



Essays on Intragroup Processes in Intergroup Relations

Citation

Romney, David Alexander. 2020. Essays on Intragroup Processes in Intergroup Relations. Doctoral dissertation, Harvard University Graduate School of Arts and Sciences.

Link

<https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37368918>

Terms of use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material (LAA), as set forth at

<https://harvardwiki.atlassian.net/wiki/external/NGY5NDE4ZjgzNTc5NDQzMGIzZWZhMGFIOWI2M2EwYTg>

Accessibility

<https://accessibility.huit.harvard.edu/digital-accessibility-policy>

Share Your Story

The Harvard community has made this article openly available. Please share how this access benefits you. [Submit a story](#)

HARVARD UNIVERSITY
Graduate School of Arts and Sciences



DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the
Department of **Government**
have examined a dissertation entitled

“Essays on Intragroup Processes in Intergroup Relations”

presented by **David A. Romney**

candidate for the degree of Doctor of Philosophy and hereby
certify that it is worthy of acceptance.

Signature Melani C. Cammett

Typed name: Prof. Melani C. Cammett (Chair)

Signature Ryan D. Enos

Typed name: Prof. Ryan D. Enos

Signature Dustin Tingley

Typed name: Prof. Dustin Tingley

Date: **August 25, 2020**

Essays on Intragroup Processes in Intergroup Relations

A dissertation presented

by

David A. Romney

to

The Department of Government

in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Political Science

Harvard University
Cambridge, Massachusetts

August 2020

©2020 — *David A. Romney*

All rights reserved.

Essays on Intragroup Processes in Intergroup Relations

Abstract

This dissertation studies the intragroup dynamics that guide the intergroup behavior and attitudes of groups in tension. The first essay develops a categorization for peace-building interventions and a theory of intragroup interventions. This essay presents an argument for why intragroup interventions have the potential to build peace in contexts where other interventions might fail. The subsequent two essays explore two very different but related contexts highlighting the importance of intragroup dynamics. In the second essay, I explore ingroup policing among Israeli Jews using two group-based experiments conducted in Haifa, Israel. The experiments explore what types of justifications for ingroup policing are most persuasive for other ingroup members, and they find that moral arguments come at a credibility cost for the one doing the policing. Lastly, in the third essay, I use a large-scale survey experiment to explore interventions to increase contributions to a collective good—drain cleaning—in slums in Delhi, India. The experiment finds that accountability mechanisms generally encourage Muslims, but not Hindus, to contribute more. I attribute this finding to Muslims' desire to maintain ingroup respectability and avoid blame from others.

| Contents

Abstract	iii
Acknowledgments	vi
1 Introduction	1
2 How Intragroup Interventions Counter Prejudice in Situations of Conflict	4
2.1 Introduction	4
2.2 A Tripartite Categorization of Conflict Interventions	5
2.3 The Limitations of Intergroup and Superordinate Interventions	7
2.4 Intragroup Interventions	10
2.5 Discussion and Conclusion	19
3 Frames for Ingroup Policing: Evidence from Two Group-Based Experiments in Israel	21
3.1 Introduction	21
3.2 Moral and Pragmatic Frames for Ingroup Policing	23
3.3 Experimental Design	27
3.4 Results	30
3.5 Conclusion	44
4 Persecuted Minorities and Prosocial Behavior: Accountability and Public Goods Provision among Hindus and Muslims in Delhi Slums	46
4.1 Introduction	46
4.2 Diversity, minority status and public goods provision	49
4.3 Hypotheses	57
4.4 Context: Hindu-Muslim relations in India	58
4.5 Data and Methods	60
4.6 Results	66
4.7 Probing the minority effect: Why Muslims contribute more	69
4.8 Conclusion	76
A Appendix to Chapter 2	79
A.1 Demographics and Treatments	79
A.2 Regressions	82
A.3 Pooled Results	89
A.4 Textual Analysis	90

B Appendix to Chapter 3	92
B.1 Survey Administration and Sampling Details	92
B.2 Survey Instrument: Key Measures and Experimental Manipulations	94
B.3 Additional Results	104
B.4 Robustness Checks	114
Bibliography	127

| Acknowledgments

So many people from so many different communities have helped me along the path to academia. For much of my life, this was not a path I imagined for myself, and it certainly was not one I knew how to navigate. I would not have successfully made it through this journey without the encouragement and collaboration of friends and mentors throughout the years.

Much of my early encouragement I owe to the relationships I forged with professors at my undergraduate institution, Brigham Young University. I am so grateful, especially, for the mentorship and guidance of Joshua Gubler, whose enthusiasm for my ideas and my abilities helped me believe in myself. I am also grateful for the help I've received over the years from Chris Karpowitz, Dan Nielsen, Ray Christensen, Dilworth Parkinson, and Meredith Nelson. Their tutelage was crucial in the early development of my intellectual curiosity and scholarly skills.

I am likewise grateful for the community of scholars who helped me find success in graduate school. I am especially grateful for the help of my committee: Melani Cammett, Ryan Enos, and Dustin Tingley, who have always been there for advice as well as for scholarly collaboration. My graduate student colleagues have been generous of their time and help, and I am especially grateful for the friendship of Angie Bautista-Chavez, Gabriel Koehler-Derrick, and Seth Soderborg. I have also benefited from feedback and conversations with a wider circle of scholars, including: Poulomi Chakrabarti, Don Green, Jennifer Hochschild, Amaney Jamal, Robert Keohane, Josh Kertzer, Gwyneth McClendon, J. Quinn Monson, Rich Nielsen, Pia Raffler, and Brandon Stewart.

I would not have been able to conduct the projects I did without the community of Israelis and Palestinians who have helped me with my field research. Daphna Cannetti and other members of her political violence lab, in particular Shani Fachter and Ohad Shaked, were indispensable in helping me carry out my project in Haifa, as were the many research assistants and other participants in my project: Rona Eed, Galia Hoz, Itzik Kushel, Eyal Marder, Iliya Ratz, Anna Reznikova, Shalev Shahar, Jessica Swerdlow, Hadas

Yehoshua, and Ishay Yemini. Thanks as well to Michael Freedman, for his feedback and assistance with Hebrew textual analysis, and Chagai Weismann, for his collaboration and help.

I owe much to the institutions that supported me throughout this process. Outside of Harvard, I am grateful for the financial support of the National Science Foundation Graduate Research Fellowship. Within Harvard, I am extremely grateful for the support of the Weatherhead Center, the Institute for Quantitative Social Science, the Inequality and Social Policy Program, the Center for Jewish Studies, and the Foundations of Human Behavior Initiative.

Lastly, and most importantly, my deepest gratitude goes out to my family. First and foremost, I thank my strong and encouraging wife, Camille, who has always believed in me and whose support means more to me than anyone's. I thank my three amazing children, who often pull me away from work but only for good reasons. I also thank my parents, whose love and support I know I can always lean on.

1 | Introduction

Underlying this entire thesis is a premise explored in the first essay—the idea that the dynamics that guide behavior and attitudes within groups is crucial to building peace as the dynamics that guide behavior between or across groups. I end up drawing two main conclusions: first, these dynamics must be taken into account when addressing intergroup relations, and, second, these dynamics can be exploited as interventions to build peace. So many peace-building interventions are designed as though each side of a conflict is going to respond in the same way to the same interventions, but in reality they do not. The way groups respond depends both on how the micro-level factors of an intervention function in a specific context as well as the macro-level context of the state or nation in which the conflict is taking place.

The project I describe in Chapter 4 is an excellent illustration of this principle, i.e. the idea that context matters in determining how groups will respond to an intervention. This paper, which is co-authored with Melani Cammett and Poulomi Chakrabarti, examines interventions to increase individuals' contributions to solving a collective action problem. In the case of our project, the collective action problem is making sure that the storm drains running along the sides of roadways and alleys are clean, allowing water to drain. Issues with the drainage system causes huge problems in Delhi slums. Especially during the rainy season, a clogged drain can cause sewage to seep into the streets and into residents houses, spreading filth and unsanitary conditions. In a survey we administered among residents of selected Delhi slums, we propose a hypothetical private-based solution. The proposed solution consists of hiring a private firm that individuals would pay to regularly clean out their drains. However, the private firm will only operate in a neighborhood if two-thirds of the residents agree to pay for the service, meaning that residents would have to coordinate to initiate a contract.

In the context of this collective action problem, we use a survey experiment to test three different mechanisms to try to increase participation in the drain cleaning program. We label the three mechanisms: horizontal accountability, vertical accountability, and ingroup underperformance. In the horizontal account-

ability treatment, individuals are told that their neighbors would know who did and did not contribute in a public list. In the vertical accountability treatment, participants are told that local informal neighborhood leaders, called pradhans, would follow up with those who did not contribute. And in the ingroup underperformance treatment, participants are presented with a hypothetical testimonial that highlights their own co-religionists, either Muslim or Hindu, not contributing enough to the program to make it work. We also explore combinations of these treatments (e.g. horizontal accountability plus ingroup underperformance).

We find that the effect of the experiments depends on the group. We targeted neighborhoods containing primarily Hindu and Muslim residents, and completed extensive fieldwork to identify some slums that were majority Hindu, some that contain both Hindus and Muslims, and some other that are majority Muslims. We find that, across the board, Muslims are more likely to respond to the experimental treatments than Hindus. Based on the fact that this effect is particularly present for participants with a strong connection to their religious community, we primarily attribute this result to Muslims' desire to preserve ingroup respectability and avoid blame; as a minority, Muslims face disproportionate blame and retribution for not contributing to communal goods, and therefore are more worried about mechanisms that will hold them accountable to their community or that highlight a lack of contribution by members of their own group.

The findings from this project illustrate an important fact: interventions designed to encourage prosocial behavior often do not operate the same for members of a minority, and this should be taken into account in similar contexts where collective action across ethnic, racial, or religious lines is important—for instance, in the context of the US. Because of minorities' unique experiences, they have ingroup concerns and norms that differ from those of the majority, and these norms cause them to behave differently in response to interventions that prime their unique concerns.

This is just one example of how ingroup processes can affect how individuals behave in intergroup contexts. In Chapter 2 of this dissertation, I explore how interventions based on intragroup processes can help build peace. Intragroup interventions are especially important because they function in contexts where other interventions—I label them intergroup and superordinate interventions—are unlikely to be as effective. I argue that segregation, asymmetry between groups, and weak unifying institutions are common factors in situations of conflict that make intragroup interventions more effective. Given the ubiquitous presence of these obstacles, we should do more to explore peace-building interventions that can take advantage of ingroup identities and institutions.

One such intragroup intervention is encouraging ingroup policing by members of groups in conflict—

by ingroup policing, I mean the condemning or punishment of ingroup bad behavior toward an outgroup. Interventions focused on ingroup policing can be effective if they lead individuals to approve of others' ingroup policing efforts, or if they lead individuals to engage in ingroup policing themselves. From current research, we know little about the micro-level factors that make ingroup policing effective. I highlight three factors that I think are important: the identity of the policer in terms of group affiliations and elite status; the identity of the audience in terms of whether the policing takes place in front of an ingroup, outgroup, or international audience; and, lastly, the frame used for policing, meaning the argument that is used to justify engaging in ingroup policing to others.

In Chapter 3, I delve into this last factor of ingroup policing in the context of the Israeli-Palestinian conflict. To explore the importance of ingroup policing frames, I conducted two group-based experiments with Israeli Jews from Haifa and surrounding areas in 2016 and 2018. I recruited participants to gather in small focus groups and discuss an incident in which an Israeli soldier was caught on film killing an incapacitated and captured terrorist. (The soldier was later arrested, went to prison, and was released after serving his sentence last year). One of the participants in these groups was an actor I trained to take certain positions depending on the treatment assignment. In the control condition, this actor took a neutral stance on the incident; in both of the treatment conditions, on the other hand, the actor opposed the Israeli soldier's killing of the Palestinian, while the frame for the actor's argument—i.e. their justification for engaging in ingroup policing—varied between the treatments. In one treatment, the actor opposes the killing based on grounds of ingroup morals. In the other condition, they oppose the soldier's actions based on a pragmatic argument based in the security concerns of Israeli Jews. These two frames cause different results in terms of outcomes measured in a survey afterward: Both frames convinced participants to either be more critical of the soldier or to support elites who condemned the incident. However, the moral frame also caused participants to rate the actor lower than others, indicating that using a moral frame can affect the credibility of someone engaging in ingroup policing.

At the end of this dissertation, two appendices provide supporting tables and figures for Chapters 3 and 4.

2 | How Intragroup Interventions Counter Prejudice in Situations of Conflict

2.1 Introduction

For decades, researchers have studied how to counter prejudice in situations of conflict, and an important aspect of this research has focused on the implementation of peace-building interventions in situations of conflict—i.e. studying the effect of specific programs that organizations, either in-country or international, can put into place to improve intergroup relations. This burgeoning field of research is rich in the variety of interventions that have been proposed and studied (Ditlmann, Samii, and Zeitzoff 2017a), and research on this topic has given us knowledge of what reduces prejudicial attitudes and behaviors and how to increase anti-conflict and anti-violent ones.

This paper builds on this research to suggest a new categorization for peace-building interventions and to propose a theoretical structure for exploring them. I argue that peace-building interventions can be categorized by the ties they have to group identities and group interactions. The conceptualization of conflict interventions in this manner leads to a tripartite categorization of interventions: (1) intergroup interventions, which focus on providing opportunities for exposure of members of an ingroup to interaction with or the perspectives of an outgroup; (2) superordinate interventions, which focus on changing attitudes and behaviors by using appeals that transcend the identities of the groups in conflict; and (3) intragroup interventions, which focus on the creation of or altering of ingroup norms of behavior toward an outgroup.

Extant research on peace-building interventions has overwhelmingly focused on the first two types of interventions, i.e. intergroup and superordinate ones. This research has established the importance of these types of interventions in countering prejudicial attitudes and behaviors in situations of conflict, and this paper does not dispute their importance to our understanding of peace-building. However, I argue that intragroup interventions should also constitute a crucial tool for countering prejudice in situations of conflict. All

three types of interventions have their advantages and limitations, but situations of conflict are often characterized by attributes that make intragroup interventions particularly important to pursue as a strategy against prejudice. Specifically, situations of conflict are often characterized by extreme segregation, asymmetrical power between groups, and weak unifying institutions, and intragroup interventions are less affected by these obstacles than intergroup and superordinate interventions are.

Different types of intragroup interventions are possible, but in this paper I highlight just one: ingroup policing. An ingroup policing intervention is one whose goal is to persuade ingroup members, organizations, and institutions to condemn bad behavior by fellow group members against an outgroup. Many recognize ingroup policing is an important way of maintaining peace in a situation of intergroup tension, but little research has been conducted on the micro-level characteristics of ingroup policing that make it more or less successful. In contrast, copious research has been conducted in the necessary conditions for positive intergroup contact, based on Allport's original conditions of equal status, common goals, intergroup cooperation, and support of authorities (Allport 1954; see reviews of such interventions in Pettigrew and Tropp 2006 and Paluck, Green, and Green 2019). In this paper, I highlight three characteristics of ingroup policing that are important to its effectiveness: the identity of the policer, the audience for an ingroup policing effort, and the framing used to justify policing ingroup members. Variations of each of these characteristics should be tested to determine how to best generate norms of ingroup policing for groups in conflict.

2.2 A Tripartite Categorization of Conflict Interventions

This paper focuses on peace-building interventions, i.e. specific programs that could be implemented by organizations to try to counter prejudice in conflict or post-conflict situations. As a result, this paper is primarily focused on interventions at the level of the individual or, at times, small groups of people. These types of programs are common in conflict settings and often take the form of tasks assigned to individuals (e.g. individual training or assignments) or to small groups (e.g. a guided group discussion or meeting, or a group-based training).

Projects evaluating such interventions often use a wide variety of metrics to measure their effectiveness. Perhaps the most common of such metrics are measures of attitudes toward the outgroup and toward peace in general, but some studies also measure pro-peace behavioral outcomes, such as the extent to which individuals choose to interact with members of another group (Mousa 2020) or come to the defense of or

support outgroup members in other ways (Ditlmann and Samii 2016). These outcomes exist on a continuum of attitudes and behaviors from those that are prejudicial to those that build peace, and the goal of these interventions is to decrease prejudicial attitudes and behaviors while increasing the peace-building ones. For the purposes of this paper, I assume that those instituting these programs are interested in improving all peace-building outcomes—i.e. they may care about particular outcomes the most but would also care if there is movement on any of the outcomes. I also assume that any given intervention is not going to have opposing effects with regard to these outcomes—e.g., for a given individual, an intervention that increases pro-peace behaviors toward outgroup members is not going to decrease warmth toward the outgroup at the same time. Thus, the most “effective” intervention is the one which produces the strongest movement on the most measures of pro-peace behaviors.

I distinguish between three types of conflict interventions: intergroup, superordinate, and intragroup. This tripartite categorization hinges on the primary mechanism behind the intervention. Intergroup interventions are ones that primarily focus on contact or exposure to the outgroup and their attitudes, and for the purposes of this categorization, this contact or exposure can be in-person or through another medium (e.g. consuming media or virtual interaction). This definition of intergroup interventions encompasses a wide variety of interventions that, admittedly, are not traditionally grouped together, such as: contact-based interventions that provide opportunities for individuals to engage in activities and institutions with members of the outgroup (Mousa 2020; Enos 2014; Scacco and Warren 2018), discussion groups that bring ingroup and outgroup members together (Yablon 2012), perspective-taking exercises (Adida, Lo, and Platas 2018), and media-based interventions that provide portrayals of outgroup members and their perspectives (Paluck 2010). Superordinate interventions, on the other hand, are ones that primarily focus on modeling behavior or changing attitudes and norms at a level that transcends the identities of the groups in conflict (Collier and Vicente 2013). Often, transcending the identities of the groups in conflict is done by appealing to a higher-level identity, such as a national identity (Charnysh, Lucas, and Singh 2015). The third category of intervention—intragroup interventions—is both the main focus of this paper and the least-researched of the three. This type of intervention is primarily concerned with changing or creating ingroup norms toward an outgroup, or using ingroup members to persuade others to refrain from prejudiced actions or engage in peace-building ones. Examples of this type of intervention include those focused on bystander intervention (Bilali, Vollhardt, and Rarick 2015) as well as studies that examine the effect of messaging by co-ethnics (Siegel and Badaan 2020).

The distinction between these three types of interventions is crucial to understanding how to mitigate negative outcomes in conflict, even though in practice there is often overlap between the three. For instance, discussion groups bring ingroup and outgroup members together but also expose individuals to the opinions of other ingroup members, who may engage in ingroup policing or otherwise attempt to affect ingroup norms. Similarly, a media intervention may at the same time provide humanizing portrayals of the outgroup, appeal to superordinate identities while encouraging specific attitudes and behaviors, and model appropriate ingroup behavior and active bystandership. This overlap is to be expected; the tripartite categorization proposed in this paper is not intended to be a description of the universe of interventions as they currently exist. Rather, the purpose of this paper is to explain the importance of categorizing interventions based on their identity appeals, examine the strengths and weaknesses of each of the three types of interventions, and provide an agenda for studying intragroup interventions that will deepen our understanding of how to prevent negative outcomes in conflict.

2.3 The Limitations of Intergroup and Superordinate Interventions

I argue that certain macro-level factors common in situations of conflict hinder the ability of intergroup and superordinate interventions to effectively counter prejudice. Below, I describe three macro-level factors: Segregation, asymmetry, and weak unifying institutions.

2.3.1 Segregation

Ethnic segregation is a common characteristic of societies in conflict and presents severe challenges to intergroup interventions. This is not to say that segregation limited to situations of conflict; it is present outside of situations of conflict and, in those contexts as well as in contexts of conflict, affects many aspects of political behavior (Enos 2017). However, segregation is particularly prevalent in situations of conflict and is an important factor in the prolonging of conflict between groups (Corvalan and Vargas 2015). Isolation between groups causes them to lose the common ground that is necessary for ending conflict.

This same aspect of segregation poses an important obstacle for intergroup interventions. Effective intergroup interventions require a common space—both in geographic as well as institutional terms—for groups to come together; heavily segregated societies lack such common spaces, and although they can be created artificially as part of a program, the lack of naturalistic settings for intergroup interaction could

make it difficult for effects to endure at the end of such a program. Additionally, heavily segregated societies are likely to produce groups that differ greatly from each other in terms of culture, perspectives, and experiences. Intergroup interactions are more difficult when the base of common experiences to draw on is small. Perhaps most extreme are groups that differ from each other not just in terms of geographic location and institutions but also in terms of language. Language is a key factor because effective communication is essential to having meaningful intergroup exposure and contact. It is true that, through translation, exposure to the perspectives of members of another group is possible when different languages are spoken. However, without a common language, an important aspect of person-to-person contact is lost.

2.3.2 Asymmetry

Many conflicts take place between groups that do not hold equal power. For instance, the Israeli-Palestinian conflict, especially after its initial years, is often cited as a classic case of asymmetric conflict because of the much larger comparative abilities of the Israeli military (Maoz 2004, 2011; Maoz and McCauley 2008). Such asymmetry poses a particular barrier to intergroup interventions because of the importance of groups interacting on an “equal status” basis—with both groups engaging as well as having similar background characteristics—in order to engender positive attitudes toward each other (Allport 1954). Without equal status, participants in an intergroup intervention may discount the experiences and perspectives of those of lower rank.

The importance of this condition is debated in the literature, with some analyses concluding that this condition is not essential to positive intergroup contact (Pettigrew and Tropp 2006) but others concluding that such factors can hinder positive contact, particularly in situations of conflict (Paluck, Green, and Green 2019). However, aside from overall effectiveness, asymmetry is important for another reason: it leads to different experiences of contact for those from advantaged and disadvantaged groups. In the context of asymmetry, contact can lead to negative side effects for those from disadvantaged groups because the power differential causes them to experience contact in a different way than their advantaged peers. For instance, when disadvantaged group members engage in intergroup contact programs, especially those focused on finding harmony or commonality, it can lead them to no longer attribute the negative experiences of ingroup members to discrimination (Saguy and Chernyak-Hai 2012). Similarly, participating in such programs can set up false expectations of equal treatment for members of disadvantaged groups, who are then confronted with a different reality in their day-to-day interactions outside of the program (Saguy et al. 2009; Saguy,

Dovidio, and Pratto 2008). This may be the mechanism that underlies some puzzling results from contact-based interventions that result in no change in or a decrease in positive attitudes toward the outgroup (see Paluck 2010 and especially Dittmann and Samii 2016, where the puzzling result is limited to Palestinian participants in a contact-based program). In such contexts, the pursuit of positive contact and avoiding conflict with other groups may come at the expense of diminishing the hardships of the disadvantaged group, thus discouraging members of disadvantaged groups from engaging in appropriate opposition to the wrongs they have suffered (Cikara and Paluck 2013).

2.3.3 Weak Unifying Institutions

The political and national institutions of a particular context are an essential factor in the success of intergroup and superordinate interventions. In some contexts of conflict, such as in Israel, the existence of these institutions helps individuals use the tools they gain through interventions to effect change (Dittmann, Samii, and Zeitzoff 2017a, pp. 68–69). In such contexts, a functioning political system and a strong national identity provides citizens with opportunities to engage in peace-building behaviors and reinforces their positive attitudes. However, in many situations of conflict, such institutions do not exist. This is a problem for at least two reasons. First, because both intergroup and superordinate interventions operate outside the realm of a single ingroup, they both need the support of institutions or an infrastructure outside of group membership in order to appropriately function.¹ Though stated somewhat differently, this is a point that Allport (1954) makes when he notes the “support of authorities” as an important condition for positive intergroup contact. Without this support, there is no way to make intergroup experiences happen or to appeal to norms that transcend the groups in conflict. Second, in contexts that lack strong overarching institutions, group-based institutions often take their place. In such contexts, not only is there no common institutions to support intergroup experiences, but the group-based institutions that exist are stronger and more exclusionary. Identities in such contexts can be more strongly tied to an ingroup than a national identity, a factor that is likely to put individuals more at risk for developing negative attitudes and behavior toward the outgroup.

¹Note that the building of these institutions is important in and of itself, and is a good justification for instituting intergroup and superordinate interventions; as noted later in the paper, one of the limitations of an intragroup approach is that it does not build these cross-cutting institutions.

2.4 Intragroup Interventions

Intragroup interventions differ from intergroup and superordinate ones in key ways, and that helps intragroup interventions overcome the limitations of the other two. Such interventions do not replace the others, but rather complement their abilities and fill in the gaps that they cannot fill. In contrast with intergroup interventions, interventions that focus on intragroup dynamics are less affected by segregation. This is true on a number of levels. Ingroup individuals are more likely to live in the same areas as their co-ethnics and, therefore, have more interaction with them. They are more likely to be members of the same institutions—social, religious, and cultural—and, therefore, be socialized together and be familiar with each others' cultural practices. In many locations and depending on the type of group, they may also be more likely to speak the same language as each other. It is easier to work within these extant institutions than it is to try to create opportunities for interaction across groups that maintain separate institutions. This is not to say that there are not divisions within groups themselves; these divisions exist and are extremely important to understanding conflict. For instance, divisions among Jews in Israel—along lines of religiosity, national origin, and economic status—have played a vital role in the Israeli-Palestinian conflict (Sprinzak 1999; Kimmerling 2001). However, in spite of the divisions that may exist, it is still more likely for individuals to maintain stronger bonds with fellow ingroup members than they have with members of other groups.

Similarly, asymmetry and weak unifying institutions are also less of a limitation for intragroup interventions. The main problem with asymmetry is that it causes a difference in power that underlies any intergroup exposure that individuals have; this difference in power results in advantaged and disadvantaged groups experiencing contact in wildly different ways and diminishing the perspectives and difficulties faced by disadvantaged groups. However, this negative side effect does not occur with intragroup interventions because they draw only on ingroup norms and experiences. It is important to note that I am *not* arguing that asymmetry is nonexistent within groups. Members of the same group can differ widely on a number of status-related factors such as wealth and occupation, but such asymmetries are not integrally tied to the intervention because they are not asymmetries between the groups in conflict. As for weak unifying institutions, they are not problematic for intragroup interventions because such interventions function in the realm of group-based institutions.

In what follows, I discuss how intragroup interventions focused on ingroup policing can increase our knowledge of what makes ingroup policing effective at the micro-level.

2.4.1 Ingroup Policing

Individuals often police fellow ingroup members in an effort to prevent what the group perceives as inappropriate behavior, but what is inappropriate can vary between groups and between contexts. Often, in a situation of conflict, a group polices their own to prevent association with the outgroup and defection—what we might call “bad” ingroup policing, in that it might punish those who express sympathy with the outgroup. However, there are also cases of “good” ingroup policing, where individuals police negative attitudes and behavior toward an outgroup. Political scientists have argued that ingroup policing has an important role in decreasing violent outcomes in situations of conflict. Most famously, Fearon and Laitin (1996a) compellingly argued that peaceful inter-ethnic equilibria are vastly more common than relationships characterized by inter-ethnic violence, and one peaceful equilibrium they highlight depending on ingroup policing, where peace is maintained via a mutual understanding between the groups in conflict that they will punish their ingroup members who misbehave. Scholars have recognized the vital role of ingroup policing in conflict—that it is an important mechanism for maintaining relative peace in a situation of ethnic tension (Fearon and Laitin 1996a; Brubaker and Laitin 1998) and that it has potential as a “bottom-up” mechanism for addressing intergroup conflict (Ditlmann, Samii, and Zeitsoff 2017a). Some have specifically tested interventions aimed at increasing dissent and ingroup policing in conflict and post-conflict contexts (Paluck and Green 2009). This research recognizes that ingroup policing can serve a strategic purpose and reduce negative outcomes in situations of conflict. However, this research has not explored the micro-level factors that determine the effectiveness of ingroup policing.

2.4.2 The Micro-level Politics of Ingroup Policing

Ingroup policing can take place in a variety of contexts and levels in society. It can be a family member reprimanding others for bad views toward the outgroup. It could be a bystander intervening when observing an ingroup member’s abuse of an individual from the outgroup. It can be a fighter discouraging brutality toward the outgroup by members of his unit. Or it can be an ethnic elite setting boundaries of appropriate group behavior. In each of these cases, through their actions or statements, individuals condemn negative behavior by their fellow ingroup members toward an outgroup.

In each of these cases, we can explore the factors that make ingroup policing more effective. As noted previously, there are a variety of measures of effectiveness for peace-building interventions, and not all

outcomes are appropriate measures for all contexts. For ingroup policing, there are three measures that are appropriate as outcomes. First, ingroup policing is effective if it persuades other ingroup members that the condemned behavior is wrong. For instance, if a member of the group condemns another ingroup member for engaging in violence against the outgroup, one metric of effectiveness for that individual's condemnation is the extent to which they convinced other ingroup members that it is wrong to engage in that type of violence against the outgroup. This would indicate persuasion to a new norm of appropriate ingroup behavior. Second, ingroup policing is effective if it causes other ingroup members to support those who engage in ingroup policing—for instance, if it convinces an individual that condemning ingroup behavior is appropriate to do. Lastly, ingroup policing is effective if it causes individuals to engage in ingroup policing themselves by speaking out or standing up to behavior that violates norms of appropriate behavior toward outgroup members.

In the context of asymmetry between groups, it is especially important to study ingroup policing among members of advantaged groups. This is not to say that ingroup policing is unimportant among disadvantaged groups. For instance, in the context of the Israeli-Palestinian conflict, Kydd and Walter (2002) and de Mesquita (2005) argue that Palestinian extremists have been able to effectively sabotage peace agreements by engaging in violence unchecked by their co-ethnics at strategic points in time. When moderates are perceived as being strong, extremist violence challenges that perception and convinces those in the opposing group that the outgroup moderates are unable to police their own. However, there are reasons to believe that ingroup policing among the advantaged group is as important or more important. Because of their relative power, a failure of ingroup policing by advantaged group members could result in violence or injustices that go unchecked because of the disadvantaged group's inability to prevent it—to use an example from the same context, the relatively unabated level of Jewish settlement in the Palestinian territories over the last approximately 40 years can be seen as a failure of ingroup policing on the Israeli side of the conflict.

In order to know what makes ingroup policing more effective at the micro-level, we need to break down ingroup policing by the set of actors and other pertinent factors involved. In any isolated instance of ingroup policing there is involvement by two main actors, each of which could consist of one individual, a group of individuals, or an organization. First, there is the actor condemning ingroup misbehavior, or the policer—the one standing up to misbehavior by a member of their own group. Second, there is the audience—the witness to the instance of ingroup policing. In each case, who the actor is and what the actor does determines how the instance of ingroup policing is received by other ingroup members. Additionally, in any instance of

ingroup policing there is also a justification, explicit or implied, for the policing that takes place, a factor with I label the frame.

Below I explore each of these three factors—the policer, the audience, and the frame—and discuss how an intragroup intervention focusing on ingroup policing might address it.

Identity of the Policer

For the policer, credibility is key. A policer who lacks credibility cannot discourage bad behavior by fellow ingroup members, and the identity of the policer is integral to how credible other ingroup members view them. Two primary identity characteristics of policers deserve the attention of those studying ingroup policing. First, other ingroup members can perceive the policer as identifying with affiliations that either attract or repel them. Of course, by definition all ingroup members belong to the same ethnic group as the ingroup policer, but subgroup affiliations can serve as an important indicator of credibility to other ingroup members. For instance, in the context of Israel and of Jewish-Israelis in particular, identities based on national origin (i.e. Ashkenazi versus Mizrahi Jews) and religion (i.e. secular to orthodox adherents of Judaism) can cause affinities even within the ingroup. Although within-group tensions have never exceeded those between Jews and Arabs, they have certainly led to confrontation and hold meaning to many (Sprinzak 1999). Other non-subgroup identities may also play an important role in making some ingroup policers more credible than others, most importantly partisan identity, which research has established as important in the development of group affinities (Green, Palmquist, and Schickler 2004). Those who belong to the same subgroup or are copartisans are more likely to find fellow subgroup members or fellow partisans more convincing than others.

Perhaps the most important identity factor is not affiliation but rather elite status. Dominant explanations for ethnic conflict emphasize the importance that rhetoric, and in particular elite rhetoric, holds in the instigation and perpetuation of inter-ethnic violence. In the most extreme form of this argument, political entrepreneurs and opportunists care little about ethnicity outside of its use as an instrument to gain what they desire, usually political or economic power. Achieving this power is the only reason they use the rhetoric at all, and systems that provide elites with the proper incentives will lead them to produce hateful rhetoric (Glaeser 2005). But even constructivist (and other) theories of ethnic conflict emphasize that rhetoric can heighten ethnic tensions, change the salience of various ethnic and religious identities, or even aid in the creation of new identities. Looking at Hindu-Muslim tensions in India, Wilkinson (2004) argues that rhetoric

encouraging ethnic violence is particularly common in electoral systems where politicians need additional votes from the majority ethnic group to win. Posner (2004) examines ethnic relations in Zambia and Malawi and attributes ethnic violence in Malawi to the fact that ethnic groups in the country are large enough for politicians to mobilize for political purposes. Petersen (2002), looking at the Baltics, Czechoslovakia, and Yugoslavia, argues that the main explanation for violence lies in the ability of entrepreneurs (some elites, some not) to manipulate or provoke the resentment and rage of citizens, mobilizing them for large-scale acts of violence. Similarly, In summarizing a separate set of case studies on ethnic conflict, Fearon and Laitin (2000) note that “if there is a dominant or most common narrative. . . it is that large-scale ethnic violence is provoked by elites seeking to gain, maintain, or increase their hold on political power.”

In spite of this research on the role of elites in inciting violence, there has been little research on the role of elites in doing the opposite—i.e. in helping citizens develop positive norms of behavior toward other groups. Ethnic elites hold positions that allow them to set norms for other ingroup members, provide instructions that others are to follow, and model behavior that others are supposed to imitate. Research in non-conflict settings indicate that those in elite positions in a network hold an outsize ability to influence others to follow positive group norms. For instance, Paluck, Shepherd, and Aronow (2016) find that those whom they call “social referents,” i.e. those who attract more student attention, are more effective at creating a norm for publicly taking a stand against student bullying. A similar effect likely exists for elites in this context.

What can we take from this in designing an intragroup intervention focused on ingroup policing? First off, when designing ingroup policing interventions, we should measure the affiliations of those engaging in ingroup policing. Salient affiliations will vary based on the context, but universally pertinent affiliations are the policer’s partisan identity and their identity when it comes to salient ingroup cleavages. Additionally, ingroup policing interventions should focus on social referents and local elites who can provide an outsize effect in terms of persuading others to support norms of ingroup policing.

The Audience

The audience for ingroup policing is another factor that likely determines its effectiveness. The audience is important because of the affect it can have on the perceived appropriateness of engaging in ingroup policing. Ingroup members often take pride in the perceived positive attributes of their group, expressing their “ingroup love” through lauding the positive aspects of their group identity (Brewer 1999). The desire

to present their group in a positive light can be threatened when an ingroup member engages in policing in a public manner. For many ingroup members, discussions of ingroup failings do not belong in the public sphere. This dynamic is reflected in studies in social psychology showing that people suppress criticism of their group when in front of non-ingroup members and look down upon those who criticize their group in front of the “other” (Hornsey et al. 2005; Packer and Miners 2014). When outside groups become involved, it can cause reticence to discuss such matters. Group members do not want to air their “dirty laundry” in front of non-group members, and they do not feel it is appropriate for others to do so either.

