ADBI Working Paper Series

WHEN DOES COMMUNITY PARTICIPATION IN DECISION-MAKING IMPROVE OUTCOMES? EVIDENCE FROM A FIELD EXPERIMENT IN BANGLADESH

Md Ahasan Habib, Malgosia Madajewicz, Anna Tompsett

No. 1406
July 2023

Asian Development Bank Institute

Md. Ahasan Habib is a PhD student at the KTH Royal Institute of Technology, Stockholm, Sweden, and Director of Research, Monitoring, and Evaluation, NGO Forum for Public Health, Dhaka, Bangladesh. Malgosia Madajewicz is an associate research scientist at the Center for Climate Systems Research, Columbia University, New York, United States. Anna Tompsett is a researcher at the Beijer Institute for Ecological Economics, Royal Swedish Academy of Sciences, Sweden, and an affiliated researcher, Institute for International Economic Studies, Stockholm University, Stockholm, Sweden.

Suggested citation:

Habib, M. A., M. Madajewicz, and A. Tompsett. 2023. When Does Community Participation in Decision-Making Improve Outcomes? Evidence from a Field Experiment in Bangladesh. ADBI Working Paper 1406. Tokyo: Asian Development Bank Institute. Available: https://doi.org/10.56506/FEQW6636

Please contact the authors for information about this paper.

Email: anna.tompsett@iies.su.se

This research was funded by National Science Foundation Grant 0624256: DHB: Decentralization and Local Public Goods: How Does Allocation of Decision-Making Authority Affect Provision? We thank the NGO Forum for Public Health for partnering with us to perform this study, particularly Executive Director S.M.A. Rashid, Kabid Ahmed, Md. Shamsul Karim, Fatema Tuz-Zohora, and Md. Rofirobi Uddin provided excellent research assistance. Thanks also to Habiba Djebbari, Ana Varela Varela, and the participants of the WASH Economics Conference for helpful comments and suggestions. We are also grateful for valuable support from ADBI, including from KE Seetha Ram, senior consulting specialist; Dil Rahut, senior research fellow; Santi Setiawati, research associate; and Adam Majoe, publishing consultant.

Asian Development Bank Institute
Kasumigaseki Building, 8th Floor
3-2-5 Kasumigaseki, Chiyoda-ku
Tokyo 100-6008, Japan

Tel: +81-3-3593-5500
Fax: +81-3-3593-5571
URL: www.adbi.org
E-mail: info@adbi.org

Abstract

Development practitioners have long advocated for targeted beneficiary communities to participate in decision-making about how to provide local public goods and services. However, community participation in decision-making may have disadvantages as well as benefits, and the balance between the benefits and disadvantages may vary across different decision-making processes, interventions, and contexts. Previous studies of community participation in decision-making indeed report mixed results, but the underlying reasons for this remain uncertain. This study demonstrates that context matters. We investigate the heterogeneity of impacts in a field experiment conducted in two regions of rural Bangladesh. The experiment randomly assigns different decision-making processes to villages that receive otherwise identical interventions in the form of a program to increase access to safe drinking water. We show that a deliberative, consensus-based approach to community participation in decision-making has strongly heterogeneous effects compared to either a top-down approach or community decision-making without rules about how decisions are made. The consensus-based process doubles the program's impact in one region but barely increases it in the other. We use machine learning to identify the baseline characteristics and mechanisms that correlate most strongly with impact. The results suggest that the consensus-based process yields better outcomes when there is more at stake: specifically, when the community has fewer pre-existing safe sources of drinking water and thus more to gain from the intervention. The results are consistent with the view that inclusive participatory approaches to decision-making can increase program impacts but fully engaging in these processes is costly. When less is at stake, communities may not fully engage, and the advantages of inclusive participatory decision-making may not be realized.

Keywords: community participation, consensus-based approach, public goods and services, heterogeneity of impacts

JEL Classification: D02, H41, O17, O20

1 Introduction

A long-standing tenet of development practice is the idea that participatory decision-making improves outcomes in projects to provide local public goods and services. Participatory decision-making may also have disadvantages, however, and the balance between the benefits and disadvantages of a given decision-making process may vary in different circumstances. Policy-makers need to know not only which decision-making processes perform best on average but also whether different decision-making processes perform differently in different contexts, when this happens, and why. To make progress on answering these questions, we use machine learning techniques to investigate the heterogeneity of effects in a randomized field experiment.

The experiment provides safe-drinking-water wells under different decision-making approaches in two regions of Bangladesh. The two regions differ in important respects hypothesized to affect the success of participatory approaches and, more generally, collective action, including pre-intervention access to safe drinking water, cohesiveness, community size, religious fractionalization, and inequality. On average, the use of safe drinking water increased more with a deliberative, consensus-based approach to community participation in decision-making than with either a top-down approach to decision-making or unrestricted community decision-making (Madajewicz, Tompsett, and Habib 2021). This paper investigates heterogeneity in how the decision-making processes affected the program's impact.

First, we evaluate *whether* decision-making processes perform differently in different contexts. We compare the relative performances of the three decision-making processes in the two different regions. The consensus-based process, which we refer to as **regulated community participation**, performed best in both regions, but the differences in performance vary starkly across the two regions. In one region, regulated community participation doubled the program's impact compared to the other two decision-making processes, while in the other region, the differences in the program's impact across decision-making processes are very small and statistically indistinguishable from zero. The difference in performance between the regulated process and the other two processes is significantly larger in the first region than in the second ($p = 0.067$). **Community participation** without rules outperforms the **top-down** approach in one region, while the reverse is true in the second region; however, in both cases, the differences in performance are small. Because the same staff implemented the same decision-making process for the same intervention in each region, the differences in the performances of the decision-making processes must be attributable to differences in the contexts.

Second, we evaluate *when* or under which circumstances the decision-making processes perform better. We use machine learning to study which pre-intervention community characteristics correlate most strongly with the relative performances of different decision-making processes. Although we observe a rich set of baseline covariates—including access to safe drinking water, community cohesion, and leader quality—it is difficult to accurately predict the relative performance of different decision-making approaches using baseline characteristics. The performance may depend on characteristics that are difficult to observe or measure, or we may simply have limited statistical power to detect these differences in our sample. With this caveat in mind, the regulated community participation process consistently exceeds or equals the performance of the community participation process and outperforms the top-down process. The community participation process may outperform or be outperformed by the top-down process in different communities, but in general, the differences in performance are small. Two groups of baseline characteristics are strongly correlated with higher performance of the regulated community participation process relative to the community participation process: 1) having more at stake, in terms of lower baseline access to safe drinking water, and 2) having a relatively low degree of community cohesion, perhaps precisely because highly cohesive communities had already solved the collective action problem themselves before our intervention.

