=1) and its interactions with other experimentally induced variations using the Two-Limit Tobit Regression technique. Column (1) presents aggregated decisions, and estimates. Column (2) captures the estimates of Immediate Decision (=1) along with demographic variables. Column (3) represents the results of the effect of decision ordering and its interaction with Immediate Decision (=1). Column (4) represents the results of the effect of the decision ordering and its interaction with Immediate Decision (=1), controlling for demographic variables.

A2: Additional Two-Limit Tobit Regression Analysis

Table 1.9 provides robustness results for the non-structural estimation specifications discussed in the paper. Using the intertemporal individual decisions, the results provide the comparison of the estimates, controlling for the demographic variables.

A3: Additional Individual vs. Group Analysis

Table 1.10 shows the association of additional individual characteristics (mentioned in the demographic section) with the group-estimated present-bias parameter. The results signify the fact that beyond discount factor heterogeneity there is no association between group present-bias and differences in groups' individual members' characteristics per se. Table 1.11 presents the robustness test of the point estimates obtained in Table 8. Controlling for

Table 1.10: Additional Individual vs. Group Regression Analysis

Dependent variable:	$\hat{\beta}_G$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\lvert\Delta Age\ in\ years\rvert$	-0.06							-0.06
	(0.14)							(0.14)
$\lvert\Delta Gender\rvert$		-0.10						-0.20
		(0.41)						(0.45)
$\lvert\Delta Outside\ Class\ Study\ Hrs\rvert$			-0.12					-0.12
			(0.08)					(0.07)
$\lvert\Delta On\ Campus\ Job\rvert$				-0.61				-0.82
				(0.37)				(0.51)
$\lvert\Delta Family\ Income\ in\ Log\rvert$					-0.17			-0.18
					(0.21)			(0.24)
$\lvert\Delta Past\ Savings\ Acc.\rvert$						0.22		0.19
						(0.49)		(0.92)
$\lvert\Delta Curr.Savings\ Acc.\rvert$							-0.59	1.05
							(0.37)	(0.76)
Constant	1.08***	1.04***	1.212***	1.06***	1.11***	0.97***	1.14***	1.42***
	(0.19)	(0.18)	(0.23)	(0.12)	(0.18)	(0.10)	(0.17)	(0.38)
# of Groups	122	120	120	120	110	122	122	110
R^2	0.00	0.00	0.01	0.01	0.00	0.00	0.01	0.06

Notes: $^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Standard errors are clustered at the group level. The table presents the estimates of other important demographic variables' differences on groups' present-biased estimated variable.

the other important demographic variables mentioned in the empirical literature, the point estimates of variation in individual member discount factors and *Acquaintance Duration* between them remain the same. F stats also indicate that given there are no variations in individual members' discount rates and in effort cost parameters, the group's allocation decisions would represent the time-consistent pattern even if the individuals are exponential discounters with no *Acquaintance Duration*.

Table 1.11: Individual vs. Group Regression Analysis

Dependent variable:	$\hat{\beta}_G$					
	(1)	(2)	(3)	(4)	(5)	(6)
$\lvert\Delta\hat{\delta}_{IND}\rvert$	-3.84**	-3.47***	-3.76***	-3.86***	-3.26***	-3.53***
	(1.28)	(1.18)	(1.12)	(1.20)	(1.16)	(1.27)
$\lvert\Delta\hat{\gamma}_{IND}\rvert$	0.20	0.24	0.23	0.32	0.23	0.37
	(0.22)	(0.30)	(0.22)	(0.25)	(0.22)	(0.25)
$\hat{\beta}_{min} \approx 1_{IND} = 1$	0.16	0.11			0.16	0.11
	(0.30)	(0.45)			(0.36)	(0.54)
$\hat{\beta}_{max} \approx 1_{IND} = 1$	-0.23	-0.34			-0.06	-0.04
	(0.13)	(0.23)			(0.12)	(0.16)
Acquaintance Duration			0.63**	0.63***	0.64**	0.66**
			(0.27)	(0.23)	(0.27)	(0.27)
$\lvert\Delta\hat{\delta}_{IND}\rvert \times Acq\ Duration$			-6.29**	-6.64**	-6.37*	-6.60**
			(2.70)	(2.89)	(2.75)	(2.94)
Constant	1.23***	1.54***	1.13***	1.29***	1.05***	1.11
	(0.21)	(0.49)	(0.14)	(0.30)	(0.14)	(0.30)
Control for Demographic Variables	No	Yes	No	Yes	No	Yes
# of Groups	122	110	122	110	122	110
R^2	0.04	0.10	0.23	0.28	0.34	0.39
$H_0 : Constant = 1$						
$p\text{-}value =$	0.26	0.27	0.35	0.32	0.71	0.69

Notes: $^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Standard errors are clustered at the group level. Column (1) again represents the formal test of Jackson and Yariv's (2015) main hypothesis. Column (5) represents the robustness of the results obtained in column(1), controlling for differences in demographic variables mentioned in the Table above. Columns (3) and (5) are the same as in Table 7, and columns (4) and (6) represent the corresponding robustness check of the results obtained.

A4: Experiment Protocol

Instructions

Thank you for participating in our experiment. We will begin shortly.

Eligibility:

To be in this study, you need to meet the following criteria:

- You must be willing to participate for three consecutive Fridays. Participation will require your presence on specific days as outlined.
- You will need at least one hour and at max three hours on Friday 13th March, Friday 20th March, and Friday 27th March.

Informed Consent:

Placed in front of you is an informed consent form to protect your rights as a subject. Please read it. If you choose not to participate in the study you are free to leave at this point, deciding to leave later would seriously harm our resources allocated to this study. If you have any questions, we can address those now. We will collect the forms after the main points of the study are discussed.

Anonymity:

Your anonymity in this study is assured. All the information we acquire will be used only for the purpose of communication with you. After the study, your email information will be destroyed and will not be connected to your responses in the experiment.

Venue:

- Venue for Friday 20th March will be the same.
- For Friday 27th March, you do not have to be present physically. You can work from anywhere remotely, given that you have an internet connection.

Rules:

- Please turn off your own cell phones.
- If you have a question at any point, just raise your hand.
- There will be a short survey once we are finished with the instructions.

- During the process of reviewing your answers in your survey, if we find your responses in violation of any of the instructions, you might get removed from the experiment.
- You will receive Rs.500 as a participation fee. Participation means showing up on the first two Fridays.
- If you complete the assigned tasks on all required days of participation as instructed, a completion payment of Rs.1500 will be provided.
- You may receive additional earnings during the experiment if you participate in potential survey games.
- If you choose to end your participation before the completion of the experiment, please report this to study administrators at the mentioned email address.
- All payments will be made on 1st April in IGC office room 161. You will return the phones given to you for experiment purposes to IGC to receive this payment.

Task:

In this study, there is only one task. This task will be completed over time. Some portion of the task may be completed sooner, and some portion of the task can be completed later depending on your choices and chance.

This task will consist of taking a specific number of pictures of books via cell phones. Remember, your phone has a unique IMEI number. Once you take a picture, you need to upload the picture using the application on the phone. Make sure your pictures are clear and the numbers are legible. If the numbers are not legible, they will not be counted. Some portion of the task may be completed on the second Friday, and some portion of the task can be completed on the third Friday. You will practice using the phone application before the actual task starts.

Task Rates:

The allocation decisions across two weeks depend on the task rate. The task rate will vary across your decisions. On the target-setting page of the application (installed the cell phones given to you), every slider bar corresponds to a specific task rate. For example, in the first slider bar, the task rate is 1:0.8, such that every task you allocate to the second Friday reduces the number of tasks allocated to the third Friday by 0.8. For simplicity, the task

rates will always be represented as 1:X, and you will be fully informed of the value of X when making your decisions.

The Experiment Timeline:

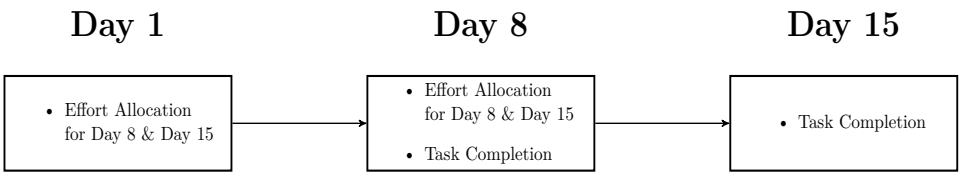


Figure 1.5: Timeline

Notes: Three weeks experimental timeline figure provided to all participants

Before explaining the activities to be done on each Friday you need to have an overall picture of the timeline.

First Friday (13th March 2015):

- First, all of the Subjects will be required to fill out different survey forms.
- After the Survey forms have been completed and collected, you will be asked to make a series of three decisions for task distribution as an individual.
- Once you have made decisions for individual task distribution, again you will be required to make three decisions and distribute the task as a group.
- Keep in mind that your decision today is for the task you will be doing on the 20th and 27th of March. This applies to both Individual and Group decisions.
- In each decision you are free to allocate your tasks as you choose.

Second Friday (20th March 2015):

- During our second session here in the very same venue, again you will be asked to make three Decisions (both individually and as a group) as you did during the first session.
- By this time we will have 12 decisions from every subject (3 Individual + 3 Group on the 13th and 3 Individual + 3 Group on the 20th).
- Exactly one of your 12 total decisions will be implemented randomly.
- We will discuss how this allocation decision is chosen during our training session.

- We refer to this allocation decision as the "decision-that-counts." The tasks you allocated yourself for the 20th and 27th in the decision-that-counts must be completed.
- If you do not complete the tasks according to the decision-that-counts, you will not receive the completion amount of Rs. 1500 and will receive the participation fee of Rs. 500.
- In order for your tasks on the second or third Friday to be counted, they must be completed between 9:30 pm and 10:30 pm of that Friday.
- Surveys will be conducted, which will give you a chance to earn more money.

  Third Friday (27th March 2015):

  - You will have to complete your tasks for this day according to your decision-that-counts.
  - You can do this remotely from anywhere.

  How we will choose the decision-that-counts:

  The process of selecting the decision-that-counts is simple probability. *Three stages to determine the decision-that-counts:*

  1. First, you will be allocated either 13th March Decisions or 20th March Decisions according to a 20% and 80% chance, respectively.
  2. Once you have been allocated to a specific date (13th or 20th March) either you will be given an Individual Task or a Group Task with a 33% and 66% chance, respectively.
  3. After both of the steps given above are complete, you will receive one of the three decisions you made for that specific date and specific task type with equal chance.

  EACH DECISION COULD BE THE DECISION-THAT-COUNTS, SO TREAT EACH DECISION AS IF IT WAS THE ONE DETERMINING YOUR TASKS.

  Short Survey: Please answer the following questions:

  1. How many weeks do we require you to participate?

2. In which of the three weeks are you asked to participate remotely and not come to this venue?

3. What is the percent chance that one of your 20th March allocations will be implemented?

4. If you face a 1:2 task rate for allocations between weeks 2 and 3, every task you allocate to week 2 decreases by how many numbers of tasks you allocate to week 3?

A5: Experiment App

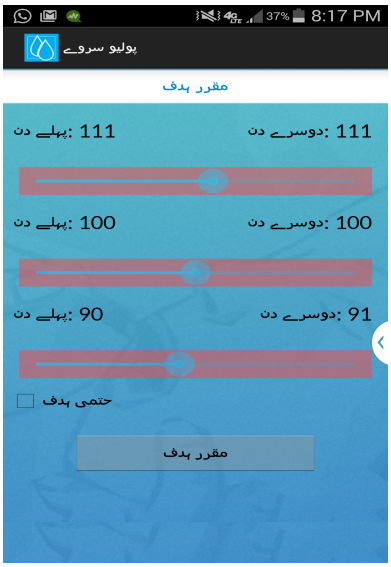


Figure 1.6: Slider Bar Used to Capture Task Allocations

Notes: The above slider bars is for individual task allocation decisions over the two weeks, i.e., $V = 200$. This is the original version is Urdu whose English translation is shown in the main body.

2 — Mobilizing Women Voters: Experimental Evidence from Pakistan

This chapter is joint work with Karrar Hussain. It has been published as: Chaudhry, Zain, Karrar Hussain, and Attique Ur Rehman. "Mobilizing women voters: Experimental evidence from Pakistan." Oxford Economic Papers 75, no. 2 (2023): 444-459.

Abstract: We provide the first estimate of a door-to-door political campaign by an incumbent politician targeting women on electoral outcomes. Women voters are informed of the public service delivery work undertaken by the incumbent in his tenure. The campaign was randomized at the precinct level, allowing us to use official electoral data on vote shares and gender-disaggregated turnout. Our results suggest that in a highly competitive campaign, the vote share of the campaigning incumbent increased by 5 percentage points. This increase was primarily driven by women who were campaigned independently of their male relatives. In precincts where both men and women were mobilized, the effect is not statistically significant. However, women's turnout in the election was unaffected.

2.1 Introduction

In most developing countries, politics is clientelistic in nature (Stokes et al. 2013) and dominated by men (Bjarnegård 2013): male politicians (patrons) deliver patronage in terms of public service work to male household heads, who in turn are expected to ensure that all their household members (men and women) vote for that politician. While a lack of focus on the female electorate's needs is not confined to clientelistic setups, we believe that two specific things happen in such settings that marginalize women: firstly, the female vote is never directly courted by politicians because the male client is expected to deliver the female vote and secondly, due to a lack of communication and interaction with female voters, the public service work demands and preferences that get communicated to the politicians are those of men, not women.[11] These two forces start a cycle of female exclusion from all stages of politics. Politicians do not campaign to seek the votes of women, resulting in a suppressed female turnout, which in turn emboldens the belief that women are not to be seen as a separate voting bloc, further reducing their importance in electoral politics. If politicians were to directly court female voters, it is possible that they would reward them in elections. In this paper, we present the first estimates of a women-focused political campaign using official electoral data. We evaluate a randomly implemented political campaign that provides information about public service to women to understand what happens when a politician directly courts female votes in 151 precincts in Kasur, Pakistan.