In the context of this attribute of ingroup policing, we can distinguish between four different types of audiences. The first is an audience that consists of only a small set of ingroup members, maybe for instance those involved in an incident and their immediate colleagues or counterparts (such as the squad of a soldier accused of misbehavior). Many ingroup members would feel comfortable with this level of policing, even if they spurn policing that involves a larger group of individuals. A second type of ingroup policing might target an audience of only fellow ingroup members but a much larger one. For instance, a public condemnation to other Jewish Israelis. As outgroup members start to enter the audience, some ingroup members will no longer feel that ingroup policing is appropriate. The third and fourth types of audiences fall into this category: an audience of outgroup members (e.g. Palestinians in our running example) or outsiders who are not involved in the conflict (i.e. international parties). Interventions should explore the extent to which each type of audience is deemed appropriate by ingroup members.

It is important to take these factors into account when designing an ingroup policing intervention. Ingroup policing takes place in locations or institutions that may have varying levels of ethnic homogeneity, and all else held equal, individuals should judge ingroup policing as more appropriate when the location or institution is relatively homogeneous. For instance, a school-based program in a heterogeneous context that trains individuals to speak up against ingroup prejudice may be less successful than an otherwise similar program in a homogeneous context.² Additionally, the public involvement of international parties is a factor that should be carefully considered, since ingroup policing in front of international parties may also be frowned upon.

²Of course, the “otherwise similar” condition here is especially important, since it is highly likely that individuals and their families select into homogeneous or heterogeneous locations.

The Frame

The policer needs an argument to make their case to other ingroup members, and the more convincing the case the more effective the ingroup policing. As noted in the discussion of elites as policers, inciteful rhetoric by elites is widely recognized as an important factor in intergroup violence, but little has been done to explore the role of rhetoric in ingroup policing. When such rhetoric promotes ethnic violence, this rhetoric serves the purpose of framing and justifying this violence for others. Given the high levels of hatred and resentment in many situations of conflict, one could imagine that justifying outgroup violence is a fairly easy task. It is likely much more difficult to convince ingroup members to go along with condemning their fellow group members than it would be to convince them to engage in violence against the outgroup. Because of this relative difficulty, the rhetorical framing of ingroup sanctioning—the way that it is justified to other group members—is in some ways more important to understand than violence-promoting rhetoric. What sort of rhetoric can ingroup member can use when promoting ingroup policing? Are some types of rhetoric more effective than others?

One rhetorical characteristic of ingroup policing is the frame that is used in a particular appeal. Framing is the “process by which people develop a particular conceptualization of an issue or reorient their thinking about an issue” (Chong and Druckman 2007, p. 104; see also Entman 1993), and frames are efforts in political communication to affect this process. Because ingroup policing is inherently tied to the identity of the ingroup, I focus in particular on frames that are rooted in the culture or experience of the ingroup. Under this rubric, I argue that two types of frames are of primary importance in the context of ingroup policing: *moral* frames,³ which appeal to purportedly ingroup-based moral values, and *pragmatic* frames, which appeal to the strategic interests of the ingroup. This contrast between moral and pragmatic frames has parallels in other political science research, such as the distinction between pragmatic and principled issue domains in discussing party policy shifts (Tavits 2007) and the discussion of moral and pragmatic arguments against torture by Leidner, Kardos, and Castano (2018). Another parallel exists in classic texts on social movement theory, where McCarthy and Zald (1977) discuss how social movements have to make different arguments to appeal to their “conscience” and “beneficiary” constituents.

Moral frames can be effective at countering prejudice because they build on the ingroup’s traditions and

³Sometimes referred to as a moral (Leidner, Kardos, and Castano 2018), values (e.g. Brewer 2002; Brewer and Gross 2005), or principled (Tavits 2007) frame/issue.

values in framing ingroup policing. Walzer (1987) refers to this concept, which he outlines as an approach to moral philosophy, as “connected criticism.” Walzer writes in response to the moral philosophies of Rawls (1971) and Habermas (1979), who argue that systems of morality can be constructed only by stepping back from context and culture. What these philosophies miss, according to Walzer, is that social critics are most effective when they can appeal to a set of traditions and values that “best accords with our everyday experience” (Walzer 1987, p. 3). Thus a ingroup moral critic applies his own interpretation of shared values to societal problems, moved to do so by his own attachment to those very values. This is similar in concept to “constructive patriotism” (Staub 1997; Schatz, Staub, and Lavine 1999), where patriotic sentiments take the form of criticism seeking to effect positive change.

In contrast, a pragmatic frame is in line with the motivation that Fearon and Laitin (1996a) ascribe for ingroup policing, i.e. to prevent outgroup retaliation and a spiral into violence that might result if ingroup policing is not enforced. It is also in line with Packer and Chasteen’s (2010) model of “loyal deviance,” which argues that group interests can motivate highly-identifying group members to criticism if they feel that something their group is doing threatens those interests. One can imagine a variety of pragmatic interests that can serve as motivation for redirecting your group’s actions, but in the context of intergroup conflict security interests are particularly salient.⁴ A security interest frame appeals to the ingroup’s desire to remain physically and economically secure and stable in the face of conflict. This is particularly true in cases of intractable conflict like Israel’s. One aspect of the ethos of intractable conflict is the position of importance that security concerns occupy. Because of constant conflict, security becomes one of the main metrics by which important decisions are made (Bar-Tal 1998, p. 27, 33). And controversial actions by representatives of Israel have often directly led to protests and violence by Palestinians; for instance, politician Ariel Sharon’s controversial visit to the Temple Mount in September 2000 is often viewed as the proximate cause of the Second Intifada, or Palestinian uprising, in the early 2000s. Using such an argument, a policer might effectively argue that criticized ingroup behavior constitutes a threat to the security of Israelis, and Israeli Jews in particular.

Both moral and pragmatic frames have attributes that make them likely to be effective—in both cases, they draw on the context and experiences of the ingroup to make a case in favor of ingroup policing. How-

⁴In reality, many types of pragmatic interests may motivate ingroup members. Aside from security interests, ingroup members care about the economic well-being of their group. They may also care about their reputational interests—saving face in front of others.

ever, an additional factor to explore in the case of frames is the way that the use of a particular frame reflects on the policer. Some frames might be more costly, meaning that they reflect poorly on the credibility of the policer in the same way that the identity factors outlined above weigh on individuals' judgments of the policer. The literature on framing has little to say about how a frame reflects on the communicator. However, it is reasonable to assume that framing can affect how a communicator is viewed and not just how an issue is interpreted. In the same way that frames in communication can cause for shifts in public opinion by allowing for evaluations of an issue along new lines, a frame can also allow for new evaluations of the one communicating with the frame. In fact, in their discussion of an ingroup policing equilibrium, (Fearon and Laitin 1996a, p. 723) note that their conclusion only holds if one assumes that condemning violence towards the outgroup is relatively costless behavior and the equilibrium falls apart if ingroup members band together to make condemning bad actors in the ingroup becomes costly.

Out of the two frames highlighted here—pragmatic and moral—there is good reason to believe that a moral frame is costly. This conclusion is grounded in psychological research on dissent within groups, which examine how approval of dissent varies by the characteristics of the dissenter and the way in which they dissent. This literature documents particularly high levels of resentment toward those who rebel for purportedly moral reasons, labeled “moral rebels” in the research. The explanation for this finding revolves around the way in which we adopt group identities into our self concept. When a policer morally objects to the behavior of someone in an ingroup, it can be interpreted as an indication that the policer would reject us (Monin, Sawyer, and Marquez 2008; Minson and Monin 2012; O'Connor and Monin 2016). Pragmatic frames, on the other hand, hold little to be offended about. A pragmatic frame involves less of a value judgment toward ingroup members; even though such frames are still arguing in favor of condemning specific types of ingroup misbehavior, the justification for condemnation is based on protecting group interests rather than on the morally objectionable actions of the ingroup member. For this reason, a pragmatic frame is likely to not incur the same costs as a moral one.

All of this entails that careful attention be paid to the arguments used in ingroup policing interventions. If such an intervention were, for instance, training individuals to call out ingroup members when they misbehave, those designing the intervention should pay attention to how they instruct individuals to justify their position. Instructing them to take a moral stance may come at a cost to that individual's credibility that does not necessarily accompany other types of arguments. Additionally, the two types of frames discussed here is just one way of categorizing frames; exploring additional ways of conceptualizing frames that are

specific to particular contexts may be useful in implementing ingroup policing interventions.

2.4.3 The Limitations of Intragroup Interventions

Intragroup interventions can overcome key difficulties of other types of interventions; however, their key strength—that they build on the already-present and strong ingroup ties of individuals—is also their key limitation. Because such interventions only work through ingroup ties, they do not aid in the creation of societal institutions that cut across ethnic divides in a society the same way that intergroup and superordinate interventions do. This should not limit the adoption of these interventions; rather, what it illustrates is that these three types of institutions complement each other. They work through different group identities and dynamics and, therefore, are tailored to affect different sets of outcomes. Different interventions affect different “peace-building behaviors, as noted by (Ditlmann, Samii, and Zeitzoff 2017a, pp. 49–52), and we should adopt multiple interventions that affect the behaviors of most interest to us in a given context.

2.5 Discussion and Conclusion

This paper is a call for researchers and those interested in peace building to design interventions that focus on intragroup processes, in particular ingroup policing. Situations of conflict are often characterized by deep segregation, group asymmetry, and weak unifying institutions. In such a context, it may be effective to design an intervention that uses ingroup institutions rather than solely focusing on interventions that are based on intergroup or superordinate mechanisms. Ingroup policing is particularly promising as a type of intragroup intervention; however, those designing an ingroup policing intervention should carefully consider the context in which they are implementing their intervention. The most effective ingroup policing appeals are those with credible ingroup policers, who use frames that resonate with other ingroup members, and which are presented to an audience that the ingroup deems appropriate.

Ingroup policing is just one possible intragroup intervention; researchers should explore other interventions that exploit ingroup dynamics. One potential intervention not discussed in this paper is the promotion of norms of protest and dissent among disadvantaged groups. A burgeoning literature in American politics examines the effects of racial protest on political outcomes (Enos, Kaufman, and Sands 2019; Mazumder 2018; Wasow 2020), and peace-building interventions could be designed around using ingroup-based institutions or appeals to encourage participation in non-violent demonstrations and protests. Although protests

are not inherently limited to intragroup dynamics the way that ingroup policing is, intragroup dynamics can play a vital role in helping people turn out and get involved.

Future research in this area should focus on testing the implications of the micro-level theory of ingroup policing laid out in this paper, but also more generally on the interaction between different types of peace-building interventions. For the most part, this paper considers each type of intervention in isolation, when in fact they can affect and be affected by each other. Is intergroup contact more effective in contexts of asymmetry when the advantaged group has had training in ingroup policing beforehand? Does positive intergroup contact make individuals more willing to public engage in ingroup policing because of increased warmth toward the outgroup? These questions are natural follow-up questions that can be explored in future research.

3 | Frames for Ingroup Policing: Evidence from Two Group-Based Experiments in Israel

3.1 Introduction

On March 24, 2016, Elor Azaria went from being like every other Israeli soldier to a household name. In the aftermath of an attack on an Israeli soldier in Hebron, which ended with one of the Palestinian attackers dead and the other wounded, Elor was videotaped as he shot and killed the incapacitated attacker.¹ Israeli society was at the time, and remains to this day, extremely divided over the incident and Elor's subsequent arrest. On the one hand, many supported the soldier's actions. In surveys soon after the incident, a majority said the soldier should never have been arrested;² some citizens created petitions and crowdfunding drives on his behalf;³ and many politicians defended him.⁴ Even after Elor's release following 9 months in prison, politicians have featured him in campaign ads and sought his endorsement in the recent Israeli elections.⁵ On the other hand, a large portion of the Israeli population supported the soldier's arrest, and three leading political and military leaders—the Prime Minister, the Defense Minister, and the Chief of Staff—all came out publicly against the killing at the time. But such statements were not without a cost. The Defense

¹See the New York Time's description of the incident at the following link: <http://www.nytimes.com/2016/03/25/world/middleeast/video-shows-israel-soldier-shooting-palestinian.html>.

²When asked, in a representative survey commissioned by Israeli Channel 2 News in the days following the incident, whether or not the soldier should have been arrested and a criminal investigation opened, 57% said it wasn't necessary, while only 32% said it was. See the following link for the survey results (in Hebrew): http://www.mako.co.il/news-military/security-q1_2016/Article-2f2e33ae6e8b351004.htm.

³See this petition (in Hebrew): <https://www.atzuma.co.il/citation>; see also this fundraising drive (in Hebrew): <https://web.archive.org/web/20180221202850/https://www.headstart.co.il/project.aspx?id=19464>.

⁴For instance, Jewish Home politician Naftali Bennett on Facebook (in Hebrew): <https://www.facebook.com/NaftaliBennett/posts/1086178811403786>.

⁵See the Times of Israel's description of the campaign ad here: <https://www.timesofisrael.com/likud-mk-enlists-hebron-shooter-azaria-to-back-his-reelection-campaign/>.

Minister resigned soon after the incident,⁶ and the Prime Minister was forced to backtrack on his initial condemnation to the point of suggesting he be pardoned after he had been tried and found guilty in military court.

The minority response to this incident is an example of ingroup policing—when ingroup members condemn aggression or misbehavior by their own group members toward an outgroup.⁷ Scholars have recognized the vital role of ingroup policing in conflict—that it is an important mechanism for maintaining relative peace in a situation of ethnic tension (Fearon and Laitin 1996a; Brubaker and Laitin 1998) and that it has potential as a “bottom-up” mechanism for addressing intergroup conflict (Ditlmann, Samii, and Zeitzoff 2017a). Some have specifically tested interventions aimed at increasing dissent and ingroup policing in conflict and post-conflict contexts (Paluck and Green 2009; Ditlmann and Samii 2016). This research recognizes that ingroup policing can serve a strategic purpose and reduce negative outcomes in situations of conflict.

The current paper builds on this research by examining the nature of ingroup policing at the micro-level and by exploring two crucial but previously unaddressed aspects of ingroup policing. First, this study recognizes that ingroup policing is an issue of persuasion. Those who seek to condemn ingroup misbehavior must persuade others in their group that this is the appropriate course of action, often against strong biases in favor of justifying ingroup behavior. Because of this, factors that make persuasion more effective can also make ingroup policing more effective. In this study specifically, I look at types of persuasive *frames* for ingroup policing—moral vs. pragmatic—and examine how effective they are at convincing others to adopt more critical attitudes and behavior toward ingroup misbehavior. Second, this study highlights the fact that frames used to justify ingroup policing differ not just in their persuasive ability but also in how they reflect on the one doing the policing—referred to as the “critic.” Some frames may engender a positive or negative reaction to the critic among ingroup members, and this positive or negative reaction can function independent of how convincing the frame is.

Using a unique research design, these aspects of ingroup policing were evaluated in two group-based experiments in 2016 and 2018. Israeli Jews from Haifa, Israel, were invited to participate in 1–1.5 hour

⁶See the New York Time’s description of his resignation: <https://www.nytimes.com/2016/05/21/world/middleeast/moshe-yaalon-israeli-defense-minister-resigns.html>.

⁷Note that this differs from definitions in some literatures, where ingroup policing is defined as condemning those who cooperate with the outgroup to prevent “defection” (e.g. Laitin 1995).

long focus groups, where they discussed the Hebron incident described at the beginning of this paper. A confederate participant administered treatment through the positions he took and the frames he used to justify those positions. Attitudes and participant behavior were measured through surveys administered after the experiment as well as through the content of participants' speech. These experiments produced an intriguing finding: both moral and pragmatic frames for ingroup policing can be convincing, but a moral frame causes participants to negatively evaluate the confederate. These frames are relatively equal in their persuasive power, albeit on different metrics, but they differ in how they reflected on the critic.

This research project expands our understanding of ingroup policing in important ways. Ingroup policing is a promising bottom-up mechanism for addressing intergroup conflict (Ditlmann, Samii, and Zeitzoff 2017a), but in spite of that it has been mostly unexplored in such a context. Most research on prejudice reduction and conflict resolution has focused on intergroup mechanisms, such as the contact hypothesis (Allport 1954; Pettigrew 1998; from political science, see Green and Wong 2009; Samii 2013; Scacco and Warren 2018). Instead, this research focuses on *intragroup* mechanisms, building on research that recognizes the importance of intragroup processes in conflict (Fearon and Laitin 1996a; Lyall 2010; Kalyvas 2008; Bakke, Cunningham, and Seymour 2012; Cunningham 2011). Additionally, the unique group-based research design used in this study provides a new and compelling template for studying interpersonal persuasion. Importantly, this design allows me to parse out a frame's persuasive effects from its effects on attitudes toward the one using the frame.

In this paper, I first provide a theory of frames for ingroup policing in a situation of conflict. I then describe my research design and analysis and present results from the two experiments.

3.2 Moral and Pragmatic Frames for Ingroup Policing

Ingroup Policing

Ingroup policing occurs when members of a given group criticize their own group's actions toward an outgroup. This paper focuses on cases where this criticism is given in front of an ingroup audience that does not include the ingroup member whose actions are being criticized. This excludes instances where the perpetrator of the criticized action is directly addressed (e.g. bystander intervention) or the audience for criticism includes outgroup members. Instead, the focus is on ingroup members discussing the appropriate boundaries for their group's behavior toward the outgroup.

Israel, the site of this project, is in the midst of perhaps the world's most famous intractable conflict—a conflict that is protracted, viewed as unsolvable, and violent, among other characteristics (Bar-Tal 2013, p. 37). This context is important for a number of reasons. In intractable conflict, groups are more likely to develop rules of engagement that guide combatants' actions, so discussions of appropriate boundaries for group behavior can be (and are) discussed formally, similarly to the discussions that form the basis for the experiments in this project. Additionally, ingroup policing may be particularly important in intractable conflict because intergroup relations are so fraught. Typical intergroup contact-based interventions may be problematic in the context of intractable conflict and imbalanced power relations (Maoz 2004; Saguy, Dovidio, and Pratto 2008; Saguy et al. 2009; Maoz 2011; Saguy and Chernyak-Hai 2012). By contrast, individuals generally respond more positively to criticism that originates from ingroup members, especially if that criticism is directed only to other ingroup members (Hornsey and Imani 2004; Hornsey et al. 2005; Sutton, Elder, and Douglas 2006; Packer and Chasteen 2010; Packer and Miners 2014; Packer 2014). Thus, ingroup policing offers a mechanism for conflict reduction in contexts where typical intergroup contact-based mechanisms might fall short—i.e., it may be more effective at persuading members of a group to be critical of ingroup misbehavior. How *effective* ingroup policing is constitutes the first aspect of ingroup policing that this study addresses.

However, ingroup policing is not costless. Ingroup members exhibit well-known biases toward their own group; all else equal, this bias means that they are likely to justify ingroup misbehavior. Because of this, ingroup policing usually requires expressing a point of view that dissents from or conflicts with the views of others (Cikara and Paluck 2013). For those engaging in ingroup policing, there is a potential cost they have to bear because their criticism may challenge more dominant viewpoints. How costly ingroup policing is constitutes the second aspect of ingroup policing that this study addresses.

Framing

What determines how effective or costly ingroup policing is? This project focuses on one important factor: the frames used to justify ingroup policing.

Framing is the “process by which people develop a particular conceptualization of an issue or reorient their thinking about an issue” (Chong and Druckman 2007, p. 104; see also Entman 1993). Frames are efforts in political communication to affect this process. Frames are used in all types of political communication, including interpersonal communication of the type that is the focus of this study. And, in spite of the

focus of much of the framing literature on elites and mass communication, the frames used in interpersonal communication can also be effective in guiding individuals' thinking about an issue and how they interpret frames from other sources (Druckman and Nelson 2003; Walsh 2010). Depending on a variety of factors, framing can have small or large effects on individuals' attitudes toward particular political topics—in the terms used in this paper, some frames on some issues can be more effective at persuading people to a particular point of view than others, although what makes a frame strong is still highly contested (Chong and Druckman 2007, pp. 116–117).

This project's focus on framing effects differentiates it from much scholarly work on ethnic conflict, which has in large part ignored this factor. However, some explanations of ethnic conflict do emphasize the importance of rhetoric, in particular elite rhetoric. These theories generally argue that elite rhetoric can heighten ethnic tensions, change the salience of various ethnic and religious identities, or even aid in the creation of new identities. Looking at Hindu-Muslim tensions in India, Wilkinson (2004) argues that rhetoric encouraging ethnic violence is used by politicians who need additional votes from the majority ethnic group to win. Similarly, Petersen (2002), looking at the Baltics, Czechoslovakia, and Yugoslavia, argues that the main explanation for violence lies in the ability of entrepreneurs (some elites, some not) to manipulate or provoke the resentment and rage of citizens, mobilizing them for large-scale acts of violence. This type of argument is also echoed in Glaeser (2005), which outlines a theory of supply and demand for hate-creating stories with applications to the US, Europe, and the Middle East. This project expands on this research by examining the opposite end of the rhetorical spectrum—instances where rhetoric can support norms against violence.

Moral and Pragmatic Frames

I argue that two types of frames are of primary importance in the context of ingroup policing: *moral* frames,⁸ which appeal to individuals' moral values, and *pragmatic* frames, which appeal to their interests. Moral frames are seen as important and widely-used across a variety of political issues (Ryan 2014; Feldman 1988). Pragmatic frames are less common, and are often described merely as nonmoral (Ryan 2014, p. 380); however, other scholars have recognized pragmatic frames, often in juxtaposition with moral frames, as a useful way of characterizing issues or frames (Leidner, Kardos, and Castano 2018; Tavits 2007).

⁸Sometimes referred to as a moral (Leidner, Kardos, and Castano 2018), values (e.g. Brewer 2002; Brewer and Gross 2005), or principled (Tavits 2007) frame/issue.

Moral and pragmatic frames both provide relevant arguments in favor of ingroup policing in a situation of conflict. For example, a moral frame is relevant because the violence condemned by ingroup policing appeals often violates the moral values and image of the ingroup. In a situation of intractable conflict, a positive moral image of the ingroup is often strongly emphasized because of a need to counter the negative effects of violence and aggression on the group's image (Bar-Tal 1998, p. 28). In the context of Israel, some of this work is done through the military, the Israeli Defense Forces. This is possible in Israel because of the connection the military has to citizens' Jewish identity. The IDF's Jewish origins (formed out of Jewish pre-independence paramilitary groups), the Jewish nature of its code of ethics,⁹ and conscription practices that primarily see Jewish citizens serve in the military—together, these factors establish the military as an ingroup institution for primarily Israeli Jews. Additionally, a range of Israeli political and military leaders from former Prime Minister Ehud Olmert¹⁰ to former Defense Minister Ehud Barak¹¹ to current Prime Minister Benjamin Netanyahu¹² regularly laud the IDF as the “most moral army in the world.” The need to uphold the military as a moral force in Israeli society is an important aspect of developing positive group image for Israeli Jews.

On the other hand, a pragmatic frame is in line with the motivation that Fearon and Laitin (1996a) ascribe for ingroup policing, i.e. to prevent outgroup retaliation and a spiral into violence that might result if ingroup policing is not enforced. Packer and Chasteen's (2010) model of “loyal deviance” argues that group interests, conceptualized generally, can motivate highly-identifying group members to criticism if they feel that something their group is doing threatens those interests. One can imagine a variety of pragmatic interests that can serve as motivation for redirecting your group's actions. In the context of Israel, a security interest frame is especially compelling. A security interest frame appeals to the ingroup's desire to remain physically and economically secure and stable in the face of conflict. Indeed, a lack of security and stability can be one of the main characteristics and results of conflict, but particularly intractable conflict. One aspect of the ethos of intractable conflict is the position of importance that security concerns occupy. Because of constant conflict, security becomes one of the main metrics by which important decisions are made, with

⁹One of the origins for the IDF's doctrine of ethics, referred to as “The Spirit of the IDF,” is “The tradition of the Jewish People throughout their history.” See <https://www.idfblog.com/about-the-idf/idf-code-of-ethics/>.

¹⁰See <http://www.ynetnews.com/articles/0,7340,L-3261303,00.html>.

¹¹See <http://www.ynetnews.com/articles/0,7340,L-3692383,00.html>.

¹²See <http://mfa.gov.il/MFA/PressRoom/2014/Pages/PM-Netanyahu-leaves-on-working-visit-to-the-US-28-Sep-2014.aspx>.

Israel no exception to this rule (Bar-Tal 1998, p. 27, 33). And controversial actions by representatives of Israel have often directly led to protests and violence by Palestinians; for instance, politician Ariel Sharon's controversial visit to the Temple Mount in September 2000 is often viewed as the proximate cause of the Second Intifada, or Palestinian uprising, in the early 2000s. In this vein, an ingroup critic might effectively argue that criticized ingroup behavior constitutes a threat to the security of Israelis, and Israeli Jews in particular.

However, moral and pragmatic frames are likely to reflect differently on the communicator, i.e. they are likely to differ in how *costly* they are. Research on dissent within groups documents high levels of resentment toward those who rebel for purportedly moral reasons, i.e. "moral rebels." We negatively evaluate moral rebels, especially when they are rebelling against something we've participated in ourselves and we anticipate that the rebel would reject us (Monin, Sawyer, and Marquez 2008; Minson and Monin 2012; O'Connor and Monin 2016). This is likely to be an issue in the context of intractable conflict, and is particularly salient for Israel given its mandatory conscription.

In summary, there are reasons to expect both moral and pragmatic frames to effectively convince ingroup members to support ingroup policing, but at the same time there is good reason to suspect that a moral frame could result in a backlash against the critic. The experiments, outlined in the following sections, address these claims.

3.3 Experimental Design

The Context

My project takes place in Israel, where studying intergroup conflict in a lab setting is common (e.g. Zeitzoff 2014; Enos and Gidron 2016). For this experiment, I used the Hebron Incident referred to in the introduction. I chose this incident because it was relevant to my research agenda, well-known, and would provoke lively discussion, as well as providing a hard test of the strength of ingroup policing because of how divided Israeli society was over the incident. Although the incident occurred in 2016, it remained relevant throughout 2017 and 2018 as Elor went to trial, was sent to prison, and was then released just months before the replication experiment began.

Recruitment

Using *Mahshov*, a well-known survey company, I recruited Israeli Jews from the general population living within 15 kilometers of Haifa to participate in both of the experiments.¹³ Participants took an online, unpaid intake survey and received NIS 150 (approximately US \$40) if they were selected for a “political focus group.” A variety of demographic measures were taken to aid with block randomization (political ideology and gender were used for this purpose) and compare my sample to the general population.

From the subject pool recruited in this manner, I block randomized sets of 4–8 participants to each focus group. Note that this selection process was independent of treatment assignment, which was done at the group level. After two pilot focus groups on July 10 and 14, 2016, the project took place over a 4-week period from July 17 to August 12. I invited 149 individuals to participate in a focus group, of which 81, or 54%, attended. Similarly, 164 individuals were invited to participate in 2018, of which 86, or 52%, attended. Overall, treatment groups in both samples were well-balanced on observables and somewhat representative of Israeli Jews as a whole, although the sample skews younger, more female, less religious, and more centrist politically than the population.

Throughout the experiment, subjects were told that the focus groups were being run under the auspices of an initiative called “Mikud 2016” (Focus 2016 - labeled Focus 2018 in the second experiment). They were told that Mikud was a non-partisan initiative to study public opinion on current political issues and that they could contact a representative for more information (none did). This deception was used to mask the fact that a foreign researcher was running the project and to avoid demand effects. Participants were debriefed after the experiment concluded

Focus Groups

Focus groups took place on weekday evenings (Sunday–Thursday) and Friday mornings. Participants were welcomed by a male moderator. They were given instructions, signed consent was obtained (including consent for recording the discussion), and then the focus group began. As an ice-breaker, participants stated their names, occupations, and a hobby. The moderator then introduced the Hebron incident, which was followed by three questions. First, participants were asked whether they felt military and political leaders

¹³I also recruited through references from participants (note that participants were *already* assigned to a focus group when references were sought, and individuals who knew each other were not allowed to participate in the same group) and my own recruitment using flyers in local neighborhoods.

should have publicly condemned the soldier's actions. Second, they were asked to describe what the criteria for public condemnation should be. Lastly, they were asked to describe what they would have done as Defense Minister. Each person was given 1–2 minutes to respond, after which a moderated discussion took place for 10–15 minutes per question.

The moderator and confederate were given careful instructions. Both were told to maintain consistent dress and grooming. The moderator was instructed to let the confederate answer each question first, with the confederate always assigned a seat to the right of the moderator to make this more natural. The confederate's answers followed a detailed script, in addition to a detailed set of arguments to counter likely responses from other participants. The script and arguments were developed with research assistants in response to the proceedings of pilot focus groups. The confederate was instructed to enter the conversation at least twice per question outside of his initial statement. This let the confederate be as consistent as possible while still adapting as necessary to each focus group. The confederate was also instructed to maintain the same, somewhat firm and aggressive tone in all conditions so that differences in tone would not be responsible for treatment effects. The confederate took one of three positions depending on the treatment assignment: a neutral position, a position in favor of condemning the soldier using a moral frame, and a position in favor of condemning the soldier using a pragmatic frame (see Table A.3 in the Appendix).

One of the only differences between the two experiments is the gender of the confederate, who was female in the 2018 experiment. Extra care was taken to help this confederate mimic the arguments of the confederate in the 2018 experiment.

Outcome and Analysis

Participants provided outcome measures as the focus group was taking place as well as in a survey administered afterwards. In a survey, they provided answers to questions that compose six separate indices as well as participated in a letter writing exercise. Participants also consented to a recording of the focus group proceedings, which was used to transcribe each participants' statements as an additional outcome measure. All of these outcome measures are outlined in greater detail in the Results section.

For the survey outcomes and the letter writing exercise, I use multiple regression to estimate treatment effects. There is no need for clustering at the group level because participants were randomly assigned to groups. Even though the confederate's assignment varied at the group level, the treatment assignment of the individuals in a given group are independent from each other because of the random assignment to groups.

Because of this, the potential outcomes of any given individual are not correlated with those of any other individual in the same group. For analyzing the discussion group transcriptions, I use a structural topic model, which allows one to estimate the relationship between textual metadata and topics present in the texts. In the corresponding sections of my results, I describe this model as well as processing done on the text in greater detail. For pooled results, which appear in discussing the results of the survey and the letter writing exercise, I use an inverse variance fixed effect model, as is commonly used in meta-analyses.

As a part of my pre-analysis plan, I also registered models with control variables for analysis. Results change little, if at all, as a result of including controls, but these models can be found in the Appendix.

3.4 Results

Omnibus balance tests (see Appendix) indicate that randomization was successful and no major imbalances appear in the data. In each of the following sections, I address individual outcomes from the survey, letter writing exercise, and textual data. Outcomes are shown in plots; table summaries of the regressions and pooled analysis can be found in the Appendix.

Oppose Soldier

Did the confederate convince others to be more critical toward the actions of the soldier? To measure this outcome, participants were asked their level of agreement on a scale of 1–7 (from “Completely agree” to “Completely disagree”) with four statements about the soldier’s actions. Two of the statements were in favor of condemnation (e.g. “The soldier Elor Azaria made a mistake when he shot the terrorist”), while two were against. The four items (two reverse-coded) about the soldier form a single index, “Oppose Soldier,” that provides a measure of the extent to which the treatments caused participants to adopt critical attitudes toward the soldier that misbehaved. Higher values indicate opposition to the soldier’s behavior.

The results for this and all other survey results are shown in Figure 4.1, and the top-most panel of Figure 4.1 provides the results for the “Oppose Soldier” outcome. Overall, there is evidence that the moral treatment increases critical attitudes, but we fail to reject the null of no effect for the pragmatic treatment. The effect of the moral treatment is significant at 0.92 (0.06, 1.79) in the first experiment and null at 0.48 (-0.46, 1.40) in the second. The pooled analysis gives a significant effect estimate of 0.72 (0.09, 1.35), equivalent to an approximately 12% increase in opposition to the soldier’s actions. By contrast, the prag-

matic treatment produces effect sizes of 0.70, -0.16, and 0.30 in the first experiment, second experiment, and pooled analysis respectively; none of these estimates are statistically significant at the 95% level.

Approve Leaders

How do the treatments affect how people view the leaders' condemnation of the incident? To measure this outcome, participants were asked their level of agreement on a scale of 1–7 (from “Completely agree” to “Completely disagree”) with two statements about the leaders who condemned the soldier's actions. The leaders were not specified by name, with the text instead referring to the fact that a “number of leaders” had come out in condemnation of the incident. However, during the group discussion, the three highest-ranking officials who condemned the incident—Prime Minister Netanyahu, Defense Minister Moshe Ya'alon, and IDF Chief-of-staff Gadi Eizenkot—were specified by name. One of the statements was in favor of condemnation (i.e. “The political and military leaders' coming out publicly against the soldier's actions was justified”), while the other was against. The two items (one reverse-coded) about the leaders' condemnation form a single index, “Support Leaders,” where higher values indicate support for the leaders' condemnation. This provides a measure of support for elite condemnations, which are one of the most common, immediate, and important responses to ingroup misbehavior.

The second panel of Figure 4.1 shows that there is evidence that the pragmatic treatment increases support for elite condemnations, while also indicating that we cannot reject the null of no result for the moral treatment. The effect of the pragmatic treatment is significant and estimated at 1.23 (0.39, 2.07) in the first experiment, null and estimated at 0.10 (-0.66, 0.87) in the second, and significant at 0.66 (0.10, 1.23) in the pooled results. The pooled result is equivalent to an 11% increase in support of the leaders' condemnations. By contrast, the moral treatment produces effect sizes are 0.80, -0.40, and 0.21 in the first, second, and pooled analyses, with marginally significant effects in only the first.

Attitudes toward Confederate

How do the treatments affect how people view the confederate? This variable measures the costliness of engaging in ingroup policing in the interpersonal setting of these group discussions. To measure this outcome, participants rated each of the other focus group participants by noting their agreement on a scale of 1–7 (from “Completely agree” to “Completely disagree”) with two statements: “The participant's arguments were convincing” and “The participant contributed to the discussion.” To aid in making this question easy to

answer, participants wrote their answers to these questions in designated spots next each chair in a diagram of the room. The moderator also drew a version of this diagram on the whiteboard with the name of each participant next to the chair they were sitting in. The ratings that each participant gave to the confederate were reverse-coded, added together, and then subtracted from the reverse-coded average rating of the other participants for a measure called “Confederate vs. Others.”¹⁴

The third panel of Figure 4.1 indicates that we can only reject the null of no effect for the moral treatment with this outcome. The moral confederate has a significant average difference of -0.99 (-1.89, -0.09) points in the first experiment, a null difference of -0.46 (-1.14, 0.23) in the second, and significant difference of -0.75 (-1.30, -0.19) in the pooled analysis. Thus, in general, it appears that those engaging in moral ingroup policing in similar scenarios would risk being down-rated by three-fourths of a point compared to their peers. Compared to the moral treatment, the estimate for the pragmatic treatment is -0.22, -0.31, and -0.28 in the first, second, and pooled analyses respectively, and it is not significant in any of them.

The finding that the confederate is viewed more negatively in the moral condition is in line with findings from the “moral rebels” literature in social psychology. Combined with the results from the previous two sections, which measured post-treatment attitudes toward the soldier’s behavior and the elites who condemned his actions, the findings thus far provide evidence for the fact that support and costliness are distinct features of ingroup policing. In the moral condition, participants indicated that the moral confederate was unconvincing; however, in spite of this, their attitudes changed based on his behavior. This particularly interesting finding illustrates the importance of differentiating between sentiment toward the one engaging in ingroup policing or the frame they use on the one hand and the ability of the criticism to change attitudes and behavior on the other. It also bodes poorly for the critic’s long-term ability to convince others, since the critic’s credibility may decrease over time in a situation of repeated interaction.