We then use machine learning to identify which mechanisms are most correlated with success under each of the three decision-making processes. In other words, we evaluate what else happens differently when the project yields greater impacts under each decision-making process. We consider how many wells communities install, who contributes towards a community contribution requirement, the characteristics of the locations chosen, who uses the wells, and what households later report about the decision-making process. These analyses confirm that the program's impact is strongly heterogeneous under the regulated community participation process. The use of safe drinking water barely increases in the least affected quartile, while it increases by 50 percentage points in the most affected quartile. Mechanism variables are less correlated with the program's impact for the top-down process and uncorrelated with impact for the unregulated community participation process. When the regulated community participation approach works well, it yields decisions that are more likely to be praised as fair and less likely to be characterized by elite capture.

We cannot definitively say *why* we observe these patterns but our results are consistent with a view of participatory decision-making that suggests that it is *costly* for communities to participate in and influence decision-making processes. The consensus-based process appears to create an opportunity for communities to influence decision-making and achieve better outcomes. However, participating in the process in a way that changes outcomes may increase the risk of social costs, in terms of conflict, stress, and strain on relationships. Communities may only be willing to incur these costs when enough is at stake. Consistent with this view, the consensus-based participatory process appears to yield the greatest improvement in outcomes when there is most at stake.

In this context, the consensus-based process rarely seems to lead to worse outcomes than other decision-making processes. Thus, when applied to this class of problems, at least, the consensus-based process appears to have few drawbacks in terms of program impact. The costs of the consensus-based process are also limited: time costs to communities are somewhat greater, while costs to the implementing organization are similar for the consensus-based approach and the top-down approach; costs are at most 20% lower for the unregulated community participation approach (Madajewicz, Tompsett, and Habib 2021). However, our results suggest that policy-makers should be cautious about viewing participatory decision-making processes as a quick win or a "magic bullet." In the lowest-impact quartile of villages, the consensus-based process does not yield detectable increases in the use of safe drinking water, not because fewer wells are built but because wells are poorly targeted or captured by elites. The consensus-based approach yields outcomes similar to those of the unrestricted community decision-making process when there is less at stake. This pattern of results suggests that communities may rationally only participate fully and effectively in the consensus-based process when there is enough at stake to justify the costs of participation.

Our results help make progress on understanding why studies of participation in decision-making find such strikingly heterogeneous effects. These studies fall into two groups: the first group compares what happens when the same intervention is implemented with participatory approaches and top-down approaches,[1] while the second evaluates interventions that aim to induce more or better-informed participation in public service provision.[2] Both groups include studies that find strongly positive impacts of participatory approaches or increased participation (e.g., Madajewicz, Tompsett, and Habib 2021; Björkman and Svensson 2009), as well as many studies with null (e.g., Olken 2007; Arkedis et al. 2021) or even negative (Allakulov et al. 2023) results. The reasons for this heterogeneity remain uncertain. Previous studies evaluate different approaches to participation for different interventions in different contexts, making it difficult to disentangle the roots of differences in impacts. Raffler, Posner, and Parkerson (2020) hypothesize that interventions designed to increase citizen engagement in public service provision may be more effective when baseline conditions are worse and greater potential improvements are feasible, showing that such a correlation holds across five studies in six contexts. However, these studies differ in other respects as well, implying that we cannot be certain that the heterogeneity in impacts is caused by baseline conditions. In this study, we confirm that context matters: different decision-making processes applied to the same decision-making problems yield different results in different contexts.

Demonstrating that context matters does not, however, tell us which contextual factors are most important in driving heterogeneity. Previous studies have hypothesized and investigated many different contextual factors that could affect the success of participatory approaches or, more generally, collective action (Mansuri and Rao 2013; Khwaja 2009; Banerjee, Iyer, and Somanathan 2008). These include inequality (Baland and Platteau 2018; Banerjee and Somanathan 2007), social fragmentation (Alesina, Baqir, and Easterly 1999; Miguel and Gugerty 2005), group size (Olson 1971; Cocciolo, Habib, and Tompsett 2019a), information (Khemani 2007), and geographical isolation (Galasso and Ravallion 2005). Most closely related to our study are a small number of studies that evaluate within-study heterogeneity in either the effects of interventions that induce more or better-informed participation in public service provision or the relative performance of different approaches to decision-making (Olken 2010; Arkedis et al. 2021; Björkman and Svensson 2010; Raffler, Posner, and Parkerson 2020; Alatas et al. 2012; Barr et al. 2012). Most of these

[1] See Olken (2007), Alatas et al. (2012), Alatas et al. (2019), and Madajewicz, Tompsett, and Habib (2021).

[2] See Björkman and Svensson (2009), Nyqvist, Walque, and Svensson (2017), Barr et al. (2012), Pradhan et al. (2014), Mohanan et al. (2020), Raffler, Posner, and Parkerson (2020), Christensen et al. (2021), Banerjee et al. (2010), Arkedis et al. (2021), BenYishay et al. (2022), and Allakulov et al. (2023).

studies focus on a few specific contextual factors that potentially explain heterogeneity. This study is, to our knowledge, unique in leveraging new approaches in machine learning to identify the contextual factors that most strongly correlate with heterogeneity in impacts. Notably, and in contrast to precedents in the literature, our approach does not identify either fragmentation or inequality as a driver of heterogeneity.

Another related body of literature evaluates heterogeneity in the impacts of participatory projects compared to no intervention, but these studies cannot distinguish between heterogeneity in the impacts of the intervention itself and heterogeneity in the impacts of participation (e.g., Fearon, Humphreys, and Weinstein 2015; Casey, Glennerster, and Miguel 2012; Humphreys, Sierra, and Windt 2019). Our results imply that it is crucial to account for heterogeneity in the impacts of an intervention to correctly understand heterogeneity in the impacts of participatory approaches.