Pakistan is a country with a wide gender gap in political participation. Female turnout in the general election of 2018, when we ran this experiment, was 47 percent compared to 58.3 percent for males.[12] Female participation in political events remains scarce, perfunctory and often dependent on men (Rai, Shah, and Ayaz 2007; Cheema, Khan, Mohmand, and Liaqat 2019). This gap is, unsurprisingly, seen at higher levels of participation too, with only 8 out of 272 parliamentarians being directly elected women, of which all but one belong to political dynasties. In such a context, which is

[11] Politicians are generally unaware of what their constituents want even when there is comparatively greater interaction (Liaqat 2019).

[12] Women are systematically disenfranchised by this exclusionary, male-focused political system at the very first step of voter registration as well. In the 2018 elections, there were 46.6 million registered female voters compared to 59.3 million male voters in the country.

equally prevalent in many other developing countries, it is important to understand the constraints on female participation in the electoral process.

Before the 2018 general elections, we partnered with an incumbent politician to evaluate his female-specific, door-to-door campaign run by female canvassers in one constituency of Kasur district. The campaign was aimed at providing information to female voters about the incumbent's record in office, delivered through a pamphlet read out by female canvassers. We argue that a campaign run by women for women, in a context where they are rarely treated as distinct voters from the men of the family, empowers female voters with information to make an informed decision about their vote choice.

Based on the randomization of 151 electoral precincts and using official electoral data, we are able to study if the campaign influenced observable female actions on the election day in terms of voter turnout and party choice. We find that the campaign did not increase female turnout: treatment areas saw a very small increase of 1.7 percentage points over control area turnout of 52.9 percent. This effect is not large enough to be statistically significant. But those who did come to the polling stations voted differently from the status-quo. The treatment increased the vote share of the campaigning incumbent by 3.6 percentage points (p-value = 0.055) in the areas that received a women-specific campaign over and above 40% in the control areas. These results suggest that women-specific campaigns are useful in providing females information to make an independent decision on who to vote for.

To further explore if it is indeed the case that male-oriented status-quo limits female access to information to make independent vote choice, polling stations were randomly assigned to either a women-only campaign or a women and men campaign. In the women and men campaign, women remain the primary target of the campaign but the information is delivered to them alongside a male member of the household. We find that the women-only campaign had the highest effect on the vote share of the incumbent, increasing it by five percentage points over the control areas. The increase in vote share of the incumbent in the women and men treatment areas was only 2.3 percentage points, and not large enough to be statistically significant. However, the difference between the two campaigns is statistically significant and together with the

first set of results, suggests that in order to improve female participation in politics as independent voters, politicians have to run specific women-targeted campaigns.

With these results, the paper relates to a long line of academic studies that are focused on getting citizens out on the election day to vote. Gerber and D. Green (2017) reviews the literature in which, with few exceptions, the cited work is focused on the U.S. and Europe, randomizes interventions at the individual or household level, and conducts non-partisan campaigns. The first seminal paper to provide detailed evidence on a nonpartisan get-out-the-vote campaign focused on women in a developing country was Giné and Mansuri (2018). They randomized the get-out-the-vote campaign based on clusters within villages (not precincts). They find that an importance-of-voting message has no effect, but, this message plus information about the voting process (such as secret balloting) increases female turnout. Recently, however, the focus has shifted to more aggregate level outcomes through precinct-level randomization which allows measurement of outcomes through official electoral data, which improves the quality of data on turnout, in addition to allowing analysis of vote shares. Pons (2018) is the first paper to randomize a door-to-door canvassing campaign at the precinct level (in France). This allows Pons (2018) to use official electoral data to measure turnout and vote shares. He finds no effect of the door-to-door campaign on turnout, but a small immediate and long-term effect on vote shares.

This paper extends all aforementioned strands of literature by providing the first rigorous evaluation of a women-focused door-to-door campaign on turnout and vote shares. Unlike most work, we target and explore the effect of the campaign on female behavior. We run the campaign in a developing country, where comparatively little get-out-the-vote work has taken place. A developing country context adds important insight to the literature because of strong norms against female empowerment and public participation. Further, our treatment is a standard partisan political campaign. This is important because as these are standard campaigns run by male politicians for men before every election, they can easily be held for women as well. Hence, the amount of effort required from politicians would be minimal.[13] If such a campaign shows positive results for the campaigner, it sends a credible signal about the importance of women as

[13]There are no additional costs to recruit women because all political parties already have female members as they are required to work as female polling agents in women-only polling booths.

a distinct bloc that politicians need to court. Finally, as we randomize our treatment at the precinct level, we can use official administrative data on female turnout and politician vote shares, thus measuring actual electoral outcomes and avoiding problems associated with self-reported voting behavior (Campbell 2010, Gelman et al. 2016).

Based on previous research, ex-ante, we expect no effect of campaigning on party choice. Using evidence from 49 rigorous field experiments in the US, Kalla and Broockman (2018) show that the persuasive effects of in-person canvassing on party choice is zero. We can expect this effect to be stronger in a developing country because unlike developed countries (i) voting decisions are often made collectively, which makes change harder (ii) women wield much less power to vote according to their preferences (N. Burns, Schlozman, and Verba 2001; Khan 2021), and (iii) voters trust information less (Algan and Cahuc 2013, Falk et al. 2018).

Furthermore, a politician directly courting the female vote may attract a backlash from male voters (Gottlieb 2016; Guarnieri and Rainer 2018) as male voters may feel threatened that their public service delivery preferences would be comparatively neglected (Chattopadhyay and Duflo 2004; Khan 2021). Gottlieb (2016) shows that a civic education program that brought women into the public sphere led to a male backlash and self-imposed constraints by women to withdraw from the public sphere in the future. In such a context, it is not unreasonable to expect that campaigns targeted at women may not deliver an increase in vote share, and potentially may lead to negative effects. In contrast, Khan (2021) shows that when get-out-the-vote campaigns engage with both men and women the couple discusses the issue within the household, which leads to better combined decisions and higher turnout for women.

There is emerging evidence of direct constraints faced by women for participation in politics. Even in developed economies, for women to be more engaged and successful in politics they have to have support from party superiors (Karpowitz, Monson, and Preece 2017). Whereas in more gender-unequal places like Pakistan, men explicitly control the political decisions of women (Cheema, Khan, Mohmand, Kuraishi, et al. 2019). This paper not only highlights the restraining nature of men's involvement in female political decisions but also provides a silver lining that at least on the margins that are not public, such as vote choice, women do make independent decisions if

provided the information that can help them make such decisions. In this sense, the paper adds to the literature on forces that help women break the restraining barriers (Prillaman 2016).

This paper makes three distinct contributions to the field of gender participation in politics. First, it shows that even traditional political campaigns can take the leap toward courting female voters. Our partnership with an incumbent politician to study an organic effort on courting women voters is a testament to that. Second, women respond to the information provided through such campaigns by making more independent vote choices, even if such efforts are not enough to get more of them to the polling stations. And lastly, it highlights that such a leap will not have any benefits unless the design of the campaign is such that it either directly addresses the norms restricting female participation or helps females circumvent the restraining forces.

In the rest of the paper, we first describe the experiment and lay out the implementation details. This is followed by a section on results and the associated discussion. Lastly, we report heterogeneous effects to tease out who may have responded to the treatment.

2.2 Context

The political campaign was run by Rana Hayat Khan, the incumbent in the constituency NA-140, in Kasur district, which is an hour away from the second-largest city in Pakistan - Lahore. As a parliamentary democracy, one national assembly constituency elects one member to the national assembly in Pakistan. Hence, the politician and performance are held constant as we randomize precincts within this one constituency.

Kasur is part of the most populous and developed province of Pakistan - Punjab. On average, Kasur is not too dissimilar to the rest of the country. In terms of literacy, 59 percent of people in Kasur are considered literate which is almost the same as the national literacy rate of 60 percent. Health indicators present a more diverging picture as the district lags behind the rest of the country on some indicators while it leads on others. Immunization rates for children under the age of two are 60 percent across the country but nearly 86 percent of children in Kasur have received full vaccinations. But it lags behind the country in terms of water infrastructure with only 15 percent of the households in Kasur having access to tap water compared to 27 percent of households

across the country. Hence, economic development in Kasur is similar to the national average.

In terms of political context, the constituency is quite typical for Punjab, Pakistan. While there are around a dozen competitors in the field, the real race is between two big parties: PML-N's Rana Hayat Khan (the incumbent) versus PTI's Sardar Talib Nakai (the challenger). It is divided into 408 precincts, where votes get tallied before the constituency-wide aggregation.

As the policy intervention we evaluate is focused on treating women as equal participants in the political process, it is helpful to explain how women are treated in general vis-a-vis politics and their typical information set. Through surveys around the 2018 election in Lahore, Cheema, Khan, Liaqat, et al. (2023) show that a third of the respondents - both men and women - believe that the very act of discussing politics is solely a man's job. Similarly, in another large city in Pakistan, Khan (2021) found that only about 53 percent of female respondents feel comfortable disclosing their support for a candidate that others in their household did not favor. This puts into focus our policy intervention that directly targets women and experiments with two arms: one in which only women are treated and the other in which they are treated in the presence of men.

Another feature of the policy intervention we evaluate is that it's a door-to-door campaign. This is important because women are mostly not able to participate in political gatherings (Rai, Shah, and Ayaz 2007). Cheema, Khan, Mohmand, and Liaqat (2019) show that women do not participate in political gatherings because their husbands/fathers do not consider it appropriate. The non-acceptance of women in the public sphere is put into focus by the fact that these women specifically mention that if a political gathering was to be covered by the media and the woman could be seen on television, it would ruin their reputation.

2.3 Research Design

From a total of 408 precincts in the constituency, the research team randomly chose 151 precincts. This was done using national census data: we dropped the smallest precincts (less than 250 households) and then used stratified randomization to choose 151 precincts. This sample comprised of five strata based on precinct population. The

incumbent politician implemented the campaign in 103 precincts randomly chosen by the research team, with the remaining 48 precincts serving as a pure control group.

Treatments The treatments focused on canvassing women voters with the help of female canvassers that provided information about the public service delivery performance of the incumbent politician. This was achieved through a door-to-door political campaign for women in 151 precincts, in which non-partisan female canvassers delivered brochures to female household members in treated precincts. We were able to reach 22,426 households across our treatment precincts. The brochures listed the actual public service delivery undertaken by the incumbent over the last five years in the whole constituency (which was verified by the research team). They covered public service delivery such as women-only parks, vocational courses for women, provision of natural gas in homes for cooking, construction of schools, provision of safe drinking water through water filtration plants, stable provision of electricity, and better sewerage systems. While the brochures were written in Urdu, the national language of Pakistan, due to many people not being able to read, the brochures were also verbally explained. For purposes of clarity and to emphasize the focus on women, they featured photos for all these services to make comprehension easier and the photos showed only women engaging in related work. A copy of the brochure, including its English translation, is provided in the appendix in section 2.6.5.

The canvassers randomly visited households in the treatment group to deliver the information. They were sent to a few selected precincts in the morning (along with monitoring teams). They would begin work from the centre of the precinct and move into all directions to ensure that they covered all areas within the precinct. Within these areas, the households were visited randomly. The canvassers visited approximately 22,426 households in the month before the election. The campaign started on 1 July and ended on 19 July. The election was held on 25 July.

The treatment group was divided into two groups: women-only and women and men campaigns. In the women-only campaign, information was provided only to women. Whereas, in the women and men campaign, information was provided to women alongside men. In both cases, the canvassers would introduce themselves as members of the incumbent's campaign team. The respondent had to be a woman who wielded

influence in the household (neither too young, nor too old). A short survey is conducted first (in treatment areas only), and then the brochure is handed over and explained to the respondent. The canvassers and the brochure clearly mention the public service delivery as work that was conducted in the last five years by the incumbent and not future promises. Beyond providing this information to respondents, the canvassers do not engage in discussion with the respondents. As the canvassers themselves were all women, we faced no difficulty in providing information to women alone. For the women and men treatment group, the survey is still conducted only with the women, as they remain the focus of the campaign. However, the brochure is explained to the women alongside a man from the household (household head or someone who wields influence). In the situation that no man is present at home, the woman is asked to share the brochure with him when he gets home (this was rare: one of the father, husband or adult son was nearly always present).

The purpose of differentiating the campaign on whether women receive the information alone or alongside a man was to be able to ascertain whether women felt constrained on future, independent political decision-making when given information in the presence of a male household member. A substantial literature discusses the lack of agency of women in developing country contexts for a range of decisions such as employment, health and politics (Beaman, Pande, and Cirone 2012; Begum and Sen 2005; Jensen 2012; Field et al. 2019; Khan 2021).

Finally, the analysis is performed at the precinct level because our treatment and outcomes are both at the precinct level. This allows us to observe aggregate voter behavior through official electoral data, hence we can measure the effect of the political campaign on voter behavior without facing the difficulties of self-reported surveys (Campbell 2010; Gelman et al. 2016). By randomizing at the precinct level, we can use official electoral data to observe turnout (dis-aggregated by gender) and party/candidate vote share.

2.3.1 Sample Details and Summary Statistics

In May 2018, two months before the general election, polling schemes for 272 national assembly seats were made available to the public by the Election Commission of Pakistan

Table 2.1: Summary Statistics

	# of Obs	Mean	Standard Deviation	Median
Pre Election				
Male voters	107	1853.72	1336.98	1384
Female voters	147	759.26	511.30	678
Households (HH)	151	861.72	700.23	618
Female Polling Stations	151	0.40	0.49	0
City	151	0.33	0.47	0
Coverage intensity	151	0.46	0.29	0.40
Post Election				
PML-N vote share	147	0.43	0.10	0.39
Female turnout	143	0.54	0.09	0.53
Male turnout	90	0.66	0.08	0.67

Notes: Pre & post 2018 election summary statistics.

(ECP). A polling scheme (Section 2.6.6 in the appendix) contains information about each national assembly seat and the associated polling stations, including details such as the number of voters by gender and types of polling stations (female-only, male-only, and combined). The experimental assignment took place after the voter registration period had ended, which ensures that registration is orthogonal to the treatment assignment.