Spillover Effects

Outside of effects on the support for and costliness of ingroup policing, these frames may have effects on more general attitudes. How does exposure to someone with controversial views affect one’s willingness

¹⁴Note that this differs slightly from my pre-analysis plan for the first experiment, where I stated that I would use the confederate’s ratings by themselves as an outcome. I used the difference measure because subtracting others’ scores provides a more precise measure due to some participants giving out low/high ratings across the board. Using the outcome measure in the pre-analysis plan for the first experiment provides results that are in the same direction but not significant. The pre-analysis plan for the second experiment specifies the outcome measure used here.

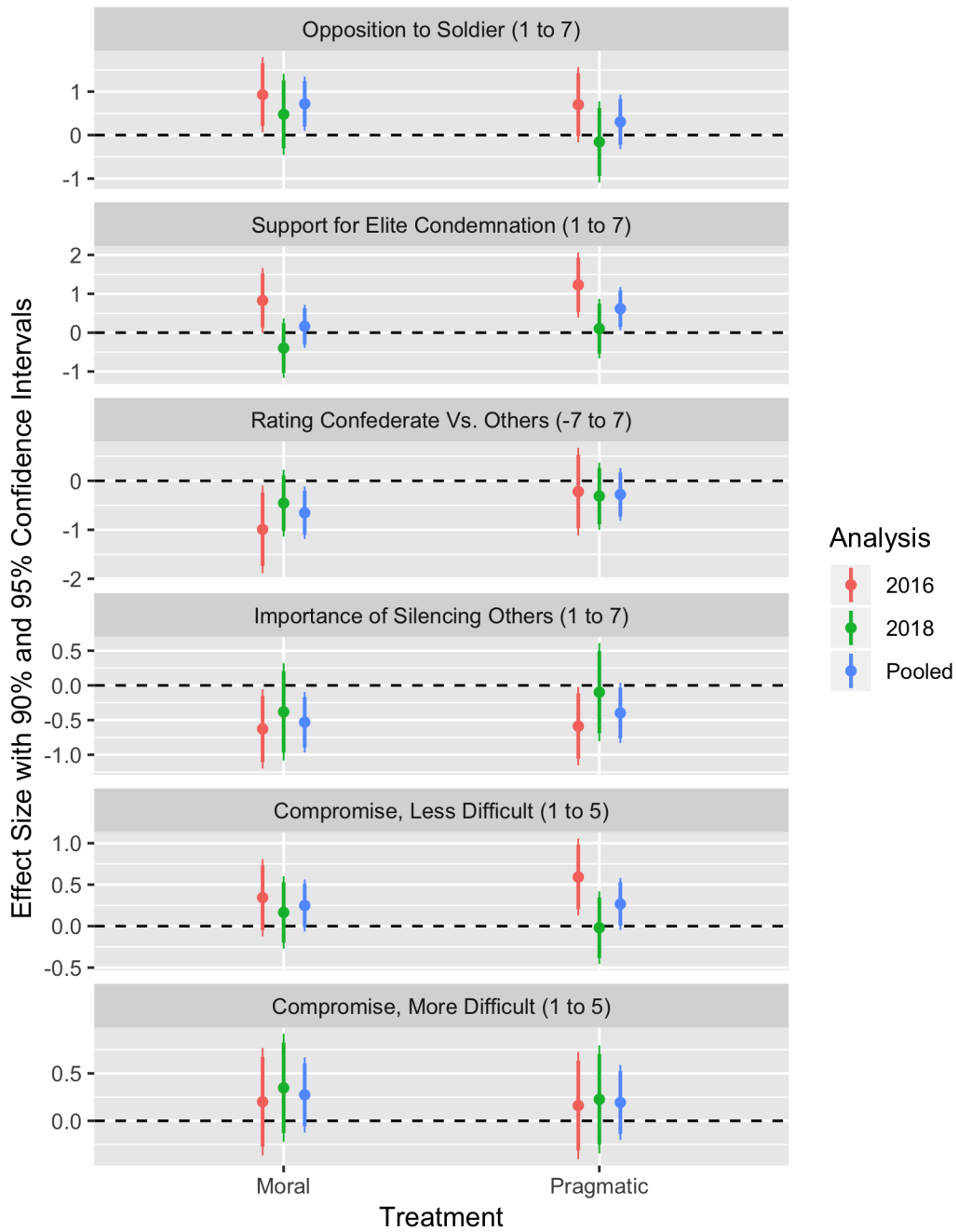


Figure 3.1: Treatment effects by outcome and analysis. Effect estimates are with the neutral condition as reference level.

to listen to those with different views more generally? Do the treatments have any affect on attitudes toward the Israeli-Palestinian conflict at large? Three additional outcomes in the survey examined both of these questions.

First, general attitudes toward censoring views one disagrees with were measured using the “belief in the importance of silencing others” index (results are in the fourth panel of Figure 4.1). Participants were asked about the importance of silencing those with dangerous or illegitimate views by noting their agreement on a scale of 1–7 (from “Completely agree” to “Completely disagree”) with eleven statements (e.g. “There are situations when it is more important to silence certain views than letting everybody express themselves”). These questions come from Tsfati and Dvir-Gvirsman’s (2018) Belief in the Importance of Silencing Others (BISO) scale. The eleven items (all reverse-coded) form an index called “Silence Others.”

Overall, there is some evidence that both treatments decrease participants’ desire to censor views they disagree with. This result is significant in the pooled results for the moral treatment with an effect size of -0.50 (-0.95, -0.05) and marginally significant at the 90% level for the pragmatic treatment with an effect size of -0.33 (-0.78, 0.11). Both treatments are significant in the first experiment and null in the second. In relative terms, the pooled effect size is equal to a 8.3% decrease in beliefs in the importance of silencing others.

Second, general attitudes toward compromise with the outgroup were measured using two indices. The first index measures attitudes toward less difficult compromise issues, and the results are found in the fifth panel of Figure 4.1. For this index, participants were asked to indicate their level of support on a scale of 1–5 (from “Completely oppose” to “Completely support”) for five statements. Four of these statements focus on compromises or aggressive policies on contemporary issues (e.g. “Israel needs to act aggressively and make Palestinians pay a meaningful price for the harm they inflict on Israeli citizens”, reverse-coded), while the final item addresses risk-taking in compromise. The second index measures attitudes toward more difficult compromise issues, and the results are found in the sixth panel of Figure 4.1. For this index, participants were asked to indicate their level of support on the same scale for two statements addressing compromise over land swaps and the appropriate level of force to be used against large-scale Palestinian demonstrations in the occupied territories.

There is little evidence for treatments affecting participants’ willingness to compromise over less difficult issues. In the first experiment, the pragmatic treatment has a significant and positive effect on this outcome with an effect size of 0.59 (0.13, 1.06). However, neither treatment is significant outside of this

result, and the pooled effects for both treatments are statistically insignificant and hover at around 0.25, equivalent to only a 6.25% change in the outcome. There are no significant effects when examining the second set of more difficult compromise issues.

Letters

To provide a behavioral measure of change in support for elite condemnation of the incident, participants were also asked to take part in a letter writing assignment as a part of the survey. Participants were asked to write a letter to the leaders who condemned the soldier's behavior. As with the survey question addressing support for elite condemnation, these leaders were not specified by name; instead, participants were asked to write a letter to "the leaders who condemned the incident." They were given the option of writing a letter of support or opposition to these leaders' condemnation, and they were provided three prompts to encourage their writing of the letter. The first prompt asked them to explain their position with their strongest points. The second prompt asked them to provide an explanation of why their position was personally important to them. The final prompt asked them to explain the type of response they would have wanted from the leaders. This template was based on form letters that are common in letter writing campaigns by activist organizations.

To provide a numerical measure of support of elite condemnation, research assistants examined and coded participants' responses. Two research assistants coded these letters on a five-point ordinal scale (from "greatly oppose" to "greatly support") based on the position the participant took. This range of coding values reflected the range of responses that participants offered in their letters; most letters (almost three quarters) fell into the range of "opposing" or "greatly opposing" elite condemnation. Several themes were common in such letters. Many of them focused on the difficulty of a soldier making snap decisions in the field; others discussed the importance of elites reserving judgment in public settings until the case could be brought to court; yet others focused on the guilt of the Palestinian who was killed, arguing that he got what was coming to him as a terrorist or stating that he could have still posed a threat. Many of these arguments are illustrated in the following letter from the 2016 experiment that was rated as greatly opposing elite condemnation by both coders:

"In my opinion, there was absolutely no reason to condemn the actions of the soldier who shot the Palestinian. The Palestinian is a terrorist who came to kill, so getting killed was always a

possibility and a part of the risk [*he was taking*]. The soldier did the right thing because he suspected that the terrorist still constituted a threat to others. In cases of killings of Palestinians, one has to act definitively because such cases can end differently—he could have gotten back up and continued to stab. The leaders should have been more restrained, publicly condemning the incident from the get-go is damaging for the country. There should have been an explanation to the soldiers, but it is unequivocally important to understand that 'if someone comes to kill you, rise up and kill him first' [*quote from the Talmud, a primary text of rabbinical Judaism*]. The soldier deserves all the backing we can give him; he behaved according to how he felt at the time of the incident.”

On the other hand, over a fifth of these letters supported elite condemnation. Again, several themes were common in these letters, but the most common was an appeal to the code of ethics of the Israeli Defense Forces. Many argued that, in such a clear case with video evidence showing a violation of these rules, there could be no other choice aside from publicly condemning the soldier’s actions. Some made such an argument dispassionately, while others’ letters were filled with emotion. The following letter, from the 2018 experiment and rated as greatly supporting elite condemnation by both coders, clearly falls into the second camp:

“I support every condemnation, and it’s a tragedy that they weren’t even stronger. We’re talking about a lawless soldier who defied his orders and flagrantly violated the IDF’s regulations, in violation of the rules of engagement and against the morals of the army. The issue is important to me primarily because of the need to educate the rising generation. A helping hand against society’s descent into savagery. Woe to the country for whom such are its heroes. Take the state’s entire budget and invest it in the ministry of education instead!”

In rare cases where the coders’ ratings differed by two points (they never differed by more than two), the author adjudicated between them. Otherwise, the ratings were averaged to provide a single outcome measure. Inter-coder reliability was high, and overall there was very little disagreement. In 62% of cases there was no difference in categorization, in 34% they differed by one point, and in the remaining 4% they differed by two points (quadratic weighted $\kappa = 0.8$). During the coding process, coders were blind to the treatment assignment of the individual who wrote the letter as well as the rating that the other coder had given the document.

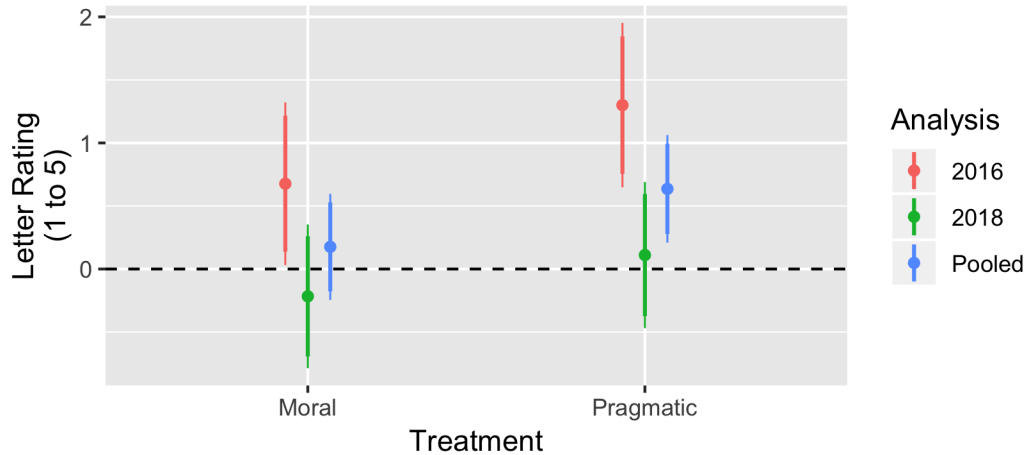


Figure 3.2: Treatment effect on letter ratings by analysis. Effect estimates are with the neutral condition as reference level

This coded outcome measure was analyzed using regression, as with the other outcomes, and the results of this regression can be found in Figure 3.2. The results are generally similar to the survey measure of the treatments' affect on support for elite condemnation of the incident. As with the survey measure, higher values indicate more support for elite condemnation. The effect of the pragmatic treatment is estimated at 1.30 (0.65, 1.95) in the first experiment, 0.11 (-0.47, 0.69) in the second experiment, and 0.64 (0.21, 1.06) in the pooled results. The pooled result is equivalent to a 16% increase in support of the leaders' condemnation as a result of the pragmatic treatment. With regards to the moral treatment, it has a marginally significant effect of 0.68 (0.03, 1.32) in the first experiment. However, this effect becomes negative and insignificant in the second experiment with an estimate of -0.22 (-0.79, 0.35), and the pooled results are also null with an estimate of 0.18 (-0.24, 0.60). These results mirror the findings in the survey and demonstrate some evidence that the pragmatic treatment affects participants' approval of elite condemnation but that a similar effect for the moral treatment is either weak or non-existent.

Textual Analysis

Aside from the letter writing task, the most prominent additional measured behavior of participants in response to the treatments is the content of their responses to the moderator's questions and the group discussion. Their speech is a unique and useful insight into how they reasoned with and reacted to what the confederate was saying. A research assistant transcribed the proceedings of every focus group, marking

who spoke and the time at which they began speaking for every utterance in the group.¹⁵ Because of this, I have both the words participants spoke as well as how long they spoke.

I use computational textual analysis to analyze the transcriptions of these focus groups, treating as each “text” the entirety of a participant’s speech during their assigned focus group. The model I use is a structural topic model,¹⁶ which allows me to estimate the relationship between treatment assignment and what people said in the focus groups. Using this method requires specifying the number of topics present in the texts being analyzed. There is no “right” or “wrong” answer to this; for simplification and ease of interpretation, I run models with ten topics for both the 2016 and 2018 experiments. The topics are computationally determined independently for each experiment and therefore direct comparisons of specific topics between the two experiments are not possible. However, I note cases where there are similarities between topics across the two.

As is standard in approaches using computational textual analysis, some pre-processing of the texts was done prior to analysis. I limited analysis of the transcripts to the portion that occurred after the moderator introduced the topic of discussion. I also removed numbers, punctuation, and a set of Hebrew stop words. Lastly, I used textual processing software available through the Knowledge Center for Processing Hebrew, an initiative of the Israel Ministry for Science and Technology, to “stem” the texts.¹⁷ This process involves first “tokenizing” the text (splitting it into meaningful units like words, sentences, and paragraphs) and then conducting morphological analysis to make a best guess at the root of each word as well as the part of speech, gender, definiteness, etc. for the word. Completing this pre-processing greatly reduces the number of unique elements that characterize a given text, making computational analysis more feasible.

After pre-processing, I complete the textual analysis in three steps. First, I prepare the documents by removing any words that appear in only 10 or fewer of the texts.¹⁸ Second, I select a model. I explore models with various numbers of topics (ranging from 2 to 20) to establish that a model with 10 topics is satisfactory. Then, I explore 50 possible 10-topic models, choosing a model for each experiment that balances exclusivity and semantic coherence (see the Appendix for more detail). Third, I estimate a regression with topic-

¹⁵Provide some detail here about what happened in areas where the research assistant couldn’t understand what was going on.

¹⁶From the `stm` package in R (Roberts et al. 2014).

¹⁷See their website and the software I used at <http://www.mila.cs.technion.ac.il/about.html>; I am extremely grateful and indebted to Michael Freedman for pointing me in the right direction with this software.

¹⁸This step does not result in the removal of any texts from the dataset.

proportions as the outcome variable and treatment assignment as the predictor for all 10 topics within each experiment.

The results of the topic-proportion regressions can be found in Figure 3.3, which shows on the x-axis the effect estimate (which is an estimate of the effect of treatment on a given topic's proportion in those texts) and 90/95% confidence intervals and treatment assignment on the y-axis. The labels for each topic are labels that I created based on the words that are both exclusive to and appear frequently in a given topic; the appendix lists the set of words in English and Hebrew for each topic that were used to determine the labels. Note that the labels for each topic are not the same between the 2016 and 2018 experiments; because these models are estimated separately, the topics necessarily differ as well. I matched up these topics to the best of my ability, so for many of the rows the most similar topics from 2016 and 2018 are paired, but direct comparisons cannot be made between specific topics. Figure 3.3 displays, from top to bottom, the set of topics with the strongest estimated positive effect of the treatments, then topics with the strongest negative effect, and finally the topics with the smallest estimated effect sizes.

The 2016 results indicate a significant effect of the treatments on the extent to which participants discussed several topics. For the moral treatment, there are two topics particularly affected. The first of these, "Quick Life-or-Death Decisions," is characterized by words such as "kill," "come," "weapon," and "all-of-a-sudden." The model indicates that the proportion of the texts in the moral condition that fall into this topic is 0.14 (0.04, 0.24) greater than the proportion for the neutral condition. Inspection of texts that are highly associated with this topic provides greater insight into its significance. These texts focus on the fact that difficult decisions have to be made quickly by soldiers in the field and that the soldier Elor Azaria either faced or helped prevent a real threat of additional killings or injuries (note that this is not supported by video of the incident, in which Elor's killing occurs while dozens of other Israeli personnel are milling about the scene). The greater prevalence of this topic in response to the moral treatment indicates a sort of backlash against the moral confederate's position—the confederate's position leads participants to increase their efforts to provide justifications for the soldier's actions. The following excerpt from a participant in the 2016 experiment highlights many of the arguments typical of texts that feature this topic:

"... The man came to kill. And he needed to take into account that he might die. I think that perhaps the soldier had a lapse in judgment, but I [also] think that, at the moment, he felt threatened, at the moment that - because if, God forbid, the incident had ended differently and

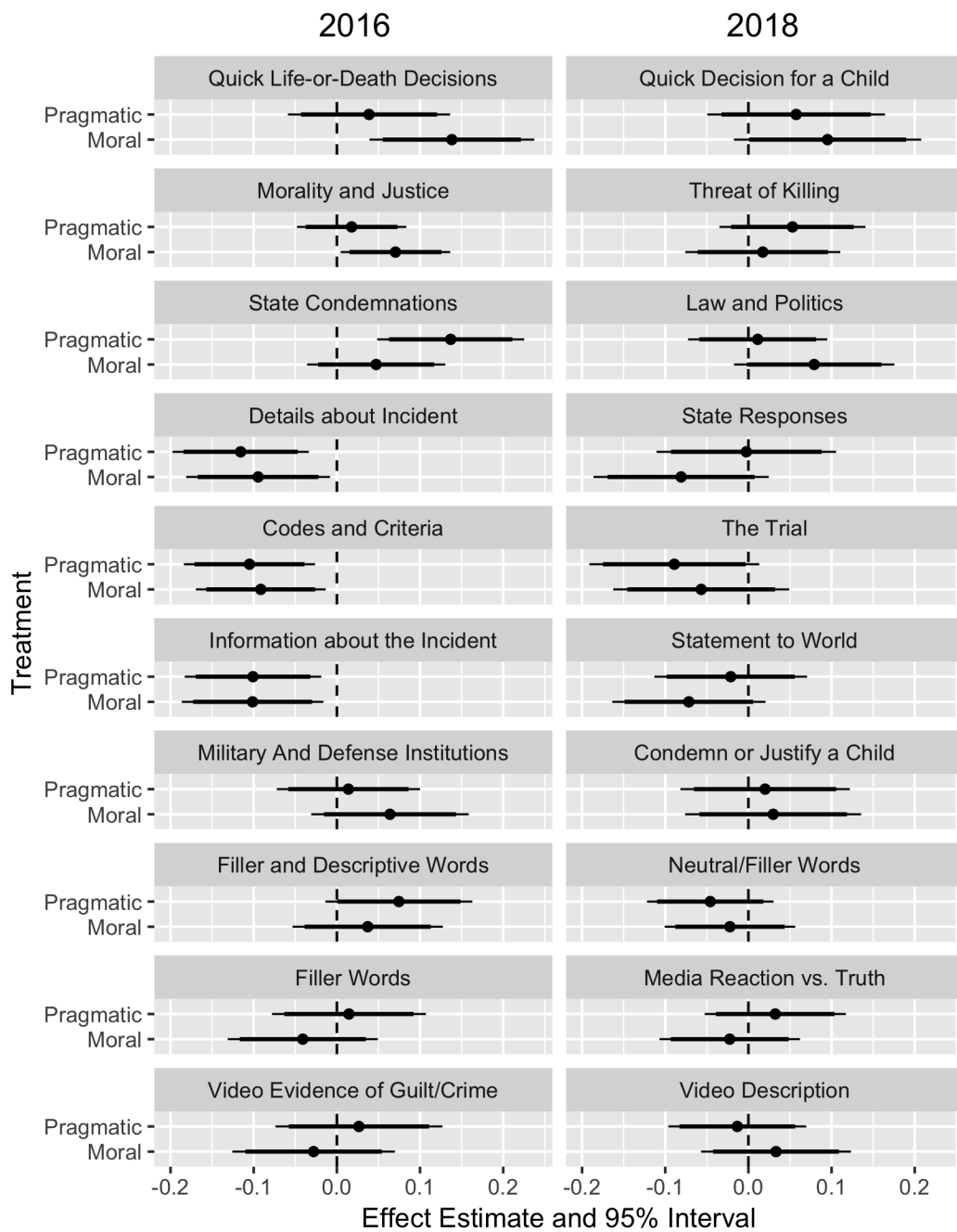


Figure 3.3: Treatment effect on words spoken in focus groups, from textual analysis of transcripts from the focus groups.

this neutralized Palestinian somehow got back up - I mean who knows - and wounded or killed people, then suddenly everyone would say 'hey, where were the soldiers?' ...”

The second topic affected by the moral treatment is one I've labeled “Morality and Justice.” This topic is characterized by words such as “moral,” “justice,” “opinion,” and “moment.” The model indicates that the proportion of the texts in the moral condition that fall into this topic is 0.07 (0.01, 0.14) greater than the proportion for the neutral condition. Texts that feature this topic prominently are not necessarily characterized by backlash but instead feature rich discussions about what is necessary to defend the purported morals of the armed forces. For instance, in one such text, the participant seems convinced that the actions of the soldier were not justifiable according to the values of the IDF but still hesitates about the propriety of military and political elites publicly condemning the incident:

“... As defense minister, I would have said: I can't condemn the incident because a ruling hasn't been issued. There must be an investigation and judgment. When that's over, then I can condemn the incident, once the truth is known. After justice is served, a judgment is issued. And aside from that, we don't agree on the role of the military. Why has an investigation been opened? Because there is suspicion. This means that I personally, as defense minister, don't agree with incidents like this that damage military morals and the values of the IDF.”

The pragmatic treatment, on the other hand, affected how much the discussion focused on a topic I've labeled “State Condemnations.” This topic is characterized by words such as “condemnation,” “leader,” “state,” “law,” and “statement.” The model indicates that the proportion of texts in the pragmatic condition that fall into this topic is 0.14 (0.05, 0.23) greater than the proportion for the neutral condition. The meaning of this topic is fairly intuitive and directly related to the pragmatic condition, with participants spending a greater amount of time discussing whether the country's interests were hurt as a result of this incident as well as what responsibility Israel's leaders have to condemn these types of incidents. For some participants, such as the one who provided the following comments, the moderator's position greatly resonated with them:

“My thinking is just like Eyal's [*the confederate*] ... [*speaking about the immediate aftermath of the attack*] before this, he had already decided 'I want to shoot him.' This really did harm to us, because it has come to - indeed its aftermath has brought us to this debate and all people are

divided. Judgments, yelling, threats - this is not what we need in this country at all, we don't need this thing. Killing terrorists - that's necessary. But doing this? There are other ways."

Aside from these effects, the treatments also lead to a decrease in participants' discussions of certain topics. These include "Details about Incident" and "Information about the Incident," both of which seem to discuss generic details of what happened. The topic "Codes and Criteria" is related to the second discussion prompt used in the focus groups, which asked participants to indicate what types of incidents they felt would warrant a condemnation from the country's leaders. It would appear that more time was spent on that prompt in groups assigned to the neutral condition. Additional categories from the 2016 experiment, most of which also focus on descriptions of the incident or filler words, were unaffected by the treatments.

Explaining the Difference between 2016 and 2018 Results

Overall, effects trend in the same direction in the first and second experiments. The moral treatment tends to positively affect how critical participants are toward the soldier's actions, negatively affect how participants view the confederate, and decrease participants' support for silencing those with controversial views. The pragmatic treatment tends to positively affect participants' support for elite condemnation of the incident and may increase participants' support for compromise in the conflict.

In spite of these overall trends, effects are much weaker in the second experiment compared to the first. Great care was taken to make these two experiments equivalent, including using the same recruitment method and same experimental instruments. For this reason, the two samples do not differ greatly on observable covariates, as noted in Table A.2 in the Appendix. The second experiment has a slightly greater proportion of female participants (65% as opposed to 58%), and income and education distributions vary slightly between the two. But, for covariates that are likely to matter the most, such as political ideology and military service, show only marginal differences between 2016 and 2018. It is unlikely that differences in the sample itself contribute to the difference in outcomes.

There are also matters of experimental implementation that differed between the two experiments. Some of these differences are small, but they may provide an explanation for the difference. First, group sizes were on average just under 1.5 people larger in the second experiment than the first. Larger groups resulted in longer sessions (on average 53.5 minutes in the 2018 experiment, as opposed to 48 minutes in the first), which by itself might indicate stronger treatments the second time around. However, because of the greater

number of people, individuals' participation decreased in spite of the longer time. Participants each spent 13 minutes talking on average in the first experiment, as opposed to 9 minutes in the second. As instructed, the confederate participated as much time or more than others talking in the group; however, the amount of time the confederate spoke also decreased, from 15 minutes in the first experiment to 11.5 minutes in the second. Lastly, the gender of the confederate differed between the experiments, with a female confederate administering the treatments in the second experiment. Participants may exhibit a bias against a female confederate, being less convinced by her arguments. Ultimately, I cannot rule these out as an explanation for the difference in my findings from 2016 to 2018.

The political context surrounding the incident, which changed over the two years between experiments, may have also been a factor. During the administration of the 2016 experiment, the incident was fresh and constantly in the news. The soldier's trial, which began in May of that year, extended throughout the summer and fall, with closing arguments taking place six months later in November, 2016. While participants discussed the incident in my focus groups, several important developments in Azaria's case took place, including his cross-examination by the prosecutors in late July. But this is not to say that the incident had disappeared from public consciousness by the time the 2018 experiment was administered. After serving a nine-month sentence, Azaria was released from prison in May 2018, just two months before the second experiment began. He remained relevant enough after the administration of the experiment that one politician recruited Azaria to appear in his campaign ads as a part of the Israeli elections held in February, 2019. Perhaps more importantly, however, decreasing salience is not a strong explanation for diminishing effect sizes over time; one could easily make the argument that it would be easier to change individuals' minds after tempers had cooled and the incident had decreased in its importance.

There is an additional potential explanation that I consider more powerful: with two years having passed since the incident, participants in the 2018 experiment may be more solidified in their opinions than those in the first experiment. One indication of this lies in the correlation between political ideology and the outcomes for those in the control condition. This correlation is often stronger in the second experiment than it is in the first, and this pattern is especially true for outcomes regarding support for elite condemnation. For instance, in the control condition, the correlation between political ideology and the survey index measuring support for elite condemnation in the first experiment is -0.19, while in the second it is -0.44. Similarly, the correlation with the position participants took in their letters is -0.36 in the first experiment and -0.63 in the second. It may be that the interpersonal persuasion my experiment tests is strongest in the immediate

aftermath of incidents like the Elor Azaria case, and that the effectiveness of any attempts at persuasion decrease over time as peoples views calcify.

3.5 Conclusion

This study assessed ingroup policing, an important potential mechanism for conflict reduction in situations of intractable conflict. Of the two frames for ingroup policing that I evaluated, some outcomes were affected by both while others were not. Both were effective at increasing critical attitudes toward an ingroup member, although estimates were higher for the moral condition. However, the moral confederate received relatively negative ratings by other participants, indicating a resistance to the confederate as a “moral rebel” even if his message was effective.

This project provides two important contributions to the study of prejudice reduction and conflict resolution. First, the project shows that ingroup policing, in the form of critical ingroup dialogue, can provide an effective alternative to intergroup-focused conflict resolution strategies, such as intergroup dialogue. At least one of the critical frames used worked at moving outcomes in each of the three categories of outcomes that I examined, each of which is substantively important to conflict resolution. This does not mean that ingroup policing should replace intergroup dialogue; there is value in intergroup contact that cannot be achieved through ingroup deliberation. However, ingroup policing holds great potential for complementing intergroup dialogue programs. For instance, in a given conflict, there are some who are unwilling to engage with the other side, even in the relatively safe spaces provided by intergroup dialogue programs. These same individuals may be willing to, instead, hold a dialogue with other members of their own group. Even if such a program does not achieve the same outcomes as intergroup dialogue, it may still be effective at moving participants on certain important outcomes, as illustrated by the findings of this project. This is just one example of a scenario where ingroup policing can effectively complement existing programs.

However, in considering the application of ingroup policing, certain conditions should hold to obtain results like those in this study. First, all those participating in these discussions should be ingroup members, as the presence of outsiders is likely to change the willingness of participants to fully debate their own group’s actions toward an outgroup. Second, the presence of an articulate ingroup critic who is dominant in the conversation is necessary. The confederate in my study got to answer every moderator question first, and played a very active role in the discussion. Ingroup dialogue programs without a dominant critic are

unlikely to elicit the same reactions as this study. And, third, the discussion should focus on the appropriate boundaries for the behavior of ingroup members who are not participating in the dialogue, rather than on the actions of those present in the discussion. Otherwise, ingroup policing is likely to feel like an attack against particular participants.

Second, this project illustrates that there is a very real set of tradeoffs that confront individuals or organizations who wish to take critical stances in a situation of conflict. A critical stance that is explicitly based in the moral values of the society and its institutions can be powerful, as demonstrated by the extent to which participants were moved by the moral frame to adopt more critical attitudes toward the Hebron incident. However, a stance based in more instrumental concerns can be effective at changing other outcomes, in particular in this study outcomes that are security-related. And, most importantly, a moral frame resulted in the lowest relative confederate ratings, indicating that such a frame comes at a cost to the perceived legitimacy of the critic's position. In the long-term, then, such a frame may prove counter-productive if its message comes at the price of de-legitimizing the critic, whether an individual or organization, in the eyes of those it is trying to persuade.

Generally, I limit the generalizability of this study's findings to other situations of intractable conflict, but one could think about applying a similar design to study the effectiveness of ingroup policing in the context of the U.S. or other non-conflict zones. In such contexts, it would be important to consider which critical frames are likely to resonate. And even in situations of conflict, future research should look at a variety of frames to see how different types of frames can affect these outcomes.

4 | Persecuted Minorities and Prosocial Behavior: Accountability and Public Goods Provision among Hindus and Muslims in Delhi Slums

This chapter is derived from co-authored work of the same title with Melani Cammett and Poulomi Chakrabarti.

4.1 Introduction

The governing Bharatiya Janata Party (BJP), which first gained national power in 2014 and won a resounding victory in the 2019 national elections in India, ran its campaign on a strong Hindu nationalist agenda, with an emphasis on preserving national security imbued with anti-Muslim rhetoric.¹ Within weeks of reassuming office, the BJP government removed the special autonomy of the Muslim-majority state of Jammu and Kashmir and passed a new citizenship law designed to fast-track Indian citizenship for migrants from neighboring countries of all religions except Islam, provoking massive demonstrations throughout India. Protests in Delhi escalated into the worst episode of Hindu-Muslim riots in three decades, with widespread accusations that the police, which are controlled by the central government, aided Hindu mobs (Gettleman and Abi-Habib 2020) and some describe the Hindu nationalist program of the current government as nothing short of genocide (Filkins 2019; The Wire 2020; Subramanian 2020; Cockburn 2020). As a result, Muslims are a persecuted minority in India and are particularly vulnerable at this historical juncture. How does the experience of marginalization affect behavior in day-to-day life at the local level? Does membership in a persecuted group shape interactions with community members in realms of everyday life

¹The Prime Minister, Narendra Modi, and the party President, Amit Shah, have faced criticism for not condemning, and on several occasions, even encouraging widespread violence against Muslims throughout their political careers. Both spent their formative years in the Rashtriya Swayamsevak Sangh (RSS), the Hindu-nationalist organization regarded as the ideological backbone of the BJP, and they began their political careers in the state of Gujarat. The 2002 Gujarat Pogrom, which killed more than 2,000 Muslims, took place under Modi's Chief Ministership.

that ostensibly have little to do with politics?

These questions have implications for the dynamics of local public goods provision in heterogeneous communities, particularly among residents of unequal social status. A large body of work in the social sciences explores the relationship between ethnicity² and development outcomes. Across developed and developing countries alike, studies overwhelmingly find that ethnic diversity impedes public goods provision (Alesina and Ferrara 1999; Banerjee, Iyer, and Somanathan 2005; Khwaja 2009; Miguel and Gugerty 2005; Putnam 2007). In fact, the negative relationship between social cleavages and development has been described as the “most powerful hypothesis in political economy” (Banerjee, Iyer, and Somanathan 2005, p. 639). Most research in this vein, however, focuses exclusively on the role of ethnic diversity, while the role of hierarchy or status of ethnic groups and exclusion remains largely overlooked. Efforts to identify the mechanisms underlying the negative association between diversity and public goods provision point to the role of social norms and networks in coordinating ingroup collective action to supply or lobby for public goods and to sanction non-contributors (Habyarimana et al. 2009). In this paper, we explore differences in the role of such informal social accountability measures in driving cooperation around public goods provision in majority versus minority groups. Do members of a persecuted minority group respond differently to norms-based accountability measures around contributions to local public goods provision?

Based on a survey experiment in slum settlements in Delhi, we test the effects of two distinct types of social accountability mechanisms on people’s willingness to cooperate in improving drainage and sewage in their neighborhoods. These include horizontal accountability, understood as accountability among citizens of relatively equal standing through measures like public shaming and gossip, and vertical accountability, or enforcement of social norms through pressure from local elites, such as community leaders and chiefs. In particular, we present respondents with a hypothetical neighborhood initiative to hire a private firm to clean and maintain drains in the community. First, the experiment randomly exposes respondents to a favorable testimonial about the hypothetical drainage scheme by a nearby resident, varying the treatment to assess the effect of perceived ingroup underperformance, or information indicating that people with a recognizable Hindu or Muslim name from the respondent’s own community failed to contribute to the initiative. Second, the experiment presents distinct ways of promoting accountability for potential contributions (or lack

²In line with standard social science applications, we use the term “ethnicity” to encompass diverse “identity” based cleavages, including religious or sectarian, and understand it to refer to “descent-based” attributes that may be malleable in the long term but structure politics and social interactions in meaningful ways (Chandra 2006; Chandra and Wilkinson 2008).

thereof), whether through the prospect of publicizing the status of individual contributions to the community, thereby facilitating neighborhood gossip about non-contributors (i.e., horizontal accountability), or follow up by local leaders (i.e., vertical accountability).

Existing theories predict that cooperative behavior is more likely to arise in homogenous areas: Majority Hindu and majority Muslim neighborhoods should exhibit greater willingness to contribute as compared to diverse areas. However, we find that mechanisms of horizontal and vertical accountability are operative among Muslims but are not effective among Hindus. These results hold even after accounting for a host of other relevant factors, including levels of caste and religious diversity at the neighborhood level, the strength of social ties, gender, and socioeconomic status, among other factors. Our findings suggest a new, distinct mechanism that affects compliance with and the enforcement of social norms regarding public goods provision—the role of *minority status*.