The remainder of the paper proceeds as follows. Section 2 describes the two different contexts in which we implemented the experiment. Section 3 briefly describes the experiment, with a more detailed treatment provided in Madajewicz, Tompsett, and Habib (2021). Section 4 outlines a simple model that helps motivate why having "more at stake" might drive heterogeneity in the impacts of participatory decision-making. Section 5 demonstrates how the results differ in the two areas in which we implemented the project. We then use machine learning to identify the village-level characteristics that are associated with larger impacts under different decision-making processes (section 6) and the mechanisms that are most strongly correlated with larger impacts under each process (section 7). Section 8 reviews the conclusions we can draw about when participatory approaches are most likely to improve outcomes.

2 Context

We implemented our experiment in two study regions, Gopalganj and Matlab. Both regions are severely affected by arsenic contamination, with the large majority of wells contaminated with arsenic before our intervention (Madajewicz, Tompsett, and Habib 2021). However, the two regions differ in many other important respects. In particular, they differ in many respects that previous studies have associated with the success of participatory approaches or collective action.

At the time of our intervention, Matlab had been the focus of more national and international efforts to improve access to safe drinking water. In Matlab, as compared to Gopalganj, households report considerably lower distances to the nearest safe water source, more households report that they use a safe water source, and more households report that they have switched from arsenic-contaminated sources to safe sources in the last five years. Possibly as a consequence of the adoption of safe community wells, fewer households report owning or using their own private well in Matlab than in Gopalganj.

Gopalganj and Matlab also differ in many social respects. Villages in Gopalganj are generally larger than those in Matlab, and they appear to be considerably less cohesive: households in Gopalganj report knowing of and participating in fewer community associations and collective actions. Villages in Gopalganj are poorer and more unequal, they are much more likely to have community leaders who inherited their status by birth and less likely to have democratically elected leaders, and a much smaller share of households report participating in community decision-making. Villages in Matlab are overwhelmingly majority Muslim, while villages in Gopalganj are more fractionalized. Figure 1 illustrates some of the key differences between the two regions.

Importantly for the latter part of our analysis, although the average differences between Gopalganj and Matlab are significant, there is considerable overlap between the two regions, implying that we can identify villages with similar observable characteristics in both regions.

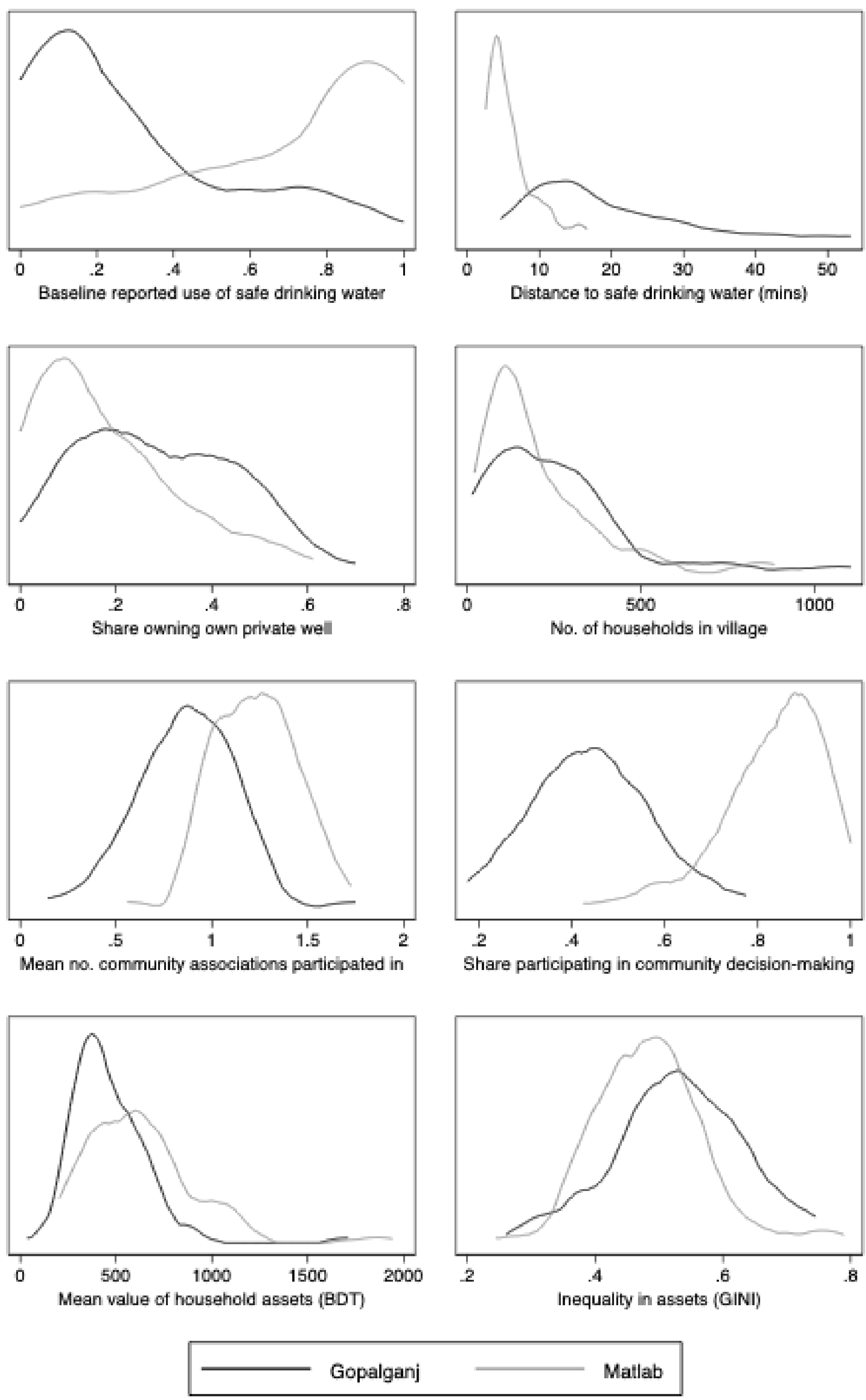
3 The experiment

3.1 The intervention

The intervention that we study is a program of subsidies and technical advice designed to improve access to safe drinking water through the construction of safe-drinking-water wells. We implemented the intervention in partnership with an experienced Bangladeshi non-governmental organization (NGO), the NGO Forum for Public Health (NGOF). We provide a summary of the intervention's features here and refer the reader to Madajewicz, Tompsett, and Habib (2021) for more details. Although the intervention offered a suite of technologies to provide safe drinking water, we follow Madajewicz, Tompsett, and Habib (2021) and focus here only on communities in which tubewells, the most common and widely accepted technology for accessing drinking water, were feasible. Where tubewells were not feasible, most communities simply declined to install water sources at all, and the use of safe drinking water did not change (Madajewicz, Tompsett, and Habib 2021).