We restrict our sample in two ways: (i) as we aimed to evaluate a women-centric campaign, we focus on female-only and combined (mixed-gender) polling stations only and (ii) we selected on population size (as measured by the Pakistan Census 2017) by choosing precincts that had at least 250 households. The restrictions were necessary due to logistical and time constraints as going to many small far-flung areas would have been very resource-consuming. Additionally, after creating this lower bound, we still had a lot of variance in household numbers across areas, and as size can determine political dynamics to a certain extent, we stratified on size by creating five equal strata.

Our effective sample was lower than the original sample of 151 precincts. We are missing data on female turnout from 8 precincts and vote shares from 4 precincts due to administrative issues faced by the Election Commission of Pakistan (ECP). These data were not reported by the ECP itself. Furthermore, due to a communication error in the field, we treated four precincts that were assigned to the control group. Due to this, in the main body,we present Intent-to-Treat (ITT) estimates, and in the appendix present instrumental variable results as well as ITT estimates when we drop the four precincts from our sample.

Table 2.1 summarizes our sample based on data from the Pakistan Census 2017, ECP polling schemes and official electoral results. There were 198,348 male and 145,471 female registered voters across 98,408 households. The average number of female voters per precinct is 759. The percentage of female-only polling stations is 40 percent. Only 33 percent of the precincts were urban. In the election in 2018, the incumbent politician's party vote share (PML-N) averaged 43 percent. In our sample, 54 percent women and 66 percent men turned out to vote.

To confirm that our randomization procedure worked as expected, we show balance tests for our full sample of 151 precincts in Table 2.2. We compare two treatment and one control group by looking at the number of male voters in a precinct, female voters, number of households, number of female-only polling stations, non-missing 2013 data, rural areas, and vote share of the incumbent in the last election of 2013. We conduct three tests, with the shown p-values corresponding to a joint test of equality and equality tests between the two treatment arms and the control groups. We find that randomization worked well and there are no statistically significant differences between the three groups. We show balance tables for pooled treatments, as well as other modifications, in the Appendix Section 2.6.2.

Table 2.2: Balance Table: Full Sample

	Male Voters (1)	Female Voters (2)	Total Households (3)	Female Polling Stations (4)	Non-Missing 2013 Data (5)	Rural Areas (6)	Vote Share, Incumbent '13 (7)
Women Only	3407.620	2486.639	1837.375	0.501	0.915	0.872	0.541
	(244.429)	(174.176)	(108.971)	(0.102)	(0.047)	(0.073)	(0.039)
Women + Men	3530.685	2562.242	1927.157	0.414	0.955	0.874	0.531
	(296.274)	(219.487)	(147.148)	(0.098)	(0.054)	(0.078)	(0.037)
Control Areas	3261.222	2364.433	1825.121	0.384	0.968	0.847	0.496
	(253.066)	(179.819)	(112.023)	(0.103)	(0.055)	(0.080)	(0.035)
Hypothesis tests p-values							
Joint orthogonality p-value(A=B=C)	0.31	0.27	0.42	0.48	0.68	0.95	0.45
A-C=0	0.33	0.24	0.83	0.24	0.39	0.78	0.25
B-C=0	0.14	0.13	0.20	0.76	0.83	0.77	0.35
# Areas	151	151	151	151	151	151	138

Notes: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. This table shows balance for the full sample of 151 precincts. Uses robust standard errors and block fixed effects.

2.4 Empirical Strategy and Results

2.4.1 Empirical Strategy

We use official data from the Election Commission of Pakistan for the general elections of 2013 and 2018 in NA-140 (Kasur) at the precinct level (one sample is shown in Section 2.6.7). We have electoral data on voter turnout (dis-aggregated by gender) and vote shares of each party (not dis-aggregated by gender). Both randomization and analysis are done at the precinct level. With the randomization built into the experiment, we show below the two main regression specifications that we use to analyze the impact of our treatments on voting behavior.

We start with the pooled specification to analyze the combined effect of both types of the political campaign on female voter turnout and the incumbent and his challenger's vote shares by estimating the following model:

$$Y_i = \alpha + \beta_1 \, Treatment_i + Block\ Fixed\ Effects_i + \epsilon_i, \tag{2.1}$$

where Y_i is the outcome of interest in precinct i. We analyze the effect of the political campaign on female turnout and vote shares for the campaigning incumbent and the challenger. We pool both treatments together as we provided information about the incumbent's public service delivery to women in both treatment groups. Here, $Treatment_i$ is a binary variable for either treatment, which is 1 when any campaign was conducted in the area and 0 where no campaign was conducted.

Then, we estimate the dis-aggregated specification to analyze separately the effect of the women-only and the women and men treatments on female voter turnout and the incumbent and his challenger's vote shares by estimating the following model:

$$Y_i = \alpha + \beta_1 (Treatment)_{1i} + \beta_2 (Treatment)_{2i} + (Block\ Fixed\ Effects)_i + \epsilon_i, \tag{2.2}$$

where Y_i is the outcome of interest (female turnout, incumbent vote share, challenger vote share) for precinct i. Here, $Treatment_{1i}$ is a binary variable for the women-only treatment, which is 1 for precincts where the women-only campaign was implemented

and 0 otherwise. Similarly, $Treatment_{2i}$ is a binary variable for the women and men campaign, which is 1 for precincts where a combined women and men campaign was implemented and 0 otherwise. We include block fixed effects because as explained, we stratified on precinct population size in the design phase.

2.4.2 Results

We show below how the campaigning incumbent's women-centric political campaign affected female turnout and candidate (party) vote shares. All tables below present intention-to-treat estimates because we suffer from a small amount of non-compliance. However, in the appendix, we present results wherein we drop the four non-compliant precincts, as well as an instrumental variable specification, and the results are qualitatively the same.

Table 2.3: Effect on Turnout and Vote Shares (ITT)

	Turnout, Male	Turnout, Female	Vote Share, Incumbent	Vote Share, Challenger
Combined Campaigns	0.000	0.017	0.036*	-0.024
	(0.018)	(0.016)	(0.019)	(0.024)
	[0.017]	[0.018]	[0.023]	[0.026]
Women Only Campaign	-0.011	0.018	0.050**	-0.031
	(0.022)	(0.018)	(0.021)	(0.026)
	[0.023]	[0.019]	[0.021]	[0.025]
Women + Men Campaign	0.010	0.017	0.023	-0.017
	(0.019)	(0.018)	(0.021)	(0.027)
	[0.019]	[0.017]	[0.022]	[0.027]
P-Value (Difference in Treatments)	0.309	0.942	0.097	0.462
Control Mean	0.660	0.529	0.409	0.454
Number of Precincts	90	143	147	147
Block Fixed Effects	X	X	X	X

Notes: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. The dependent variables are male and female turnout and vote shares of the campaigning incumbent and the main challenger - all from official electoral data. The independent variables are the assignment of precincts to women only or women and men treatment (two binary variables in dis-aggregated specification) or the assignment to either of the two campaigns (one binary variable in the pooled specification). We are missing data on female turnout from 8 precincts and vote shares from 4 precincts due to administrative issues faced by the Election Commission of Pakistan, which didn't report these results. Hence, the samples are of 143 and 147 precincts respectively. We report robust standard errors in parentheses and bootstrapped standard errors in brackets. Uses block fixed effects. 'P-Value (Difference in Treatments)' performs a t-test to test whether the two treatments, women only campaign and women and men campaign, are statistically different from each other. We are missing data on female turnout from 8 precincts and vote shares from 4 precincts due to administrative issues faced by the Election Commission of Pakistan, which didn't report these results. Hence, the samples are of 143 and 147 precincts respectively. Uses robust and bootstrapped standard errors and block fixed effects.

In Table 2.3, we show the Intent to Treat (ITT) estimates for the campaign with the pooled specification in the top panel and the dis-aggregated specification in the bottom panel, reporting robust standard errors in parentheses and bootstrapped standard errors in brackets. For female voter turnout, while the point estimates are positive, they are statistically indistinguishable from zero. The estimates are imprecise and we are unable to confirm whether this is a precise zero, because we lack the power to detect such a small effect. We have calculated the minimum detectable effect (MDE) using the level of uncertainty of the point estimates and the MDE is higher than the calculated estimates (the MDE is 0.045). Hence, we are unable to confirm whether an effect exists.

However, we see a statistically significant and economically substantial effect of the campaign on the incumbent's vote share: in areas where the campaign provided information to women alone, the incumbent's vote share increased by a substantial five percentage points. These are very large gains for any political campaign, particularly one focused on women in developing countries. Kalla and Broockman (2018) have shown in the US context that the persuasive effects of in-person canvassing on party choice are zero. While this is not always true, it is a rigorous finding in many settings. Secondly, we also estimate a positive effect of the women and men campaign, however, the point estimate is half the size of the women-only campaign and is not statistically significant. The difference between the two estimates is, however, indeed statistically significant. Hence, for reasons we discuss below, women are reacting much more strongly to the campaign when they are campaigned alone, i.e. without male presence. When information is received alone it clearly leads women to different decisions compared to when information is received in men's presence. Similarly, we find a positive and statistically significant effect for the pooled specification: the campaigns (women-only or women and men) together increase the vote share of the incumbent by 3.7 percentage points.

Additionally, as vote share is not dis-aggregated by gender, the fact that the difference between the women-only and women and men treatments is statistically significant is important as it establishes that women, the primary target of the campaign, were responsible for the increase in the vote share of the incumbent (see Table 2.3). Had the five percentage point increase in vote share been driven by male voters, we would not have seen half the effect for the treatment arm in which men are treated alongside

women. To further test this, we run the same regressions on the female-only polling stations sample (where any change is by design due to women) and find results in the same direction (Appendix Table 2.10).

In any information campaign, spillovers are a relevant concern. However, with the given data, it is not possible for us to measure spillovers. The spillovers could have occurred within and between precincts. Within a precinct, our campaign was visible, and would have led to discussions among people. In aggregate, this means that our treatment reached a larger number of potential voters. The within precinct spillovers would mean that our treatment was more intense than expected. This is reasonable because many precincts, particularly rural, tend to be small and cohesive, where people will often meet and discuss events in the area. These spillovers do not bias our results because we simply estimate the effect of the campaign on a treated region in a binary sense without any differences between intensities.

The spillovers could also have occurred between precincts. Unfortunately, with such a geographically close sample, we cannot estimate the spillover effect between precincts. However, if between precinct spillovers do exist, this would mean that the effects we estimate on the vote share of the campaigning incumbent were a lower bound, and the real effects of the campaign may even be higher.

These results show that a political campaign focused on courting the female vote led to a substantial increase in the vote share for the campaigner, particularly when this information was delivered directly and solely to women. Hence, undertaking public service delivery and campaigning to inform women directly of this work can yield substantial returns for politicians. Based on the prior work (Gottlieb 2016; Guarnieri and Rainer 2018), we may have expected a backlash from men due to a rare political campaign focused primarily on women. However, while we cannot directly detect whether a backlash occurred, it is clear that the incumbent did not lose electorally in any way as his vote share rose significantly. This increase happened in an environment where we would not have expected such results as this is an environment where (i) women are deemed generally to have low agency to make decisions and (ii) they are deemed generally to be less interested in politics. These effects are consistent with

the idea of women being sensitive to campaigns and that they reward politicians who deliver public services and campaign to inform women of them.

The campaign could potentially operate through two main channels. It may have provided information to women that they did not previously possess (or made it more salient before the election). This is an information effect that tells the women about the candidate's public service performance. It may have made women feel elevated to a level more equal to men, or made them see the campaign as a signal that the candidate cares about them - they were for the first time the direct target of a door-to-door campaign. This is an information effect that tells the women about the candidate's esteem for women. Both effects are plausible based on what we know about the context - women are less politically informed compared to men and are not treated as equals in the political sphere (Cheema, Khan, Mohmand, and Liaqat 2019). However, we believe that we can rule out the first effect. In both treatments, the same information is delivered in the same way to women. The only difference is that in the women and men treatment, the information is delivered in the presence of men. Hence, the information set remains constant in both cases. The only reason for it to differ would be if the brochure was taken by the men and women were unable to read it later. However, considering the low level of literacy and the that fact that it's unlikely to be hard to keep the brochure away from the women if they wish to read it, we find it highly unlikely that in this setting the information set differed much. We ensured that in the men and women treatment, it was the women who were handed the brochure and the women to whom the information was given. However, it is reasonable to believe that the presence of men may have led women to see the men and women campaign as not exclusively focused on women and hence a weaker or non-existent signal of the candidate's respect for them.

2.4.3 Channels

To understand better what factors drive the campaign's effect on voter behavior, we analyze how the treatment effects are heterogeneous along social norms. Our hypothesis is that a political campaign that targets women in a context where voting costs are high and the intra-household bargaining power of women is low, a campaign will succeed where women can be more independent. We know that both education and income can

empower women (Field et al. 2019; Duflo 2012). Hence, we look at treatment effect heterogeneity for low female unemployment and high female education. Both these variables allow us to measure different types of female empowerment, which can drive female behavior.

To measure education and income, we use a large dataset from the Benazir Income Support Program (BISP), which is Pakistan's largest cash transfer program. The government runs a census-like survey to collect data on incomes and assets to create a poverty index on which it bases its decision for cash transfer eligibility. However, while not everyone is surveyed because the target is poor, the data collection is so extensive that we estimate that we have data for 90 percent of the population in our sample. We use the individual-level data to create precinct-level dummies, which are one if an area has higher than median education or income and zero otherwise.

Table 2.4: Heterogeneous Effects: Female Unemployment

	Turnout, Female	Vote Share, Incumbent	Vote Share, Challenger
Women Only	0.025	0.023	-0.002
	(0.025)	(0.028)	(0.032)
Low Female Unemployment	-0.015	0.002	-0.027
	(0.047)	(0.052)	(0.067)
Women Only * Low female Unemployment	-0.003	0.119*	-0.158*
	(0.052)	(0.063)	(0.080)
Women + Men	0.041*	0.011	-0.011
	(0.023)	(0.028)	(0.033)
Women + Men * Low Female Unemployment	-0.021	-0.041	0.071
	(0.053)	(0.072)	(0.081)
Control Mean	0.529	0.409	0.454
Number of Precincts	97	98	98
Block Fixed Effects	X	X	X

Notes: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. The dependent variables are female turnout and vote shares (not dis-aggregated by gender) of the campaigning incumbent and the main challenger from official electoral data. The independent variables are assignment of precincts to either treatment (women only or women and men treatments and their interactions with whether a precinct has higher than median unemployment). We are missing data because of difficulties in perfectly matching locations in the Benazir Income Support Program (BISP) and ECP (Election Commission of Pakistan) datasets. Data on female turnout is a further unit smaller because the Election Commission of Pakistan didn't report separate female turnout for a particular unit due to administrative issues. Uses robust standard errors and block fixed effects.