We explore several possible mechanisms that may underlie the higher propensity of Muslims to contribute to public goods. First, as members of a persecuted minority, a version of the “*politics of respectability*” (Higginbotham 1993, p. 187) may compel Muslims to adopt more prosocial behavior than their Hindu neighbors in order to avoid potential opprobrium and negative attention in their communities. Second, given the history of anti-Muslim discrimination, Muslims are less likely to *trust* government institutions in India. As a result, they may be more likely to rely on in-group networks or non-state solutions for access to services and public goods, making them more willing to cooperate on issues of drainage that bypass the state. Third, by virtue of their minority status, Muslims are more likely to develop stronger *social networks* as a means of self-protection as compared to the majority group. Denser social ties facilitate cooperative behavior. Finally, we explore the idea of *cultural differences* between Hindus and Muslims, rooted in distinct religious norms and practices, as a possible explanation for differential preferences for cooperation around drainage and sanitation provision.

In the next section, we discuss the role of social accountability mechanisms in public goods provision in general and in minority communities more specifically. We then introduce the geographic context and sectoral focus of the research – slum settlements in Delhi and drainage – as well as demographic characteristics of the settlements where our survey was conducted. The subsequent section outlines our data and methods. We then discuss the results and introduce suggestive evidence aimed at disentangling potential mechanisms. Finally, we summarize the key theoretical contributions and outline possible extensions of the study.

4.2 Diversity, minority status and public goods provision

An extensive body of research focuses on the effects of diversity and public goods provision. Reviews of this work highlight its main findings, which hold across diverse settings, and push its boundaries by probing the origins of politicized ethnoreligious cleavages the conditions under which it hampers social and economic investment and outcomes (Singh and vom Hau 2016; Weber 1976; Wimmer 2015). Yet few studies explore how the nature of intergroup relations, and particularly group *status* may affect public goods provision. Status has been broadly conceptualized as “a comparative social ranking of people, groups, or objects in terms of social esteem, honor, and respect accorded to them” (Ridgeway 2019). We focus on the subordinate status of a group, which we understand to be a function of its relative size or demographic weight as well as the politicization of its identity, in turn generating a conflictual relationship with the majority group. Many ethnic minorities, including Muslims in India, are subject to discrimination and sometimes violence. In this section, we briefly review the mechanisms posited to undergird the general negative association between diversity and public goods provision. We then turn to the ways in which hierarchical social relations among groups living side by side might differentially affect these mechanisms, focusing on the behavior of persecuted minority groups in particular.

4.2.1 Diversity, social norms, and public goods provision

In response to the apparent conventional wisdom that diversity has a negative impact on development outcomes (Easterly and Levine 1997), a subsequent generation of scholarship has aimed to identify the specific mechanisms underlying this relationship. Recent studies zero in on the role of distinct types of social norms and networks in governing how both within-group and intergroup relations shape relevant outcomes.

Social norms are “standards of behavior that are based on widely shared beliefs about how individual group members ought to behave in a given situation” (Bernhard, Fehr, and Fischbacher 2006, p. 217) and have been shown to be powerful motivators both of political behavior, such as voter turnout (Gerber, Green, and Larimer 2008), as well as social outcomes (Björkman and Svensson 2009; Olken, Onishi, and Wong 2014). As forms of pro-social motivators, social norms can be viewed as a type of accountability mechanism: When people are held accountable for their behavior to those whose opinions they care about, they behave more in line with those individuals’ expectations (Lerner and Tetlock 2002). An important aspect of

social norms is that they are group specific – that is, they emerge through interactions in groups, apply to interactions within groups, and are enforced by group members. Historically, ethnicity, defined in terms of ascriptive group identities, including race, language, religion, tribe and caste, has been the most important characteristic of distinct social groups in many contexts, and collective norms are intertwined with the cultural traits of the group (Horowitz 1970). Thus, social networks are critical for disseminating and sustaining social norms, which tend to be shared and upheld by people in the same groups.

For at least two reasons, studies show, public goods provision may be superior in homogenous areas. First, shared social networks, which tend to be denser among coethnics, facilitate cooperation and coordination, which in turn is essential for supplying or demanding many types of public goods (Baldwin and Huber 2010; Deutsch 1953; Habyarimana et al. 2009; Miguel and Gugerty 2005). Second, norms of reciprocity and social sanctioning are more likely to be enforced within a homogenous community, where community members are better able to identify shirkers and hold them accountable (Fearon and Laitin 1996b; McClendon 2014; Panagopoulos 2010; Sinclair 2012).³

Existing social science literature suggests that social norms – including those that might affect contributions to public goods provision – can be enforced through three distinct mechanisms. First, a body of work in psychology suggests that ingroups are more likely to cooperate under conditions of *intergroup competition*. In general, members of a group view their own group more positively than others, and strive to behave in a way that reinforces this view. Social identity theory characterizes this as a need for the “positive distinctiveness” of one’s group (Tajfel and Turner 1986). Although not all agree that positive distinctiveness motivates individuals to advance perceptions of ingroup superiority, political behavior to protect group status has been observed across many contexts — from the US in the form of white reactions to “racial threat” (Enos 2016; Key 1949) to the Middle East in response to attempts to humanize an outgroup (Gubler, Halperin, and Hirschberger 2015). We posit that, even in the absence of any external social sanctions and other enforcement mechanisms, provoking intergroup competition will result in improved coordination among members of the ingroup.

A second way of enforcing social norms is via horizontal social networks, through actions such as public shaming and gossip among community members. We call this mechanism *horizontal accountability* since it

³Another potential mechanism focuses on variation in tastes or policy preferences across distinct ethnic groups (Lieberman and McClendon 2013), which may arise because of systematic marginalization of some groups, the varied impact of policies across ethnic lines, or distinct cultural practices, among other factors. We explore this explanation in the next section in discussing how variation in cultural beliefs and practices might give rise to differential preferences regarding local public goods provision.

operates among citizens of relatively equal standing. Horizontal social relationships may be especially critical for providing resources in the context of low-income settings, where resources are constrained. Public shaming, for example, has been shown to motivate improved service delivery, community coordination and, potentially, social or economic outcomes (Björkman and Svensson 2009; Duflo, Dupas, and Kremer 2015; Fung et al. 2008; Haley and Fessler 2005; Hibbard, Stockard, and Tusler 2005; Olken, Onishi, and Wong 2014; Panagopoulos 2010), while gossip among community members of relatively equal social standing is an important informal channel for the spread of information about issues related to public health (Banerjee et al. 2019). In the case of voting behavior, which can also be conceived of as a non-excludable good, research demonstrates that the promise of public shaming produces a stronger get-out-the-vote effect than other mechanisms such as highlighting one's civic duty (Gerber, Green, and Larimer 2008). The effect of horizontal networks, however, is contextually dependent. Some scholars, for example, have argued that the findings of numerous studies testing forms of horizontal social accountability are inconsistent, with some associated with improved outcomes and others finding no effect (Kosack and Fung 2014; Pritchett and Sandefur 2013). Furthermore, the strength of horizontal accountability mechanisms may be mediated by the nature and density of social ties (Lust and Rakner 2018). Incentives to contribute to the common good should be higher for people who are more embedded in local social networks due to norms of obligation and reciprocity underpinning social trust (Costa and Kahn 2003; Granovetter 1973; Ostrom 1990; Putnam 1993; Woolcock 1998).

Third, local elites, like community leaders and chiefs, may be instrumental in driving collective action around public goods provision at the local level. Such vertical accountability mechanisms have been shown to improve the supply of public goods in a variety of contexts. Studies of accountability and provider effort in Jordan and Morocco, for example, indicate that elite participation in local health centers can improve provider accountability and mobilize greater community engagement (Belkâab and Cammett 2014; Brix, Lust, and Woolcock 2015; Rabie et al. 2014). Similarly, Baldwin (Baldwin 2015) finds that hereditary chiefs in Zambia can mobilize resources and generate improved service delivery, and even Members of Parliament rely on local chiefs to use their social influence to “unlock” their community's contribution to local development proposals. Elites in influential positions with formal roles, such as elected officials, or informal roles, such as local leaders (pradhans) in Indian slums (Auerbach 2016; Thachil 2017) and religious authorities, play a central role in their communities. For better or for worse, these types of local elites serve as brokers for citizen access to public and private resources, distribute patronage, facilitate access to

economic or other opportunities, command the respect of citizens, and constitute a focal point for claim-making, among other functions. Elites are therefore well positioned to influence community participation in cooperative initiatives to supply public goods.⁴ The effectiveness of vertical accountability mechanisms may therefore depend on the strength and nature of social ties between slum residents and local leaders. Denser vertical social ties, which may increase the propensity of residents to contribute to public goods, may be manifested in family and social relationships between community members and local elites and in the frequency and conditions under which people interact with and seek assistance from local authority figures.

4.2.2 Persecuted minorities and local public goods provision

The impact of social norms and networks, whatever form they take, may operate differently for groups with distinct social status. Members of the majority group tend to enjoy social dominance and privileged access to resources, whereas members of the minority group are often disadvantaged in social, economic and political terms and may even face threats to their safety and well-being. With respect to the focus of our study, we envision several potential channels through which accountability mechanisms might differentially affect the willingness of minority group members to contribute to public goods at the local level. The first three relate to social and political dynamics that are likely to vary across minority and majority groups: members of minority groups, especially those who feel persecuted, may feel greater pressure to present themselves as model citizens, have lower trust in political institutions, or exhibit denser social ties. A fourth potential mechanism centers on distinct cultural practices, resulting in varied group preferences and behavior among Hindus and Muslim around public goods such as drainage that relate to cleanliness and hygiene.

The Politics of Respectability

The concept of the “politics of respectability” (Higginbotham 1993, p. 187) may help to explain why members of a persecuted minority. Because they are the target of negative stereotypes and increasingly feel threatened, Muslims in Indian slum communities may push for behavior by fellow community members to counteract these stereotypes. Building on this concept in his study of attitudes towards welfare policies

⁴To be sure, hierarchical social relations are often associated with clientelism and elite capture of public service delivery, resulting in unequal provision and inferior quality public goods (Bardhan and Mookherjee 2012; Chen and Cammett 2012; Khemani 2015). Elites can also capture rents from projects aimed at promoting public goods, even in the face of participatory planning and implementation (Pan and Christiaensen 2011; Sheely 2015).

among Black Americans, Jefferson (Jefferson 2018) examines individual variation in *policing* the behavior of fellow Black Americans, who are perceived to conform with negative stereotypes. Monitoring and sanctioning of behavior that reinforces such negative stereotypes is a means of convincing the dominant group that the in-group deserves full inclusion and equality.

The politics of respectability aim to challenge negative stereotypes, which are context specific. In the US, this translates to promoting "temperance, cleanliness of person and property, thrift, polite manners, and sexual purity" as key properties of Black Americans (Higginbotham 1993, p. 193). In the case of India, and of relevance for our focus on contributions to improved neighborhood public goods, Muslims might aim to present themselves as model neighbors who contribute to the well-being of the community to avoid negative opprobrium by residents from the majority community. As a result, members of this persecuted minority may be more keen to avoid blame for poor outcomes, feel increased concern about the failure of ingroup members to contribute to community welfare, and more likely to try to compensate for the perceived failures of fellow Muslims.

The desire to promote or uphold a favorable image of the minority group is likely to be especially pronounced among those who feel greater attachment to the ingroup. As a result, strong ingroup identifiers may be more likely to police the behavior of "black sheep," or ingroup members who deviate from normative behavior. Ingroup members react in one of two ways to the behavior of black sheep. On the one hand, they can try to rationalize the bad behavior, even to the point of justifying the misbehavior because someone "psychologically close" has engaged in it (Gino and Galinsky 2012). On the other hand, ingroup members can attempt to compensate for ingroup misbehavior out of concern for the potentially negative reputational consequences for the ingroup. Importantly, the nature of the black sheep effect depends on the composition of the audience (Gino, Gu, and Zhong 2009). When outgroup members can observe the negative behavior, then knowledge of ingroup misbehavior leads to guilt and compensatory behavior. In slum communities in Delhi, the context of our study, Muslims live in close proximity to members of the majority Hindu group, implying that the black sheep effect would induce contributions to the drainage scheme by Muslim respondents.

Empirical research from other contexts suggests that Muslims in Delhi slums may have reason to be concerned about upholding a reputation for being model neighbors - or at least may be held to higher standards than members of the majority community. For example, a field experiment in Germany assessed the likelihood of natives and minorities to reprimand or express disapproval of study confederates who

littered on the streets of ethnically diverse neighborhoods in Bonn and Cologne. Despite carrying out the same actions violating prevailing social norms, perpetrators who appeared to be migrants were sanctioned at higher rates across the board (Winter and Zhang 2018, p. 2724).

In the Indian context, prevalent negative stereotypes depict Muslims as unpatriotic or prone to violence (Centre for Regional Political Economy 2019; Gayer and Jaffrelot 2012). While these stereotypes do not center on contributions to public goods per se, members of a persecuted minority have an incentive to present their group in the most favorable light in realms beyond the explicit focus of predominant stereotypes, including as positive contributors to local infrastructure and advancing the well-being of their residential areas. As Mac Ginty's (Mac Ginty 2014, p. 554) analysis of "everyday peace" in conflict-affected societies holds, people "may go out of their way to avoid giving offence and to be deliberately polite to out-group members" in order to maintain a calm, including by policing behavior within the group. We expect this is all the more applicable to members of persecuted minorities, who are likely to perceive higher risks of dissent or deviating from social norms in intergroup situations (Ditlmann, Samii, and Zeitzoff 2017b, p. 50), including with regard to improving the public welfare of diverse communities where they reside.

Political Trust

A second potential mechanism underlying our results relates to differential levels of political trust among members of majority and minority communities. A history of experiencing discrimination reduces trust in the state, particularly when public officials and agencies have not worked to counter bias against minorities and have even actively pursued policies leading to marginalization. The exclusion of minority groups from the state can exacerbate ethnic divisions and reduce trust in political institutions among the marginalized (Wimmer, Cederman, and Min 2009). Findings from diverse contexts confirm this general claim. Based on a lab-in-the-field experiment in Bosnia-Herzegovina and household survey data, Lazarev and Mironova (Lazarev and Mironova 2018) argue that minority status is associated with lower levels of trust in institutions that are controlled by the majority. Black Americans' levels of trust on the government, for example, is attributed to systemic discrimination faced by the community (Nunnally 2012; Omi and Winant 2014).

Of direct relevance for our study, low trust in government may induce minorities to be favorably inclined towards non-state options for the supply of public goods in their neighborhoods. Studies from other contexts on related topics support this interpretation. Members of minority groups are reluctant to contribute to public goods via taxes as the benefits disproportionately accrue to the majority, who are members of the

outgroup (Li 2010). Similarly, ethnic minorities, such as Blacks in the U.S. and Muslims in Europe, are distrustful of law enforcement and institutions of justice (Tyler 2005), while perceptions of discrimination reduce trust in the police among the minority group (Van Craen and Skogan 2015). Low trust in political institutions is likely to increase reliance on communal networks for security as well as the provision of social services, suggesting that minority groups may be more amenable to cooperate in local public goods provision. As a result, Muslims in Delhi slums may be more likely to favor a non-state option for drainage in their neighborhoods.

Compensatory Social Ties

A third possible mechanism relates to the nature of social ties in minority communities. As longstanding victims of marginalization, members of minority groups may be especially reliant on ingroup members. While this arises in part out of mistrust in the state as a result of official neglect or discrimination, the focus here is on the nature of social networks rather than trust per se.

Ethnic minorities around the world are subject to discrimination in employment and other opportunities (Banerjee et al. 2009; Bertrand and Mullainathan 2004; Fix 1993; Lee 2018) and residential segregation of minority groups is a common feature in many multiethnic societies, including India (Bharathi, Malghan, and Rahman 2015; Susewind 2017). In some countries, including contemporary India, minorities are also perceived as a threat and, hence, are targets of violence. As a means of protection and to facilitate collective action around shared interests such as service provision, minorities are therefore more likely to develop strong *social institutions* and *social ties* within the group. Further, minority groups tend to expend greater effort and resources to maintain their cultural distinctiveness compared to majority groups. Social institutions may therefore serve to preserve cultural practices, reinforcing ingroup solidarity (Bisin and Verdier 2001).

Evidence from recent research in a similar context finds that minority groups exhibit higher levels of ingroup bias in social trust. Based on a study of Muslims in West Bengal and Hindus in Bangladesh, Gupta et al. (Gupta et al. 2018) conclude that the minority group in both contexts show positive in-group bias in trustworthiness. Further, its effects are stronger among highly religious individuals. A lab-in-the-field coordination game in India further suggests that Muslim leaders improve coordination in Muslim majority towns while Hindu leaders are not able to elicit similar levels of performance from Hindu participants (Bhalotra et al. 2018). Collectively, findings from these diverse studies suggest that persecuted minorities,

such as Muslims in Delhi slums, are likely to exhibit higher levels of social ties than their Hindu counterparts. As a result, they may be more likely to hold each other accountable for failing to contribute to local public goods.

Cultural Characteristics and Public Goods Provision

Finally, beyond the potential ramifications of minority status, distinctive cultural attributes of ethnic or religious groups may also affect the propensity of group members to take part in collective action around public goods provision. At least three possible features of Muslim versus Hindu culture in the Indian context are potentially relevant, including distinct practices around communal living and the relative lack of caste-based norms among Muslims.

First, religious practices of charitable donations may make Muslims more likely to contribute to public goods. *Zakat*, or charitable donations, is one of the five pillars of Islam in South Asia. According to a study in 2000, *Zakat* amounted to \$11.5 million (867 million rupees) in towns and cities (Gayer and Jaffrelot 2012). Community resources, including mosques and religious schools (*madrasas*) are supported by such donations. While religious charities are not unique to Muslims, longstanding practices of communal contribution can potentially support cooperative behavior in other realms, including public goods provision. Further, the rise of Hindu nationalism and violence against Muslims is believed to have heightened in-group solidarity in recent years, which too can have positive implications for cooperative behavior (Gayer and Jaffrelot 2012).

Second, unlike Hindus, the caste system among Indian Muslims is less rigid. The absence of strict norms of purity and pollution and social separation has allowed for greater social mobility within the Muslims community (Mines 1972). To the extent that horizontal social ties facilitate coordination, this relative egalitarian social structure may also provide greater scope for collective action as compared to Hindus. At the same time, for both Hindus and Muslims, caste is less consequential for cooperation around public goods in urban areas, casting doubt on the claim that differential caste structures across the two religious communities explain variation in the propensity to contribute to local public goods (Auerbach and Thachil 2018; Auerbach 2017; Chidambaram 2020; Jha, Rao, and Woolcock 2007).

Finally, Hindu caste norms of purity and pollution can have a direct bearing on sanitation practices in the two communities. Geruso and Spears (Geruso and Spears 2018) argue that ideas about purity and pollution make Hindus reluctant to use toilets or latrines in their homes since human excreta is perceived to pollute the

home. They find that Hindus are 25 percent more likely to defecate in the open, which in turn contributes towards the prevalence of pathogens and spreads water-borne diseases in majority-Hindu neighborhoods. As a result, despite lower incomes and educational attainment in the Muslim community, Muslim children in India have lower levels of infant mortality as compared to Hindus (Geruso and Spears 2018). These distinct religious practices and norms may also influence the response of the two communities to local drainage schemes, although the expected effects are ambiguous. On the one hand, Hindus might be more likely to support a community drainage initiative to better ensure the removal of human waste from the vicinity of their homes. On the other hand, they might be less concerned with local drainage systems because they tend to defecate far away from their residential areas.

4.3 Hypotheses

Our study assesses the willingness of community members to contribute to a hypothetical collective scheme to improve drainage systems in slum areas contingent on different accountability mechanisms. In particular, we focus on the variable likelihood of Hindus versus Muslims to respond to distinct accountability treatments. Literature on diversity and public goods provision and on the impact of minority status on group behavior give rise to several hypotheses that we test in this paper.⁵

First, our treatments, described in detail below, tap into a variety of ways in which accountability mechanisms might induce contributions to local public goods.

Horizontal Accountability Respondents who are told that community members will learn and gossip about those who have not contributed will be more likely to contribute than those who do not receive this information.

Vertical Accountability Respondents who are told that local leaders will contact and reprimand those who have not-contributed will be more likely to contribute than those who do not receive this information.

Black Sheep Effect In diverse contexts, respondents who receive information on ingroup underperformance are more willing to contribute than those who do not receive this information.

⁵Our hypotheses related to accountability were preregistered prior to data collection on the Experiments for Governance and Development (EGAP) platform and subsequently migrated to the Open Science Foundation (Cammett, Chakrabarti, and David 2018). In the analyses, we combine all accountability treatments because the results show no difference across treatments.

Second, for a variety of potential reasons elaborated above, we expect that Muslims are more likely to respond to accountability mechanisms.

Minority Effect Muslims are more likely than Hindus to contribute in response to diverse accountability mechanisms.

In the next section, we describe the context in which we set out to test these hypotheses.

4.4 Context: Hindu-Muslim relations in India

To explore the impact of minority status on intergroup cooperation, we take advantage of the history of Hindu-Muslim relations in the city of Delhi. Delhi has been one of the major centers of Islamic culture and politics in India for centuries. Hindu-Muslim relations in independent India is strongly shaped by the dynamics of religious nationalism in the subcontinent in the 20th century. Ideological differences between Muslim nationalists and secular nationalists led to the partition of British India and the creation of Pakistan in 1947. Millions of Hindus and Muslims were displaced in what is arguably one of the greatest migrations in human history. Delhi lost about two-thirds of its Muslim population in the process (Gayer 2012). The birth of the two nations also witnessed massive sectarian violence that killed close to 2 million people (Talbot and Singh 2009). Muslims currently comprise about 13% of Delhi's population. This proportion roughly mirrors that of India as a whole, in which roughly 14% of the population is Muslim. At 172 million, though, India's Muslim population is the third largest in the world. India is the world largest Muslim-minority country.

As the largest religious minority and with the formation of Pakistan as a Muslim homeland, Muslims are the main ideological adversary of Hindu nationalism, the political ideology espoused by the ruling BJP party. Hindu nationalism is based on the belief that Hindus should have cultural and political primacy in India (Varshney 2003). The electoral success of the BJP in recent years, as discussed earlier, has led to a systematic violence against Muslims. While the current political vitriol against Muslims is unprecedented in recent history, violence and discrimination against Muslims is by no means a new phenomenon. They are a frequent target of discrimination and violence, which has generated insecurity, displacement, segregation, and loss of property and life (Varshney 2003). Muslim loyalty to India has been continuously questioned by sections of the state, media and the political class (Gayer and Jaffrelot 2012). Their precarious position is also reflected in their socioeconomic status. On average, Muslims have lower levels of socioeconomic

outcomes as compared to Hindus – they are poorer, less educated, and less connected to the formal sector (Government of India 2006). Moreover, intergenerational mobility for Muslims has declined over time while other marginalized groups have experienced notable gains (Asher, Novosad, and Rafkin 2017). Unlike other minorities, Muslims do not benefit from affirmative action and reservation policies. They are underrepresented in elected bodies as well as the bureaucracy. Their marginalization in the police and judiciary has made them vulnerable in anti-terrorist measures that disproportionately target innocent Muslim men (Gayer and Jaffrelot 2012).

Delhi manifests all these forms of marginalization against Muslims. It has one of the highest levels of Muslim segregation in urban India, which is indicative of relatively weak intergroup social ties (Susewind 2017). At 2.3%, Muslim representation in the local police force is amongst the lowest in India (Gayer 2012). Only seven of the 124 judges appointed to the Delhi High Court since 1966 are Muslim. Muslims in Delhi, on average, are also poorer and less educated than the overall population and hence more likely to reside in slums (Gayer 2012), which house about half of Delhi's 18 million residents and are the sites of our study.⁶

Delhi's slums, like many poverty-ridden populations in the world, are characterized by inferior public goods provision. The poor quality of sanitation, garbage disposal, water supply, electricity, and drainage in these settlements has severe consequences for the health of their residents. A number of factors contribute towards the deficient quality of public services in slums, including the multiplicity of administrative jurisdictions and government agencies, low state capacity, lack of land tenure rights in slum settlements, and high levels of rural to urban migration (Banerjee, Pande, and Walton 2012; Ghertner 2008; GNCTD 2006; Heller et al. 2015; Jha, Rao, and Woolcock 2007). According to a recent survey of over 3,000 slum-dwelling households in Delhi, for example, only 14% have a private tap or toilet. About 60% of households had no specific outlet for drainage from their home. This figure is even higher for the poorest households (72%), while 90% of those with a drain say that it emits bad smells or overflows (Banerjee, Pande, and Walton 2012).

We focus on drainage for several reasons. First, drainage and sewage are among the most strained public services. According to the latest figures, storm water drainage coverage stands at less than 50% in Indian cities (Ministry of Urban Development 2012). In Delhi in particular, drainage is one of the most contentious

⁶Estimates of Delhi's slum population vary by the criteria used to define a slum. While government reports in 2008-09 estimated Delhi's slum population at 580,000 households (Banerjee, Pande, and Walton 2012), the Delhi Human Development Report uses a broader definition of slums that includes 45% of the city's population (GNCTD 2006). A recent study argues that more than 60% of Delhi's population lives in informal settlements that suffer from inadequate public service provision (Heller et al. 2015).

public goods, with residents often negotiating with municipal workers or having to collectively pay for a private organization to regularly clean them out (Heller et al. 2015). Further, access to water and sanitation was cited as the biggest source of discontent among slum dwellers (Banerjee, Pande, and Walton 2012). The issue of drainage is hence contextually relevant and important.

From a theoretical perspective, storm water drainage stands out as a case of a (near) pure public good that cannot be addressed by individual, uncoordinated solutions. Heavy rainfall and the accumulation of garbage causes drains to clog and in turn affect the well-being of the entire community, particularly because the drains are interconnected and pass through entire slum neighborhoods, including in small trenches immediately adjacent to private homes. Thus, by design, drainage requires cooperation to resolve blockages. A resident may thoroughly clear the portion of the drain crossing underneath the threshold of her home, but if she does not coordinate with her upstream neighbors, her hard work will be for naught. Indeed, our fieldwork revealed that residents sometimes band together to clean their drains on their own initiative.

Further, though drainage in slum settlements is supposed to be maintained by the municipal government, in practice the municipality is often either incapable or unwilling to carry out its official duties. During fieldwork in multiple sites, we witnessed that municipal workers had extracted and placed trash alongside the drains in order to dry it out before collection and removal at the periphery of the settlements. But the interior drains were clogged with trash and organic waste. Hence the case of drainage is closer to a common pool resource than a public good provided by the state.⁷ As compared to public goods, social accountability mechanisms play an especially critical role in cooperation around the maintenance of common resources (Björkman and Svensson 2009; Habyarimana et al. 2007; Lust and Rakner 2018; Miguel and Gugerty 2005; Ostrom 1990). The context of drainage in slum settlements in Delhi is therefore particularly well suited to study the effects of social accountability norms around cooperation.

4.5 Data and Methods

We use a survey experiment to explore our research questions. This method has been used to study citizens' willingness to contribute to communal goods, particularly in developing country contexts. Through an experimental design, treatment effects can be separated from other factors that, in an observational setting,

⁷A common pool resource refers to a system in which it is costly but possible to exclude some from making use of the benefits of its resources. In contrast to poor public goods, a common pool resource can be subject to congestion or overconsumption. Examples include irrigation systems, fishing grounds or forests (Ostrom 1990).

would confound the relationship of interest. In line with norms in the discipline, we pre-registered our design with Evidence in Governance and Politics (EGAP), noting in our discussion of the analyses and results where we may diverge from the pre-registered study.

4.5.1 Site Selection and Survey

Before the administration of our survey, a team of field researchers from a research institute in Delhi, the Centre for Policy Research (CPR), aided us in gathering qualitative data on 12 slum settlements. These settlements were selected on the basis of their ethnic composition and consist of smaller communities or *bastis*. This preliminary fieldwork, conducted for about three months, provided us with a wealth of information on the demographic characteristics, social and political life, state of public services, and nature of community relations in these areas, aiding both in the initial design of the project and in the interpretation of the results (see the appendix). Information from this stage guided the design of our instrument and site selection for the field experiment. For example, we learned that slum residents favored low-cost, private options over free, public services offered by government agencies and non-governmental organizations (NGOs). Residents complained about the deterioration in community toilet complexes (CTCs) after a new municipal policy eliminated user fees, and tended to regard NGOs as money-making schemes for well-connected locals who want to extract funds from the government and private donors. Our survey instrument hence utilizes a hypothetical private drain cleaning company. The initial fieldwork also allowed us to select five heterogeneous and homogeneous sites that were largely similar to each other in other respects such as socio-economic characteristics, age of settlements, and condition of public services. We piloted the instrument in three *bastis* to improve clarity and relevance of questions.

A second firm, Across Research and Communications (ARC), administered the survey. Together with an administrator from the survey firm and representatives from CPR, we trained a team of 25 enumerators on the instrument, recruitment strategy, and location characteristics. Because the survey focused on decisions made by a household, and because prevailing social norms in India dictate that financial decisions are made by the (typically male) head of household, enumerators were directed to interview the head of household. The enumerators used tablets to administer the survey, which allowed us to measure the GPS coordinates of each household. As discussed later, this data was used to construct a measure of ethnic diversity at the household level. Overall, the response rates varied between 30 and 70 percent across seventeen sub-settlements/*bastis*, with 5,844 residents taking the survey. For the purposes of this paper, we only use the

3,844 participants assigned to one of six treatment groups. The appendix provides details regarding the recruitment strategy and lists the settlements surveyed as well as their respective sample sizes and response rates.

Participants answered a battery of questions before the experimental manipulations were introduced. Pre-treatment questions assessed a variety of topics such as basic demographic characteristics, participants' locations of residence, religion and religiosity,⁸ perceived efficacy, connections to political and voluntary institutions and figures, the strength of social ties between different groups of people in the respondents' local communities, participation in electoral politics, and the state of the drainage system in the participants' neighborhoods. Random assignment took place via Qualtrics software and then respondents were exposed to one of six experimental manipulations, which presented a variety of accountability mechanisms that might boost the willingness of the respondent to contribute to a community drainage initiative. Five outcome questions immediately followed the administration of the treatment. The survey ended with a battery of questions unlikely to be affected by, or unnecessary for, the experimental manipulation. These included questions about education, socioeconomic status, and political leaning and knowledge. The outcome questions included the following, each measured on a 1–4 scale, with higher levels indicating more positive responses:

Benefit Would this program be beneficial for your neighborhood?

Interest Given this scenario, how interested would you be, overall, in the program?

Fee How likely would you be to pay the monthly fee?

Contract Would you be willing to sign a six-month contract for this service?

Influence How likely would you be to try to get your neighbors to sign up for the program?

4.5.2 Experimental Treatments

Each of the experimental manipulations began with the same introduction, which presented a service offered by a drain cleaning company. To avoid deception, the offer was described as hypothetical throughout the manipulation. Using the name of the sub-settlement in which the survey took place, the participant was

⁸At this point, potential respondents who were neither Muslim nor Hindu, who constituted only a handful of people, were dropped.

told that the study aims to find out if residents of that neighborhood would be interested in participating in a service like the one proposed. In addition, they were reminded twice that there were two caveats: First, enrollment would entail a (small) monthly subscription fee of Rs. 50 per household (about US \$0.70) and, second, that two-thirds of local residents had to sign up for the service in order for it to be implemented. This setup ensured a collective action problem that necessitated the coordination of a majority of the community.

Manipulation texts for each of the treatments diverged from this point. Table 4.1 depicts the treatment conditions.) In the control, the enumerator introduced a hypothetical testimony from someone whose neighborhood had been offered the service. The testimony-giver first attested to the high quality of the drain cleaning service and then stated that the service was not implemented in his neighborhood (the “underperforming” neighborhood) because some people (whom he names) did not want to contribute. He then noted that the program was implemented in an adjoining neighborhood where enough people (again, a few of them were named) decided to contribute. Crucially for the other treatments, the names mentioned by the testimony-giver in the control were half Muslim and half Hindu. After reading the testimony, the enumerator stated that all participant information would be kept private if the service were implemented in the neighborhood.

Table 4.1: Treatment Group Sizes

Treatment	N
Control	668
Horizontal Accountability	635
Vertical Accountability	649
Ingroup Underperformance	615
Hor. Acct. + Ingroup Underperformance	630
Vert. Acct. + Ingroup Underperformance	647

Two dimensions of the control text were altered to create the other treatments. The first was an accountability dimension, with treatments aimed at encouraging participation in the drainage scheme based on either horizontal or vertical accountability mechanisms. Horizontal accountability was introduced by altering the final statement of the enumerator. Instead of guaranteeing the participant’s privacy, the enumerator stated that community members would discuss who did and did not contribute, potentially singling out and shaming non-contributors. A list of six names was read as an example of people who were singled out in a different neighborhood. Similarly, the vertical accountability treatment altered the final statement to

indicate that a local leader (*Pradhan*)⁹ would find out who had not contributed and would single out and shame non-contributors.

The second dimension manipulated depictions of ingroup underperformance, which was primed by altering the testimony in a manner that depended on the religious identity of the individual taking the survey. If the participant was Hindu, then the name of the “underperforming” community and residents (and, by implication, the one giving the testimony) were made to sound obviously Hindu, whereas the community and residents who successfully implemented the service were made to sound obviously Muslim; the opposite was true if the participant was Muslim. The treatment was constructed in this manner so that the testimony-giver was implied to be a coreligionist and therefore potentially more credible. In the two treatment conditions in which accountability mechanisms and ingroup underperformance were combined, similar adjustments were made to the list of residents described as being shamed.¹⁰

On average, the treatment was administered in less than 30 seconds for every treatment condition, with differences of only 1 or 2 seconds between the different treatments. Participants seemed to be paying attention and absorbed the information: After administration of the treatment, approximately 90% of participants could correctly remember what proportion of their neighbors would need to sign up, and about 95% of participants correctly remembered the cost of the service. Participants who incorrectly answered these questions were told the correct answers afterwards. We did not exclude participants based on their response to these manipulation checks.

4.5.3 Sample characteristics

The sample includes 3,844 participants spread across the six treatments of interest. Because of the use of multiple enumerators and randomization in the field by random chance there are minor imbalances in the size of each treatment group, from a minimum of 615 to a maximum of 668 (see Table 4.1).

In spite of these slight differences in the sizes of the treatment groups, the sample was well-balanced in terms of the covariates. We performed an omnibus test of balance (Hansen and Bowers 2008) and found that we could not reject the null of a balanced sample, indicating successful randomization.¹¹

Overall, as illustrated in Table 4.2, participants in our sample demonstrated a high need for drain cleaning

⁹A *pradhan* is an informal slum leader who acts as an intermediary between the formal government and the urban poor

¹⁰A full text of these treatments is provided in the appendix.

¹¹See the Appendix for the results of these randomization tests.

services and expressed strong interest in the proposed service. On a scale of 1 to 5, participants rated the quality of drainage in their neighborhoods at a dismal 1.83. Additionally, on average, 23% of participants indicated that they had a problem with drainage in the last year that required the help of someone other than themselves. In the control condition, participants' interest in the program registered at 2.95 on a 1–4 scale, indicating that, on average, participants were interested in the program. Table 4 provides an overview of the demographic characteristics and other basic information on the sample.