Each village could install up to three safe wells as part of the intervention. Before the experiment, we implemented an information campaign to ensure that all communities were aware of the arsenic problem. Field staff then organized community meetings in all treated villages to explain the intervention and its rules. Each village had to contribute about 15% of the cost of each well that they wanted to install in cash, with up to 12 weeks to raise the required contributions. If villages raised the required contributions within the designated period, the project installed the well. Field staff provided technical support, ensuring that all project wells were installed in technically appropriate locations and confirming that the wells yielded safe water after installation. A small team of field staff implemented the project in each upazila. As with most development projects, the communities were responsible for the future repair and maintenance of any installed wells.

Figure 1: Descriptive baseline statistics



Note The empirical densities of key variables at the baseline in Gopalganj and Matlab. Source: The authors of this paper.

3.2 Decision-making processes

The key project decisions were how many, if any, of the offered wells to install and where to locate any installed wells. We randomly assigned villages to one of three decision-making processes.

The top-down (TP) process We conceptualize "top-down" decision-making as taking place when a provider who is not part of the community makes project decisions. The top-down provider in our study is our partner NGO. Field staff selected well locations based on information they gathered from the community before the community meetings. They then offered to install wells at these locations during the community meetings. Communities then raised the required contribution for any well they wanted to install. The locations selected by field staff were final and could not be changed.

The community participation (CP) process Communities made decisions. At the community meetings, field staff explained the project rules and then gave the communities up to two weeks to make decisions, using any process they chose. Field staff did not observe decision-making, but community members later reported in focus groups that small groups of influential people, usually men, made decisions in closed meetings (Madajewicz, Tompsett, and Habib 2021).

The regulated community participation (RCP) process Communities made decisions, but field staff imposed rules about how decisions were made. We designed these rules to limit potential elite capture and broaden participation in decision-making. Communities had to make decisions at the community meetings organized by the field staff, at which the field staff imposed representation requirements for women and the poor. Decisions were by consensus, and communities deliberated during the meetings until they reached consensus. In a few cases, villages held several meetings before reaching consensus. Decisions made at the meetings were binding.

3.3 Selection of study sites and assignment of treatment

The study villages were selected based on arsenic contamination. We included villages in the study if the share of village wells that had unsafe levels of arsenic was over 65% in Matlab and over 75% in Gopalganj, according to data from the Bangladesh Arsenic Mitigation Water Supply Project (BAMWSP), and if no other organization was actively working to provide new sources of safe drinking water. We originally selected 125 villages in Gopalganj and 125 villages in Matlab, later excluding 23 villages after cost increases reduced the number of villages in which we could implement the program.

Our study protocol randomly assigned 50 villages in each region to the control group, which would not receive any intervention, and all other villages to the treatment group under different decision-making processes.[3] In Matlab, the project director at the time did not follow this protocol;

[3] We used a list randomization approach that was standard practice at the time; see Madajewicz, Tompsett, and Habib (2021).

he assigned all the villages in one area of Matlab to the treatment group. The random assignment to treatment in Gopalganj, and the random assignment of decision-making processes for all treated villages, occurred as planned.

From this original sample, we follow Madajewicz, Tompsett, and Habib (2021) and exclude 17 treated villages in Gopalganj in which we could not install arsenic-safe wells. To provide arsenic-safe water, wells must penetrate an arsenic-safe aquifer—a layer of permeable earth through which water can flow—that usually lies deep below the surface. In the excluded villages in Gopalganj, a rocky layer overlies the arsenic-safe aquifer. This rocky layer cannot be penetrated using local drilling technologies, making safe well construction infeasible. Communities rejected alternative technologies in all but a tiny handful of these communities, rendering the intervention ineffective on average. The presence of this rocky layer deep below the surface is uncorrelated with village characteristics, but since it is spatially correlated, we exclude a spatially matched subset of the control villages in Gopalganj. The results are, however, almost identical if we instead include the full control group.

We lost baseline data for one control village and one village treated under the CP process, both of which were in Matlab. The final sample of treated villages with non-missing data in which tubewells were feasible comprises 56 villages in Matlab—19 under the TD process, 18 under the CP process, and 19 under the RCP process—and 51 villages in Gopalganj, where 17 villages were assigned to each of the three decision-making processes. The final sample of control villages with non-missing data that are spatially matched to the tubewell-feasible villages in Gopalganj comprises 49 villages in Matlab and 35 villages in Gopalganj. Figures 2a and 2b summarize the sample selection process.

The deviation from the study protocol in Matlab resulted in lower baseline access to safe drinking water in treated villages compared to control villages in Matlab. Madajewicz, Tompsett, and Habib (2021) confirm using statistical tests that treated villages and control villages in Gopalganj, and villages treated under different decision-making processes in both regions, are indistinguishable from one another at the baseline.

3.4 Description of data

We surveyed a random sample of households in each village in a baseline survey in 2007 and again in an endline survey in late 2010 or early 2011. In the baseline survey, we asked households about their awareness of arsenic contamination, water sources used for drinking and cooking, proxies for household wealth and income, social networks, and other community characteristics. We also conducted a village survey with key informants from each community; we asked these key informants additional questions about community characteristics. In the endline survey, we resurveyed households, primarily focusing on water sources used for drinking and cooking, as well as their opinions about the project. Attrition was very low.[4]

Field staff documented the project implementation process, including whether or not wells were

[4] Madajewicz, Tompsett, and Habib (2021) provide a full description.

installed, whether they were installed in public places or on private land, and how many households contributed to each installed well.

For this study, where our core interest is the heterogeneity of treatment effects across villages and regions, we assemble these data into a village-level panel of baseline and endline data that combines village-level data with mean or other summary values of the household-level data.

3.5 Outcome variable

The primary outcome that we study is the use of safe drinking water, as reported by the household. The outcome variable takes the value one if the household reports that their primary source of drinking water is safe from arsenic contamination and at low risk of bacterial contamination, and it takes the value zero if they report that their primary source is unsafe or of unknown safety, or if the source is particularly vulnerable to bacterial contamination, such as, for example, a dug well or surface water. We installed wells between 2008 and 2011, and we measure outcomes between 2010 and 2011, implying that we measure impacts up to about two years after well installation.