In Table 2.4, we run a disaggregated specification to explore the effect of the two different campaigns and how they interact with female empowerment measured through low female unemployment. We interact the treatment campaigns with a dummy variable

for each area where the dummy is one if the area is over median female unemployment. We find that the precincts in which female employment is higher and precincts that were treated with the women-only campaign had an increase in the campaigning politician's vote share. We find a substantial and statistically significant increase in the vote share of the campaigning politician of 11.9 percentage points. As expected, we see this mirrored in a substantial decline in the vote share of the main (but not sole) opponent.

Table 2.5: Heterogeneous Effects: Female Education

	Turnout, Female	Vote Share, Incumbent	Vote Share, Challenger
Women Only	0.034	0.030	-0.011
	(0.025)	(0.028)	(0.033)
High Female Education	0.048**	0.029	-0.022
	(0.024)	(0.047)	(0.055)
Women Only * High Female Education	-0.086***	0.127**	-0.105*
	(0.032)	(0.054)	(0.061)
Women + Men	0.053**	0.012	-0.005
	(0.023)	(0.029)	(0.034)
Women + Men * High Female Education	-0.133***	-0.021	-0.010
	(0.034)	(0.053)	(0.068)
Control Mean	0.529	0.409	0.454
Number of Precincts	97	98	98
Block Fixed Effects	X	X	X

Notes: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. The dependent variables are female turnout and vote shares (not dis-aggregated by gender) of the campaigning incumbent and the main challenger - all from official electoral data. The independent variables are assignment of precincts to either treatment (women only or women and men treatments and their interactions with whether a precinct has higher than median female education). We are missing data because of difficulties in perfectly matching locations in the Benazir Income Support Program (BISP) and ECP (Election Commission of Pakistan) datasets. Data on female turnout is a further unit smaller because the Election Commission of Pakistan didn't report separate female turnout for a particular unit due to administrative issues. Uses robust standard errors and block fixed effects.

In Table 2.5, we run a disaggregated specification to explore the effect of the two different campaigns and how they interact with female empowerment measured through high female education. As in the previous regression, we interact the treatment campaigns with a dummy variable for each area where the dummy is one if the area is over the median female education. We find that the precincts in which female education is higher and precincts that were treated with the women-only campaign, had an increase in the campaigning politician's vote share. We find a substantial and statistically significant increase in the vote share of the campaigning politician of 12.7 percentage points.

As mentioned above, we believe that this effect is driven by social norms. The effect of our treatment depended substantially on the social norms of the areas. In areas where women are more empowered either through being employed and being earners or through being more educated, our treatment has substantial returns. An alternative explanation would be that it is not social norms themselves that drive the increase in vote share, but rather, income and education directly empower women, which drives the increase in vote share. However, as we can see, in areas where female employment and female education are higher, we find null effects. It is also important to note the limitations of these freedoms. In both results, we find that it is the women-only treatment that drives an increase in vote share. Hence, when women are informed alongside men, it is likely that either through intra-household bargaining over politician choices, or more coercive means, the male members of the household can shape the behavior of the female members.

2.5 Conclusion

In this paper, we provide evidence on female responsiveness to information campaigns. The paper extends the literature on female voting participation with the aim of reducing the gender gap in voting on election day. Existing efforts have focused on non-partisan get-out-the-vote campaigns and civic education. While these are important avenues of increasing female turnout, the literature has a gap on what could be done within the organic political campaigns of parties and candidates to achieve the goal of equality in politics. In this paper, we make a contribution towards filling that one particular gap.

Women have different policy preferences than men in many developing countries. However, the conventional assumption is that they vote according to the preferences of the men of their families. This understanding latently assumes that women do so with the full knowledge of what they want and what their representatives provide. In the paper, we provide evidence that women may not be completely informed about their representatives. When they are provided information about the women-specific policy actions and development projects of the incumbent, female voters vote differently.

We consider this to be an important finding, especially from a policy perspective. It will help address at least one of the structural reasons behind the gender gap. The

politicians and parties follow the conventional wisdom that women are not independent voters and that households act as a unitary agent when deciding whether to vote or not. However, that appears not to be the case. Women do appear to make an independent decision about voting and it is not just the decision to take the costly action of traveling to a polling station, but also how they vote. If the incumbent takes actions that are generally in line with issues the women care about then highlighting them in an election campaign helps the candidate.

We believe this paper also provides an important avenue for future research. At present, one limitation of the paper is that it cannot separate the effect of campaigning from a campaign that provides information. It also provides suggestive evidence that norms matter since the effect is bigger in areas where men and women can vote in the same polling station. In all, this paper extends the important literature on bridging the gender gap in politics and also opens up avenues for potential research.

2.6 Appendix

2.6.1 Ethics, Fieldwork and Partnership

The incumbent politician's son, Rana Faisal Hayat, was invited to an event at the Lahore University of Management Sciences in Lahore, where a member of the research team met him. During the conversation, the incumbent's son let the researcher know that he was planning to run a female-focused campaign. The researcher offered to train and monitor the campaign in return for the research opportunity. This openness towards a female campaign makes the incumbent's political team different from the average politician, even though a few other politicians ran such campaigns as well. Nevertheless, in terms of public service delivery, campaigning style and female representation, he was no different from the average politician. All politicians, including our partner, focus on men by holding male-only rallies and visiting the homes of important voters to court their vote (where women are nowhere to be seen). This is reflected in public service delivery - the focus across Punjab is on the priorities of men. The park built by the incumbent is not an exceptional act and a few dozen other politicians have done so too. However, this does not change the fundamental nature of the political relationship

between the politicians and their voters: it is a male-focused relationship. This is why we believe that the results we see cannot be solely due to the incumbent's image as a pro-women politician.

A typical day in the field was as follows: the survey team, having chosen areas to campaign the day before (from the random list that the research team had created), would leave in the morning together and then work as one to four teams, in one to four areas, depending on the size of the area to be covered and the closeness of other chosen areas. The female campaigners would start work from the center of a precinct and then spread around, conducting their work, always under the supervision of one senior member of the survey team. One co-author was present permanently in the field so as to visit precincts randomly to ensure high work quality. However, due to the fact that we were managing a campaign so close to elections and one specifically for women, the campaign at times became contentious, with local leaders from opposing parties taking umbrage at such a campaign. Hence, the circumstances that the work was conducted eventually became a bit challenging in the strongholds of the opposition. Nevertheless, the work continued without any substantial problems. The campaign began on 01 July 2018 and ended on 19 July 2018. The elections were held on 25 July 2018.

The research team had sought permission from the district government to allow the campaign to be run, which is a constitutional right of every citizen. To ensure that any difficulties were handled promptly and maturely, the survey firm had senior staff on the ground and two co-authors were heavily involved in the work as well, with one permanently in the field.

The political campaign was conducted by the politician himself. The research team spent no research funds on the campaign including printing brochures, salaries of the field teams, and transport for the canvassers. The role of the research team was to train the canvassers to pick households randomly, deliver the message to the right person and deliver the same message precisely, and to monitor their performance to ensure compliance with all research protocols. The female-focused campaign would have been run irrespective of our involvement. However, the research team's involvement meant that different areas got treated compared to the counterfactual. The research team made choices randomly, while the incumbent would have selected them based

on potential electoral gain. Further, the involvement of the research team meant that instead of just one, two treatments were implemented: women-only and women and men campaigns.

2.6.2 Additional Balance Tables

We first show the balance table for the sample of 151 precincts for the pooled specification, i.e., both treatment arms put together and compared to the control group. As can be seen, we are well balanced on all variables.

Table 2.6: Balance Table (Pooled Treatments): Full Sample

	Male Voters (1)	Female Voters (2)	Total Households (3)	Female Polling Stations (4)	Missing 2013 Data (5)	Rural Areas (6)	Vote Share, Rana (13) (7)
Treatment Areas	3471.013	2525.583	1883.624	0.456	0.935	0.873	0.535
	(259.026)	(188.617)	(125.420)	(0.088)	(0.039)	(0.061)	(0.032)
Control Areas	3262.473	2365.201	1826.034	0.383	0.968	0.847	0.496
	(253.145)	(179.924)	(112.684)	(0.103)	(0.055)	(0.080)	(0.035)
Joint Orthogonality P-Value	0.14	0.11	0.32	0.39	0.53	0.74	0.21
# Areas	151	151	151	151	151	151	138

Notes: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. This table shows the balance for the full sample of 151 precincts, where the two treatments (women only and women and men treatments have been pooled together). Uses robust standard errors and block fixed effects.

Furthermore, as explained in the paper, we were not able to measure our outcomes of interest for the whole 151 precinct sample. However, for both the sub-sample of 147 areas for which vote share data are available and the sub-sample of 143 for which female turnout data are available, we find qualitatively the same results: we are well-balanced across all arms. Table 2.7 shows the balance tests for the sub-sample of 147 precincts, while Table 2.8 in the Appendix shows the same for the sub-sample of 143 precincts.

Table 2.7: Balance Table: Available Sample

	Male Voters (1)	Female Voters (2)	Total Households (3)	Female Polling Stations (4)	Missing 2013 Data (5)	Rural Areas (6)	Vote Share, Rana (13) (7)
Women Only	3421.474	2492.618	1839.243	0.505	0.949	0.877	0.552
	(250.107)	(178.363)	(111.545)	(0.103)	(0.040)	(0.074)	(0.039)
Women + Men	3528.325	2559.273	1938.663	0.430	0.977	0.869	0.538
	(309.280)	(229.205)	(153.032)	(0.101)	(0.038)	(0.081)	(0.037)
Control Areas	3248.284	2357.775	1830.427	0.398	0.991	0.837	0.505
	(258.638)	(184.040)	(114.834)	(0.104)	(0.051)	(0.081)	(0.035)
Hypothesis tests p-values							
Joint orthogonality p-value(A=B=C)	0.28	0.26	0.39	0.56	0.76	0.90	0.44
A-C=0	0.26	0.21	0.88	0.30	0.47	0.66	0.23
B-C=0	0.13	0.14	0.19	0.75	0.79	0.74	0.37
# Areas	147	147	147	147	147	147	135

Notes: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. This table shows the balance for the sample of 147 precincts for which vote share data is available. The missing data is due to administrative issues faced by the Election Commission of Pakistan. Uses robust standard errors and block fixed effects.

Table 2.8: Balance Table (Pooled Treatments): Available Sample

	Male Voters (1)	Female Voters (2)	Total Households (3)	Female Polling Stations (4)	Missing 2013 Data (5)	Rural Areas (6)	Vote Share, Rana (13) (7)
Women Only	3421.474	2492.618	1839.243	0.505	0.949	0.877	0.552
	(250.107)	(178.363)	(111.545)	(0.103)	(0.040)	(0.074)	(0.039)
Women + Men	3528.325	2559.273	1938.663	0.430	0.977	0.869	0.538
	(309.280)	(229.205)	(153.032)	(0.101)	(0.038)	(0.081)	(0.037)
Control Areas	3248.284	2357.775	1830.427	0.398	0.991	0.837	0.505
	(258.638)	(184.040)	(114.834)	(0.104)	(0.051)	(0.081)	(0.035)
Hypothesis tests p-values							
Joint orthogonality p-value(A=B=C)	0.28	0.26	0.39	0.56	0.76	0.90	0.44
A-C=0	0.26	0.21	0.88	0.30	0.47	0.66	0.23
B-C=0	0.13	0.14	0.19	0.75	0.79	0.74	0.37
# Areas	147	147	147	147	147	147	135

Notes: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. This table shows balance for the sample of 147 precincts for which vote share data is available. The missing data is due to administrative issues faced by the Election Commission of Pakistan. Uses robust standard errors and block fixed effects.

2.6.3 Turnout and Vote Share without Non-Compliant Precincts

In the table below, we drop the five precincts that were assigned to control but got treated in the field due to a communication error. The results are qualitatively similar to the ITT results shown above.

Table 2.9: Effect on Turnout and Vote Shares without Non-Compliers

	Turnout, Female	Vote Share, Incumbent	Vote Share, Challengers
Women Only Campaign	0.013	0.048**	-0.027
	(0.019)	(0.023)	(0.026)
	[0.018]	[0.019]	[0.025]
Women + Men Campaign	0.011	0.022	-0.012
	(0.019)	(0.022)	(0.027)
	[0.019]	[0.020]	[0.023]
Control Mean	0.529	0.409	0.454
Number of Precincts	139	142	142
Block Fixed Effects	X	X	X

Notes: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. The dependent variables are female turnout and vote shares of the campaigning incumbent and the main challenger - all from official electoral data. The independent variables are assignment of precincts to women only or women and men treatment (binary variables in dis-aggregated specification) or assignment to either of the two campaigns (binary variable in pooled specification). We are missing data on female turnout from 13 precincts and vote shares from 9 precincts mainly due to administrative issues faced by the Election Commission of Pakistan, which didn't report these results, and due to miscommunication within the field team. Hence, the samples are of 138 and 142 precincts respectively. Uses robust and bootstrapped standard errors and block fixed effects.

2.6.4 Turnout and Vote Share in Female-Only Polling Stations

To further test whether our finding that the increase in the incumbent's vote share is driven by women, we test what the effect of the campaign was in female-only polling stations. We find a point estimate of 2.7 percentage points for the women-only campaign. The effect is negative for the women and men campaign as expected because the increase in vote share for the incumbent is driven by the combined polling stations (as shown above).