Table 4.2: Demographic Variables

Variable	Mean	SD
Gender (1 = Female, 0 = Male)	0.20	0.40
Married (1 = Married, 0 = Other)	0.91	0.29
Age	38.43	11.77
Years Residing in Settlement	20.63	10.55
Ownership (1 = Own Residence, 0 = Other)	0.82	0.39
Number of Rooms in Residence	1.85	1.02
Employment Status (1 = Employed)	0.84	0.37
Financial Hardship Index (1 = More Hardship)	0.25	0.23
Assets Index (1-3, 3 = Most Assets)	1.90	0.43
Education Level (1-10 Scale)	3.90	2.30
Practice of Prayer (1-5 Scale)	4.25	1.08
Religious Behavior Index (1-5 Scale)	3.77	0.88
Religious ID (1 = Muslim, 0 = Hindu)	0.24	0.43
Local Political Efficacy Index (1-4 Scale)	1.92	0.78
Relation to Local Figures (1 = Any)	0.14	0.35
Involvement in Local Organizations (1 = Any)	0.06	0.24
General Social Ties (0-1 Scale)	0.89	0.20
Helpfulness Index (1-4 Scale)	3.11	0.79
Forgo Wages Index, Family (1-3 Scale)	2.79	0.35
Forgo Wages Index, Others (1-3 Scale)	2.25	0.52
% Political Activities Engaged In	0.16	0.25
Quality of Drainage (1-5)	1.82	1.06
Drainage Problem Requiring Help	0.23	0.42
Caste Level (1 = Upper, 0 = Lower)	0.63	0.48

Perceptions of community drainage needs also varied somewhat by religious identification. Muslims rated the quality of drains in their communities slightly higher, at 1.87 compared to 1.81 for Hindus. However, a larger proportion of Muslims also reported having an issue with their drains that necessitated others' help, with 32% of Muslims saying this occurred in the last two years compared to 20% of Hindus.

4.6 Results

In this section, we report findings from the main analyses, focusing on variation in the effectiveness of the accountability treatments across Hindus and Muslims. As shown below, the results are robust to a variety of alternative specifications.

4.6.1 Heterogeneous effects by religious identification

Our experimental results show that the effect of our treatments differ greatly between Hindus and Muslims. The set of panels on the left side of Figure 4.1 show the means and confidence intervals for Hindus (solid line) and Muslims (dashed line) for participants in control and the treatments, divided by each of our five outcomes of interest. There are several findings to note from this figure. First, we designed our questions to elicit a range of responses on our 1–4 scale, and these estimates show that this effort was successful. Generally, as we move from questions that would require little to no effort on the part of the participant (e.g. “Would this program benefit you?”) to those that would require more (e.g. “Would you be willing to sign a 6-month contract?”), the average response decreases by about 20%. Second, Muslims consistently have a lower baseline response to our outcome measures in the control group. For instance, when asking participants if they would be willing to pay the fee associated with the program, Hindus in the control group have an average response of 2.8, whereas for Muslims in the control group it is 2.6. This indicates that the treatments have the effect of bringing Muslims up to or slightly surpassing the responses of Hindus.¹² Third, Muslim participants respond to the treatments by saying they they would benefit more from the program, be more willing to pay the program fee and sign a contract, and be more willing to try to influence their neighbors to enroll. Hindu participants, on the other hand, show little to no change in these outcome measures in response to the treatments.¹³

For a formal test of the difference between effect size for Muslims and Hindus, and for all remaining analyses in this paper, we combine our five outcomes into a single index labeled “Index of Favorability toward Drainage Program.” This is well justified because of the similarity in patterns by religious identity

¹²One might be worried about ceiling effects in this context. However, the baseline gap between Hindus and Muslims seems to exist across our outcome measures, even those where both group averages are further away from the “ceiling” of the measure. Additionally, we run a tobit model (see the appendix) to account for floor and ceiling effects that do appear in a histogram of the outcome index. The tobit model confirms our results and actually produces larger estimates for the difference between Muslims and Hindus than we report here.

¹³A more detailed look at models with separate outcomes can be found in the appendix.

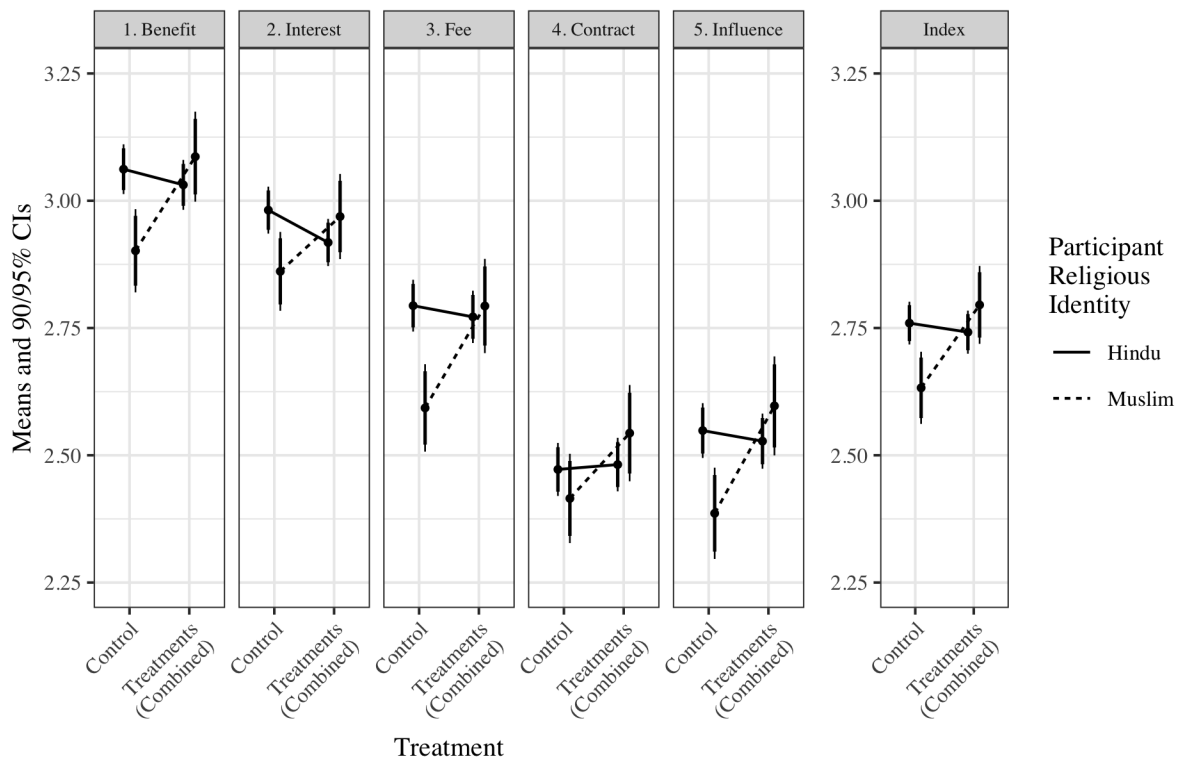


Figure 4.1: Interactive Models.

Table 4.3: Interactive Model

Dependent Variable: Index of Favorability toward Drainage Program	
Constant	2.76 (0.02)
Treatments (Combined)	-0.02 (0.03)
Muslim	-0.13 (0.04)
Treatments (Combined) x Muslim	0.18 (0.06)
Observations	3,832
R ²	0.003
Adjusted R ²	0.002
Residual Std. Error	1.16 (df = 3828)
F Statistic	3.90 (df = 3; 3828)

across different outcomes as well as by the high internal consistency of these five measures ($\alpha = 0.89$). Our main test consists of an interactive regression between an indicator for each treatment group and an indicator for whether the participant identifies as Muslim instead of Hindu. The results of the test are in the right-most panel of Figure 4.1, for comparison with each of the separate outcomes, as well as in Table 4.3. For the rest of our analysis, we use only the combined index as our outcome measure.

The estimate in Table 4.3 indicates that the effect of the treatments for Hindus is -0.02 and insignificant, whereas for Muslims the effect of the treatments is 0.18 and statistically significant ($p = 0.00$). This effect size for Muslims constitutes approximately 6.5% of the 1–4 scale used for the outcome index, which is a substantively significant effect that is similar in magnitude to other important predictors of favorability toward the drainage program. For instance, one question in our survey asks participants to rate the quality of their drainage on a 1–5 scale, with higher values meaning greater satisfaction. We might expect this question to be highly predictive of favorability toward the drainage program, in that those who rate their drainage quality as high are less likely to be in need of a drainage cleaning program. In a regression of the outcome index on the drainage quality question, a move from the 25% to 75% quantile response (i.e., from a 1 to a 3) on the drainage quality question corresponds to a negative change in our favorability index that is of similar magnitude to the treatment effect for Muslims.

4.6.2 Robustness Checks

Standard checks on our experimental procedure indicate that our manipulations were administered and received in the appropriate manner. An omnibus balance test indicates that there are no significant imbalances, either when examining the treatments separately or combined. The survey was timed and there are only small differences of approximately 1–2 seconds in the amount of time it took to administer the control and treatment conditions. A vast majority of participants passed our manipulation checks—89% correctly noted what proportion of their neighbors would need to pay the fee for the company to implement the service, and 95% correctly identified Rs. 100 as the fee amount—and results do not change if those who did not pass are excluded from analysis. Lastly, only two participants, or lower than 0.1% of our total sample, correctly identified the purpose of the experiment in a final question probing their perceived reason for the study. Our results remain the same without these participants.

A second set of robustness checks revolves specifically around the heterogeneous effects we probe. Religious identity is, of course, not randomly assigned, and other factors correlated with religious identity

could be responsible for our findings. We conduct a set of robustness checks to account for alternative explanations of our findings. First, we produce models with a large set of relevant controls, among the most important of which are an indicator for the caste of the respondent, the diversity of the participant's surrounding area (based on survey responses), and the religion of the respondent's enumerator. Our findings change little throughout each of these models, as seen in Table 4.4.¹⁴ Caste divisions, within both Hindu and Muslim communities, do not produce statistically significant coefficients. This is consistent with the existing literature on Indian slums, including in Delhi, that emphasize that while caste does influence social life (Auerbach and Thachil 2018; Chidambaram 2020; Thachil 2017), it does not create barriers for local-level collective action (Auerbach 2016; Auerbach and Thachil 2018; Auerbach 2017; Chidambaram 2020; Jha, Rao, and Woolcock 2007). Community development committees in slums tend to be diverse, in terms of both religion and caste, and slum leaders cater to multiple groups (Auerbach 2016, 2017).

Our pre-registered design considered both Muslim identity and diversity at the neighborhood level as factors to assess for heterogeneous treatments effects. We also consider the fact that religious identification is tied to our measure of diversity because, as a demographic minority, Muslims are more likely to live in diverse neighborhoods. We therefore examine the interaction of the treatments with neighborhood diversity levels.¹⁵ In our dataset, the diversity measure is 0.34 for the average Hindu respondent and 0.59 for Muslims. When we interact our treatments with measures of neighborhood diversity, the results are not compelling. While in the same directions as the results for religious identification, they are much smaller and not significant (treatment effect for Muslims: $\beta = 0.10, SE = 0.09$).¹⁶ Note as well that this result is the opposite of what would be expected. For this reason, we consider the interaction with religious identification to be the more compelling of the two potential sources of heterogeneous effects.

4.7 Probing the minority effect: Why Muslims contribute more

In this section, we explore possible reasons why Muslims respond positively to our accountability and ingroup underperformance treatments while Hindus do not. In particular, four potential mechanisms might

¹⁴See the appendix for the model with a full list of controls used and coefficient estimates for controls.

¹⁵Based on geocoded data, diversity is based on the religious identities of people within a 100 meter circle around the respondent's place of residence. The results hold if the radius is adjusted to alternative values.

¹⁶A full examination of this model is available in the appendix, which also explores how the treatment effect for Muslims varies by the composition of their neighborhoods (i.e., based on whether they live in a small Muslim minority, large Muslim minority, or Muslim majority neighborhood).

Table 4.4: Models with Controls

	Dependent Variable: Index of Favorability toward Drainage Program				
	(1)	(2)	(3)	(4)	(5)
Constant	2.76 (0.02)	1.00 (0.16)	1.01 (0.16)	0.93 (0.16)	0.94 (0.16)
Treatments (Combined)	-0.02 (0.03)	-0.001 (0.03)	0.003 (0.03)	0.002 (0.03)	0.004 (0.03)
Muslim	-0.13 (0.04)	-0.07 (0.04)	-0.06 (0.04)	-0.10 (0.04)	-0.09 (0.04)
Treatments (Combined) x Muslim	0.18 (0.06)	0.17 (0.06)	0.17 (0.06)	0.15 (0.06)	0.15 (0.06)
Demographic Controls?	No	Yes	Yes	Yes	Yes
Caste?	No	No	Yes	Yes	Yes
Diversity?	No	No	No	Yes	Yes
Enumerator Religion?	No	No	No	No	Yes
Observations	3,832	3,684	3,577	3,555	3,549
R ²	0.003	0.10	0.10	0.11	0.11
Adjusted R ²	0.002	0.09	0.09	0.10	0.10
Residual Std. Error	1.16 (df = 3828)	1.08 (df = 3658)	1.08 (df = 3550)	1.08 (df = 3527)	1.07 (df = 3520)
F Statistic	3.90 (df = 3; 3828)	15.69 (df = 25; 3658)	14.90 (df = 26; 3550)	15.45 (df = 27; 3527)	15.57 (df = 28; 3520)

compel Muslims to respond more to our treatments: i) stronger concerns about group respectability and avoiding blame, ii) lower levels of political trust, iii) stronger social ties, and iv) distinct cultural practices. While our research design does not allow us to test these mechanisms definitively, we provide evidence on the expected empirical implications for each mechanism. We find the most compelling evidence for the first explanation—the politics of respectability—weaker evidence for the trust mechanism. For these reasons, we concentrate on the first two potential mechanisms and discuss the latter two in the appendix.

4.7.1 Concern for ingroup respectability and blame avoidance

As a minority group in India, Muslims are subject to negative stereotypes and prejudice that marginalized minorities face in a variety of contexts. For instance, a recent study on twenty-four major states, including Delhi, finds that the majority of respondents perceived Muslims to be less patriotic than members of other religious communities. Muslims are also believed to be more violent than other groups. Respondents in Delhi expressed some of the most negative perceptions of Muslims in India (Centre for Regional Political Economy 2019).

As discussed earlier, some members of persecuted minority groups react to negative stereotypes about their group by behaving in accordance with the “politics of respectability” (Higginbotham 1993; Jefferson 2018). The empirical implication of this mechanism will be the following: Muslims who identify more strongly with their community should present stronger treatments. Because these individuals are more tied to their community, we should expect them to be more interested in maintaining a positive image of their group by ensuring that their members avoid blame for community issues. More importantly, this does not imply that more religious respondents are more likely to respond positively to the treatment. Rather, the politics of responsibility point to differences in the way that religious practice and intragroup ties function for Muslims, potentially moderating the effects of the accountability treatments on the outcome.

Our survey contains nine questions about individual religious practices and ties to the religious community. We used parallel analysis to determine that four factors best account for variation in responses to these questions¹⁷ and label each based on the variables that loaded most heavily onto them. The first, “Institutionally Connected,” incorporates two variables, one measuring respondents’ participation in religious organizations and the other measuring connections to local religious leaders. The second, “Obligated to

¹⁷Our appendix includes a detailed description of the methods used to obtain these four factors.

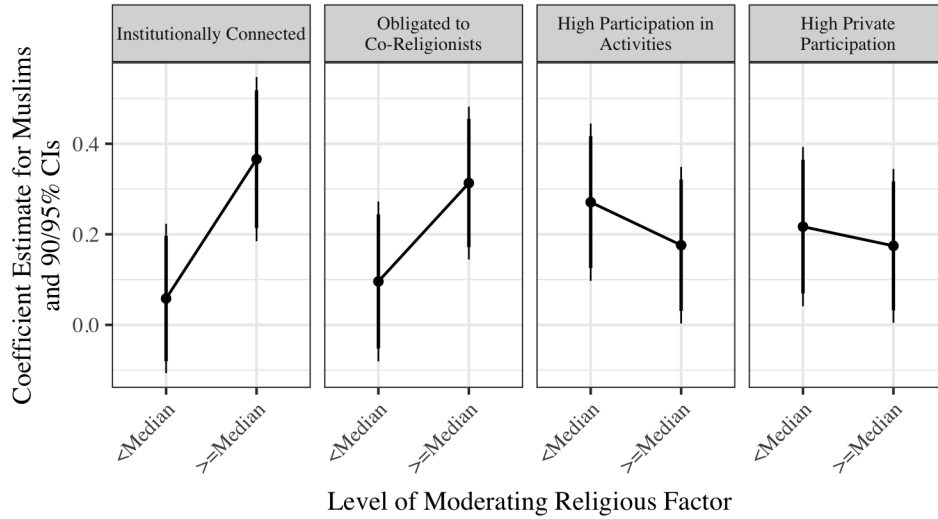


Figure 4.2: Estimates of Main Analysis Interaction Coefficient, by Religiosity Variables.

Co-religionists,” is primarily composed of two questions measuring respondents’ feelings of obligation to help co-religionists. The third, “High Participation in Activities,” consists of four questions that measure respondents’ level of participation in a variety of religious activities such as visiting houses of worship, attending services, fasting, and donating. The last, labeled “High Private Participation,” is primarily a measure of how often the respondent prays. Importantly, the politics of respectability implies that only some of these measures should matter for moderating the treatment effect for Muslims — specifically, those that entail collective rather than private religious engagement, including the “Institutionally Connected,” “Obligated to Co-Religionists,” and “High Participation in Activities” factors. Thus, moderating effects should only occur among Muslims who measure high in factors that indicate strong ties to fellow members of their religious group and investment in religious institutions.

We estimated our main model for participants with below and above Muslim respondents’ median values for each of the four factors, resulting in eight estimates for the effect of the treatments for Muslims (two for each factor). Figure 4.2 depicts the eight estimates based on levels of religious engagement factors. The results demonstrate a pattern that is mostly consistent with a concern for ingroup respectability: The effect of the treatments is particularly strong for Muslims who are connected to religious institutions (an increase of 0.31 in the coefficient estimate) and those who feel obligated to their co-religionists (an increase of 0.21). On the other hand, greater participation in religious practices by itself does not correspond with an increase in the coefficient estimates, in fact it is associated with small and insignificant decreases (of 0.09 and 0.04)

in the estimate.

These findings suggest that Muslims who may care more about the respectability of their religious group are driving the increased propensity to contribute to the drainage scheme in the face of diverse forms of accountability.

4.7.2 Low Political Trust

Another possible explanation for why our treatments make Muslims contribute more relates to political trust. According to this argument, which was pre-registered, those who are dissatisfied with the quality of service provision provided by local institutions and leaders are more likely to be open to collective action and market-related solutions and, therefore, more likely to respond to accountability measures. As discussed earlier, non-representative state institutions are likely to result in lower levels of political trust among minorities. Muslim representation in elected bodies in India has steadily declined since the early nineties with the rise of the BJP (Gayer 2012). Muslims underrepresentation in the bureaucracy and law enforcement agencies is particularly alarming in Delhi. A recent nationally representative survey finds that Muslims hold a more negative view of the police as compared to all other religious groups (Centre for Regional Political Economy 2019). The systematic targeting of Muslims since the BJP assumed power in 2014 is likely to have further reduced their trust in state institutions. Among sixteen state institutions, for example, the Prime Minister's office under Modi was the least trusted elected office among Muslims respondents (Centre for Regional Political Economy 2019). Given Muslims' negative experiences with some government institutions, it is possible that they have lower trust and, therefore, respond more to our treatments. Note that, in contrast to the mechanism in the previous section, we are not positing that political trust functions differently for Hindus and Muslims; this mechanism depends instead on differing levels of political trust between groups.

To evaluate the effects of political trust, we examine differences in levels of trust between Hindus and Muslims in specific institutions and then compare the interaction effects between the treatments and these ratings. Table 4.5 shows that the largest difference between Hindus and Muslims' rating of political institutions include the prime minister, the police, political parties, and the *pradhan*. Both communities rate the local government, or the MCD, at comparable levels. Figure 3 presents the interaction effects for the two local institutions that are involved in drainage provision—the *pradhan* and the MCD. Because of the Hindu-Muslim difference in *pradhan* ratings and the interaction effect shown in Figure 4.3, we conclude that lower trust in the *pradhan* may account for our results in the main analysis.

Table 4.5: Levels of Trust in Political Figures and Institutions, by Religion

	Pradhan		MCD		Prime Minister		Parties		Police		Chief Minister	
	\bar{X}	SD	\bar{X}	SD	\bar{X}	SD	\bar{X}	SD	\bar{X}	SD	\bar{X}	SD
Hindu	2.13	1.06	2.14	0.98	2.93	1.01	1.96	1.01	2.63	1.06	1.62	0.49
Muslim	1.83	0.99	2.12	0.97	2.07	1.01	1.65	0.89	2.19	1.02	1.62	0.49

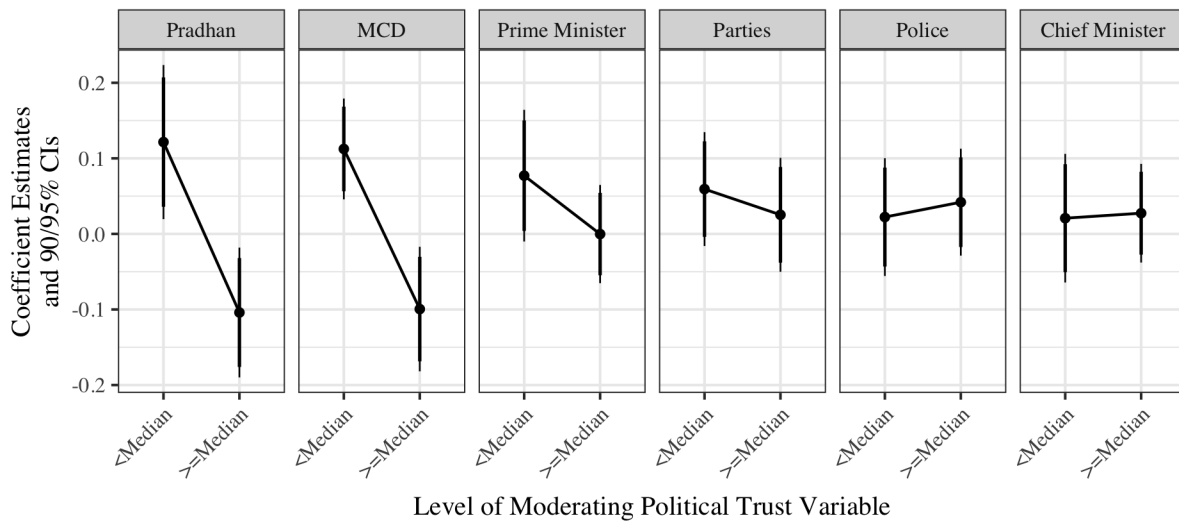


Figure 4.3: Effect of Combined Treatments, by Levels of Political Trust.

These results, however, need some qualification. Though *pradhans* have been shown to be important actors in local level collective action in Indian slums (Auerbach 2016; Auerbach and Thachil 2018; Auerbach 2017; Chidambaram 2020; Jha, Rao, and Woolcock 2007), *pradhans* in our sites were not seen as uniformly effective. Almost 40% of respondents did not rate their *pradhan* because they were not familiar enough with them. This number rises to just over 50% for Muslim respondents. Evidence from our qualitative fieldwork presents similar results. Residents in *Jawahar Mohalla*, for example, complained that they have not had a *pradhan* for more than a decade. In another area, *Daya Basti*, residents were politically aware - they knew about their state legislator but could not name their *pradhan*. We recognize that even if local slum leaders may not be effective, they are the primary channel through which the urban poor access the state in India. They yield power at the local level and we expect *pradhans* to be in a position to hold slum residents socially accountable.

4.7.3 Minority social ties

Perhaps because of their status as a minority group, Muslims may have stronger social ties with their neighbors, making them more likely to respond to the accountability treatments than Hindus. Four types of measures in the survey instrument tap into this potential mechanism, including questions about how much neighbors know each other, how often the respondent talks to their neighbors, a set of five questions regarding how often respondents' neighbors help with a set of common issues (e.g. childcare and solving disputes), and a set of nine questions on the extent to which respondents feel they are obligated to help different groups of people (e.g. a member of their immediate family, a neighbor, etc.). As with political trust, a first step empirical implication of this mechanism is a difference in levels of social ties between Hindus and Muslims. In this regard, we find no variation across the two religious groups: Across all four measures, Hindus and Muslims are virtually indistinguishable from each other, leading us to reject an explanation based these types of social ties (see the appendix for exact numbers).

4.7.4 Cultural practices and norms

We also find little evidence for cultural arguments explaining the difference between the two groups. There is no difference in caste between Muslims and Hindus in our sample. Religious cultural practices, such as participation in communal religious activities, is not associated with an interactive effect with the treatments. Additionally, Muslims showed no more concern than Hindus for drain cleaning (see appendix).

There are good reasons to be skeptical about the relevance of norms of purity and pollution in the context of urban slums. First, open defecation is largely a rural phenomenon in India. About 92% of the households without access to toilets lived in rural areas (Spears and Thorat 2019). Further, open defecation in rural areas represents choice and behavior, rather than affordability – i.e. households with the option of a working latrine chose open defecation (Coffey et al. 2014). Second, we focus on drains, not toilets. Hindus religious beliefs about impurity should be more pertinent to defecation than to drains more generally. If anything, we expect that Hindus have a strong cultural incentive to support drain cleaning initiatives in their neighborhoods. Third, a large section of the population in slums is composed of residents from lower caste backgrounds (Banerjee, Pande, and Walton 2012), yet notions of purity are strongest among upper caste members. Finally, members of different caste groups live in close proximity to each other in urban slums and ideas about purity are weaker in the city. Furthermore, in the context of poor sanitation services, it is difficult to remain faithful to caste-based notions of purity.

4.8 Conclusion

In this paper, we find that Hindus and Muslims in similarly disadvantaged communities in Delhi display divergent preferences regarding cooperation on the provision of local public goods. On average, members of the dominant Hindu community showed less willingness to contribute towards a proposed local drainage scheme as compared to members of the minority Muslim community. Based on quantitative data from our survey and qualitative data from our own fieldwork as well as existing research, we find evidence that differential group *status* shapes the differential propensity to contribute to local public goods provision. In response to efforts to promote neighborhood-level accountability, Muslims who care more about maintaining a respectable image of their religious group in society may be driving the increased propensity to contribute to a local drainage scheme.

Our work contributes to the vibrant research program on diversity and development by highlighting an additional mechanism that may moderate the relationship between diversity and local public goods provision, notably group status. To uncover the microfoundations of the “diversity deficit” hypothesis, recent scholarship has examined the role of social norms that facilitate cooperation in homogenous communities (Habyarimana et al. 2009). Among neighbors and peers, these may take the form of social sanctions, through measures like public shaming and gossip. For some, perceived violations of norms may be especially vexing

when perpetrated by ingroup members, inviting especially strong efforts to police ingroup behavior. Alternatively, social norms promoting cooperation may rely on more top-down or vertical enforcement through pressure from community leaders. The interventions in this study tap into these diverse forms of accountability mechanisms but focus on the way they may operate differently in high and low status groups.

Scholars increasingly push back against concepts of ethnic fractionalization that treat groups as equivalent, instead examining the effects of inequality between groups (Baldwin and Huber 2010). Economic differences between groups may be a stronger predictor of public goods provisions than diversity and cultural differences between groups per se. Our approach further widens that concept of intergroup differences by highlighting differences not just in socioeconomic outcomes across groups but also of relative group status. Indeed, the very existence of intergroup economic disparities may reflect historical discrimination against and even persecution of marginalized groups. In contemporary India, the gap between Muslims and Hindus is manifested not only in socioeconomic differences but also in social status, with high stakes for Muslims who are increasingly persecuted. Beyond India, these findings are likely to resonate broadly because many developing countries suffer from poor public goods provision in urban slums and encompass residents from unequal status groups.

Our findings suggest multiple areas for further study. First, the research program on intergroup politics should incorporate greater attention to status differences across groups and their potential consequences for *intragroup* dynamics. In particular, scholars should develop and test contextual valid measures that tap into the politics of responsibility, or the ways in which members of low status groups might adopt attitudes and behaviors aiming to police the behavior of fellow group members in order to enhance others' perceptions of the ingroup. Second, our study identifies and begins to explore potential mechanisms for the apparent minority effect in intergroup exchanges around local public goods. Future research should gather further qualitative and quantitative data and design tests to explore how membership in a persecuted minority shapes norms around contributions to communal goods.

From a policy perspective, our study highlights the unique challenges of public service provision for members of minority groups in informal settlements in a world characterized by rapid urbanization (United Nations 2019). Drainage and sanitation is clearly a failure of public goods in the slums of Delhi and many megacities in other developing countries, prompting residents to look beyond the state for solutions. In ethnically diverse communities, however, collective action to provide local public goods may be all the more challenging when residents must overcome status hierarchies. In our study, members of the dominant

group (Hindus) and subordinate group (Muslims) expressed relatively strong interest in the program, but only the latter responded positively to hypothetical mechanisms aimed at promoting contributions to the collective good. If collaboration across status divides exacerbates the well-known challenges of public goods provision in diverse communities, then local officials and community representatives must account for the specific circumstances facing persecuted minorities as they confront the demands of urban governance on a scale unprecedented in human history.

A | Appendix to Chapter 2

A.1 Demographics and Treatments

Table A.1: Omnibus Balance Tests

	Chi Squared	Degrees of Freedom	P-value
2016 Experiment	6.95	11	0.804
2018 Experiment	10.7	15	0.773

Table A.2: Key Survey Demographics of Focus Group Participants

Characteristic	2016		2018	
	N	%/Mean	N	%/Mean
Overall N	81		86	
Gender (1 = Female, 0 = Male)	46	58%	56	65%
Age				
18–24	7	9%	7	8%
25–34	28	35%	29	34%
35–44	19	23%	23	27%
45–54	14	17%	11	13%
55–64	10	12%	10	12%
65–74	3	4%	5	6%
75+	0	0%	1	1%
Income, per month				
Up to NIS 5,270	21	26%	22	26%
NIS 5,271–9,000	32	40%	23	27%
NIS 9,001–13,990	15	19%	21	24%
NIS 13,991–19,980	10	12%	11	13%
NIS 19,981–41,790	2	2%	1	1%
Other/refuse to answer/no income	1	1%	8	9%
Education				
High school graduate or less	15	19%	18	21%
Some college/higher education	18	22%	10	12%
Higher education with professional certification	5	6%	6	7%
Higher education with bachelor's degree	30	37%	33	38%
Higher education with master's or higher	13	16%	17	20%
Other/refuse to answer	0	0%	2	2%
Military Service (1 = Any, 0 = None)	69	85%	73	85%
Religiosity				
Secular	58	72%	61	71%
Traditional	18	22%	16	19%
Religious	3	4%	7	8%
Orthodox	2	2%	2	2%
Political Ideology (1–7, Higher = Conservative)		4.3		4.6
Certainty in Conflict Position (1–6, Higher = Certain)		4.6		4.6
Exposure to Political Violence (1 = Any, 0 = None)	27	33%	25	30%

Table A.3: A selection of the confederate’s script for treatment assignments

Neutral	Connected Moral	Pragmatic
<i>1. Was the leaders’ condemnation justified?</i>		
“I understand the positions of those on both sides... In the end, I think it is difficult to decide which of the sides is right and whether or not to condemn the soldier’s actions.”	“I understand the positions of those on both sides... However... the soldier was in the wrong when he killed someone who no longer posed a threat... when someone behaves wrongly, as happened in the case of the soldier, we as a group need to express our opinion on the matter.”	“I understand the positions of those on both sides... However... not condemning this incident would be damaging for Israel, for its security... when an incident like the shooting in Hebron occurs, that poses a threat to Israel’s interests, we as a group need to take a stand on the matter.”
<i>2. What should the criteria for condemnation be?</i>		
“Regarding criteria... every situation has its own peculiarities... we should trust out leaders... to decide.”	“Regarding criteria... the heart of the matter for me is whether or not the action was right. If the soldier broke the rules, then he deserves condemnation and punishment. The decision about whether or not to condemn should be based on the moral aspect of the matter. When someone’s actions violate moral standards, we should condemn them.”	“Regarding criteria... the heart of the matter for me is not whether or not the action was right, in spite of the fact that I think the soldier broke the rules; rather, I think that the decision... should be based on what is best for the state of Israel. When someone’s actions threaten all of us, we should condemn them.”
<i>3. What would you have done if you were Prime Minister?</i>		
“First thing I would... make sure an investigation had begun... However, beyond this, I don’t think I would provide a response to the incident.”	“First thing I would... make sure an investigation had begun... [and] beyond this, I would condemn the incident... by connecting it to important moral standards. I would note that the soldier broke standards that we expect soldiers to obey.”	“First thing I would... make sure an investigation had begun... [and] beyond this, I would condemn the incident... by connecting it to the security of Israel. I would note that in order to keep the state of Israel strong and secure, we need to condemn incidents like these.”