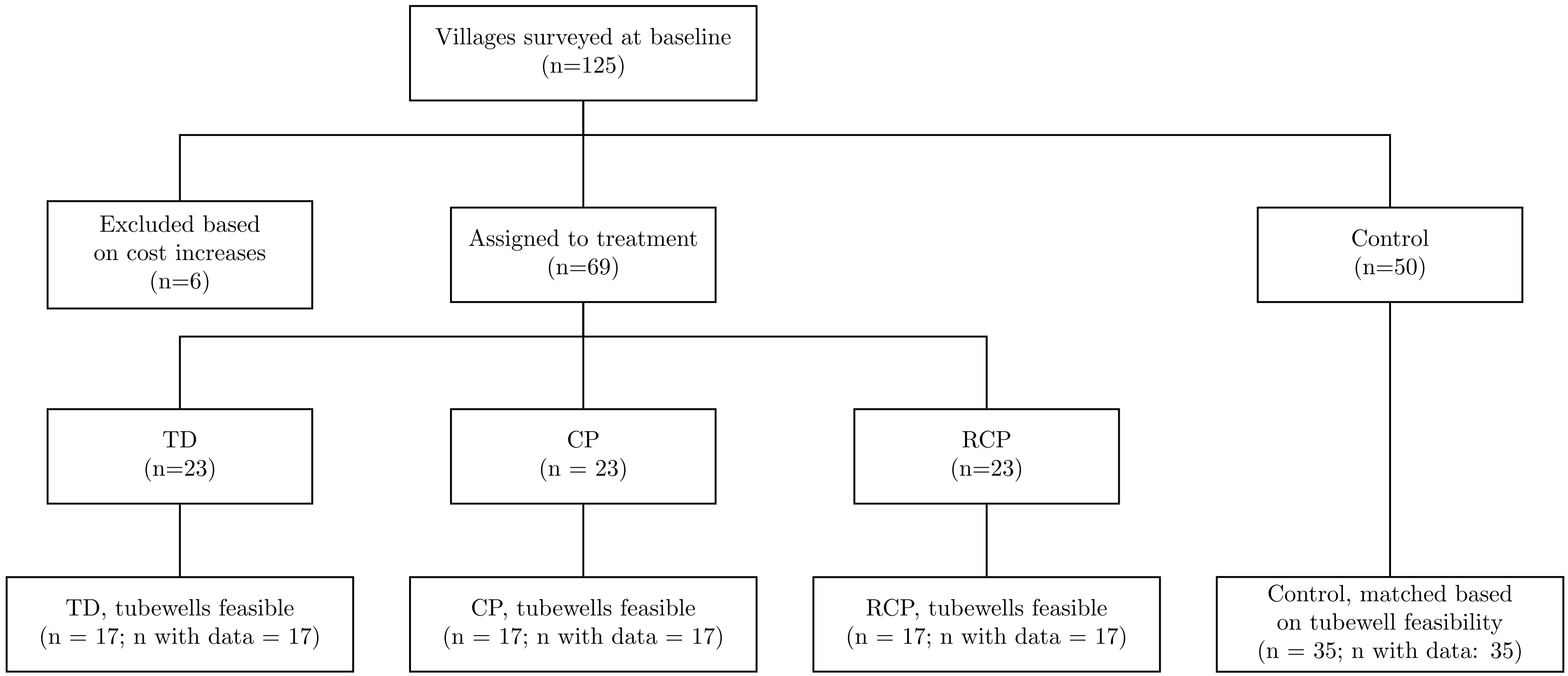
One might be concerned that households might overreport the use of safe sources, especially after exposure to an intervention designed to increase the use of safe drinking water (Ahuja, Kremer, and Zwane 2010). However, the reported measure was largely consistent with enumerators' own assessments of the safety of the drinking water source (Madajewicz, Tompsett, and Habib 2021) and, in a similar context, tests using multiple measures of drinking water safety suggested very limited reporting bias (Cocciolo et al. 2021; Cocciolo et al. 2023).

4 A simple model for heterogeneity in the effects of consensus-based decision-making

We use a highly simplified model to highlight conditions under which the consensus-based decision-making process improves outcomes relative to unrestricted decision-making by communities. We focus on these two processes since we find the strongest evidence for heterogeneity in their effects. The model focuses on the decision of where to locate wells, expanding and extending elements from a similar model in Madajewicz, Tompsett, and Habib (2021). The model sheds light on why how much households stand to gain from obtaining access to safe drinking water might drive heterogeneity in the impacts of consensus-based decision-making, helping to rationalize our results.