Table 2.10: Effect on Turnout and Vote Shares: Female-Only Polling Stations

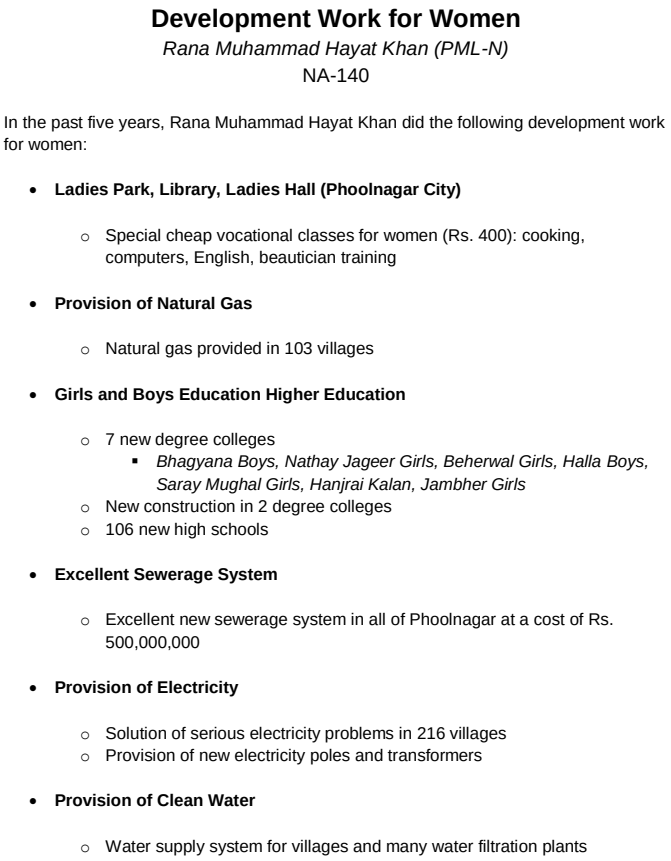
	Turnout, Female	Vote Share, Incumbent	Vote Share, Challengers
Women Only Campaign	-0.010	0.027	-0.000
	(0.022)	(0.019)	(0.024)
Women + Men Campaign	-0.022	-0.040*	0.044
	(0.022)	(0.022)	(0.033)
Control Mean	0.529	0.409	0.454
Number of Precincts	56	60	60
Block Fixed Effects	X	X	X

Notes: $^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. The dependent variables are female turnout and vote shares (not dis-aggregated by gender) of the campaigning incumbent and the main challenger - all from official electoral data. The sample is limited to female-only polling stations, so that the estimates reflect a change in women's behavior only. The independent variables are assignment of precincts to either treatment (women only or women and men treatments. The sampler is smaller as we have excluded all combined polling station precincts. Uses robust standard errors and block fixed effects.

2.6.5 Brochure

The brochure was designed by the politician's campaign team that was delivered to women in the door-to-door campaign. All information is correct. A translation in English is also provided.

Figure 2.1: Brochure used in door-to-door campaigning



Development Work for Women

Rana Muhammad Hayat Khan (PML-N)

NA-140

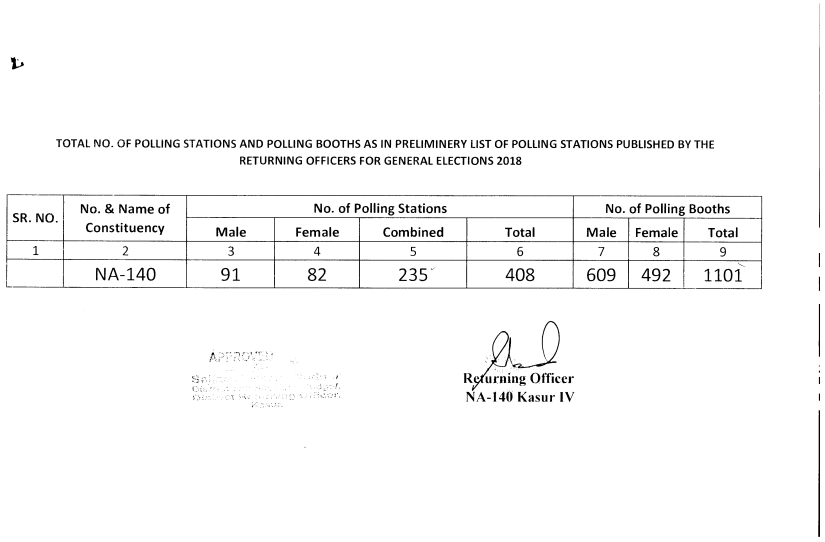
In the past five years, Rana Muhammad Hayat Khan did the following development work for women:

- **Ladies Park, Library, Ladies Hall (Phoolnagar City)**
  - Special cheap vocational classes for women (Rs. 400): cooking, computers, English, beautician training
- **Provision of Natural Gas**
  - Natural gas provided in 103 villages
- **Girls and Boys Education Higher Education**
  - 7 new degree colleges
    - *Bhagyana Boys, Nathay Jageer Girls, Beherwal Girls, Halla Boys, Saray Mughal Girls, Hanjrai Kalan, Jambher Girls*
  - New construction in 2 degree colleges
  - 106 new high schools
- **Excellent Sewerage System**
  - Excellent new sewerage system in all of Phoolnagar at a cost of Rs. 500,000,000
- **Provision of Electricity**
  - Solution of serious electricity problems in 216 villages
  - Provision of new electricity poles and transformers
- **Provision of Clean Water**
  - Water supply system for villages and many water filtration plants

Figure 2.2: Brochure used in door-to-door campaigning: Translation

2.6.6 Polling Scheme: NA-140

This is the first page of the polling scheme for NA-140, which lists which area is served by which polling station and whether they are combined or gender-segregated polling stations.



TOTAL NO. OF POLLING STATIONS AND POLLING BOOTHS AS IN PRELIMINERY LIST OF POLLING STATIONS PUBLISHED BY THE RETURNING OFFICERS FOR GENERAL ELECTIONS 2018

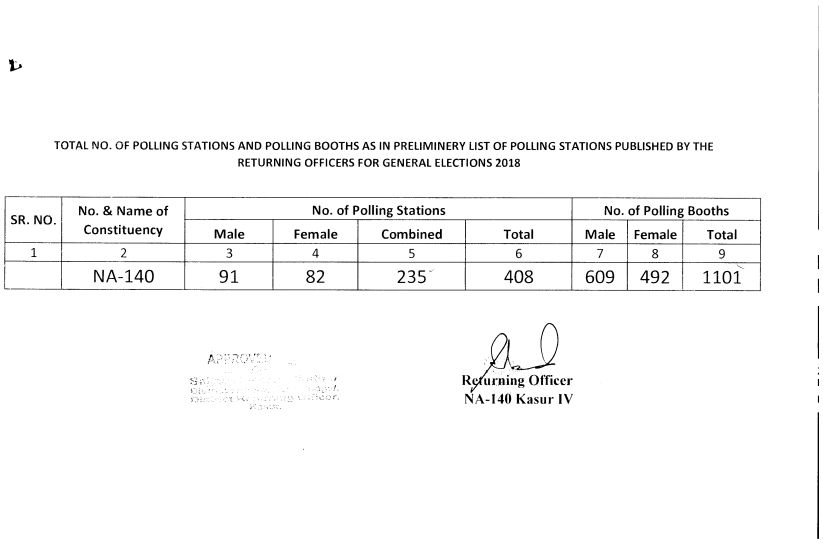
SR. NO.	No. & Name of Constituency	No. of Polling Stations				No. of Polling Booths		
		Male	Female	Combined	Total	Male	Female	Total
1	2	3	4	5	6	7	8	9
	NA-140	91	82	235	408	609	492	1101

Returning Officer
NA-140 Kasur IV

Figure 2.3: Polling Scheme (Main Page): NA-140

2.6.7 Form-45: NA-140

This is the form filled out by the officials of the Election Commission of Pakistan for each precinct, listing the names of all the contesting candidates and the number of valid votes polled for each candidate. Furthermore, it separates this total polled vote number by gender. This forms the basis on the precinct-level data that we use in our analysis.



TOTAL NO. OF POLLING STATIONS AND POLLING BOOTHS AS IN PRELIMINERY LIST OF POLLING STATIONS PUBLISHED BY THE RETURNING OFFICERS FOR GENERAL ELECTIONS 2018

SR. NO.	No. & Name of Constituency	No. of Polling Stations				No. of Polling Booths		
		Male	Female	Combined	Total	Male	Female	Total
1	2	3	4	5	6	7	8	9
	NA-140	91	82	235	408	609	492	1101

Returning Officer
NA-140 Kasur IV

Figure 2.4: Form 45 (One Polling Station): NA-140

3 The Economic Effects of Inter-Sectarian Contact

This paper is single-authored. It has not yet been published.

Abstract: I conduct a field experiment to analyze the effect of contact and leadership in reducing prejudice between discordant sectarian groups in Pakistan. We randomly assign 32 mosques with 423 worshipers to four groups in which (i) we send volunteer worshipers to mosques of the opposite sect to pray, or (ii) have the head of the mosque make an announcement in favor of inter-sectarian harmony, or (iii) both volunteer visits and announcements, or (iv) a pure control group. We find that the combined treatment reduced prejudiced beliefs and in an incentivized experiment makes respondents choose the services of a worker from the opposite sect.

3.1 Introduction

The Shia-Sunni sectarian divide within Islam has recently become of great importance economically and politically, not just in the Middle East but globally. At a geo-political level, this has led to competition between states such as Saudi Arabia and Iran with proxy effects in Pakistan, Lebanon, Yemen, Syria, and elsewhere. Further, the existence of distrust between two important societal groups creates the potential for conflict in every country with substantial Shia and Sunni populations. In Pakistan, the country of our focus, this has led to terrorism that has killed thousands over the past few decades. Even without terrorism, a lack of trust between societal groups can lead to lower public good provision (Alesina, Baqir, and Easterly 1999) and lower productivity (Hjort 2014). I study whether exogenously generated contact inside a mosque and support for harmony from religious leaders induce changes in beliefs and economic interactions.

Today, substantial Shias and Sunnis hold incorrect beliefs about the religiosity of the other group. One such misconception is that the other group does not pray - formal daily prayers are a fundamental religious ritual for both sects. It is important to note that Shias and Sunnis do not differ on any other relevant dimension such as ethnicity.

Pakistan, where I conduct our research, is particularly relevant because the country has the world's largest share of Muslims after Indonesia and the world's largest share of Shias after Iran. In addition, the Shia-Sunni intersectarian relationship has been difficult since the 1980s when both Saudi Arabia and Iran waged a proxy war against each other in Pakistan. While the violence has died down at times, the underlying beliefs and preferences that led to the violence continue to exist. Less than four percent of Sunnis express some level of agreement with statements about openness to sectarian intermarriage and support to the other side in the event that one's community initiates sectarian violence (Kalin and Siddiqui 2014).

I explore the impact of contact and leadership motivated by the "contact hypothesis" (Allport, Clark, and Pettigrew 1954) about how certain types of interaction between groups can reduce prejudice, with the effects dependent on having common goals, equal

status in a particular situation, intergroup cooperation and the support of authorities, law or custom.

I conduct the experiment in Haripur district, Pakistan - an area in North West Pakistan, which has been affected substantially by the War on Terror since 2002, including a wave of terrorism against the Shias. I randomized 32 mosques across 10 different towns and villages - 23 mosques belonged to the majoritarian Sunni sect and nine to the minority Shia sect. I surveyed 423 worshipers frequenting these mosques.

I have three treatments and one pure control group. The first treatment group consists of sending trained volunteer worshipers to mosques of the other sect, thus exposing worshipers of the mosque to worshipers of the opposite sect. This happened once every day for twelve days. This is a very natural setting - there is no forced interaction. The presence and actions of the minority group are visible to everyone without any effort because the number of worshipers during daily prayers is quite low (on average 13.7). Further, the visibility is enhanced by the fact that it occurs within a religious, sacred place and conveys information about the religiosity of the minority group. The religiosity of the minority group is easily visible because both sects pray differently in visibly different ways (the act of praying involves physical movement). I carefully trained the volunteer worshipers so that they would not engage in any additional activity. In contrast to an information treatment, this is a more believable and concrete demonstration of the minority group's religiosity.

In our second treatment group, the leader of the mosque, the Imam, delivers a message of inter-sectarian harmony shortly before the commencement of prayers. This announcement includes a simple verse from the holy book of Muslims, both Sunnis, and Shias, the Quran: "Hold fast together to the cable of Allah and be not divided." This is a famous verse from the Quran that everyone knows well and focuses on unity and firmness in belief in Allah among Muslims and not being divided into groups. In our third treatment group, I combine our two treatments.

Our first set of findings considers our real world economic activity experiment in which I offer discounted vouchers for plumbing services from two plumbers whose names are clearly Shia or Sunni. I show how our treatments affect whether our respondents choose plumbers of the opposite sect (a binary variable). I find economically and

statistically significant effects for our combined announcement and volunteer treatment which increases demand for a plumber of the opposite sect by 0.18 points against a mean control group demand of 0.153. I show that this behavior is driven by the majoritarian sect, Sunnis, as I observe a substantial reduction in demand for Sunni plumbers. It is important to note that Shias and Sunnis do not differ on any other metric including competence for plumbing.

Using our survey measures, I also show how contact affects potential hiring decisions. I find that our respondents are more open to hiring members of the opposite sect when I analyze how their endline beliefs differ compared to their baseline beliefs. Our combined treatment group increases willingness to hire a member of the opposite sect by 0.240 points.

While Allport, Clark, and Pettigrew (1954) led to a huge empirical literature by social psychologists, I still lack a deep understanding of what makes contact useful and under which conditions it lowers prejudice (Paluck, S. Green, and D. Green 2019). Paluck, S. Green, and D. Green (2019)'s reviews 418 experiments on the contact hypothesis and finds that only 27 studies randomized contact, of which only six focused on adults over twenty-five years of age. Only one study took place in a developing country and only one measured outcomes using experiments. All the remaining studies use self-reported surveys, or different types of tracking on prejudiced beliefs and actions, not economic activity. Finally, most of these studies look at racial or ethnic prejudice, which is the predominant concern in the country where these studies took place (USA).