A.2 Regressions

Table A.4: Regression of opposition to soldier on treatment assignment

	<i>Dependent variable:</i>			
	2016	2018	2016, Controls	2018, Controls
	Opposition to Soldier (1 to 7)			
Moral	0.928** (0.435)	0.477 (0.470)	0.881** (0.375)	0.361 (0.416)
Pragmatic	0.698 (0.435)	-0.156 (0.470)	0.568 (0.378)	0.040 (0.414)
Political Ideology			-0.608*** (0.121)	-0.639*** (0.141)
Religious ID			-0.089 (0.243)	-0.310 (0.239)
Constant	3.952*** (0.290)	4.298*** (0.344)	6.772*** (0.601)	7.661*** (0.691)
Observations	81	86	81	80
R ²	0.061	0.025	0.322	0.315
Adjusted R ²	0.037	0.001	0.287	0.278

Note:

*p<0.1; **p<0.05; ***p<0.01

Table A.5: Regression of support for elite condemnation on treatment assignment

	<i>Dependent variable:</i>			
	Support for Elite Condemnation (1 to 7)			
	2016	2018	2016, Controls	2018, Controls
Moral	0.828* (0.420)	-0.397 (0.386)	0.844** (0.414)	-0.327 (0.347)
Pragmatic	1.228*** (0.420)	0.103 (0.386)	1.241*** (0.417)	0.409 (0.346)
Political Ideology			-0.294** (0.133)	-0.352*** (0.118)
Religious ID			0.266 (0.268)	-0.305 (0.199)
Constant	1.952*** (0.281)	2.481*** (0.282)	2.854*** (0.663)	4.333*** (0.576)
Observations	81	86	81	80
R ²	0.104	0.024	0.159	0.217
Adjusted R ²	0.081	0.00002	0.115	0.175

Note:

*p<0.1; **p<0.05; ***p<0.01

Table A.6: Regression of rating confederate vs. others on treatment assignment

	<i>Dependent variable:</i>			
	Rating Confederate Vs. Others (-7 to 7)			
	2016	2018	2016, Controls	2018, Controls
Moral	-0.991** (0.450)	-0.455 (0.342)	-0.981** (0.457)	-0.393 (0.351)
Pragmatic	-0.221 (0.450)	-0.312 (0.345)	-0.201 (0.461)	-0.113 (0.351)
Political Ideology			0.019 (0.147)	-0.027 (0.120)
Religious ID			0.068 (0.296)	-0.273 (0.203)
Constant	0.789** (0.301)	0.239 (0.251)	0.605 (0.732)	0.699 (0.584)
Observations	81	85	81	79
R ²	0.062	0.022	0.063	0.045
Adjusted R ²	0.038	-0.002	0.014	-0.007

Note:

*p<0.1; **p<0.05; ***p<0.01

Table A.7: Regression of importance of silencing others on treatment assignment

	<i>Dependent variable:</i>			
	Importance of Silencing Others (1 to 7)			
	2016	2018	2016, Controls	2018, Controls
Moral	-0.628** (0.286)	-0.381 (0.353)	-0.624** (0.282)	-0.449 (0.357)
Pragmatic	-0.587** (0.283)	-0.097 (0.355)	-0.563** (0.281)	-0.206 (0.358)
Political Ideology			0.197** (0.090)	0.061 (0.122)
Religious ID			-0.045 (0.181)	0.248 (0.206)
Constant	4.158*** (0.189)	4.063*** (0.258)	3.354*** (0.447)	3.479*** (0.594)
Observations	80	85	80	79
R ²	0.076	0.016	0.134	0.061
Adjusted R ²	0.052	-0.008	0.088	0.010

Note:

*p<0.1; **p<0.05; ***p<0.01

Table A.8: Regression of compromise over less difficult issues on treatment assignment

	<i>Dependent variable:</i>			
	Compromise, Less Difficult (1 to 5)			
	2016	2018	2016, Controls	2018, Controls
Moral	0.343 (0.234)	0.166 (0.219)	0.286* (0.165)	0.109 (0.191)
Pragmatic	0.591** (0.234)	-0.019 (0.219)	0.454*** (0.166)	-0.003 (0.190)
Political Ideology			-0.398*** (0.053)	-0.279*** (0.067)
Religious ID			-0.272** (0.107)	-0.222** (0.110)
Constant	2.665*** (0.156)	2.800*** (0.159)	4.827*** (0.264)	4.460*** (0.322)
Observations	81	84	81	78
R ²	0.077	0.011	0.555	0.324
Adjusted R ²	0.054	-0.013	0.532	0.287

Note:

*p<0.1; **p<0.05; ***p<0.01

Table A.9: Regression of compromise over more difficult issues on treatment assignment

	<i>Dependent variable:</i>			
	Compromise, More Difficult (1 to 5)			
	2016	2018	2016, Controls	2018, Controls
Moral	0.201 (0.284)	0.347 (0.286)	0.186 (0.249)	0.195 (0.235)
Pragmatic	0.161 (0.284)	0.227 (0.286)	0.106 (0.251)	0.110 (0.234)
Political Ideology			-0.399*** (0.080)	-0.435*** (0.082)
Religious ID			0.065 (0.162)	-0.292** (0.135)
Constant	2.839*** (0.190)	2.308*** (0.208)	4.505*** (0.399)	4.824*** (0.396)
Observations	81	84	81	78
R ²	0.007	0.018	0.261	0.415
Adjusted R ²	-0.018	-0.006	0.222	0.383

Note: *p<0.1; **p<0.05; ***p<0.01

Table A.10: Regression of letter rating on treatment assignment

	<i>Dependent variable:</i>			
	Letter Rating (1 to 5)			
	2016	2018	2016, Controls	2018, Controls
Moral	0.677** (0.324)	-0.217 (0.287)	0.667** (0.299)	-0.297 (0.270)
Pragmatic	1.300*** (0.327)	0.111 (0.291)	1.280*** (0.305)	0.237 (0.274)
Political Ideology			-0.376*** (0.096)	-0.345*** (0.094)
Religious ID			0.132 (0.193)	-0.128 (0.154)
Constant	1.783*** (0.218)	2.300*** (0.212)	3.255*** (0.482)	4.051*** (0.455)
Observations	79	83	79	77
R ²	0.173	0.018	0.317	0.226
Adjusted R ²	0.151	-0.007	0.280	0.183

Note:

*p<0.1; **p<0.05; ***p<0.01

A.3 Pooled Results

Table A.11: Pooled Results

Outcome and Treatment	Effect	SE	Z	P
Opposition to Soldier (1 to 7)				
Moral	0.72	0.32	2.26	0.02
Pragmatic	0.30	0.32	0.95	0.34
Support for Elite Condemnation (1 to 7)				
Moral	0.16	0.28	0.57	0.57
Pragmatic	0.62	0.28	2.17	0.03
Rating Confederate Vs. Others (-7 to 7)				
Moral	-0.65	0.27	-2.39	0.02
Pragmatic	-0.28	0.27	-1.02	0.31
Importance of Silencing Others (1 to 7)				
Moral	-0.53	0.22	-2.38	0.02
Pragmatic	-0.40	0.22	-1.79	0.07
Compromise, Less Difficult (1 to 5)				
Moral	0.25	0.16	1.56	0.12
Pragmatic	0.27	0.16	1.67	0.10
Compromise, More Difficult (1 to 5)				
Moral	0.27	0.20	1.36	0.17
Pragmatic	0.19	0.20	0.96	0.34
Letter Rating (1 to 5)				
Moral	0.18	0.21	0.82	0.41
Pragmatic	0.64	0.22	2.92	0.003

A.4 Textual Analysis

2016 Experiment Topic Label	English Words (Translated)
Quick Life-or-Death Decisions	kill, come, murder, killed, consideration, weapon, all of a sudden
Morality and Justice	again, moment, opinion, connect, moral, incident, justice
State Condemnations	damage, condemnation, leader, state, law, citizen, statement
Details about Incident	soldier, army, person, exit, life, detail, conclusion
Codes and Criteria	criteria, determine, speak, to you, code, true, incident
Information about the Incident	information, situation, again, story, statement, from, issue
Military And Defense Institutions	framework, minister, occur, security, say, army, prison
Filler and Descriptive Words	day, want, one, manner, happen, moral, here
Filler Words	like, end, difficult, person, very, know, attack
Video Evidence of Guilt/Crime	video, justify, guilt, yes, crime, burn, agree

Table A.12: Top FREX words from the textual analysis for the 2016 experiment.

2018 Experiment Topic Label	English Words (Translated)
Quick Decision for a Child	child, leader, for us, situation, here, place, all of a sudden
Law and Politics	punishment, receive, citizen, law, politician, month, exit
Threat of Killing	kill, killed, okay, come, see, show, knife
The Trial	trial, law, fact, many, lawyer, to you, danger
State Responses	respond, make a mistake, matter, security, things, minister, problem
Statement to World	defend, statement, in reality, state, need, give, world
Neutral/Filler Words	agree, yes, exactly, think, very, neutral, arrive
Condemn or Justify a Child	weapon, child, condemnation, justified, position, army, want
Media Reaction vs. Truth	truth, much, too, media, less, think, actually
Video Description	like, remember, again, video, simply, jewish, person

Table A.13: Top FREX words from the textual analysis for the 2018 experiment

B | Appendix to Chapter 3

B.1 Survey Administration and Sampling Details

B.1.1 Site Selection

We selected 12 candidate slum settlements in different parts of Delhi based on their ethnic composition. Our initial goal was to identify three types of slum communities for the study, including “heterogeneous” communities with a close to 50/50 Hindu-Muslim split, “homogeneous” areas with an overwhelmingly Muslim majority, and “homogeneous” sites with an overwhelmingly Hindu majority. We relied on publicly available booth-level electoral lists, collected from the website of the Chief Electoral Officer of Delhi, to estimate the proportion of the Muslim population in each settlement. The names in the electoral lists were digitized and matched against a list of common South Asian Muslim names to arrive at an estimate of Muslim population in each *basti*. While *bastis* displayed greater variation in the proportion of Muslim population, settlements at the city level tended to be largely Muslim or Hindu. Mixed communities were identified at the *basti* level from the subset of the following Hindu or Muslim settlements. The electoral lists also provided information on the number of voters in each *basti* and hence allowed us to generate a rough estimate of the population of settlements.

B.1.2 Guidelines for Collecting Settlement-Level Data

What follows were the instructions given to the Centre for Policy Research when doing initial site selection.

—

In order to begin to explore the public services in detail, some background about the sites is important. This will draw upon a combination of observations and information obtained through interviews. *Please note the sources of information in the field notes—who are you interviewing and the selection criteria (why*

did you choose that particular person to interview, i.e. why do you think that they are a reliable source of information).

Approach and location of the settlement:

Prepare a rough map of the settlement with respect to surrounding roads and landmarks, approach roads/major flyovers and main drainage networks.

Focus on the following public services:

Drainage, Local Lanes, Garbage Collection, Community Toilets, Water Supply.

Drainage:

How does the drainage connect to the city-wide network?

What is the state of the drainage – is it clogged/ looks like it's well maintained?

Who is in charge of maintaining the drainage?

Ask local residents about issues with drainage—how often is it clogged? Does the MCD maintain it? If not, do residents attempt to clean it themselves? Are NGOs or other community organizations involved?

Under the current system, how many people have to coordinate in order to successfully maintain drainage?

Which blocks/communities are connected to the drainage network? What is the demographic profile of those communities? *How are you finding out the demographic characteristics?*

- Mostly Hindu

- Mostly Muslim

- Mixed. If mixed, roughly what proportion of the two?

Local residents and public service delivery:

Are the residents aware of paid arrangements for a third-party service (either the government or a private company), either in their own settlement or in others, for taking care of the good?

How open would they be, personally, to having a third party take care of this particular thing, for a reasonable price, and what would a reasonable price be?

If they are open to a third party taking care of the public good/utility, then would they be more responsive to an NGO, CGO or for-profit firm?

Local Demographics:

In addition to the demographic profile of the community, find out what the salient cleavage along caste/religious line is? If there are any conflicts on public services or other local issues, is it along certain communities?

–

Report after the first site visit to decide if the settlement is suitable for the study, based on two criteria: The extent to which these public services are important in the settlement and the extent to which they can be administered by a hypothetical service.

B.1.3 Administration

Enumerators worked in teams of 3 to 5 individuals, each of which were assigned to portions of the settlement communities where they carried out the survey. Because of the extreme heat during the survey period and the work schedules of residents of the settlement communities, the enumerators mainly carried out surveys in the morning and early afternoon. They were instructed to approach every third household in rounds until all households in an area had been approached. Enumerators were instructed to interview only one person per household. When approaching a residence, they first introduced their company and the topic of the survey. They then asked to speak to the person in the house responsible for financial and other major household decisions. If this individual was not available, they asked for someone in the house who could make such decisions in the absence of the head of household. If neither was available, then the enumerator attempted to come back a different time. Because of this approach, our sample is overwhelmingly (80%) male.

B.2 Survey Instrument: Key Measures and Experimental Manipulations

B.2.1 Outcome Measures

We use five questions to measure outcomes from our experiment, all of which are answered on a 1–4 scale (with 4 indicating more interest in the drainage cleaning program):

- Would this program be beneficial for your neighborhood?
- How interested would you be in the program?
- How likely would you be to pay the monthly fee?

- Would you be willing to sign a six-month contract for this service?
- How likely would you be to try to get your neighbors to sign up for the program?

We combine these measures into an index by averaging a respondent's answers across the five questions. For those who do not answer all of the questions, we average the questions they did respond to. Figure 5 in our robustness checks shows the distribution of this index, which is trimodal because of small floor and ceiling effects (which we address in our robustness checks). These five questions have a Cronbach's alpha of 0.89, indicating a high internal consistency for this index.

B.2.2 Measures of Religion and Religiosity

There are eleven questions related to religion and religiosity in our survey; see Figure B.1 for a graphical representation of these questions. The first two are simple questions about religious affiliation:

- What is your religion?
- *If respondent is a Muslim:* To which sect do you belong, Shia or Sunni?

For the first question, respondents are asked to pick from the following list: Hindu, Muslim, Christian, Buddhist/Neo-Buddhist, Jain, or no religion. They are also able to specify their religion if it is not on the list. Because of our choice of neighborhoods, the vast majority of all respondents approached was either Hindu or Muslim. Those who responded with anything else were dropped from the survey. For the second question, respondents were asked to pick Shia or Sunni (almost all indicated they were Sunni).

After this, respondents were asked five questions about religious activities and practices:

- Please tell me, how often do you practice these religious activities?
 - Prayer/puja/namaz
 - Visiting temple/mosque/church/gurudwara
- And what about these, how often do you practice them?
 - Participating in kathas, sangats, bhajan-kirtans, jalsas, church services, etc.
 - Giving donations for religious activities

- Keeping fast, rozas, etc.

For the first two of these questions, respondents were asked to respond with one of the following: daily, 1–2 times a week, on festivals, during periods of distress, or never. For the last three, they were asked to respond with one of the following: Whenever I get a chance, sometimes, rarely, during periods of distress, or never.

Elsewhere in the survey, respondents were asked a series of questions about local social ties, and these included the following two questions measuring religious ties:

- We would like to understand how people interact in your area. How obligated are you to help the following people, even if it costs you a day's wages?
 - A member of your religion
 - Someone associated with the day-to-day functioning of their temple/mosque/gurudwara

Respondents stated their level of obligation on a three-point scale: Very obligated, somewhat obligated, or not obligated at all.

Lastly, respondents were also asked about their involvement in local organizations and with local elites with two questions:

- Now I will ask you about some important people. Please indicate if you are related to or a close friend of the following.
 - Religious leader in your area
- Have you or a member of your household participated in any of the following voluntary organizations?
 - Religious organization

For the first of these two questions, respondents indicated that they were the relative of, close friend of, both relative and close friend of, or had no relationship with a local religious leader. For the second, they responded with a simple yes or no.

For the nine questions about religious practice and activities, religious social ties, and religious involvement, we wanted to divide these questions into factors based on different facets of religious belief and

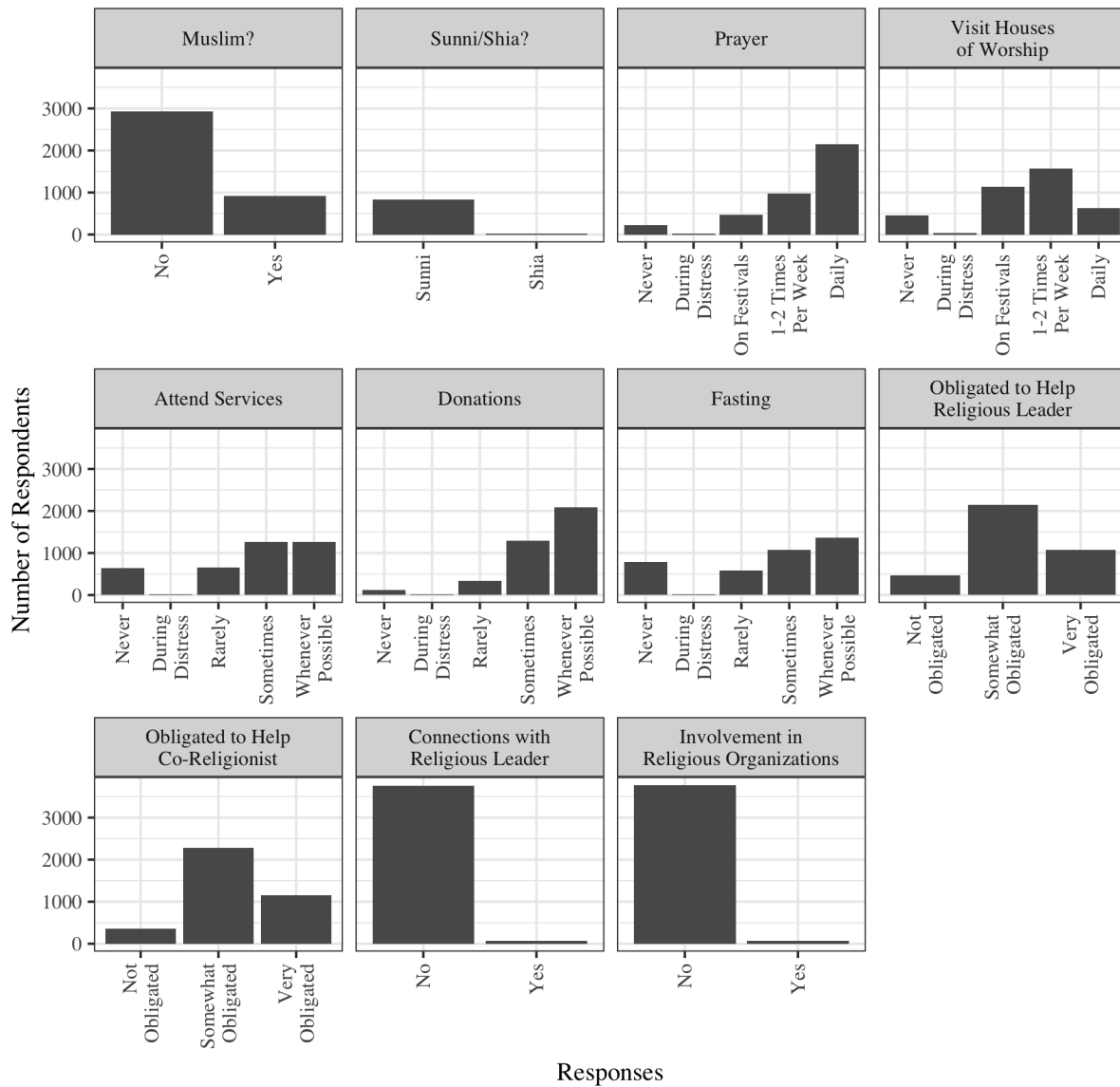


Figure B.1: Measures of Religion and Religiosity (low total count in the “Sunni/Shia?” panel is due to the fact that only Muslims responded to that question).

practice. We used parallel analysis to determine the number of factors, using a heterogeneous correlation matrix because of the dichotomous variables on religious involvement. This process indicated that four factors were appropriate for the data, which when generated have the loadings shown in Table B.1.

Table B.1: Factor Loadings for Questions on Religiosity

	Factor 1	Factor 2	Factor 3	Factor 4
	Institutional Connectedness	Obligation to Co-Religionists	Participation in Activities	Private Participation
Prayer	0.01	0.02	0.28	0.96
Visit Houses of Worship	0.12	-0.01	0.48	0.19
Attend Services	0.15	-0.01	0.78	0.002
Donations	-0.02	0.20	0.43	0.14
Fasting	0.06	-0.02	0.63	0.06
Obligated to Help Religious Leader	0.10	0.95	0.07	0.01
Obligated to Help Co-Religionist	0.12	0.81	0.01	0.01
Connections with Religious Leader	0.39	0.06	0.11	0.001
Involvement with Religious Organization	0.99	0.10	0.05	0.02

These factors align pretty clearly with the different sets of questions we asked. Factor 1, which we label as “Institutional Connectedness,” loads primarily on the questions regarding respondents’ connections with religious leaders and their involvement in religious organizations. Note that, in contrast to the other factors, this one has an especially skewed distribution because only a small proportion of respondents affirmatively responded to these two questions (see the final two panels in Figure 1). Factor 2, which we label as “Obligation to Co-Religionists,” loads primarily on questions regarding respondents’ feelings of obligation to help religious leaders or other co-religionists. It also to a lesser extent loads on a measure of how often the respondent donates to their religion. Factor 3, which we label as “Participation in Activities,” loads primarily on questions about attending houses of worship and services, donating, and fasting, with smaller loading on the question about frequency of prayer. Factor 4, which we label “Private Participation,” loads primarily on just the question about prayer.

These measures are used to test for one of our treatment mechanisms, as outlined in a subsequent section of the appendix.

B.2.3 Measures of Political Trust

There are six measures of political trust that we use in the survey. All of the political trust questions are in the same section of the survey and follow a common introduction:

- I'm going to name a number of institutions. For each one, please tell me how well you think they perform their responsibilities?
 - Prime Minister and Central Government
 - Chief Minister and Delhi Government
 - The MCD
 - Political Parties
 - The Police
 - Your Local Pradhan

Respondents rated these institutions on a 1–4 scale: Not well at all, somewhat well, mostly well, and extremely well.

In our control models, we include an index of all six of these measures. However, it should be noted that an index of these measures does not have very good internal consistency (Cronbach's alpha = 0.57). We also explored the creation of factors from these items; however, the number of factors suggested by parallel analysis (two) produced factors that were not substantively useful. In exploring mechanisms, we therefore look at measures individually, particularly focusing on local institutions that are of importance to drain cleaning and maintenance: evaluations of the MCD and the local Pradhan. However, it is important to note that there is a high degree of missingness for the Pradhan question. Our other questions only have between 25 (the MCD) and 266 (political parties) missing responses; the Pradhan question has 1,518 (~40%) missing responses. The cause of this missingness is that many respondents did not know enough about their Pradhan to feel comfortable evaluating them.

B.2.4 Measures of Social Ties

There are sixteen total questions in our survey that relate to respondents' social ties to other in their neighborhood. The first of these is an independent question that is not a part of a particular index:

- In your neighborhood, would you say:

With a three-point scale of possible responses: Few people know each other, some people know each other, or most people know each other.

The second social ties question is an independent question about the frequency with which respondents talk to their neighbors:

- How often do you talk to your neighbors?

With a five-point scale of possible responses: Never, only when necessary, rarely, sometimes, or daily.

Respondents are then asked to answer a set of five questions that form an index measure:

- When asked for help, do your neighbors help you with the following issues?
 - Childcare
 - Finances (help with small loans, etc)
 - Obtaining government services or documents
 - Solving disputes among family members or neighbors
 - Keeping the area clean

Respondents could answer one of the following on a four-point scale: Never, rarely, sometimes, or always. This index had a high internal consistency, with a Cronbach's alpha of 0.86.

Lastly, we included one additional nine-question index of social ties, this one measuring individuals' feeling of obligation to others in their area:

- We would like to understand how people interact in your area. How obligated are you to help with the following people, even if it costs you a day's wages?
 - A member of your immediate family (father, mother, children)
 - A member of your extended family (uncle, aunt, grandparent)
 - A neighbor
 - A distant relative
 - A member of your caste

- A member of your religion
- People who come from the same town/village
- A fellow work mate/co-worker
- Someone associated with the day-to-day functioning of their temple/mosque/gurudwara

Respondents could answer one of the following on a three-point scale: Not obligated at all, somewhat obligated, and very obligated. This index had a high internal consistency, with a Cronbach’s alpha of 0.93.

Unlike some other variables, these four measures—the two standalone questions as well as the two indices—varied little by the religious identity of respondents, as can be seen in Table B.2.

Table B.2: Social Ties by Respondent Religion

	Hindus		Muslims	
	\bar{X}	SD	\bar{X}	SD
Neighborhood Knows Each Other	0.86	0.27	0.90	0.23
Talk to Neighbors	0.90	0.19	0.93	0.16
Neighbors Help Me	3.11	0.79	3.12	0.78
Obligation to Others	2.38	0.46	2.35	0.42

B.2.5 Measure of Caste

We used an open-ended survey question to gather data on respondents’ caste identification. A closed-ended question would have been too limited for the variety of castes (e.g. our enumerators entered over 800 unique text strings to record responses to this question!). After administering the survey, a researcher identified the caste level (high vs. low) of all the unique responses. This resulted in 1,368 respondents identified as belonging to a low caste and 2,366 identified as belonging to a high caste. Proportion of caste membership based on this dichotomous categorization did not significantly vary by respondents’ religious identification; approximately 37% of Hindus and 36% of Muslims were categorized as low caste.

B.2.6 Geocoordinates and Diversity Measure

Because we were interested in how local diversity affected the treatments in our survey, we included two methods for collecting geolocation data from participants as the survey was administered. First, the tablets used to administer the survey automatically recorded the enumerator’s geolocation at the beginning

of the survey. Additionally, we instructed enumerators to use Google Maps to find their current location and record their coordinates in the notes section at the end of the survey. In piloting the survey, we found that simultaneously using both approaches was the best way to get reliable geolocation data. In assigning geolocations to respondents, we default to the manually-entered Google Maps coordinates and use the automatically-recorded coordinates as a fallback.

However, cellular service can be spotty in Delhi. This sometimes resulted in inaccurate geolocation coordinates, with the enumerator either being unable to access the Google Maps application and record their location or the coordinates providing the location of the closest cell phone tower or the center of town rather than the enumerator's precise location. Approximately 7% of our surveys had evidence that this occurred. For these cases, there were often coordinates available from the previous or next respondent for that same enumerator (or the same device) on the same day. We averaged these to provide approximate coordinates for most of the 7% of respondents who did not have reliable coordinates through our primary data collection. We felt it was better to provide an approximate estimate rather than eliminating these data. For five respondents, none of these methods provided reliable coordinates. Results that depend on geolocation data do not significantly vary if we exclude manually-edited coordinates.

B.2.7 Experimental Manipulations

The text of the experimental manipulations was as follows:

1: Control *Our basis for comparison with the treatments.*

This service was offered in and around the neighborhood of a resident of Delhi. He shares his experience: “Everything they told us about the service is true. Because of them, the drains are in a much better condition and there are no blockages due to garbage. The company is quick to respond in case people have any maintenance needs. But the most difficult aspect of this service is to get enough people to contribute. There were many residents in my neighborhood—among them, Sunil and Abdul—who did not contribute, and because of this the service could not be implemented in my neighborhood. But in the adjoining neighborhood, many residents did contribute, such as Salim and Mahesh, and the service was implemented as a result.”

If this service were implemented in your neighborhood, all information about you, including your name, address, and whether or not you contribute would remain anonymous and will not reported to the other

residents in your area.

2: Horizontal Accountability *Same as control until the final paragraph.*

... Once this service is started, community members would discuss amongst themselves who contributed and who did not pay the monthly fee. For example, in one of the neighborhoods that recently adopted this service, community members singled out and shamed the following non-contributors:

Javed, Vikas, Anwar, Gaurav, Quayum, and Sachin

3: Vertical Accountability *Same as control until the final paragraph.*

... Once this service is started, the local leader would inform local residents of the benefits of hiring the company and would follow up personally with some of those who did not contribute money to the startup fee. For example, in one of the neighborhoods that recently adopted this service, the local leader personally visited the following non-contributors and shamed them.

Javed, Vikas, Anwar, Gaurav, Quayum, and Sachin

4: Ingroup Underperformance (Hindu/Muslim) *Differs from control in the italicized portions of the following paragraphs. Portions of the treatment vary based on the religious identification of the participant; In the paragraphs below, religious-dependent portions are displayed in the format Hindu/Muslim.*

This service was offered in and around the neighborhood of a resident of Delhi. He shares his experience:

“Everything they told us about the service is true. Because of them, the drains are in a much better condition and there are no blockages due to garbage. The company is quick to respond in case people have any maintenance needs. But the most difficult aspect of this service is to get enough people to contribute. There were many residents in my neighborhood *of Saraswatinagar/Muhammadpur*—among them, *Sunil/Salim* and *Mahesh/Abdul*—who did not contribute, and because of this the service could not be implemented in my neighborhood. But in the adjoining neighborhood *of Muhammadpur/Saraswatinagar*, many residents did contribute, such as *Salim/Sunil* and *Abdul/Mahesh*, and the service was implemented as a result.”

If this service were implemented in your neighborhood, all information about you, including your name, address, and whether or not you contribute would remain anonymous and will not reported to the other residents in your area.

5: Horizontal Accountability and Ingroup Underperformance (Hindu/Muslim) *Same as the ingroup underperformance treatment until the final paragraph.*

... Once this service is started, community members would discuss amongst themselves who contributed and who did not pay the monthly fee. For example, in one of the neighborhoods that recently adopted this service, community members singled out and shamed the following non-contributors:

Prakash/Qasam, Gaurav/Jamal, Sachin/Quayum, Manish/Javed, Vikas/Iqbal, Arun/Anwar

6: Vertical Accountability and Ingroup Underperformance (Hindu/Muslim) *Same as the ingroup underperformance treatment until the final paragraph.*

... Once this service is started, the local leader would inform local residents of the benefits of hiring the company and would follow up personally with some of those who did not contribute money to the startup fee. For example, in one of the neighborhoods that recently adopted this service, the local leader personally visited the following non-contributors and shamed them.

Prakash/Qasam, Gaurav/Jamal, Sachin/Quayum, Manish/Javed, Vikas/Iqbal, Arun/Anwar

B.3 Additional Results

B.3.1 Models without Interactions

Table B.3 shows the results of the experiment on the outcome index and each of the separate outcomes, with no interaction effects. For this table, the treatments have been combined as they were for all analyses in the paper. Table B.4, on the other hand, shows the results with each separate treatment.

These tables show that, for our sample as a whole, there are uniformly null effects of the treatments for the combined treatments as well as for the separate treatments.

Table B.3: Model with Combined Treatments and No Interactions

	Dependent Variable:					
	Benefit	Interest	Fee	Contract	Influence	Index
Constant	3.02 (0.02)	2.95 (0.02)	2.74 (0.02)	2.46 (0.02)	2.51 (0.02)	2.73 (0.02)
Treatments (Combined)	0.02 (0.03)	-0.02 (0.03)	0.04 (0.03)	0.04 (0.03)	0.04 (0.03)	0.03 (0.03)
Observations	3,801	3,818	3,804	3,769	3,718	3,832
R ²	0.0002	0.0001	0.0003	0.0004	0.0004	0.0003
Adjusted R ²	-0.0001	-0.0001	0.0001	0.0001	0.0001	0.0000
Residual Std. Error	1.34 (df = 3799)	1.27 (df = 3816)	1.41 (df = 3802)	1.43 (df = 3767)	1.47 (df = 3716)	1.17 (df = 3830)
F Statistic	0.64 (df = 1; 3799)	0.48 (df = 1; 3816)	1.21 (df = 1; 3802)	1.42 (df = 1; 3767)	1.31 (df = 1; 3716)	1.11 (df = 1; 3830)

Table B.4: Model with Separate Treatments and No Interactions

	Dependent Variable:					
	Benefit	Interest	Fee	Contract	Influence	Index
Constant	3.02 (0.04)	2.95 (0.03)	2.74 (0.04)	2.46 (0.04)	2.51 (0.04)	2.73 (0.03)
Horizontal Accountability	0.04 (0.05)	-0.01 (0.05)	0.05 (0.06)	0.04 (0.06)	0.07 (0.06)	0.04 (0.05)
Vertical Accountability	0.03 (0.05)	-0.01 (0.05)	0.05 (0.05)	0.06 (0.06)	0.06 (0.06)	0.04 (0.04)
Ingroup Underperformance	0.03 (0.05)	0.001 (0.05)	0.04 (0.06)	0.04 (0.06)	0.04 (0.06)	0.03 (0.05)
Ingroup Und. + Hor. Acct.	0.01 (0.05)	-0.05 (0.05)	-0.02 (0.06)	0.001 (0.06)	-0.03 (0.06)	-0.01 (0.05)
Ingroup Und. + Vert. Acct.	0.02 (0.05)	-0.02 (0.05)	0.06 (0.05)	0.06 (0.06)	0.05 (0.06)	0.04 (0.04)
Observations	3,801	3,818	3,804	3,769	3,718	3,832
R ²	0.0002	0.0004	0.001	0.001	0.001	0.001
Adjusted R ²	-0.001	-0.001	-0.0004	-0.001	-0.0001	-0.001
Residual Std. Error	0.94 (df = 3795)	0.90 (df = 3812)	0.99 (df = 3798)	1.00 (df = 3763)	1.02 (df = 3712)	0.81 (df = 3826)
F Statistic	0.15 (df = 5; 3795)	0.31 (df = 5; 3812)	0.68 (df = 5; 3798)	0.44 (df = 5; 3763)	0.95 (df = 5; 3712)	0.55 (df = 5; 3826)

B.3.2 Interactive Model with Separate Treatments and Outcome Measures

In the main analysis in the paper, we opt to combine our five treatment groups. In the heterogeneous effects model, which we present in Table B.5 and Figure B.2, all treatments tend to have a same effect magnitude and direction, although there are exceptions. In particular, the Ingroup Underperformance + Vertical Accountability treatment has a particularly large effect on individuals' willingness to try to influence others to join the program. However, we think it is more fruitful to examine the overall pattern rather than one particular effect. If hypothesis testing, statistical significance would vary across the treatments and outcomes in this analysis, but point estimates for Muslims are always positive while for Hindus there is a clear null effect.

Table B.5: Interaction Model with Separate Treatments

	Dependent Variable:					
	Benefit	Interest	Fee	Contract	Influence	Index
Constant	3.06 (0.04)	2.98 (0.04)	2.79 (0.04)	2.47 (0.05)	2.55 (0.05)	2.76 (0.04)
Muslim	-0.16 (0.08)	-0.12 (0.08)	-0.20 (0.09)	-0.06 (0.09)	-0.16 (0.09)	-0.13 (0.07)
Horizontal Accountability	-0.02 (0.06)	-0.05 (0.06)	-0.01 (0.06)	-0.02 (0.06)	0.03 (0.07)	-0.01 (0.05)
Hor. Acct. x Mus.	0.25 (0.12)	0.12 (0.12)	0.23 (0.13)	0.23 (0.13)	0.16 (0.13)	0.19 (0.10)
Vertical Accountability	-0.03 (0.06)	-0.07 (0.06)	-0.04 (0.06)	0.01 (0.06)	-0.001 (0.07)	-0.02 (0.05)
Vert. Acct. x Mus.	0.20 (0.12)	0.24 (0.12)	0.34 (0.13)	0.21 (0.13)	0.24 (0.13)	0.23 (0.10)
Ingroup Underperformance	-0.04 (0.06)	-0.05 (0.06)	-0.02 (0.06)	0.02 (0.07)	-0.03 (0.07)	-0.02 (0.05)
Ing. Und. x Mus.	0.27 (0.12)	0.19 (0.12)	0.23 (0.13)	0.04 (0.13)	0.26 (0.13)	0.20 (0.11)
Ing. Und. + Hor. Acct.	-0.03 (0.06)	-0.10 (0.06)	-0.07 (0.06)	-0.01 (0.07)	-0.07 (0.07)	-0.05 (0.05)
Ing. Und. + Hor. Acct. x Mus.	0.17 (0.12)	0.19 (0.11)	0.19 (0.13)	0.06 (0.13)	0.14 (0.13)	0.14 (0.10)
Ing. Und. + Vert. Acct.	-0.03 (0.06)	-0.05 (0.06)	0.02 (0.06)	0.04 (0.06)	-0.04 (0.07)	0.003 (0.05)
Ing. Und. + Vert. Acct. x Mus.	0.18 (0.12)	0.12 (0.12)	0.13 (0.13)	0.06 (0.13)	0.38 (0.13)	0.15 (0.11)
Observations	3,801	3,818	3,804	3,769	3,718	3,832
R ²	0.002	0.002	0.003	0.002	0.004	0.002
Adjusted R ²	-0.001	-0.001	0.0002	-0.001	0.001	-0.0004
Residual Std. Error	0.94 (df = 3789)	0.90 (df = 3806)	0.99 (df = 3792)	1.00 (df = 3757)	1.02 (df = 3706)	0.81 (df = 3820)
F Statistic	0.68 (df = 11; 3789)	0.66 (df = 11; 3806)	1.06 (df = 11; 3792)	0.78 (df = 11; 3757)	1.31 (df = 11; 3706)	0.85 (df = 11; 3820)

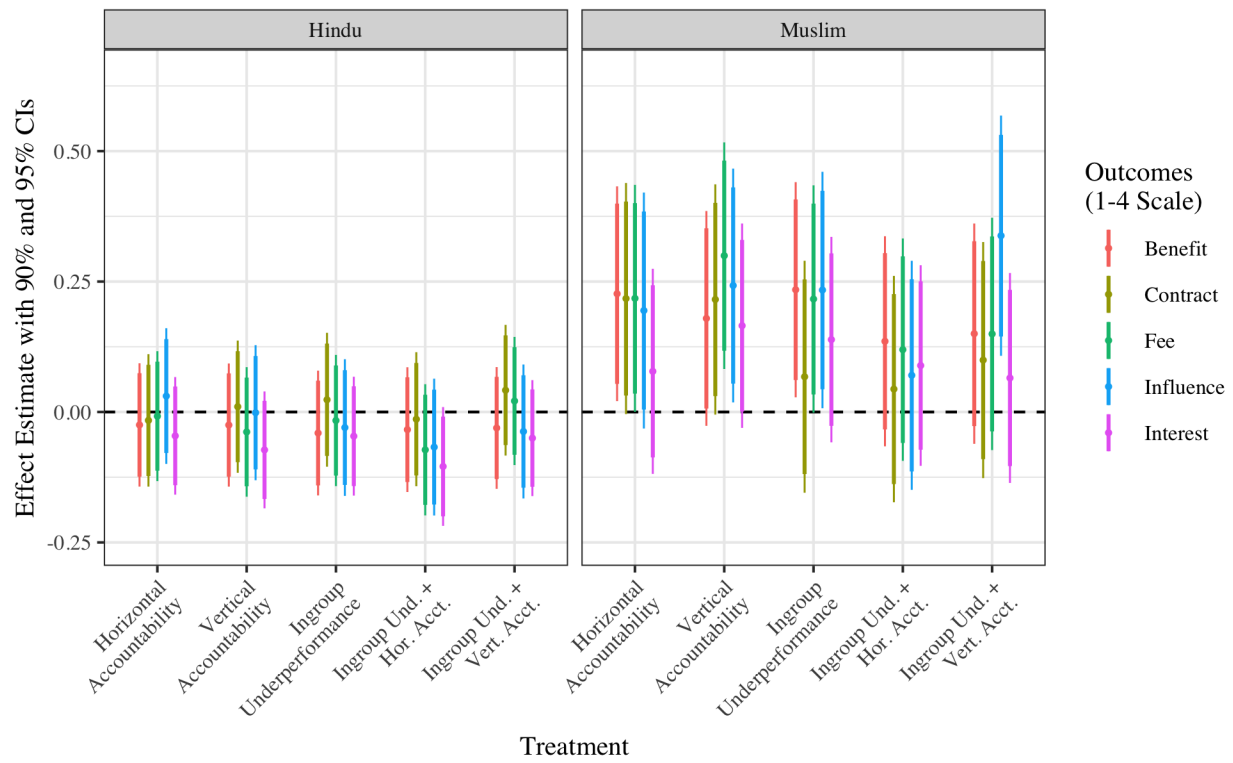


Figure B.2: Interaction Model with Separate Treatments

B.3.3 Full Control Model

In the paper, we provide a condensed version of our model with controls. In Table B.6 below, we provide the full model. Note that the coefficients on the combined treatment and the interaction terms vary only marginally between these specifications and are statistically significant at the 0.05 level in all of them.