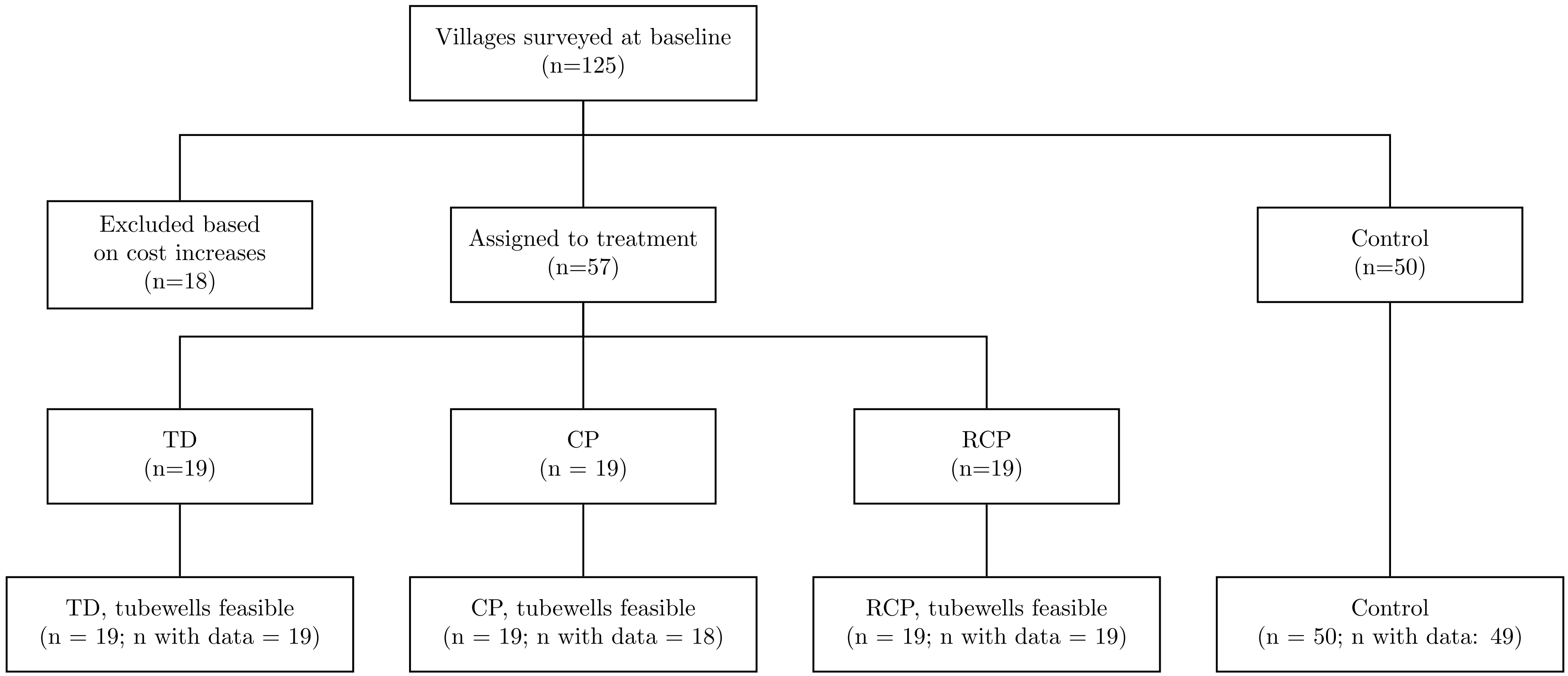
The model imagines a simplified community geography, consisting of a line of length d (Figure 3). An elite household e resides at one vertex, and a group N of non-elite households (ne) resides at the other. Each household owns a plot of land, and households can restrict access to wells built on their own land. Under the status quo, each household uses a private well on their own land. Each private well has safety s, which we assume is constant for all houses within a village, reflecting the spatial correlation in arsenic contamination. Each household's plot of land could also accommodate a community well.

Figure 2a: Experimental design and sample selection in Gopalganj



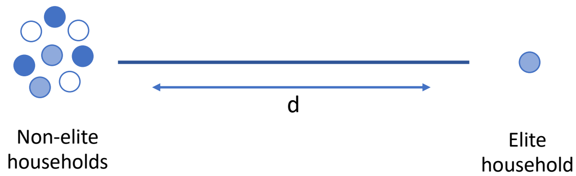
Notes TD: Top-down. CP: Community participation. RCP: Regulated community participation. Source: The authors of this paper.

Figure 2b: Experimental design and sample selection in Matlab



Notes TD: Top-down. CP: Community participation. RCP: Regulated community participation. Source: The authors of this paper.

Figure 3: Idealized community geography



Notes Source: The authors of this paper.

As in the real-world project, communities decide where to locate a well, with or without the rules designed to restrict elite capture. Landowners decide whether or not to allow other households to use a well built on their land. Households decide whether or not to use the well, if landowners allow them access. We solve the model backwards. First, we describe which households use a well in a given location. Second, we investigate how restrictions on elite capture affect the location and thus the overall outcome.

As in Madajewicz, Tompsett, and Habib (2021), installing a safe well yields the following payoff for each household i:

$$V_i = \alpha_i - \delta_i - \gamma_i + \theta_i \sum_{j \neq i} U_j, \tag{1}$$

where

$$
\begin{aligned}
&U_j = \alpha_j - \delta_j - \gamma_j,\\
&\alpha = \begin{cases} a \geq 0, & \text{if a household adopts a safe well, where } \frac{da}{ds} < 0,\\ 0, & \text{otherwise,} \end{cases}\\
&\delta = \begin{cases} d > 0, & \text{if a household uses a well that is located at the other vertex,}\\ 0, & \text{otherwise,} \end{cases}\\
&\gamma = \begin{cases} g > 0, & \text{if others use a well located on the household's own land,}\\ 0, & \text{otherwise,} \end{cases}\\
&\theta_i \geq 0.
\end{aligned}
$$

The term α captures the benefits of adopting a safe well relative to the status quo. Naturally, the benefits of adopting a safe well are higher when the status quo well is more unsafe. The assumption that all private wells in the community have equal safety implies that all households have an equal payoff a from adopting a safe well. The parameter δ measures the travel cost to the well, and the parameter γ captures a disutility cost that households incur if other households use a well installed on their land. If $\theta_i > 0$, the household i cares about the utility of other households in the same community, denoted by j. We will refer to θ_i as an altruism parameter, although it could also capture other, more strategic interests in other households' welfare. For tractability, we assume that altruistic preferences only extend to the private dimensions of other households' payoff functions, denoted U_j.

If a household uses a well built at the other vertex, their payoff is $a-d$. We assume that $a>d$, so that households always adopt a safe well to which they have access. We also assume that $d>g$, implying that a purely selfish household prefers to install a well on their own land and incur the disutility cost of allowing access instead of walking to a well located at the other vertex.

If a well is built on a household's own land, they allow other households to use the well if their utility gains from allowing access exceed the disutility costs of allowing access.[5] We assume that at least one non-elite household is altruistic enough to allow access to a well that is built on their land.

Each household knows the payoff functions of the other households, how altruistic every other household is, the safety of the community's wells, and the geography of the community.

The location of the well depends on the decision-making process. Each outcome is associated with a payoff to each household and with a value of social welfare (Table 1). The welfare function considers only the private dimensions of household welfare (Johansson 1992) and is given by

$$W=\sum_i U_i. \tag{2}$$

Under the two decision-making processes that we consider, there are three possible outcomes. The outcome that maximizes social welfare is to place the well on land belonging to a non-elite household that is altruistic enough to allow access to the well.

Table 1: Payoff matrix

Outcome	Land	Accessible	V_e	V_{ne}	W
1	Elite	Yes	$a-g+$ $\theta_e N(a-d)$	$a-d+$ $\theta_i(Na-(N-1)d-g)$	$(N+1)a-Nd-g$
2	Elite	No	a	$\theta_i a$	a
3	**Altruistic non-elite**	**Yes**	$a-d+$ $\theta_e(Na-g)$	$a\{-g\}+\theta_i(Na-d[-g])$	$(N+1)a-d-g$

Notes: The term that appears in curly parentheses only applies to the non-elite household on whose land the well is constructed; the term that appears in square brackets only applies to non-elite households on whose land the well is not constructed. Source: The authors of this paper.