However, more recently, economists and political scientists have furthered our understanding of the contact hypothesis substantially. Lowe (2021) explores inter-caste contact in India by randomizing cricket teams and shows that cooperative, not adversarial contact leads to a short-term reduction in prejudice against out-group members. Corno, La Ferrara, and J. Burns (2022) exploit random allocation of interracial roommates in a South African university and show that White students' negative stereotypes towards Black students go down and Black students improve their GPA and drop out at lower rates. Scacco and Warren (2018) randomize contact between Christians and Muslims and offers educational training for sixteen weeks but find no changes in

prejudice, though mixed-class subjects discriminate significantly less against out-group members than subjects in homogeneous class groups.

This work has substantially advanced our understanding of the contact hypothesis. However, I believe our work advances the literature in three important ways. First, our project randomizes contact among adults of all ages and socio-economic backgrounds in a developing country. Much other work focuses on smaller sub-samples of the populace and is often undertaken in developed countries. Second, I explore intra-religious differences with substantial cross-country geo-political considerations, which have never been explored before. This is an important element not only due to its economic and political importance, but also because while intra-religious differences are in many ways similar to issues of race and caste, they nevertheless provide a commonality between the two groups. Third, I measure outcomes using surveys but also incentivized lab-in-the-field experiments and a real-world economic activity. Fourth, a lot of the previous work does not conduct baseline data collection, which is needed to understand how the contact created by the researchers interacts on top of daily societal interaction.

The rest of the paper proceeds as follows. Section 3.2 explains the context of the inter-sectarian clashes. Section 3.3 presents the details of the experimental design and the data I collected. Section 3.4 describes our econometric specification and the main regression and heterogeneity analysis. Section 3.5 concludes.

3.2 Context: Inter-Sectarian Relationships

The Shia-Sunni division began just a few decades after the advent of Islam, with theological differences becoming larger over time. In modern times, the relationship between Shias and Sunnis has worsened substantially due to proxy wars fought between Saudi Arabia and Iran, which led both countries to support extremist elements abroad through charitable funding and ideological propaganda, which eventually led to a wave of terrorism in countries such as Pakistan. The propaganda created two groups that hold at least very exclusionary beliefs and preferences, sometimes not even seeing the other group as being Muslim, as well as more extreme beliefs that support killing each other. The support for more extreme actions is partially driven by incorrect beliefs propagated by Saudi Arabia and Iran: that the other sect does not follow Islamic

religious rituals and thus are either bad Muslims who should be looked down upon or not Muslim at all. As the offering of daily prayers is one of the most important Islamic rituals (for both sects), incorrect beliefs exist on this question too. It is important to note that while both sects' daily prayers have the same content (which is read in silence), the precise physical movements differ, from which one can easily infer the sect of the worshiper. Our experiment exploits the existence of this incorrect belief about the religiosity of the other group, the importance of prayer to both sects and the observable nature of the difference in prayer movements to create contact between the two groups and provide information about the religiosity of the other sect. It is important to note that Shias and Sunnis do not differ on any other important and relevant characteristic such as ethnicity.

3.3 Experimental Design

I explain below our sample, treatments, and data collection exercise that took place between May 2022 and February 2023.

3.3.1 Sample

I conduct the experiment in Haripur district, Khyber Pakhtunkhawa, Pakistan. Haripur is one of the 34 districts in the third largest province of Pakistan - Kyber Pakhtunkhawa. KP is in North West Pakistan along the Afghan border. Due to its proximity to Afghanistan, it was the most affected province in the War on Terror, with the Shia a particular target of terrorist organizations. In Pakistan, during the War on Terror from 2002-2013, 2971 civilians were killed with 5517 civilian casualties and specifically 1956 Shias killed and 3781 injured. There were 432 targeted attacks on the Shia and their religious places. This unfortunate violent history makes KP particularly relevant to our question of interest.

I select 32 mosques for our experiment from ten different towns and villages in Haripur, of which 23 mosques belong to the majoritarian Sunni sect and 9 to the minority Shia sect. I surveyed 423 regular worshipers at baseline and endline.[14]

[14]Our baseline sample size is 457, however, I were not able to conduct endline data collection with 30 respondents. I compare the attrited group to the non-attrited group and show that the two groups do not differ in any visible characteristic.

Below, I show a balance table that confirms that I randomized correctly and that I have similar respondents across all our treatment groups.

Table 3.1: Balance Table

	Age	Marital Status	Income	Employment	Wife, Same Sect	Business, Opp Sect	Trust, Opp Sect
Announcement Only (A)	37.111	1.757	1.966	1.906	0.936	3.891	3.870
	(4.474)	(0.097)	(0.287)	(0.314)	(0.049)	(0.164)	(0.229)
Prayer Volunteers Only (B)	38.602	1.849	2.058	2.260	0.925	3.919	3.918
	(3.923)	(0.060)	(0.243)	(0.326)	(0.050)	(0.133)	(0.199)
Announcement and Volunteers Both (C)	39.593	1.863	2.310	2.374	0.915	3.974	3.999
	(3.640)	(0.063)	(0.304)	(0.351)	(0.056)	(0.160)	(0.184)
Control (D)	38.628	1.772	2.385	2.177	0.959	4.012	4.083
	(2.680)	(0.037)	(0.157)	(0.245)	(0.039)	(0.107)	(0.131)
Hypothesis tests p-values							
Joint orthogonality p-value(A=B=C=D)	0.86	0.33	0.20	0.38	0.58	0.58	0.51
A-D=0	0.67	0.87	0.10	0.16	0.43	0.34	0.26
B-D=0	0.99	0.15	0.14	0.67	0.41	0.30	0.21
C-D=0	0.67	0.10	0.79	0.41	0.30	0.76	0.47
Number of Regular Worshipers	427	428	423	427	327	428	427

Notes: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. This table shows the balance for the full sample of 151 precincts. Uses robust standard errors and block fixed effects.

3.3.2 Treatment

I have three treatment arms and one pure control group. Our first treatment arm consists of sending trained volunteer worshipers to mosques of the other sect. Thus the volunteers expose the local worshipers to the other sect simply through their presence. They are trained to not initiate any conversation, or engage in any other activity, other than acting like any normal worshiper (i.e., pray in the mosque and leave). I send two volunteers every day over a twelve-day period during the second last prayer of the day (the most frequented prayer out of the five daily prayers as it is around sunset and thus right after work) to every mosque. The volunteers' presence is thus visible because (i) except for congregational prayers on Friday, the number of worshipers in mosques is small (an average of 13.7 in our sample) and (ii) there are clear, known, visible differences in how both sects pray.

This interaction inside a mosque is very natural. There is no legal, moral, or religious reason for the sects not to pray in the same mosque and nobody stopped our volunteers from praying in these mosques. There is no forced interaction between the sects. The information a visit of the opposite sect provides to the usual worshipers is subtle - it is different from direct information and occurs inside a religious, sacred place.

In our second treatment group, the leader of the mosque, the Imam, delivers a message of inter-sectarian harmony shortly before the commencement of prayers. This announcement includes a simple verse from the holy book of Muslims, both Sunnis, and Shias, the Quran: "Hold fast together to the cable of Allah and be not divided." This is a famous verse from the Quran that everyone knows well and focuses on unity and firmness in belief in Allah among Muslims and not being divided into groups.

In our third treatment group, I combined our first and second treatments. Our final fourth group is a pure control group.

3.3.3 Data Collection

First, I measured the number of mosque-goers who frequently come to the mosque to offer the Maghrib prayer (second-last prayer of the day). Then, based on these numbers, I invited those mosque-goers to be part of our survey.

I collected data at baseline and endline with all 423 worshipers in each mosque. Through surveys, I measured the worshipers' demographics, religiosity, beliefs, and preferences about their own and the opposite sect. I also had enumerators collect visible information about religiosity from clothes and accessories worn by the worshipers.

I conduct two incentivized experiments to measure the respondents' beliefs and preferences for the opposite sect and their economic interaction with them.

First, at baseline and endline, I offer every respondent a voucher to buy one of four (heavily discounted) books. I provided discounts of 80 percent on the purchase of a book about the opposite sect and 20 percent for a book on one's own sect. I show the voucher below (Figure 3.1). The top row in both conditions shows Sunni books with the discount percentage (retail prices are PKR 80 and PKR 180 respectively). The bottom row in both conditions shows Shia books with the discount percentage (retail prices are PKR 120 and PKR 135 respectively). The voucher is worth PKR 100. I carefully selected these four books in consultation with a religious scholar who is an authority on the subject. I selected two books each from each sect about daily ritual prayers and their narration of early Islamic history.

At endline, I also offer every respondent a voucher worth PKR 1,000 for plumbing services. The voucher offers the services of two plumbers who are named on the voucher

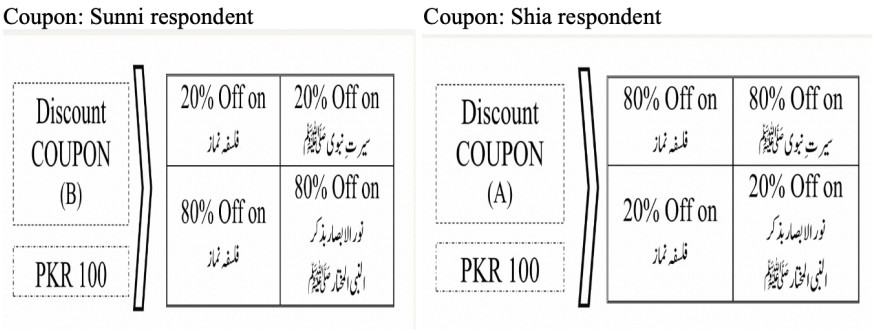


Figure 3.1: Book Vouchers: Four books are offered - two from the perspective of Sunnis and two from the perspective of Shias. The names of the books are mentioned in Urdu.

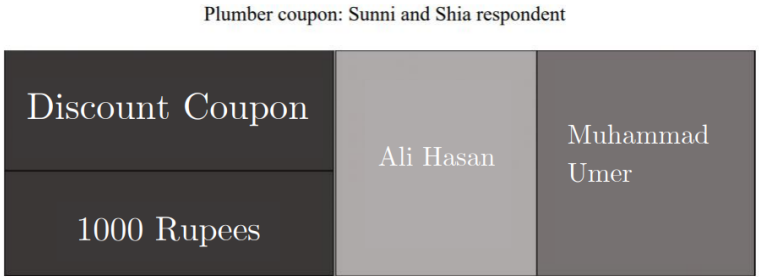


Figure 3.2: Plumber Vouchers: One obviously Sunni and Shia name each is visible in this voucher as is the value of the coupon.

(Figure 3.2). The names are clearly Shia or Sunni names and the respondents have to choose one of the two plumbers. Eventually, 10 percent of our respondents were randomly chosen to be provided the plumbing services. I arranged for the plumber to carry out repairs in the respondent's house.

3.4 Results

3.4.1 Econometric Specification

I estimate the following regression specification to analyse the effect of our treatments on multiple outcomes related to the beliefs and behavior of members of the opposite sect:

$$Y_i = \alpha + \beta_1 (Treatment)_{1i} + \beta_2 (Treatment)_{2i} + \beta_2 (Treatment)_{3i} + (Block\ Fixed\ Effects)_i + \epsilon_i, \tag{3.1}$$

where Y_i is the outcome of interest for individual i. Here, $Treatment_i$ is a binary variable for each of our three treatments. I include block fixed effects based on age, education, income, religiosity, and beliefs about Shias and Sunnis.

3.4.2 Results

Our first result is for the incentivized, real-world plumbing experiment in which I offer discounted plumbing services from two plumbers of clearly different sects (the sect is known from plumber names)[15]. I offer this choice to every participant at endline, however, it is made clear that only a random selection of respondents will be provided the plumbing service. The use of plumbing services of the opposite sect is an important signal of openness for two reasons. First, in many developing countries, kin networks matter greatly. It is highly usual to hire workers and service members from your own caste, religion or linguistic group (whichever identity is most salient). Hence, hiring members outside your network would be an important signal of openness. Second, plumbing services at home are not just a short, impersonal purchase made at a small

[15]There are certain personalities in Islamic history that have over time become very partisan, e.g., a caliph who fought against and killed the person that Shia today revere. Thus, this caliph's name, while still used by Sunnis, is never used by Shias.

kiosk/local grocery store. This is an invitation to come inside one's home, where one's family resides, and hence it is something that is not easily done in South Asian culture.

Table 3.2: The Effect of Contact on Plumber Choices

	Plumber, Opp Sect	Plumber, Sunni
	(1)	(2)
Announcement Only	-0.074	0.133***
	(0.052)	(0.034)
Prayer Volunteers Only	0.017	-0.011
	(0.023)	(0.048)
Announcement and Volunteers Both	0.180**	-0.376***
	(0.065)	(0.058)
Control Mean	0.153	0.435
Number of Respondents	423	423
Number of Mosques	32	32
Block Fixed Effects	X	X

Notes: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. The dependent variables are (i) a binary variable which is 1 when any respondent, either from the Sunni (majoritarian) sect or the Shia (minority) sect, chooses discounted plumbing services from a member of the opposite sect (the names of the plumbers allow for clear sectarian identification) and 0 otherwise and (ii) a binary variable which is 1 if a Sunni plumber's discounted services are chosen by respondents of any sect and 0 otherwise. Both outcomes are based on incentivized lab-in-the-field experiments. The independent variables are the assignment of mosques to the prayer volunteer visits treatment (where volunteer worshipers are sent to mosques of the opposite sect), the mosque leader announcement treatment (where the leader of the mosque makes a religious statement about inter-sectarian harmony), or the combined treatment. I use block fixed effects and cluster errors at the strata level.

In column 1 of Table 3.2, I show the effect of different types of contact on opposite-sect plumber choice. Our outcome variable is a binary variable which is 1 when any respondent, either from the Sunni (majoritarian) sect or the Shia (minority) sect, chooses discounted plumbing services from a member of the opposite sect and 0 otherwise. I find economically and statistically significant effects for our combined announcement and volunteer treatment which increases demand for a plumber of the opposite sect by 0.18 points against a mean control group demand of 0.153. This is a substantial increase, which I contextualize later. In column 2, I try to understand which sect is driving this behavior. Our outcome variable is a binary variable which is 1 if a Sunni plumber's discounted services are chosen by respondents of any sect and 0 otherwise. I can see a substantial reduction in demand for Sunni plumbers which indicates that Sunnis reduced their demand for their own sect substantially and chose Shia plumbers.