Readers will note that we do not include neighborhood or enumerator fixed effects in any of these models. If we were running a non-interactive model, including these fixed effects as control variables is not problematic. However, our main model interacts the treatments with religious identification, which makes these fixed effects problematic for our coefficient estimates of interest. As noted in the paper and elsewhere in this appendix, areas differ greatly in their religious composition, such that religious identification is highly correlated with neighborhoods. Enumerator fixed effects are problematic for a related reason: enumerators were not randomly assigned to areas, so this issue of high correlation with specific neighborhoods also affects enumerator fixed effects as a control. Evidence that this is problematic includes wild swings in coefficient estimates and significance when fixed effects are introduced, as well as a very large (>50) variance inflation factor. We feel that, given this limitation, the best way to account for important neighborhood and enumerator characteristics is to focus on the characteristics that are important to our analysis. For this reason, our robustness checks include an examination of how our main model estimates vary based on the religious composition of neighborhoods as well as an examination of how enumerator religious identification affects our analysis.

Table B.6: Main Model with Controls

	Dependent Variable: Index of Favorability toward Drainage Program				
	(1)	(2)	(3)	(4)	(5)
Constant	2.76 (0.02)	1.00 (0.16)	1.01 (0.16)	0.93 (0.16)	0.94 (0.16)
Treatments (Combined)	-0.02 (0.03)	-0.00 (0.03)	0.00 (0.03)	0.00 (0.03)	0.00 (0.03)
Muslim	-0.13 (0.04)	-0.07 (0.04)	-0.06 (0.04)	-0.10 (0.04)	-0.09 (0.04)
Treatments (Combined) x Muslim	0.18 (0.06)	0.17 (0.06)	0.17 (0.06)	0.15 (0.06)	0.15 (0.06)
Gender (1 = Female, 0 = Male)		-0.01 (0.04)	-0.003 (0.04)	-0.01 (0.04)	-0.01 (0.04)
Married (1 = Married, 0 = Other)		0.02 (0.05)	0.00 (0.05)	0.01 (0.05)	0.01 (0.05)
Age		0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)

Table B.6 (Continued)

	Dependent Variable: Index of Favorability toward Drainage Program				
	(1)	(2)	(3)	(4)	(5)
Years Residing in Settlement		-0.00 (0.00)	-0.00 (0.00)	-0.00 (0.00)	-0.00 (0.00)
Ownership (1 = Own Residence)		-0.00 (0.04)	-0.01 (0.04)	0.00 (0.04)	0.00 (0.04)
Number of Rooms in Residence		0.04 (0.01)	0.04 (0.01)	0.03 (0.01)	0.03 (0.01)
Employment Status (1 = Employed)		0.18 (0.04)	0.19 (0.04)	0.19 (0.04)	0.18 (0.04)
Fin. Hardship (1 = More Hardship)		-0.03 (0.06)	-0.02 (0.06)	-0.05 (0.06)	-0.06 (0.06)
Assets Index (1-3, 3 = Most Assets)		0.09 (0.03)	0.09 (0.03)	0.08 (0.03)	0.09 (0.03)
Education Level (1-10 Scale)		0.02 (0.01)	0.02 (0.01)	0.02 (0.01)	0.02 (0.01)
Practice of Prayer (1-5 Scale)		0.06 (0.01)	0.06 (0.01)	0.06 (0.01)	0.06 (0.01)
Religious Index (1-5 Scale)		0.05 (0.02)	0.05 (0.02)	0.04 (0.02)	0.03 (0.02)
Political Efficacy Index (1-4 Scale)		0.04 (0.02)	0.04 (0.02)	0.05 (0.02)	0.05 (0.02)
Relation to Local Figures (1 = Any)		-0.04 (0.04)	-0.04 (0.04)	-0.06 (0.04)	-0.05 (0.04)
Involve in Local Org. (1 = Any)		-0.02 (0.05)	-0.02 (0.05)	-0.03 (0.05)	-0.02 (0.05)
General Social Ties (0-1 Scale)		0.25 (0.08)	0.23 (0.08)	0.20 (0.08)	0.22 (0.08)
Helpfulness Index (1-4 Scale)		0.02 (0.02)	0.01 (0.02)	0.01 (0.02)	0.00 (0.02)
Forgo Wages Index, Family (1-3 Scale)		0.11 (0.04)	0.11 (0.04)	0.13 (0.04)	0.16 (0.05)
Forgo Wages Index, Others (1-3 Scale)		0.20 (0.03)	0.21 (0.03)	0.21 (0.03)	0.18 (0.03)
% Political Activities Engaged In		0.15 (0.05)	0.16 (0.05)	0.18 (0.05)	0.18 (0.05)
Quality of Drainage (1-5)		-0.10 (0.01)	-0.10 (0.01)	-0.09 (0.01)	-0.09 (0.01)
Drainage Problem Requiring Help		-0.12 (0.03)	-0.12 (0.03)	-0.12 (0.03)	-0.09 (0.03)
Caste Level (1 = Upper, 0 = Lower)			-0.01 (0.03)	0.01 (0.03)	0.01 (0.03)
Diversity				0.25 (0.05)	0.23 (0.05)
Enumerator Religion					-0.31 (0.07)
Observations	3,832	3,684	3,577	3,555	3,549
R ²	0.00	0.10	0.10	0.11	0.11
Adjusted R ²	0.00	0.09	0.09	0.10	0.10
Residual Std. Error	1.16 (df = 3828)	1.08 (df = 3658)	1.08 (df = 3550)	1.08 (df = 3527)	1.07 (df = 3520)

Table B.6 (Continued)

Dependent Variable: Index of Favorability toward Drainage Program					
	(1)	(2)	(3)	(4)	(5)
F Statistic	3.90 (df = 3; 3828)	15.69 (df = 25; 3658)	14.90 (df = 26; 3550)	15.45 (df = 27; 3527)	15.57 (df = 28; 3520)

B.3.4 Mechanisms, Respectability

In Table B.7, we present details for the models used to address the ingroup respectability/blame avoidance mechanism for our treatment. We address this mechanism by seeing how our effect estimates for Muslims and Hindus vary based on measures of religiosity. As outlined in a previous section of this appendix, we estimate levels of religiosity using four factors. For each factor, we estimate our main model for respondents above and below the median value for that factor. As reflected in the figure in the paper and in Table 9, our results indicate that the effect we find in our main analysis is especially driven by those with high religious institutional connectedness and high feelings of obligation toward co-religionists.

Table B.7: Main Analysis, by Levels of Religiosity Factors

	Dependent Variable: Index of Favorability toward Drainage Program							
	Institutional Connectedness		Obligation to Co-Religionists		Participation in Activities		Private Participation	
	< Med.	≥ Med.	< Med.	≥ Med.	< Med.	≥ Med.	< Med.	≥ Med.
Constant	2.79 (0.03)	2.78 (0.03)	2.68 (0.03)	2.87 (0.03)	2.75 (0.03)	2.86 (0.04)	2.65 (0.03)	2.88 (0.03)
Treat	0.03 (0.04)	-0.10 (0.05)	-0.02 (0.05)	-0.03 (0.04)	-0.03 (0.04)	-0.05 (0.05)	0.06 (0.05)	-0.10 (0.04)
Muslim	-0.09 (0.06)	-0.20 (0.06)	-0.07 (0.06)	-0.21 (0.06)	-0.27 (0.06)	-0.07 (0.06)	-0.18 (0.06)	-0.07 (0.06)
Treat x Mus.	0.06 (0.08)	0.37 (0.09)	0.10 (0.09)	0.31 (0.09)	0.27 (0.09)	0.18 (0.09)	0.22 (0.09)	0.17 (0.09)
Obs.	1,838	1,807	1,776	1,869	2,229	1,416	1,565	2,080
R ²	0.002	0.01	0.001	0.01	0.01	0.003	0.01	0.004
Adj. R ²	0.001	0.01	-0.001	0.01	0.01	0.001	0.01	0.002
Res. SE	1.09 (df = 1834)	1.22 (df = 1803)	1.17 (df = 1772)	1.13 (df = 1865)	1.19 (df = 2225)	1.11 (df = 1412)	1.14 (df = 1561)	1.16 (df = 2076)
F Stat.	1.42 (df = 3; 1834)	5.39 (df = 3; 1803)	0.45 (df = 3; 1772)	5.68 (df = 3; 1865)	6.82 (df = 3; 2225)	1.37 (df = 3; 1412)	6.48 (df = 3; 1561)	2.45 (df = 3; 2076)

B.4 Robustness Checks

B.4.1 Balance

In the paper, we combine treatment groups for our analyses. The lower panel in Figure B.3 shows that, in this combined analysis, our sample is well-balanced between the control group and the group of the combined treatments. Standardized differences for the pre-treatment covariates all fall within a range of -0.75 to 0.75. An omnibus balance test (Hansen and Bowers 2008), reported in the final row of Table B.8, also indicates that we cannot reject the null of a balanced sample for the combined treatments.

We also address balance by examining the treatment groups separately rather than combined. The upper panel in Figure 3 shows the standardized differences for pre-treatment covariates between the control and each of the treatment groups. Only a few variables show minor imbalances: Gender, Number of Rooms in Residence, Religious Behavior Index, Religious ID, and Forgo Wages Index. However, the omnibus balance tests in rows 1-5 of Table 10 indicate that these minor imbalances are not unexpected for our randomization and we cannot reject the null of a balanced sample.

Table B.8: Omnibus Balance Tests

Treatment	χ^2	DF	P-value
Horizontal Accountability	32.14	31	0.41
Vertical Accountability	23.91	32	0.85
Ingroup Underperformance	26.95	31	0.67
Horizontal Accountability + Ingroup Underperformance	15.67	31	0.99
Vertical Accountability + Ingroup Underperformance	39.47	31	0.14
Treatments Combined	30.57	34	0.64

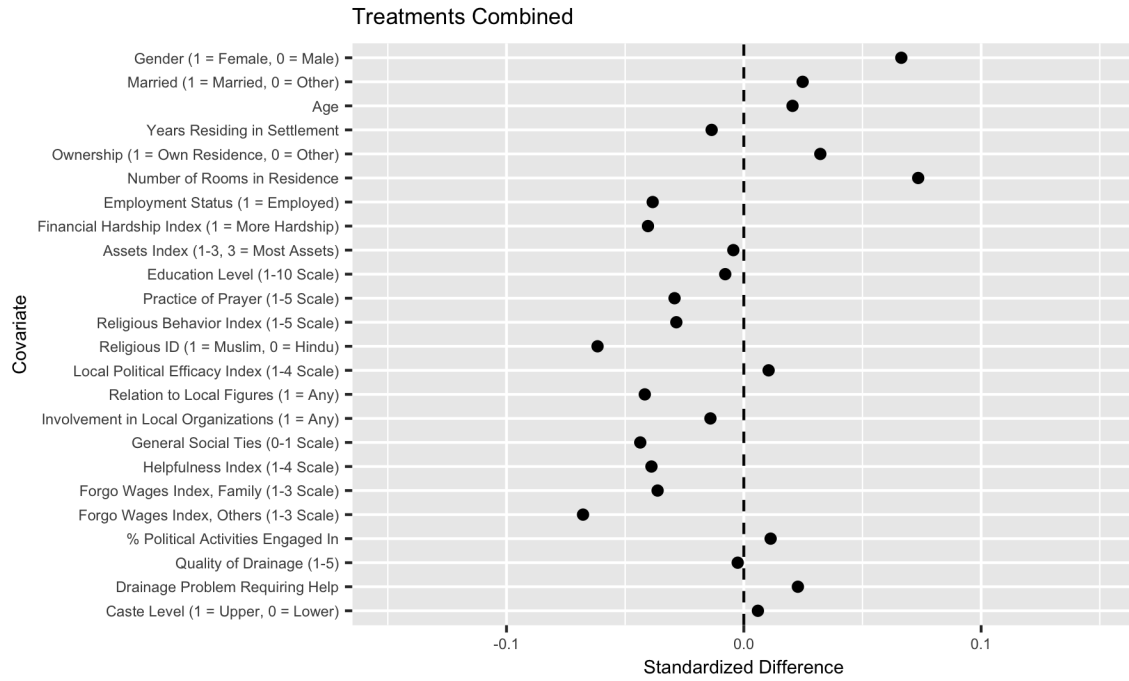
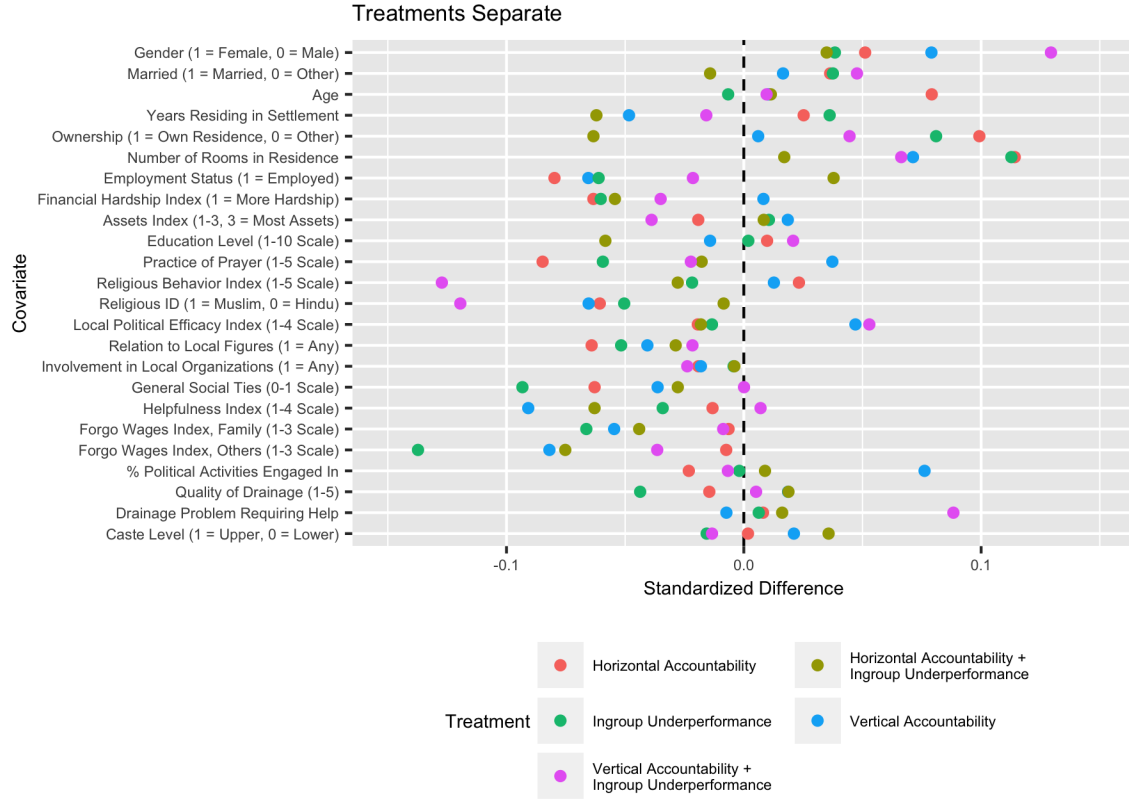


Figure B.3: Comparison of Baseline Characteristics between Control and Treatments

B.4.2 Survey Purpose and Manipulation Checks

In order to make sure that participants received and understood the information they were presented about the proposed drainage service, they were asked two questions immediately after the administration of the survey. First, participants were asked: “How many people in your neighborhood need to agree to pay the fee before the company will begin the new service?” They were read three options to choose from: (a) 1/4 or 25%, (b) 1/2 or 50%, and (c) 2/3 or 67%. These choices mirror the way these proportions were presented in the experimental manipulation itself, where participants were told that a condition for the drainage cleaning program was that “2/3 or 67% of the residents in your neighborhood indicate their agreement with this contract.” Both fractions and percentages were used because both were common ways of expressing proportions in pre-testing. A total of 3,416 respondents, or 88.9% of the sample, passed the manipulation check. Those who did not were told the correct answer after their response, to make sure that they completed the rest of the survey with the correct conditions in mind.

Second, participants were asked: “What is the monthly subscription fee required by this service?” They were read three options as answers: (a) Rs. 50, (b) Rs. 100, and (c) Rs. 150. A total of 3,651 respondents, or 95.0% of the sample, passed this manipulation check. Those who did not were told the correct answer after their response.

We present results in the paper *without* excluding those who failed the manipulation check. However, results are largely the same if we do exclude these respondents, as shown in Table B.9. If anything, the results excluding those who failed the manipulation checks are stronger, with a coefficient estimate that is 0.035 larger than the estimate in our main analysis with everyone included. This may indicate that the effects are stronger for those who absorbed the most information from the treatment administration.

Table B.9: Main Analysis, Excluding Those Who Failed Manipulation Checks

Dependent Variable: Index of Favorability toward Drainage Program	
Constant	2.84 (0.02)
Treatments (Combined)	-0.04 (0.03)
Muslim	-0.17 (0.04)
Treatments (Combined) x Muslim	0.22 (0.06)
Observations	3,387
R ²	0.005
Adjusted R ²	0.004
Residual Std. Error	1.15 (df = 3383)
F Statistic	5.22 (df = 3; 3383)

B.4.3 Timing of Survey Administration and Treatments

Because our survey was administered using tablets, we have detailed information on the amount of time it took to administer each portion of the survey. It took, on average, 13.5 minutes to administer the survey, although this average varied from a low of 9 minutes to a high of 16.5 minutes depending on the enumerator.¹

Out of the 13.5 minutes on average spent in the survey, less than half a minute was spent on the experimental manipulation. Enumerators spent just under 17 seconds on average explaining the drainage cleaning program in general, which did not vary between treatments, and 7 seconds on average explaining the portions of the program that varied between treatments. This second portion varied only slightly, by less than 0.75 seconds, between the different treatment texts (see Table B.10). This indicates that, in terms of time spent completing the experimental manipulation, there is little difference between the treatments

Table B.10: Length of Treatment Assignment

Treatment	Average Time Administering Treatment (Seconds)
Control	6.95
Horizontal Accountability	6.77
Vertical Accountability	7.32
Ingroup Underperformance	7.38
Hor. Acct. + Ingroup Underperformance	7.36
Vert. Acct. + Ingroup Underperformance	6.67

¹This calculation includes the twenty enumerators who administered approximately 93% of the surveys; it excludes the five enumerators who each administered 30 surveys or fewer, within which there was more variation (minimum of 9 minutes, maximum of 32 minutes).

B.4.4 Diversity as an Alternative Hypothesis

The interaction in our main analysis is with religious identification—whether the participant was Muslim or Hindu. However, in our pre-analysis plan, we also pre-registered an analysis examining the interaction of the treatments with religious diversity. Because of a robust literature tying diversity to poor public goods provision, we anticipated finding that the treatments would be less effective in diverse locations. We thought that participants in more diverse areas might not respond as much to our treatment mechanisms because they would not care as much about accountability to peers or to local elites.

For our measure of diversity, we use the polarization index from Montalvo and Reynal-Querol (2005). This measure, labeled as RQ below, is represented by the equation

$$RQ = 1 - \sum_{i=1}^N \left(\frac{\frac{1}{2} - \pi_i}{\frac{1}{2}} \right)^2 \pi_i,$$

where π_i represents the proportion of the population that belongs to group i . This measure captures “how far the distribution of the . . . groups is from the $(1/2, 0, 0, \dots, 0, 1/2)$ distribution (bipolar), which represents the highest level of polarization” (pp. 798). However, because our context only involves two groups, this measure simplifies to $RQ = 4\pi_1\pi_2$, where π_1 and π_2 are the group proportions, and it is almost identical to the ethnic fractionalization index for two groups ($2\pi_1\pi_2$) that is also often used to measure diversity. Using the geocoordinates collected as a part of the survey, we calculate the polarization index for each participant based on the religious identification of other participants within 100 meters of them. This provides us with a near-continuous measure of diversity at the local level.

Using this measure, we find that there is no interaction with religious diversity. Results for our analysis with a diversity interaction effect can be found in Table B.11 and Figure B.4. The coefficient on the interaction between the combined treatments and the diversity index is 0.1 ($SE = 0.09$), statistically insignificant and in the opposite direction of our hypothesis. Because diversity is a continuous variable, we also explore the interaction effect graphically in Figure B.4. This figure shows the estimate for the linear interaction effect from Table B.11, an estimate of interaction effects based on binning observations into tertiles from the diversity index, and a histogram at the bottom of the plot showing the distribution of the diversity measure (pink colored bars indicate observations in the combined treatment groups; gray bars indicate those in the control group). This figure confirms what we find the regression model, which is that there is no significant interaction effect.

As noted in the paper, we think the most likely explanation for this finding is the correlation between being a minority (Muslim) and living in a “diverse” location, which have a medium-sized correlation in our sample ($r = 0.36$). By this explanation, our finding that diversity has little interaction effect, and that what exists is in the opposite direction of our prior hypotheses and what the literature would expect, is explained by individuals’ religious identification.

Table B.11: Model with Diversity Interaction

Dependent Variable: Index of Favorability toward Drainage Program	
Constant	2.71 (0.03)
Treatments (Combined)	-0.02 (0.04)
Diversity (100m)	0.04 (0.06)
Treatments (Combined) x Diversity (100m)	0.10 (0.09)
Observations	3,810
R ²	0.002
Adjusted R ²	0.001
Residual Std. Error	1.17 (df = 3806)
F Statistic	2.19 (df = 3; 3806)

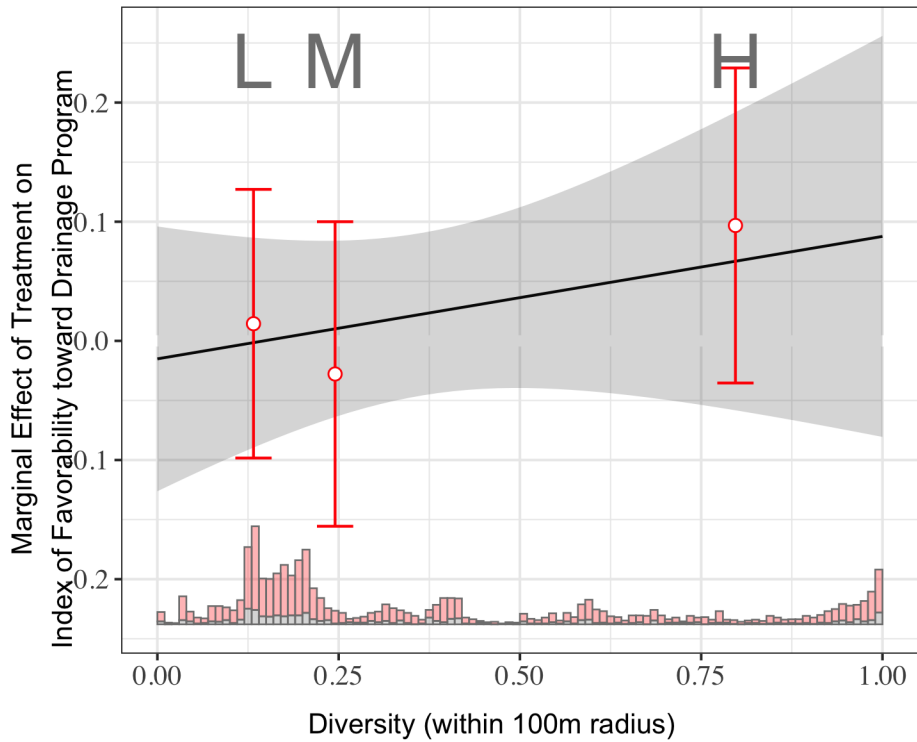


Figure B.4: Graphical Representation of Interaction with Diversity

B.4.5 Effects by Neighborhood

We explore the effect estimates of the main model and how they vary between different types of neighborhoods. This is an alternative way of exploring the interaction effect of diversity; we can see how the estimates change depending on the religious composition of each neighborhood. By design, the neighborhoods included in this study vary widely in terms of their religious composition; as noted in the section of this appendix on site selection, we used data from electoral rolls to select three types of locations: majority Hindu, mixed, and majority Muslim. We were successful in obtaining appropriate variation on this variable by neighborhood. As a reminder, we surveyed participants in sixteen neighborhoods, and these neighborhoods themselves can be divided into five areas: Kirti Nagar, Daya Basti, Adarshnagar, Jawahar Mohalla, and Seemapuri. By the estimates in our sample, the six neighborhoods in the Kirti Nagar area, as well as the neighborhood of Daya Basti, all contained a small percentage of Muslim residents (below 10%). Five of the neighborhoods—Adarshnagar, three of the neighborhoods in the Jawahar Mohalla area, and one of the Seemapuri neighborhoods—we classify as mixed because they contain more than 10% but less than 50% Muslim residents. Lastly, four neighborhoods—one from Jawahar Mohalla and three from Seemapuri—have at least 50% Muslim residents in our sample, with the three Seemapuri neighborhoods having especially high percentages of Muslim residents (ranging between 84% and 97% Muslim).

When we estimate effects separately in these three types of locations, we do find that effects for both Muslims and Hindus vary in interesting ways. Table B.12 shows the estimates from each of these models, the left column containing results for the large Hindu majority locations, the middle column containing results for the mixed locations, and the right column containing results for the majority Muslim locations. Three aspects of these estimates deserve mentioning. First, the main model's finding that Muslims respond to the treatments while Hindus do not seem to be primarily by residents in neighborhoods with either strong Hindu majorities (<10% Muslim) or Muslim majorities (>50% Muslim). The interaction effect for a respondent being Muslim in these neighborhoods is higher than the main model; at 0.26 and 0.25, respectively, the interaction estimate is approximately 1.5 times the estimate in the main model, although the standard error for these estimates is also larger. In comparison, the interaction estimate for respondents in the mixed neighborhoods (>10% but <50% Muslim) is effectively zero. Second, we also note that the main model's estimate of a lower baseline for Muslim participants in the control condition (the lower-level coefficient on Muslim in the main model is estimated at -0.13,) seems to be primarily driven by respondents in the Muslim majority neighborhoods. In these neighborhoods, Hindu participants have a very high baseline of interest (2.95, $SE = 0.09$, compared to 2.76, $SE = 0.02$, in the main model) in the drainage cleaning program. Third, we lastly note that Hindus in mixed neighborhoods do have a marginally significant effect estimate for the combined treatments (0.12, $SE = 0.07$, compared to an effect of -0.02, $SE = 0.03$ in the main model).

It is difficult to draw strong conclusions from these findings; we did not hypothesize about effects varying by these neighborhood groups prior to carrying out this project, and the differences we do find here are all of marginal

Table B.12: Main Model by Percent Muslim in Neighborhoods

	Dependent Variable: Index of Favorability toward Drainage Program		
	Low (<10%)	Medium (>10%, <50%)	High (>50%)
Constant	2.77 (0.02)	2.62 (0.05)	2.95 (0.09)
Treatments (Combined)	-0.04 (0.03)	0.12 (0.07)	-0.11 (0.13)
Muslim	-0.13 (0.10)	-0.04 (0.09)	-0.30 (0.10)
Treatments (Combined) x Muslim	0.26 (0.14)	0.08 (0.13)	0.25 (0.15)
# Hindu Respondents	2157	579	188
# Muslim Respondents	114	211	594
Observations	2,262	789	781
R ²	0.002	0.01	0.01
Adjusted R ²	0.0004	0.004	0.01
Residual Std. Error	1.09 (df = 2258)	1.20 (df = 785)	1.33 (df = 777)
F Statistic	1.30 (df = 3; 2258)	2.00 (df = 3; 785)	3.31 (df = 3; 777)

significance compared to the findings in the main model of the paper. However, if we are to draw conclusions from these varying effects by neighborhood, it would be that Muslims are least likely to be affected by our treatments in non-polarized neighborhoods.

B.4.6 Ceiling/Floor Effects (Tobit Model)

Our main outcome variable in this study is an index composed of five questions about participant interest in the drainage cleaning program. These questions are designed to elicit a range of responses on the 1–4 scale used to measure the outcomes. For the first two questions—would this program be beneficial to your neighborhood and how interested would you be in the program—respondents were much more likely to register high values, with averages of 3.02 and 2.95, respectively, for those in the control condition. In response to the third question—how likely would you be to pay the monthly fee—respondents reported slightly lower values, with an average of 2.74 for control condition respondents. Lastly, in response to the fourth and fifth questions—would you be willing to sign a six-month contract for this service and how likely would you be to try to get your neighbors to sign up for the program—respondents in control reported the lowest average responses of 2.46 and 2.51, respectively.

Table B.13: Tobit Model

Dependent Variable: Index of Favorability toward Drainage Program	
Constant	2.77 (0.02)
Treatments (Combined)	−0.02 (0.03)
Muslim	−0.13 (0.04)
Treatments (Combined) x Muslim	0.21 (0.05)
Observations	3,832
Log Likelihood	−10,601.50
Wald Test	20.98 (df = 3)

However, average responses were higher than the midpoint of the scale for all our measures, which may make some worried about ceiling effects for Hindus, who saw no effect of the treatments but who also had a higher baseline than Muslims for each of the index items. Additionally, these questions evaluate attitudes on a somewhat polarizing topic for communities, and we might be worried about a floor for respondents' attitudes as well because of strong opinions from those who oppose the program. Indeed, when we look at a histogram of the index in Figure B.5, we see that that respondents pool at the lower and upper ends of our scale, with 7.4% of respondents recording a 1 for all five outcome questions and 9.3% of respondents recording a 4 for all of them. To make sure that the results we find in our main model are not driven by this pooling in any way, we estimate a tobit model with a floor of 1 and a ceiling of 4, the results of which can be found in Table B.13. The estimates from this model are almost the same as those in the main analysis; if anything, the coefficients from the tobit model would indicate a slightly higher interaction effect for Muslims (0.21 instead of 0.18).

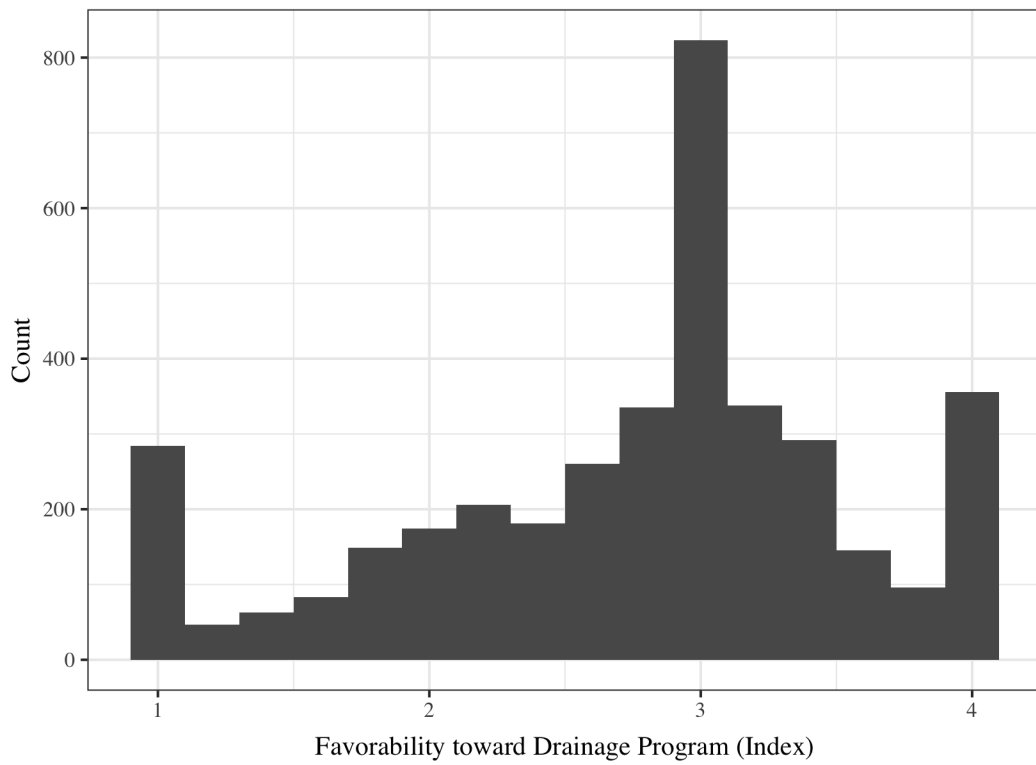


Figure B.5: Histogram of Outcome Index, All Respondents

B.4.7 Enumerator Effects

Because our survey hinges on religious identification, we may be concerned about how the religion of the survey enumerator affects our results. In Table B.14 we calculate our main analysis separately for those who had a Hindu enumerator and those who had a Muslim one. Before discussing these results, we want to note that this analysis should be viewed as highly speculative. We did not pre-register an analysis of this sort. Additionally, out of the twenty-five total enumerators who administered our survey, only two of them were Muslim. There are only 141 respondents who took the survey from a Muslim enumerator. Lastly, enumerators were not randomly assigned to settlements, so our Muslim enumerator results come primarily from just six of our sixteen locations.