Although the top-down process is not the focus of the present discussion, we note that the top-down provider cannot consistently select the welfare-maximizing outcome. Unlike the households in the community, the top-down provider does not know how altruistic each household is, and non-altruistic households have an incentive to misrepresent their type. Thus, the top-down provider risks unintentionally installing the well on a non-altruistic household's land, resulting in elite capture.

[5] The elite household allows access if $\theta_e>\frac{g}{N(a-d)}$. A non-elite household ($i\neq e$) allows access if $\theta_i>\frac{g}{Na-d}$.

We discuss in Madajewicz, Tompsett, and Habib (2021) the strategies that the top-down provider may employ in response to this risk.

The community participation process We assume that when communities make decisions in an unrestricted way, the elite household chooses the location that maximizes their own payoff. As we discuss in Madajewicz, Tompsett, and Habib (2021), this assumption, while simplistic, captures in essential respects the decision-making process in the project villages. The location the elite household chooses depends on the altruism parameter θ_e. If θ_e is high enough, the elite household places the well in the welfare-maximizing location. If θ_e is low enough, the elite household places the well on their own land and restricts access to it. For some parameter values, an intermediate range of θ_e may exist for which the elite household places the well on their own land but allows access to the non-elite households. The elite household never chooses to locate the well on the land of a non-elite household that would restrict access, even if such households exist, as this outcome is strictly worse than locating it on the land of a non-elite household that would allow access. The locations chosen are illustrated in Figure 4a.

The regulated community participation process We model the regulated decision-making process as follows. We assume that only the elite household can propose locations. Although this is a stylized assumption, it reflects what we observe in the community meetings under this decision-making process: although many households attend the meeting, only a few participate actively, and the people who do participate are overwhelmingly male and of high status. We assume that any non-elite household can veto any proposal made by the elite household, at a cost to the vetoing household of V. The cost V captures any ability the elite household has to punish the non-elite households in the future. If more than one household would be willing to use a veto, V may also capture the costs of coordinating action.

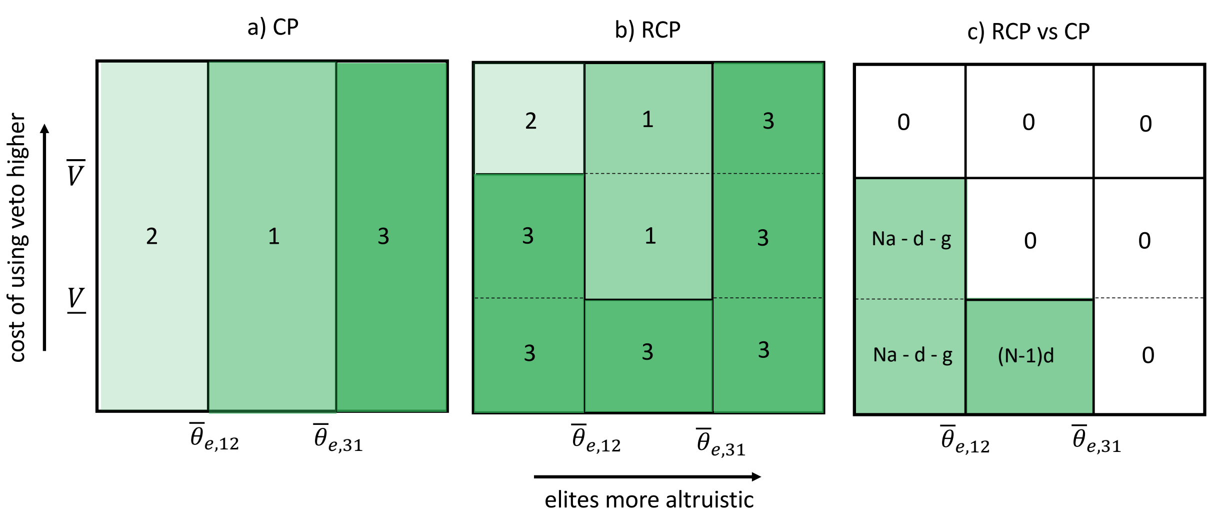
The net payoff from using a veto is the difference between the payoffs of the vetoed proposal and the alternative, minus the cost of using the veto. With all else equal, the more altruistic households have higher payoffs from using the veto, because they value the gains to the other households more highly. Under our assumptions, the threshold cost at which at least one non-elite household would be willing to use the veto is higher when the elite household is more selfish, because the non-elite households' gains from moving from outcome 2 to outcome 3 are larger than the gains from moving from outcome 1 to outcome 3.[6] Because the elite households anticipate the veto, they do not propose locations that a non-elite household would veto. The elite households thus propose the private payoff-maximizing location that no non-elite household will veto. The resulting locations chosen are shown in Figure 4b.

[6]A non-elite household will veto a proposal from a selfish elite household if for at least one household i, $V < a\{-g\} + \theta_i\left((N-1)\,a - d[-g]\right) = \overline{V}$. A non-elite household will veto a proposal from an elite household with intermediate altruism levels that would otherwise construct the well on their own land but allow access to it if for at least one household i, $V < d\{-g\} + \theta_i\left((N-2)\,d[-g]\right) = \underline{V}$. The term in curly brackets applies if the well is installed on household i's land; the term in square brackets applies if not. Under our assumptions, $\overline{V} > \underline{V}$.

The regulated community participation process thus always either increases or provides the same welfare when compared to the community participation process (Figure 4c). The difference between the outcomes of the processes is largest when i) the elites are least altruistic; ii) the cost of using the veto is lower; and iii) the gains from using the veto are highest. In the model, the gains from using the veto increase with a and d and with the maximum values of θ_i among the non-elite households. The model thus nests several hypothesized reasons for heterogeneity in the effects of decision-making processes. Most salient for this study, the model provides a rationale for why consensus-based decision-making has a larger effect when more is at stake: when the gains from avoiding elite capture are high, community members are more likely to be willing to pay the costs of using veto rights. The same logic applies more generally to the decision about whether or not to participate actively in community decision-making, given the opportunity.