It is important to note that Shias and Sunnis do not differ on any other metric including competence for plumbing.

Next, in Table 3.3, I show how contact affects potential hiring decisions using our survey. In column 1, the outcome variable is the change in openness to hiring a member of the opposite sect at the endline compared to baseline. In column 2, the outcome variable is a binary variable which is 1 when the first dependent variable is strictly positive. Both regressions show a positive effect of our combined treatment: our respondents are more open to hiring members of the opposite sect. For column 1, our control group shows a negative change from baseline to endline, but our combined treatment group increases willingness to hire a member of the opposite sect by .240 points.

Table 3.3: The Effect of Contact on Hiring Decisions

	Hiring Change	Hiring Change, Positive
	(1)	(2)
Announcement Only	0.012	0.063**
	(0.097)	(0.021)
Prayer Volunteers Only	-0.021	-0.013
	(0.102)	(0.031)
Announcement and Volunteers Both	0.240*	0.088**
	(0.113)	(0.028)
Control Mean	-0.107	0.237
Number of Respondents	421	423
Number of Mosques	32	32
Block Fixed Effects	X	X

Notes: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. The dependent variables are (i) the change in openness to hiring a member of the opposite sect at the endline compared to baseline and (ii) whether the latter variable is strictly positive. These are survey-based measures. The first outcome is an answer to the following survey question: "What do you think about recruiting Shia/Sunni workers?" The answers can range from very bad to very good in a five-point scale. The independent variables are the assignment of mosques to the prayer volunteer visits treatment (where volunteer worshipers are sent to mosques of the opposite sect), the mosque leader announcement treatment (where the leader of the mosque makes a religious statement about inter-sectarian harmony), or the combined treatment. I use block fixed effects and cluster errors at the strata level.

3.4.3 Heterogeneity

I are now interested in understanding how our treatment effect differs based on conservativeness. A fundamental individual-level impediment to such a positive change is the religious conservativeness of an individual.

To do this, I exploit even deeper groupings within the Sunni sect: Deobandi, Ahle Hadith and Barelvi. These three sub-groups make up all Sunnis within Pakistan. However, there are substantial differences between them. Deobandis and Ahle Hadith are much more conservative compared to Barelvis. I use this division to create a binary variable which is 1 if a respondent is from either of the two conservative sub-groups, and 0 otherwise.

In Table 3.4, I are interested in the triple interaction of announcement and volunteer worshiper treatment with conservativeness (1 being more conservative). I see that when more conservative worshipers are treated it reduces demand for the opposite-sect plumber - the completely opposite result to our main finding.

Table 3.4: The Effect of Contact on Plumber Choice: Heterogeneity by Sub-Sect Conservativeness

	Plumber, Opposite Sect (1)	Plumber, Sunni (2)
Annoucement	-0.108***	0.173***
	(0.019)	(0.047)
Volunteers	0.024	0.040
	(0.028)	(0.027)
Announcement x Volunteers	0.262***	-0.401***
	(0.033)	(0.052)
Announcement x Conservative	0.204***	-0.118
	(0.055)	(0.149)
Volunteers x Conservative	0.008	-0.161
	(0.047)	(0.095)
Announcement x Volunteers x Conservative (Sub-Sect)	-0.307**	0.155
	(0.091)	(0.123)
Control Mean	0.153	0.435
Number of Respondents	423	423
Number of Mosques	32	32
Block Fixed Effects	X	X

Notes: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. The dependent variables are (i) a binary variable which is 1 when any respondent, either from the Sunni (majoritarian) sect or the Shia (minority) sect, chooses discounted plumbing services from a member of the opposite sect (the names of the plumbers allow for clear sectarian identification) and 0 otherwise and (ii) a binary variable which is 1 if a Sunni plumber's discounted services are chosen by respondents of any sect and 0 otherwise. Both these outcomes are based on incentivized lab-in-the-field experiments. I define conservativeness based on differences in sub-sects within Sunni Islam: two of the main sub-sects within Sunni Islam in Pakistan are substantially more conservative than the third. This variable is a binary variable which is 1 for members of these two conservative sub-sects and 0 otherwise. The independent variables are assignment of mosques to the prayer volunteer visits treatment (where volunteer worshipers are sent to mosques of the opposite sect), the mosque leader announcement treatment (where the leader of the mosque makes a religious statement about inter-sectarian harmony) or the combined treatment. I use block fixed effects and cluster errors at the strata level.

3.5 Conclusion

I conduct a field experiment to analyze the effect of contact and leadership in reducing prejudice between discordant sectarian groups in Pakistan. The Shia-Sunni division has resulted in a wave of terrorism in countries such as Pakistan and created two groups that hold at least very exclusionary beliefs and preferences, sometimes not even seeing the other group as being Muslim, as well as more extreme beliefs that support killing each other. I explore whether such deep divisions can be healed through contact. In our field experiment, I find that when I send volunteer worshipers to mosques of the opposite sect to pray and have the leader of the mosque make an announcement in

favour of unity respondents choose the services of a plumber from the opposite sect much more in the treatment group compared to the control group.

This is work in a particular context and does not yet deeply explore changes in beliefs and preferences, nor is it clear whether when scaled it, it would lead to effects of a similar magnitude.

Bibliography

Alesina, Alberto, Reza Baqir, and William Easterly (1999). "Public Goods and Ethnic Divisions". *The Quarterly Journal of Economics* 114.4, pp. 1243–1284.

Algan, Yann and Pierre Cahuc (2013). "Trust and Growth". *Annual Review of Economics* 5.1, pp. 521–549.

Allport, Gordon Willard, Kenneth Clark, and Thomas Pettigrew (1954). "The Nature of Prejudice". *Political Psychology* 12.1, pp. 125–157.

Andersen, Steffen et al. (2008). "Eliciting Risk and Time Preferences". *Econometrica* 76.3, pp. 583–618.

Andreoni, James and Charles Sprenger (2012a). "Estimating Time Preferences from Convex Budgets". *American Economic Review* 102.7, pp. 3333–56.

— (2012b). "Risk Preferences are not Time Preferences". *American Economic Review* 102.7, pp. 3357–76.

Augenblick, Ned, Muriel Niederle, and Charles Sprenger (2015). "Working Over Time: Dynamic inconsistency in real effort tasks". *The Quarterly Journal of Economics* 130.3, pp. 1067–1115.

Banfield Edward, C (1958). *The Moral Basis of a Backward Society.*

Barrick, Murray R and Michael K Mount (1991). "The Big Five Personality Dimensions and Job Performance: a meta-analysis". *Personnel Psychology* 44.1, pp. 1–26.

Beaman, Lori, Rohini Pande, and Alexandra Cirone (Mar. 2012). *Politics as a Male Domain and Empowerment in India.* Oxford University Press, Oxford, UK, pp. 208–226.

Begum, Sharifa and Binayak Sen (2005). "Maternal Health, Child Well-being and Intergenerationally Transmitted Chronic Poverty: Does Women's Agency Matter?" *Chronic Poverty Research Centre Working Paper.*

Beshears, John et al. (2008). "How are Preferences Revealed?" *Journal of Public Economics* 92.8-9, pp. 1787–1794.

Besley, Timothy and Torsten Persson (2009). "The Origins of State Capacity: Property Rights, Taxation, and Politics". *American Economic Review* 99.4, pp. 1218–1244.

Bjarnegård, Elin (2013). *Gender, Informal Institutions and Political Recruitment: Explaining male dominance in parliamentary representation*. Springer, London, UK.

Bortolotti, Stefania et al. (2021). "Patience, Cognitive Abilities, and Cognitive Effort: Survey and Experimental Evidence From a Developing Country". *American Behavioral Scientist* 65.11, pp. 1512–1530.

Brocas, Isabelle and Juan D Carrillo (2008a). "The brain as a Hierarchical Organization". *American Economic Review* 98.4, pp. 1312–46.

— (2008b). "Theories of the Mind". *American Economic Review* 98.2, pp. 175–80.

Browning, Martin and Pierre-André Chiappori (1998). "Efficient Intra-Household Allocations: A general characterization and empirical tests". *Econometrica*, pp. 1241–1278.

Burns, Nancy, Kay Lehman Schlozman, and Sidney Verba (2001). *The Private Roots of Public Action*. Harvard University Press, Cambridge, MA.

Cadena, Brian C and Benjamin J Keys (2015). "Human Capital and the Lifetime Costs of Impatience". *American Economic Journal: Economic Policy* 7.3, pp. 126–53.

Callen, Michael et al. (2015). "Personalities and Public Sector Performance: Evidence from a Health Experiment in Pakistan". Working Paper Series 21180. DOI: 10.3386/w21180.

Campbell, James E (2010). "Explaining Politics, Not Polls: Reexamining macropartisanship with recalibrated NES Data". *Public Opinion Quarterly* 74.4, pp. 616–642.

Carbone, Enrica and Gerardo Infante (2015). "Are Groups Better Planners than Individuals? An Experimental Analysis". *Journal of Behavioral and Experimental Economics* 57, pp. 112–119.

Castillo, Marco et al. (2011). "The Today and Tomorrow of Kids: Time preferences and educational outcomes of children". *Journal of Public Economics* 95.11-12, pp. 1377–1385.

Chabris, Christopher F et al. (2008). "Individual Laboratory-Measured Discount Rates Predict Field B Behavior". *Journal of Risk and Uncertainty* 37.2, pp. 237–269.

Chattopadhyay, Raghabendra and Esther Duflo (2004). "Women as Policy Makers: Evidence from a Nationwide Randomized Experiment in India". *Econometrica* 72.5, pp. 1409–1443.

Cheema, Ali, Sarah Khan, Asad Liaqat, et al. (2023). "Canvassing the Gatekeepers: A Field Experiment to Increase Women Voters' Turnout in Pakistan". *American Political Science Review* 117.1, pp. 1–21.

Cheema, Ali, Sarah Khan, Shandana Khan Mohmand, Anam Kuraishi, et al. (2019). "Women's Political Participation in a Pakistani Metropolis: Navigating gendered household and political spaces". *The Institute of Development and Economic Alternatives*.

Cheema, Ali, Sarah Khan, Shandana Khan Mohmand, and Asad Liaqat (2019). *Invisible Citizens: Why more women in Pakistan do not vote*. IDS, McLean, VA.

Chiappori, Pierre-André (1992). "Collective Labor Supply and Welfare". *Journal of Political Economy* 100.3, pp. 437–467.

Cohen, Jonathan et al. (2020). "Measuring Time Preferences". *Journal of Economic Literature* 58.2, pp. 299–347.

Condorcet, Jean Antoine Nicolas de Caritat (1785). *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. de l'Imprimerie Royale.

Corno, Lucia, Eliana La Ferrara, and Justine Burns (2022). "Interaction, Stereotypes, and Performance: Evidence from South Africa". *American Economic Review* 112.12, pp. 3848–75.

Cubitt, Robin P and Daniel Read (2007). "Can Intertemporal Choice Experiments Elicit Time Preferences for Consumption?" *Experimental Economics* 10.4, pp. 369–389.

DellaVigna, Stefano and Ulrike Malmendier (2006). "Paying Not to Go to the Gym". *American Economic Review* 96.3, pp. 694–719.

Denant-Boemont, Laurent and Youenn Loheac (2011). "Time and Teams: An experimental study about group inter-temporal choice".

Dohmen, Thomas et al. (2010). "Are Risk Aversion and Impatience Related to Cognitive Ability?" *American Economic Review* 100.3, pp. 1238–60.

Duflo, Esther (2012). "Women Empowerment and Economic Development". *Journal of Economic Literature* 50.4, pp. 1051–79.

Ericson, Keith Marzilli and David Laibson (2019). "Intertemporal Choice". *Handbook of Behavioral Economics: Applications and Foundations 1*. Vol. 2. Elsevier, pp. 1–67.

Evans, Peter B (1995). *Embedded Autonomy: States and industrial transformation*. Princeton University Press.

Falk, Armin et al. (2018). "Global Evidence on Economic Preferences". *The Quarterly Journal of Economics* 133.4, pp. 1645–1692.

Feldstein, Martin S (1964). "The Social Time Preference Discount Rate in Cost Benefit Analysis". *The Economic Journal* 74.294, pp. 360–379.

Field, Erica et al. (2019). "On Her Own Account: How Strengthening Women's Financial Control Impacts Labor Supply and Gender Norms". *Yale University Economic Growth Center Discussion Paper* 1068.

Frederick, Shane (2005). "Cognitive Reflection and Decision Making". *Journal of Economic Perspectives* 19.4, pp. 25–42.

Frederick, Shane, George Loewenstein, and Ted O'donoghue (2002). "Time Discounting and Time Preference: A critical review". *Journal of Economic Literature* 40.2, pp. 351–401.

Fudenberg, Drew and David K Levine (2006). "A Dual-Self Model of Impulse Control". *American Economic Review* 96.5, pp. 1449–1476.

— (2011). "Risk, Delay, and Convex Self-Control Costs". *American Economic Journal: Microeconomics* 3.3, pp. 34–68.

Gelman, Andrew et al. (2016). "The Mythical Swing Voter". *Quarterly Journal of Political Science* 11.1, pp. 103–130.

Gerber, Alan and Donald Green (2017). "Field Experiments on Voter Mobilization: An overview of a burgeoning literature". *Handbook of Economic Field Experiments*. Vol. 1. Elsevier, pp. 395–438.

Giné, Xavier, Jessica Goldberg, et al. (2018). "Revising Commitments: Field evidence on the adjustment of prior choices". *The Economic Journal* 128.608, pp. 159–188.

Giné, Xavier and Ghazala Mansuri (2018). "Together We Will: Experimental evidence on female voting behavior in Pakistan". *American Economic Journal: Applied Economics* 10.1, pp. 207–235. ISSN: 19457790. DOI: 10.1257/app.20130480.