With these caveats in mind, the analysis in Table B.14 indicates that Muslim enumerators may elicit very different effects from participants in a number of ways, especially in terms of the “baseline” support for the program in the control group. For instance, the baseline in the control group for Hindu respondents is much lower when they have a Muslim enumerator (2.15, $SE = 0.11$) than when they have a Hindu enumerator (2.78, $SE = 0.02$). On the other hand, the opposite is true for Muslim respondents; the baseline for Muslims is much higher with a Muslim enumerator (3.17, $SE = 0.16$) than it is with a Hindu enumerator (2.61, $SE = 0.04$). Lastly, from this higher baseline, the treatment effect for Muslim respondents with Muslim enumerators is now estimated to be negative (-0.54, $SE = 0.31$), as opposed to the positive effect with Hindu enumerators and in the main analysis.

As already noted, we do not think we can draw much from these models. The estimates in the right column of Table B.14 are entirely dependent on two Muslim enumerators who only administered the survey to 141 respondents who are not representative of our sample. However, examining these results is useful for considering the scope and generalizability of this research project. Because the vast majority of our enumerators are Hindu, our findings are primarily applicable to situations where programs like ours would be instituted by members of the majority religious/ethnic group.

Table B.14: Main Analysis, by Religion of Enumerator

	Dependent Variable: Index of Favorability toward Drainage Program	
	Hindu Enumerators	Muslim Enumerators
Constant	2.78 (0.02)	2.15 (0.11)
Treatments (Combined)	-0.01 (0.03)	-0.06 (0.15)
Muslim	-0.17 (0.04)	1.02 (0.20)
Treatments (Combined) x Muslim	0.20 (0.06)	-0.54 (0.31)
Observations	3,685	141
R ²	0.01	0.21
Adjusted R ²	0.004	0.19
Residual Std. Error	1.16 (df = 3681)	1.06 (df = 137)
F Statistic	6.30 (df = 3; 3681)	12.13 (df = 3; 137)

Bibliography

- Adida, Claire L., Adeline Lo, and Melina R. Platas. 2018. "Perspective taking can promote short-term inclusionary behavior toward Syrian refugees." *Proceedings of the National Academy of Sciences* 115 (38): 9521–9526.
- Alesina, Alberto, and Eliana La Ferrara. 1999. "Participation in Heterogeneous Communities." *NBER Working Paper No. 7155*.
- Allport, Gordon W. 1954. *The nature of prejudice*. Cambridge, MA: Addison-Wesley Pub. Co.
- Asher, Sam, Paul Novosad, and Charlie Rafkin. 2017. Estimating intergenerational mobility with coarse data: A nonparametric approach. Technical report World Bank.
- Auerbach, Adam Michael. 2016. "Clients and Communities: The Political Economy of Party Network Organization and Development in India's Urban Slums." *World Politics* 68 (January): 111–148.
- Auerbach, Adam Michael. 2017. "Neighborhood Associations and the Urban Poor: India's Slum Development Committees." *World Development* 96 (April): 119–135.
- Auerbach, Adam Michael, and Tariq Thachil. 2018. "How Clients Select Brokers: Competition and Choice in India's Slums." *American Political Science Review* 112 (November): 775–791.
- Bakke, Kristin M., Kathleen Gallagher Cunningham, and Lee J. M. Seymour. 2012. "A plague of initials: Fragmentation, cohesion, and infighting in civil wars." *Perspectives on Politics* 10 (June): 265–283.
- Baldwin, Kate. 2015. *The Paradox of Traditional Leaders in Democratic Africa*. New York: Cambridge University Press.
- Baldwin, Kate, and John Huber. 2010. "Economic versus Cultural Differences: Forms of Ethnic Diversity and Public Goods Provision." *The American Political Science Review* 104 (4): 644–662.
- Banerjee, Abhijit, Arun G Chandrasekhar, Esther Duflo, and Matthew O Jackson. 2019. "Using gossips to spread information: Theory and evidence from two randomized control trials." *The Review of Economic Studies* 86 (6): 2453–2490.
- Banerjee, Abhijit, Lakshmi Iyer, and Rohini Somanathan. 2005. "History, Social Divisions, and Public Goods in Rural India." *Journal of the European Economic Association* 3 (2/3): 639–647.
- Banerjee, Abhijit, Marianne Bertrand, Saugato Datta, and Sendhil Mullainathan. 2009. "Labor Market Discrimination in Delhi: Evidence from a Field Experiment." *Journal of Comparative Economics* 37 (1): 14–27.
- Banerjee, Abhijit, Rohini Pande, and Michael Walton. 2012. Delhi's Slum-Dwellers: Deprivation, Preferences and Political Engagement among the Urban Poor. Working Paper S-5003-INC-1 IGC International Growth Centre London: .
- Bar-Tal, Daniel. 1998. "Societal beliefs in times of intractable conflict: The Israeli case." *International Journal of Conflict Management* 9 (January): 22–50.
- Bar-Tal, Daniel. 2013. *Intractable conflicts: Socio-psychological foundations and dynamics*. Cambridge: Cambridge University Press.

- Bardhan, Pranab, and Dilip Mookherjee. 2012. Political Clientelism and Capture Theory and Evidence from West Bengal, India. Working Paper 97 UNU-WIDER.
- Belkâab, Nejoua, and Melani Cammett. 2014. Mission Report: Governance and the Quality of Primary Health Care in Morocco. Technical report World Bank Washington, DC: .
- Bernhard, Helen, Ernst Fehr, and Urs Fischbacher. 2006. "Group Affiliation and Altruistic Norm Enforcement." *The American Economic Review* 96 (2): 217–221.
- Bertrand, Marianne, and Sendhil Mullainathan. 2004. "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *American Economic Review* 94 (September): 991–1013.
- Bhalotra, Sonia, Irma Clots-Figueras, Lakshmi Iyer, and Joseph Vecci. 2018. Leader Identity and Coordination. Working Paper S-89211-INC-1 IGC International Centre London: .
- Bharathi, Naveen, Deepak Malghan, and Andaleeb Rahman. 2015. "Neighbourhood-Scale Residential Segregation in Indian Metros." *Economic and Political Weekly* 54 (June): 7–8.
- Bilali, Rezarta, Johanna Ray Vollhardt, and Jason Ray David Rarick. 2015. "Assessing the impact of a media-based intervention to prevent intergroup violence and promote positive intergroup relations in Burundi." *Journal of Community & Applied Social Psychology* 26 (3): 221–235.
- Bisin, Alberto, and Thierry Verdier. 2001. "The Economics of Cultural Transmission and the Dynamics of Preferences." *Journal of Economic Theory* 97 (April): 298–319.
- Björkman, Martina, and Jakob Svensson. 2009. "Power to the People: Evidence from a Randomized Field Experiment of a Community-Based Monitoring Project in Uganda." *Quarterly Journal of Economics* 124 (2): 735–769.
- Brewer, Marilyn B. 1999. "The psychology of prejudice: Ingroup love or outgroup hate?" *Journal of Social Issues* 55 (3): 429–444.
- Brewer, Paul R. 2002. "Value words and lizard brains: Do citizens deliberate about appeals to their core values?" *Political Psychology* 22 (1): 45–64.
- Brewer, Paul R., and Kimberly Gross. 2005. "Values, framing, and citizens' thoughts about policy issues: Effects on content and quality." *Political Psychology* 26 (6): 929–948.
- Brixi, Hana, Ellen Lust, and Michael Woolcock. 2015. Trust, Voice and Incentives: Learning from Local Success Stories in Service Delivery in the Middle East and North Africa. Technical report World Bank Washington, DC: .
- Brubaker, Rogers, and David D. Laitin. 1998. "Ethnic and nationalist violence." *Annual Review of Sociology* 24: 423–452.
- Cammett, Melani, Poulomi Chakrabarti, and Romney David. 2018. Pre-Analysis Plan: Accountability in Intergroup Relations and Public Goods Provision in Indian Slums. EGAP Registration ID: 20180613AA EGAP.
- Centre for Regional Political Economy. 2019. Politics and society between elections. Technical report Azim Premji University.
- Chandra, Kanchan. 2006. "What is ethnic identity and does it matter?" *Annual Review of Political Science* 9: 397–424.
- Chandra, Kanchan, and Steven Wilkinson. 2008. "Measuring the Effect of 'Ethnicity'." *Comparative Political Studies* 41 (4-5): 515–563.
- Charnysh, Volha, Christopher Lucas, and Prerna Singh. 2015. "The ties that bind: National identity salience and pro-social behavior toward the ethnic other." *Comparative Political Studies* 48 (March): 267–300.
- Chen, Yen-Ting, and Melani Cammett. 2012. "Informal Politics and Inequity of Access to Health Care in Lebanon." *International Journal for Equity in Health* 11 (23): 1–8.

- Chidambaram, Soundarya. 2020. "How Do Institutions and Infrastructure Affect Mobilization around Public Toilets vs. Piped Water? Examining Intra-Slum Patterns of Collective Action in Delhi, India." *World Development* 132 (August): 1–14.
- Chong, Dennis, and James N. Druckman. 2007. "Framing theory." *Annual Review of Political Science* 10: 103–126.
- Cikara, Mina, and Elizabeth Levy Paluck. 2013. "When going along gets you nowhere and the upside of conflict behaviors." *Social and Personality Psychology Compass* 7 (August): 559–571.
- Cockburn, Patrick. 2020. "While Muslims are being murdered in India, the rest of the world is too slow to condemn." *The Independent* (February).
- Coffey, Diane, Aashish Gupta, Payal Hathi, Dean Spears, Nikhil Srivastav, and Sangita Vyas. 2014. "Revealed Preference for Open Defecation: Evidence from a New Survey in Rural North India." *Economic and Political Weekly* 49 (September): 43–55.
- Collier, Paul, and Pedro C. Vicente. 2013. "Votes and violence: Evidence from a field experiment in Nigeria." *The Economic Journal* 124 (574): F327–F355.
- Corvalan, Alejandro, and Miguel Vargas. 2015. "Segregation and conflict: An empirical analysis." *Journal of Development Economics* 116 (C): 212–222.
- Costa, Dora L., and Matthew E. Kahn. 2003. "Civic engagement and community heterogeneity: An economist's perspective." *Perspectives on Politics* 1 (1): 103–111.
- Cunningham, Kathleen Gallagher. 2011. "Divide and conquer or divide and concede: How do states respond to internally divided separatists?" *American Political Science Review* 105 (May): 275–297.
- de Mesquita, Ethan Bueno. 2005. "Conciliation, counterterrorism, and patterns of terrorist violence." *International Organization* 59 (1): 145–76.
- Deutsch, Karl Wolfgang. 1953. *Nationalism and Social Communication: An Inquiry Into the Foundations of Nationality*. Published jointly by the Technology Press of the Massachusetts Institute of Technology, and Wiley, New York.
- Ditlmann, Ruth K., and Cyrus Samii. 2016. "Can intergroup contact affect ingroup dynamics? Insights from a field study with Jewish and Arab-Palestinian youth in Israel." *Peace and Conflict: Journal of Peace Psychology* 22 (4): 380–392.
- Ditlmann, Ruth K., Cyrus Samii, and Thomas Zeitzoff. 2017a. "Addressing violent intergroup conflict from the bottom up?" *Social issues and policy review* 11 (1): 38–77.
- Ditlmann, Ruth K., Cyrus Samii, and Thomas Zeitzoff. 2017b. "Addressing Violent Intergroup Conflict from the Bottom Up?" *Social Issues and Policy Review* 11 (1): 38–77.
- Druckman, James N., and Kjersten R. Nelson. 2003. "Framing and deliberation: How citizens' conversations limit elite influence." *American Journal of Political Science* 47 (4): 729–745.
- Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2015. "School Governance, Teacher Incentives, and Pupil–Teacher Ratios: Experimental Evidence from Kenyan Primary Schools." *Journal of Public Economics* 123: 92–110.
- Easterly, William, and Ross Levine. 1997. "Africa's Growth Tragedy: Policies and Ethnic Divisions." *The Quarterly Journal of Economics* 112 (4): 1203–1250.
- Enos, Ryan D. 2014. "Causal effect of intergroup contact on exclusionary attitudes." *Proceedings of the National Academy of Sciences* 111 (March): 3699–3704.

- Enos, Ryan D. 2016. "What the Demolition of Public Housing Teaches Us about the Impact of Racial Threat on Political Behavior." *American Journal of Political Science* 60 (1): 123–142.
- Enos, Ryan D. 2017. *The space between us*. Cambridge: Cambridge University Press.
- Enos, Ryan D., Aaron R. Kaufman, and Melissa L. Sands. 2019. "Can violent protest change local policy support? Evidence from the aftermath of the 1992 Los Angeles riot." *American Political Science Review* 113 (4): 1012–1028.
- Enos, Ryan D., and Noam Gidron. 2016. "Intergroup behavioral strategies as contextually determined: Experimental evidence from Israel." *The Journal of Politics* 78 (May): 851–867.
- Entman, Robert M. 1993. "Framing: Toward clarification of a fractured paradigm." *Journal of Communication* 43 (4): 51–58.
- Fearon, James D., and David D. Laitin. 1996a. "Explaining interethnic cooperation." *American Political Science Review* 90 (4): 715–735.
- Fearon, James D., and David D. Laitin. 2000. "Violence and the social construction of ethnic identity." *International Organization* 54 (October): 845–877.
- Fearon, James, and David Laitin. 1996b. "Explaining Interethnic Cooperation." *American Political Science Review* 90 (4): 715–735.
- Feldman, Stanley. 1988. "Structure and consistency in public opinion: The role of core beliefs and values." *American Journal of Political Science* 32 (2): 416–440.
- Filkins, Dexter. 2019. "Blood and Soil in Narendra Modi's India." *The New Yorker* (December).
- Fix, Michael. 1993. *Clear and Convincing Evidence: Measurement of Discrimination in America*. Washington, DC: Urban Institute Press.
- Fung, Constance H., Yee-Wei Lim, Soeren Mattke, Cheryl Damberg, and Paul G. Shekelle. 2008. "Systematic Review: The Evidence That Publishing Patient Care Performance Data Improves Quality of Care." *Annals of Internal Medicine* 148 (2): 111–123.
- Gayer, Laurent. 2012. "Safe and Sound: Searching for a "Good Environment" in Abul Fazl Enclave, Delhi." In *Muslims in Indian Cities: Trajectories of Marginalisation*, ed. Laurent Gayer and Christophe Jaffrelot. New York, NY: Columbia University Press.
- Gayer, Laurent and Christophe Jaffrelot, eds. 2012. *Muslims in Indian Cities: Trajectories of Marginalisation*. New York, NY: Columbia University Press.
- Gerber, Alan S., Donald P. Green, and Christopher W. Larimer. 2008. "Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment." *American Political Science Review* 102 (1): 33–48.
- Geruso, Michael, and Dean Spears. 2018. "Neighborhood Sanitation and Infant Mortality." *American Economic Journal: Applied Economics* 10 (April): 125–162.
- Gettleman, Jeffrey, and Maria Abi-Habib. 2020. "In India, Modi's Policies Have Lit a Fuse." *The New York Times* (March).
- Ghertner, D. Asher. 2008. "Analysis of New Legal Discourse behind Delhi's Slum Demolitions." *Economic and Political Weekly* 43 (May): 57–66.
- Gino, Francesca, and Adam D. Galinsky. 2012. "Vicarious Dishonesty: When Psychological Closeness Creates Distance from One's Moral Compass." *Organizational Behavior and Human Decision Processes* 119 (1): 15–26.
- Gino, Francesca, Jun Gu, and Chen-Bo Zhong. 2009. "Contagion or Restitution? When Bad Apples Can Motivate Ethical Behavior." *Journal of Experimental Social Psychology* 45 (6): 1299–1302.

- Glaeser, Edward L. 2005. "The political economy of hatred." *Quarterly Journal of Economics* 120 (1): 45–86.
- GNCTD. 2006. Delhi Human Development Report 2006. Technical report Government of National Capital Territory of Delhi/Oxford University Press New Delhi: .
- Government of India. 2006. Social, Economic and Educational Status of the Muslim Community of India (Sachar Committee Report). Technical report Government of India New Delhi: .
- Granovetter, Mark S. 1973. "The Strength of Weak Ties." *American Journal of Sociology* 78 (6): 1360–1380.
- Green, Donald P., Bradley Palmquist, and Eric Schickler. 2004. *Partisan hearts and minds: Political parties and the social identities of voters*. New Haven, CT: Yale University Press.
- Green, Donald P., and Janelle S. Wong. 2009. "Tolerance and the contact hypothesis: A field experiment." In *The political psychology of democratic citizenship*, ed. Eugene Borgida, Christopher M. Federico, and John L. Sullivan. Oxford University Press.
- Gubler, Joshua Ronald, Eran Halperin, and Gilad Hirschberger. 2015. "Humanizing the Outgroup in Contexts of Protracted Intergroup Conflict." *Journal of Experimental Political Science* 2 (1): 36–46.
- Gupta, Gautam, Minhaj Mahmud, Pushkar Maitra, Santanu Mitra, and Ananta Neelim. 2018. "Religion, Minority Status, and Trust: Evidence from a Field Experiment." *Journal of Economic Behavior & Organization* 146: 180–205.
- Habermas, Jürgen. 1979. *Communication and the evolution of society*. Boston: Beacon Press.
- Habyarimana, James, Macartan Humphreys, Daniel N. Posner, and Jeremy M. Weinstein. 2007. "Why does ethnic diversity undermine public goods provision?" *American Political Science Review* 101 (November): 709–725.
- Habyarimana, James, Macartan Humphreys, Daniel N. Posner, and Jeremy M. Weinstein. 2009. *Coethnicity: Diversity and the Dilemmas of Collective Action*. New York: Russell Sage Foundation.
- Haley, Kevin J., and Daniel M.T. Fessler. 2005. "Nobody's Watching? Subtle Cues Affect Generosity in an Anonymous Economic Game." *Evolution and Human Behavior* 26 (3): 245–256.
- Hansen, Ben B., and Jake Bowers. 2008. "Covariate balance in simple, stratified, and clustered comparative studies." *Statistical Science* 23 (2): 219–236.
- Heller, Patrick, Partha Mukhopadhyay, Subhadra Banda, and Shahana Sheikh. 2015. "Exclusion, Informality, and Predation in the Cities of Delhi An Overview of the Cities of Delhi Project." Working paper.
- Hibbard, Judith H., Jean Stockard, and Martin Tusler. 2005. "Hospital Performance Reports: Impact on Quality, Market Share, and Reputation." *Health Affairs* 24 (4): 1150–1160.
- Higginbotham, Evelyn Brooks. 1993. *Righteous Discontent: The Women's Movement in the Black Baptist Church, 1880–1920*. Cambridge, MA: Harvard University Press.
- Hornsey, Matthew J., and Armin Imani. 2004. "Criticizing groups from the inside and the outside: An identity perspective on the intergroup sensitivity effect." *Personality and Social Psychology Bulletin* 30 (March): 365–383.
- Hornsey, Matthew J., Petra de Bruijn, Jennifer Creed, James Allen, Amarina Ariyanto, and Alicia Svensson. 2005. "Keeping it in-house: How audience affects responses to group criticism." *European Journal of Social Psychology* 35 (May): 291–312.
- Horowitz, Donald. 1970. *Ethnic Groups in Conflict*. University of California Press.
- Jefferson, Hakeem. 2018. "Policing norms: Punishment and the politics and respectability among Black Americans." Working paper.

- Jha, Saumitra, Vijayendra Rao, and Michael Woolcock. 2007. "Governance in the Gullies: Democratic Responsiveness and Leadership in Delhi's Slums." *World Development* 35 (February): 230–246.
- Kalyvas, Stathis N. 2008. "Ethnic defection in civil war." *Comparative Political Studies* 41 (August): 1043–1068.
- Key, V.O. 1949. *Southern Politics in State and Nation*. New York: A.A. Knopf.
- Khemani, Stuti. 2015. "Buying votes versus supplying public services: Political incentives to under-invest in pro-poor policies." *Journal of Development Economics* 117 (C): 84–93.
- Khwaja, Asim Ijaz. 2009. "Can Good Projects Succeed in Bad Communities?" *Journal of Public Economics* 93 (August): 899–916.
- Kimmerling, Baruch. 2001. *The invention and decline of Israeliness: State, culture and military in Israel*. Berkeley: University of California Press.
- Kosack, Stephen, and Archon Fung. 2014. "Does Transparency Improve Governance?" *Annual Review of Political Science* 17: 65–87.
- Kydd, Andrew, and Barbara F. Walter. 2002. "Sabotaging the peace: The politics of extremist violence." *International Organization* 56 (2): 263–296.
- Laitin, David D. 1995. "Marginality: A microperspective." *Rationality and Society* 7 (1): 31–57.
- Lazarev, Egor, and Vera Mironova. 2018. "The Economic Consequences of Political Alienation: Ethnic Minority Status and Investment Behavior in a Post-Conflict Society." *World Development* 106: 27–39.
- Lee, Alexander. 2018. "Ethnic Diversity and Ethnic Discrimination: Explaining Local Public Goods Provision." *Comparative Political Studies* 51 (10): 1351–1383.
- Leidner, Bernhard, Peter Kardos, and Emanuele Castano. 2018. "The effects of moral and pragmatic arguments against torture and demands for judicial reform." *Political Psychology* 39 (1): 143–162.
- Lerner, Jennifer S., and Philip E. Tetlock. 2002. "Bridging Individual, Interpersonal, and Institutional Approaches to Judgment and Decision Making: The Impact of Accountability on Cognitive Bias." In *Emerging Perspectives on Judgment and Decision Research*, ed. Sandra L. Schneider and James Shanteau. Cambridge: Cambridge University Press.
- Li, Sherry Xin. 2010. "Social Identities, Ethnic Diversity, and Tax Morale." *Public Finance Review* 38 (2): 146–177.
- Lieberman, Evan S., and Gwyneth H. McClendon. 2013. "The Ethnicity–Policy Preference Link in Sub-Saharan Africa." *Comparative Political Studies* 46 (May): 574–602.
- Lust, Ellen, and Lise Rakner. 2018. "The Other Side of Taxation: Extraction and Social Institutions in the Developing World." *Annual Review of Political Science* 21 (1): 277–294.
- Lyall, Jason. 2010. "Are coethnics more effective counterinsurgents? Evidence from the Second Chechen War." *American Political Science Review* 104 (February): 1–20.
- Mac Ginty, Roger. 2014. "Everyday Peace: Bottom-up and Local Agency in Conflict-Affected Societies." *Security Dialogue* 45 (6): 548–564.
- Maoz, Ifat. 2004. "Coexistence is in the eye of the beholder: Evaluating intergroup encounter interventions between Jews and Arabs in Israel." *Journal of Social Issues* 60 (June): 437–452.
- Maoz, Ifat. 2011. "Does contact work in protracted asymmetrical conflict? Appraising 20 years of reconciliation-aimed encounters between Israeli Jews and Palestinians." *Journal of Peace Research* 48 (January): 115–125.
- Maoz, Ifat, and Clark McCauley. 2008. "Dehumanization and support for retaliatory aggressive policies in asymmetric conflict." *Journal of Conflict Resolution* 52 (1): 93–116.

- Mazumder, Soumyajit. 2018. "The persistent effect of U.S. civil rights protests on political attitudes." *American Journal of Political Science* 62 (4): 922–935.
- McCarthy, John, and Mayer Zald. 1977. "Resource mobilization and social movements: A partial theory." *American Journal of Sociology* 82: 1212–1241.
- McClendon, Gwyneth H. 2014. "Social Esteem and Participation in Contentious Politics: A Field Experiment at an LGBT Pride Rally." *American Journal of Political Science* 58 (2): 279–290.
- Miguel, Edward, and Mary Kay Gugerty. 2005. "Ethnic Diversity, Social Sanctions, and Public Goods in Kenya." *Journal of Public Economics* 89 (December): 2325–2368.
- Mines, Mattison. 1972. "Muslim Social Stratification in India: The Basis for Variation." *Southwestern Journal of Anthropology* 28 (December): 333–349.
- Ministry of Urban Development. 2012. Service levels in urban water and sanitation sector - a status report (2010–2011). Technical report Government of India New Delhi: .
- Minson, Julia A., and Benoît Monin. 2012. "Do-gooder derogation: Disparaging morally motivated minorities to defuse anticipated reproach." *Social Psychological and Personality Science* 3 (March): 200–207.
- Monin, Benoît, Pamela J. Sawyer, and Matthew J. Marquez. 2008. "The rejection of moral rebels: Resenting those who do the right thing." *Journal of Personality and Social Psychology* 95 (July): 76–93.
- Mousa, Salma. 2020. "Building social cohesion between Christians and Muslims through soccer in post-ISIS Iraq." *Science* 369 (6505): 866–870.
- Nunnally, Shayla C. 2012. *Trust in Black America: Race, Discrimination, and Politics*. NYU Press.
- O'Connor, Kieran, and Benoît Monin. 2016. "When principled deviance becomes moral threat: Testing alternative mechanisms for the rejection of moral rebels." *Group Processes & Intergroup Relations* 19 (September): 676–693.
- Olken, Benjamin A., Junko Onishi, and Susan Wong. 2014. "Should aid reward performance? Evidence from a field experiment on health and education in Indonesia." *American Economic Journal: Applied Economics* 6 (4): 1–34.
- Omi, Michael, and Howard Winant. 2014. *Racial Formation in the United States*. New York, NY: Routledge.
- Ostrom, Elinor. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge; New York: Cambridge University Press.
- Packer, Dominic J. 2014. "On not airing our dirty laundry: Intergroup contexts suppress ingroup criticism among strongly identified group members." *British Journal of Social Psychology* 53 (March): 93–111.
- Packer, Dominic J., and Alison L. Chasteen. 2010. "Loyal deviance: Testing the normative conflict model of dissent in social groups." *Personality and Social Psychology Bulletin* 36 (January): 5–18.
- Packer, Dominic J., and Christopher T. H. Miners. 2014. "Tough love: The normative conflict model and a goal system approach to dissent decisions." *Social and Personality Psychology Compass* 8 (July): 354–373.
- Paluck, Elizabeth Levy. 2010. "Is it better not to talk? Group polarization, extended contact, and perspective taking in Eastern Democratic Republic of Congo." *Personality and Social Psychology Bulletin* 36 (9): 1170–1185.
- Paluck, Elizabeth Levy, and Donald P. Green. 2009. "Deference, dissent, and dispute resolution: An experimental intervention using mass media to change norms and behavior in Rwanda." *American Political Science Review* 103 (November): 622–644.
- Paluck, Elizabeth Levy, Hana Shepherd, and Peter M. Aronow. 2016. "Changing climates of conflict: A social network experiment in 56 schools." *Proceedings of the National Academy of Sciences* 113 (3): 566–571.

- Paluck, Elizabeth Levy, Seth A. Green, and Donald P. Green. 2019. "The contact hypothesis re-evaluated." *Behavioural Public Policy* 3 (2): 129–158.
- Pan, Lei, and Luc Christiaensen. 2011. Who Is Vouching for the Input Voucher? Decentralized Targeting and Elite Capture in Tanzania. Technical report World Bank Washington, DC: .
- Panagopoulos, Costas. 2010. "Affect, Social Pressure and Prosocial Motivation: Field Experimental Evidence of the Mobilizing Effects of Pride, Shame and Publicizing Voting Behavior." *Political Behavior* 32 (3): 369–386.
- Petersen, Roger Dale. 2002. *Understanding ethnic violence: Fear, hatred, and resentment in twentieth-century Eastern Europe*. New York: Cambridge University Press.
- Pettigrew, Thomas F. 1998. "Intergroup contact theory." *Annual Review of Psychology* 49 (1): 65–85.
- Pettigrew, Thomas F., and Linda R. Tropp. 2006. "A meta-analytic test of intergroup contact theory." *Journal of Personality and Social Psychology* 90 (May): 751–783.
- Posner, Daniel N. 2004. "The political salience of cultural difference: Why Chewas and Tumbukas are allies in Zambia and adversaries in Malawi." *American Political Science Review* 98 (November): 529–545.
- Pritchett, Lant, and Justin Sandefur. 2013. "Context Matters for Size: Why External Validity Claims and Development Practice Do Not Mix." *Journal of Globalization and Development* 4 (2): 161–197.
- Putnam, Robert D. 1993. *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton: Princeton University Press.
- Putnam, Robert D. 2007. "E Pluribus Unum: Diversity and Community in the Twenty-First Century The 2006 Johan Skytte Prize Lecture." *Nordic Political Science Association* 30 (2): 137–174.
- Rabie, Tamer Samah, Ellen Lust, Cari Clark, Melani Cammett, and Hannah Linnemann. 2014. Improving Quality of Care Against All Odds: A Local Success Story in Jordan. Technical report World Bank Washington, DC: .
- Rawls, John. 1971. *A theory of justice*. Cambridge, MA: Belknap Press of Harvard University Press.
- Ridgeway, Cecilia L. 2019. *Status: Why Is It Everywhere? Why Does It Matter?* 1 edition ed. New York, New York: Russell Sage Foundation.
- Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson, and David G. Rand. 2014. "Structural topic models for open-ended survey responses." *American Journal of Political Science* 58 (October): 1064–1082.
- Ryan, Timothy J. 2014. "Reconsidering moral issues in politics." *The Journal of Politics* 76 (2): 380–397.
- Saguy, Tamar, John F. Dovidio, and Felicia Pratto. 2008. "Beyond contact: Intergroup contact in the context of power relations." *Personality and Social Psychology Bulletin* 34 (March): 432–445.
- Saguy, Tamar, and Lily Chernyak-Hai. 2012. "Intergroup contact can undermine disadvantaged group members' attributions to discrimination." *Journal of Experimental Social Psychology* 48 (May): 714–720.
- Saguy, Tamar, Nicole Tausch, John F. Dovidio, and Felicia Pratto. 2009. "The irony of harmony: Intergroup contact can produce false expectations for equality." *Psychological Science* 20 (January): 114–121.
- Samii, Cyrus. 2013. "Perils or promise of ethnic integration? Evidence from a hard case in Burundi." *American Political Science Review* 107 (August): 558–573.
- 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 (August): 654–677.
- Schatz, Robert T., Ervin Staub, and Howard Lavine. 1999. "On the varieties of national attachment: Blind versus constructive patriotism." *Political Psychology* 20 (March): 151–174.

- Sheely, Ryan. 2015. "Mobilization, Participatory Planning Institutions, and Elite Capture: Evidence from a Field Experiment in Rural Kenya." *World Development* 67: 251–266.
- Siegel, Alexandra A., and Vivienne Badaan. 2020. "#No2Sectarianism: Experimental approaches to reducing sectarian hate speech online." *American Political Science Review* 114 (3): 837–855.
- Sinclair, Betsy. 2012. *The Social Citizen: Peer Networks and Political Behavior*. Chicago: University of Chicago Press.
- Singh, Prerna, and Matthias vom Hau. 2016. "Ethnicity in Time: Politics, History, and the Relationship between Ethnic Diversity and Public Goods Provision." *Comparative Political Studies* 49 (September): 1303–1340.
- Spears, Dean, and Amit Thorat. 2019. "The Puzzle of Open Defecation in Rural India: Evidence from a Novel Measure of Caste Attitudes in a Nationally Representative Survey." *Economic Development and Cultural Change* 67 (July): 725–755.
- Sprinzak, E. 1999. *Brother against brother: Violence and extremism in Israeli politics from Altalena to the Rabin assassination*. New York: The Free Press.
- Staub, Ervin. 1997. "Blind versus constructive patriotism: Moving from embeddedness in the group to critical loyalty and action." In *Patriotism in the lives of individuals and nations*, ed. Daniel Bar-Tal and Ervin Staub. Chicago, IL: Nelson-Hall Publishers.
- Subramanian, Samanth. 2020. "How Hindu Supremacists Are Tearing India Apart." *The Guardian* (February).
- Susewind, Raphael. 2017. "Muslims in Indian Cities: Degrees of Segregation and the Elusive Ghetto." *Environment and Planning A: Economy and Space* 49 (June): 1286–1307.
- Sutton, Robbie M., Tracey J. Elder, and Karen M. Douglas. 2006. "Reactions to internal and external criticism of outgroups: Social convention in the intergroup sensitivity effect." *Personality and Social Psychology Bulletin* 32 (May): 563–575.
- Tajfel, Henri, and John C. Turner. 1986. "The Social Identity Theory of Intergroup Behavior." In *Psychology of Intergroup Relations*, ed. Stephen Worchel and William G. Austin. Chicago: Nelson-Hall.
- Talbot, Ian, and Gurharpal Singh. 2009. *The Partition of India*. New York, NY: Cambridge University Press.
- Tavits, Margit. 2007. "Principle vs. pragmatism: Policy shifts and political competition." *American Journal of Political Science* 51 (1): 151–165.
- Thachil, Tariq. 2017. "Do Rural Migrants Divide Ethnically in the City? Evidence from an Ethnographic Experiment in India." *American Journal of Political Science* 61 (4): 908–926.
- The Wire. 2020. "'State-sponsored genocide': Mamata Banerjee slams BJP over Delhi riots." (March).
- Tsfati, Yariv, and Shira Dvir-Gvirsman. 2018. "Silencing fellow citizens: Conceptualization, measurement and validation of a scale for measuring the belief in the importance of actively silencing others." *International Journal of Public Opinion Research* 30 (3): 391–419. Working paper.
- Tyler, Tom R. 2005. "Policing in Black and White: Ethnic Group Differences in Trust and Confidence in the Police." *Police Quarterly* 8 (September): 322–342.
- United Nations. 2019. *World Urbanization Prospects: The 2018 Revision*. Department of Economic and Social Affairs, Population Division, United Nations.
- Van Craen, Maarten, and Wesley G. Skogan. 2015. "Differences and Similarities in the Explanation of Ethnic Minority Groups' Trust in the Police." *European Journal of Criminology* 12 (May): 300–323.
- Varshney, Ashutosh. 2003. *Ethnic Conflict and Civic Life: Hindus and Muslims in India*. New Haven, CT: Yale University Press.

- Walsh, Katherine Cramer. 2010. *Talking about politics: Informal groups and social identity in American life*. Chicago, IL: University of Chicago Press.
- Walzer, Michael. 1987. *Interpretation and social criticism*. Cambridge, MA: Harvard University Press.
- Wasow, Omar. 2020. "Agenda seeding: How 1960s Black protests moved elites, public opinion and voting." *American Political Science Review* 114 (3): 638–659.
- Weber, Eugen. 1976. *Peasants into Frenchmen: The Modernization of Rural France, 1870-1914*. Stanford, CA: Stanford University Press.
- Wilkinson, Steven. 2004. *Votes and violence: Electoral competition and ethnic riots in India*. Cambridge: Cambridge University Press.
- Wimmer, Andreas. 2015. "Is Diversity Detrimental? Ethnic Fractionalization, Public Goods Provision, and the Historical Legacies of Stateness." *Comparative Political Studies*: 0010414015592645.
- Wimmer, Andreas, Lars-Erik Cederman, and Brian Min. 2009. "Ethnic Politics and Armed Conflict: A Configurational Analysis of a New Global Data Set." *American Sociological Review* 74 (April): 316–337.
- Winter, Fabian, and Nan Zhang. 2018. "Social Norm Enforcement in Ethnically Diverse Communities." *PNAS* 115 (11): 2722–2727.
- Woolcock, Michael. 1998. "Social Capital and Economic Development." *Theory and Society* 27 (2): 151–208.
- Yablon, Yaacov B. 2012. "Are we preaching to the converted? The role of motivation in understanding the contribution of intergroup encounters." *Journal of Peace Education* 9 (3): 249–263.
- Zeitsoff, Thomas. 2014. "Anger, exposure to violence, and intragroup conflict: A "lab in the field" experiment in southern Israel." *Political Psychology* 35 (June): 309–335.