Glätzle-Rützler, Daniela, Philipp Lergetporer, and Matthias Sutter (2021). "Collective Intertemporal Decisions and Heterogeneity in Groups". *Games and Economic Behavior* 130, pp. 131–147.

Gollier, Christian and Richard Zeckhauser (2005). "Aggregation of Heterogeneous Time Preferences". *Journal of Political Economy* 113.4, pp. 878–896.

Golsteyn, Bart HH, Hans Grönqvist, and Lena Lindahl (2014). "Adolescent Time Preferences Predict Lifetime Outcomes". *The Economic Journal* 124.580, F739–F761.

Gottlieb, Jessica (2016). "Greater Expectations: A Field Experiment to Improve Accountability in Mali". *American Journal of Political Science* 60.1, pp. 143–157. ISSN: 15405907. DOI: 10.1111/ajps.12186.

Guarnieri, Eleonora and Helmut Rainer (2018). "Female Empowerment and Male Backlash". *CESifo Working Paper Series* 7009.

Gul, Faruk and Wolfgang Pesendorfer (2001). "Temptation and Self-Control". *Econometrica* 69.6, pp. 1403–1435.

Harrison, Glenn W, Morten I Lau, and Melonie B Williams (2002). "Estimating Individual Discount Rates in Denmark: A field experiment". *American Economic Review* 92.5, pp. 1606–1617.

Hausman, Jerry A (1979). "Individual Discount Rates and the Purchase and Utilization of energy-using durables". *The Bell Journal of Economics*, pp. 33–54.

Hjort, Jonas (2014). "Ethnic Divisions and Production in Firms". *The Quarterly Journal of Economics* 129.4, pp. 1899–1946.

Holcomb, James H and Paul S Nelson (1992). "Another Experimental Look at Individual Time Preference". *Rationality and Society* 4.2, pp. 199–220.

Jackson, Matthew O and Leeat Yariv (2014). "Present Bias and Collective Dynamic Choice in the Lab". *American Economic Review* 104.12, pp. 4184–4204.

— (2015). "Collective Dynamic Choice: The necessity of time inconsistency". *American Economic Journal: Microeconomics* 7.4, pp. 150–78.

Jensen, Robert (2012). "Do Labor Market Opportunities Affect Young Women's Work and Family Decisions? Experimental Evidence from India". *The Quarterly Journal of Economics* 127.2, pp. 753–792.

John, Oliver P, Laura P Naumann, and Christopher J Soto (2008). *Paradigm Shift to the Integrative Big Five trait taxonomy: History, measurement, and conceptual issues.* The Guilford Press.

Kalin, Michael and Niloufer Siddiqui (2014). *Religious Authority and the Promotion of Sectarian Tolerance in Pakistan.* Tech. rep. United States Institute of Peace.

Kalla, Joshua L and David E Broockman (2018). "The Minimal Persuasive Effects of Campaign Contact in General Elections: Evidence from 49 field experiments". *American Political Science Review* 112.1, pp. 148–166.

Karpowitz, Christopher F, J Quin Monson, and Jessica Robinson Preece (2017). "How to Elect More Women: Gender and candidate success in a field experiment". *American Journal of Political Science* 61.4, pp. 927–943.

Khan, Sarah (2021). "Count Me Out: Gendered preference expression in Pakistan". *Working Paper.*

Koopmans, Tjalling C (1960). "Stationary Ordinal Utility and Impatience". *Econometrica: Journal of the Econometric Society*, pp. 287–309.

Laibson, David (1997). "Golden Eggs and Hyperbolic Discounting". *The Quarterly Journal of Economics* 112.2, pp. 443–478.

Laibson, David, Andrea Repetto, and Jeremy Tobacman (2007). *Estimating Discount Functions with Consumption Choices over the Lifecycle.*

Lawrance, Emily C (1991). "Poverty and the Rate of Time Preference: Evidence from panel data". *Journal of Political economy* 99.1, pp. 54–77.

Liaqat, Asad (2019). "No Representation Without Information: Politician Responsiveness to Citizen Preferences".

Lowe, Matt (2021). "Types of Contact: A field experiment on collaborative and adversarial caste integration". *American Economic Review* 111.6, pp. 1807–44.

Mankiw, N Gregory, David Romer, and David N Weil (1992). "A Contribution to the Empirics of Economic Growth". *The Quarterly Journal of Economics* 107.2, pp. 407–437.

Manser, Marilyn and Murray Brown (1980). "Marriage and Household Decision-Making: A bargaining analysis". *International Economic Review*, pp. 31–44.

Marglin, Stephen A (1963). "The Social Rate of Discount and the Optimal Rate of Investment". *The Quarterly Journal of Economics*, pp. 95–111.

Mazzocco, Maurizio (2007). "Household Intertemporal Behaviour: A collective characterization and a test of commitment". *The Review of Economic Studies* 74.3, pp. 857–895.

McElroy, Marjorie B and Mary Jean Horney (1981). "Nash-Bargained Household Decisions: Toward a generalization of the theory of demand". *International Economic Review*, pp. 333–349.

Metcalfe, Janet and Walter Mischel (1999). "A Hot/Cool-System Analysis of Delay of Gratification: Dynamics of willpower." *Psychological Review* 106.1, p. 3.

Migdal, Joel S (1988). *Strong Societies and Weak States: State-society relations and state capabilities in the Third World*. Princeton University Press.

Milch, Kerry F et al. (2009). "From Individual Preference Construction to Group Decisions: Framing effects and group processes". *Organizational Behavior and Human Decision Processes* 108.2, pp. 242–255.

Noor, Jawwad (2011). "Temptation and Revealed Preference". *Econometrica* 79.2, pp. 601–644.

Noor, Jawwad and Norio Takeoka (2022). "Optimal Discounting". *Econometrica* 90.2, pp. 585–623.

O'Donoghue, Ted and Matthew Rabin (1999). "Doing it Now or Later". *American Economic Review* 89.1, pp. 103–124.

— (2001). "Choice and Procrastination". *The Quarterly Journal of Economics* 116.1, pp. 121–160.

Paluck, Elizabeth Levy, Seth Green, and Donald Green (2019). "The Contact Hypothesis Re-Evaluated". *Behavioural Public Policy* 3.2, pp. 129–158.

Perez-Arce, Francisco (2017). "The Effect of Education on Time Preferences". *Economics of Education Review* 56, pp. 52–64.

Pons, Vincent (2018). "Will a Five-Minute Discussion Change Your Mind? A Country-wide Experiment on Voter Choice in France". *American Economic Review* 108.6, pp. 1322–1363. ISSN: 00028282. DOI: 10.1257/aer.20160524.

Prillaman, Soledad Artiz (2016). "Strength in Numbers: How Women's Networks Close India's Political Gender Gap".

Rae, John (1905). *The Sociological Theory of Capital: being a complete reprint of the new principles of political economy, 1834*. Macmillan.

Rai, S, Nafisa Shah, and Aazar Ayaz (2007). "Achieving Gender Equality in Public Offices in Pakistan". *Islamabad: United Nations Development Programme. Available online at: http://undp. org. pk (accessed 26 July 2008)*.

Samuelson, Paul A (1937). "A Note on Measurement of Utility". *The Review of Economic Studies* 4.2, pp. 155–161.

Scacco, Alexandra and Shana S Warren (2018). "Can Social Contact Reduce Prejudice and Discrimination? Evidence from a field experiment in Nigeria". *American Political Science Review* 112.3, pp. 654–677.

Schaner, Simone (2015). "Do Opposites Detract? Intrahousehold Preference Heterogeneity and Inefficient Strategic Savings". *American Economic Journal: Applied Economics* 7.2, pp. 135–74.

Sprenger, Charles (2015). "Judging Experimental Evidence on Dynamic Inconsistency". *American Economic Review* 105.5, pp. 280–85.

Stokes, Susan C et al. (2013). *Brokers, Voters, and Clientelism: The puzzle of distributive politics*. Cambridge University Press, Cambridge, UK.

Sunde, Uwe et al. (2022). "Patience and Comparative Development". *The Review of Economic Studies* 89.5, pp. 2806–2840.

Thaler, Richard (1981). "Some Empirical Evidence on Dynamic Inconsistency". *Economics Letters* 8.3, pp. 201–207.

Thaler, Richard and Hersh Shefrin (1981). "An Economic Theory of Self-control". *Journal of Political Economy* 89.2, pp. 392–406.

Van der Linden, Dimitri, Jan te Nijenhuis, and Arnold B Bakker (2010). "The General Factor of Personality: A meta-analysis of Big Five intercorrelations and a criterion-related validity study". *Journal of Research in Personality* 44.3, pp. 315–327.

Warner, John T and Saul Pleeter (2001). "The Personal Discount Rate: Evidence from military downsizing programs". *American Economic Review* 91.1, pp. 33–53.

Wooldridge, Jeffrey M (2002). "Econometric Analysis of Cross Section and Panel Data MIT Press". *Cambridge, MA* 108.2, pp. 245–254.

Acknowledgments

Many people have contributed to this dissertation. First of all, I am deeply thankful to Axel Dreher for his advice and support at every juncture of the doctorate. The amount of space and support he gives to his doctoral students to pursue their own research agendas is admirable. I am very grateful to him for pushing us to do ambitious, important work. I am also grateful for his constant efforts to ensure his international students feel at home, including strategically delivered nods to colleagues to speak in English when I was around. While a PhD is no easy task, it was made much easier due to this intellectual, emotional and social support. I am thankful to Alexandra Avdeenko for being highly supportive and making the time for discussions whenever necessary at short notice. I am thankful to Reimut Zohlnhöfer for immediately accepting to serve as the third examiner of this dissertation.

In addition to my supervisors, I am thankful for many great and supportive colleagues in Heidelberg. From my first officemate, Sarah, to colleagues with whom I started: Angelika, Gerda, Katha, Lenny, John, Johannes, Lukas, Sven, Tobi. I am in awe of their amazing work and their humility. I am grateful to the seniors among them who taught me a lot and made me feel at home. I feel particularly privileged to have had exceptional female colleagues. To the Mannheim group: Christian, Robbert, Tilmann and particularly Michael and Paula, who were always there to help, to give advice and to listen. I am grateful to them all for the support when moving in, the noodle group, lunches, coffee breaks, birthday celebrations, motivational videos, riverside walks, exotic recipes and mindless laughter - for the great effort to ensure we function well as a group. This is no easy thing, and unfortunately too often neglected. Without this community, there would be infinitely less joy in our work. To those who joined the chair and department as juniors: Annika, Charlotte, Jingke, Robin and Teresa. To

those affiliated with the chair or the department: Andi, Christoph, Kafeel, Marina, Min, Rajesh, Theo, and Valentin. To other professors in the department, particularly Dietmar Fehr and Stefan Trautmann, who were always exceptionally supportive of my work, happy to give feedback on my early-stage ideas and immensely supportive and understanding during my job search. To the seminar goers in Heidelberg who gave good, tough comments to help improve my work. I always appreciated the engagement. To the research assistants at the chair who worked diligently. To Barbara Neef, Freya, Ulrike and particularly Heike for supporting and helping me deal with the university administration, and managing the myriad changes in immigration statuses and grants. To the IT team for always being responsive and supportive.

My research profited from many gifted co-authors from whom I learned a lot. I thank Karrar Hussain for being such an engaged and supportive co-author. Finding great co-authors is not the easiest task in our field and I feel incredibly lucky. I thank Tobi for being always available and pushing forward every idea as if his world did not contain 24 hours in a day. I also thank the DFG for its generous funding for my field experiments and doctoral positions. In a world where we see a substantial reduction in research funds in many developed countries, the DFG has remained impressive in its support of research in economics.

The thesis would not have been possible without the continuous support of my family. I thank my parents who lived an unbelievably honest life in a tough country, my siblings for always being happy to talk and their immense love, and my wife and son for their support, tolerating the late nights and frequent travel and bringing the necessary joy in one's private life.

Furthermore, I would like to acknowledge the following colleagues for their input to the individual chapters:

Chapter 1: We are grateful to Michael Callen, Asim Khwaja, Aprajit Mahajan, Charlie Sprenger, Henrik Orzen, Peter Duersch and participants in seminars and conferences in Heidelberg, NEUDC, Pacific Conference in Development Economics (PacDev), Naples Workshop on Networks and Development, EUDN PhD Workshop, Advances in Field Experiments for their feedback. We are grateful to the Center for Economic Research Pakistan (CERP), Sohaib Athar, Zara Salman and Mehroz

Alvi for excellent logistical support. We are indebted to Dr. Turab Hussain and Muhammad Hussain for championing this project. Rao Hashim, Danish Raza, Shanila Parveez provided excellent research assistance. We received Institutional Review Board approvals from Harvard and UC San Diego.

Chapter 2: We are thankful to Tahir Andrabi, Parnab Bardhan, Christine Binzel, Axel Dreher, Muhammad Farooq Naseer, Stefan Klonner, Andreas Landmann, Melissa Lee, Katharina Michaelowa, Jim Snyder, Galina Zudenkova, and participants in seminars and conferences at LUMS, Heidelberg, Advances in Field Experiments (Chicago), Annual Experimental Political Science Conference (NYU), AEL Berlin (DIW/WZB), and Field Days (INSEAD) for helpful comments, to Turab Hussain and Syed Zahid Ali for supporting this project and Rana Sikandar Hayat Khan (District Chairman, Kasur) for field support. Kiran Khalid and Alisha Yazdani provided excellent research assistance.

Chapter 3: I are grateful to Christine Binzel, Eoin McGuirk and Andreas Landmann for their feedback. I thank Rashid Langrial (Federal Secretary, MoE Pakistan), Ejaz Sarwar (Deputy commissioner, Usta Muhammad district), and Saqlain Saleem (Assistant Commissioner, Abbottabad Tehsil) for championing the project. The grant was ably managed by the Consortium for Development and Policy Research Economic Research Pakistan (CDPR) and Lahore University of Management Sciences (LUMS). Muhammad Talha Hassan provided excellent research assistance. We received Institutional Review Board approvals from LUMS. The project was funded by the International Growth Centre and Innovations for Poverty Action.