The model also incidentally helps illustrate why the causes of heterogeneity in the impacts of different decision-making processes might be difficult to detect in aggregate data. Whether at least one household is willing to use the veto depends implicitly on the distribution of altruism preferences among the non-elite households and in particular on the *maximum* value of the altruism parameter among the non-elite households. In the real world, this corresponds to whether or not the community includes one individual willing to stand up for the common good. Anecdotal evidence from the field suggests that the presence of such an individual matters critically for the success or failure of participatory projects, but aggregate data would be unlikely to capture the presence or absence of such an individual.

Figure 4: How outcomes and impacts on welfare vary with community characteristics under different decision-making processes



Notes a) Locations chosen under the CP process; b) Locations chosen under the RCP process; c) Difference in welfare. Source: The authors of this paper.

5 Do the effects of community participation in decision-making vary by context?

On average, the consensus-based RCP process outperformed the two other decision-making processes: the use of safe drinking water increased by 27 percentage points under the RCP process, while it increased by 14 percentage points under the TD process and an essentially identical 15 percentage points under the CP process (Madajewicz, Tompsett, and Habib 2021). This section evaluates whether the absolute and relative effects of the decision-making processes vary in the two study regions.

5.1 Methodology

We estimate the following, fully saturated equation by ordinary least squares (OLS):

$$\Delta Y_v = \alpha + \alpha_G \mathbb{1}_{G,v} + \sum_{p \in P} \beta_p \mathbb{1}_{p,v} + \sum_{p \in P} \beta_{p,G} \mathbb{1}_{p,v} \times \mathbb{1}_{G,v} + \epsilon_v, \tag{3}$$

where ΔY_v is the change in the outcome variable between the baseline and endline in village v, $\mathbb{1}_{G,v}$ is an indicator that takes the value one if v is in Gopalganj and zero if v is in Matlab, and $\mathbb{1}_{p,v}$ is an indicator variable that takes the value one if v was treated under decision-making process $p \in P = [RCP, CP, TD]$ and zero otherwise. The constant α captures the mean change between the baseline and endline in control villages in Matlab and α_G captures the difference between the mean changes in control villages in Gopalganj and Matlab. The coefficients of interest are β_p, the effect of implementing the program under decision-making process p in Matlab, and $\beta_{p,G}$, the difference in the effects of decision-making process p between Gopalganj and Matlab. We estimate and report heteroskedasticity-robust standard errors, following Long and Ervin (2000).

The random assignment of treated villages to decision-making processes implies that the differences between any two coefficients β_p, or between any two coefficients $\beta_{p,G}$, have a causal interpretation. Because villages were not randomly assigned to treatment in Matlab, the coefficients β_p do not themselves automatically have a causal interpretation. However, they are consistent and unbiased estimates of treatment effects under two assumptions: 1) any differences between treatment and control villages in Matlab affect the use of safe drinking water additively; and 2) any such differences do not change between the baseline and endline (Madajewicz, Tompsett, and Habib 2021). Comparing results from samples that include and exclude the villages in which random treatment assignment was not implemented correctly suggests that these assumptions most likely hold (Madajewicz, Tompsett, and Habib 2021).

We note that the model we estimate here diverges from that in Madajewicz, Tompsett, and Habib (2021), which uses household-level data in a two-period model with village and time fixed effects and stratification controls. We estimate the simpler model using village-level data and first differences here for consistency with the rest of the analysis in this paper. In practice, the simpler model yields very similar results to the two-period household-level model. The average effects we obtain from the simpler model are nearly identical to the effects we report in Madajewicz, Tompsett, and Habib (2021), differing by at most 1 percentage point for both coefficients and standard errors (Appendix Table C1).

5.2 Results

The RCP process increases the proportion of households that use safe water much more than the other two processes primarily in Gopalganj. Column 1 of Table 2 reports the effects of the program in Gopalganj, which we obtain using the linear combination $\beta_p + \beta_{p,G} \; \forall \; p \in P$. Treatment under the RCP process increases the proportion of households that use safe water by 42 percentage points, while treatment under the CP process increases this proportion by 19 percentage points and treatment under the TD process increases it by 21 percentage points. In Gopalganj, the difference between the RCP process and the other two processes combined, as well as the pairwise differences between the RCP process and the other two processes, are statistically significant. The difference between the effects of the CP and the TD processes is not. Figure 5 shows empirical cumulative distribution functions for the three processes in each region, which confirm that the RCP process clearly dominates in Gopalganj.

The three processes produce more similar results in Matlab, as we show in column 2 of Table 2. While treatment under the RCP process also has the largest impact on the use of safe water in Matlab, treatment increases the proportion of households that use safe water by very similar amounts under all three processes: this proportion increases by 14 percentage points under the RCP process, 12 percentage points under the CP process, and 10 percentage points under the TD process (see column 2 of Table 2). None of the differences between the processes are statistically significant in Matlab.

The difference between the treatment effects under the RCP process and the other two processes combined in Gopalganj is statistically significantly different from the same difference in Matlab. We also reject the null hypothesis that the pairwise difference between the RCP and CP processes is the same in Gopalganj and Matlab. The p value for the comparison of the differences between the RCP and TD processes in the two regions is narrowly outside the range of statistical significance (p=0.147). The CP process performs better than the TD process in Matlab but performs worse than the TD process in Gopalganj; however, the differences are small and not statistically significant. All p values are given in Table 2.

One might reasonably be concerned that multiple hypothesis testing might result in some of the individual tests in Table 2 being rejected due to chance, given that we report the results of 27 correlated hypothesis tests. However, among these tests, we reject four at the 1% level, a further six at the 5% level, and two more at the 10% level. For the six tests of the null hypothesis that the differences in impacts between decision-making processes are equal in the two regions, we reject two at the 10% level, and two more p values lie between 0.1 and 0.2. Overall, this is a pattern of results that is unlikely to have arisen due to chance.

The results are not simply explained by proportional differences in how effectively the program increases the use of safe drinking water under the different decision-making processes: the RCP process yields treatment effects that are three times higher in Gopalganj compared to Matlab. These effects are 2.1 and 1.6 times as high under the TD and CP processes, respectively.

We interpret these results as implying that differences in the performance of the decision-making processes must be attributable to differences in the contexts, given that our study holds the intervention constant and that the same staff implemented all three processes in each region. One alternative explanation that we cannot rule out is that the different teams in the two regions differed in their relative skill in implementing the decision-making processes. However, this alternative explanation seems unlikely to hold: some staff members left and others joined the project over the two-year implementation period, and there are no systematic trends over time in differences between the decision-making processes (Appendix Figures C1a and C1b).

5.3 Robustness

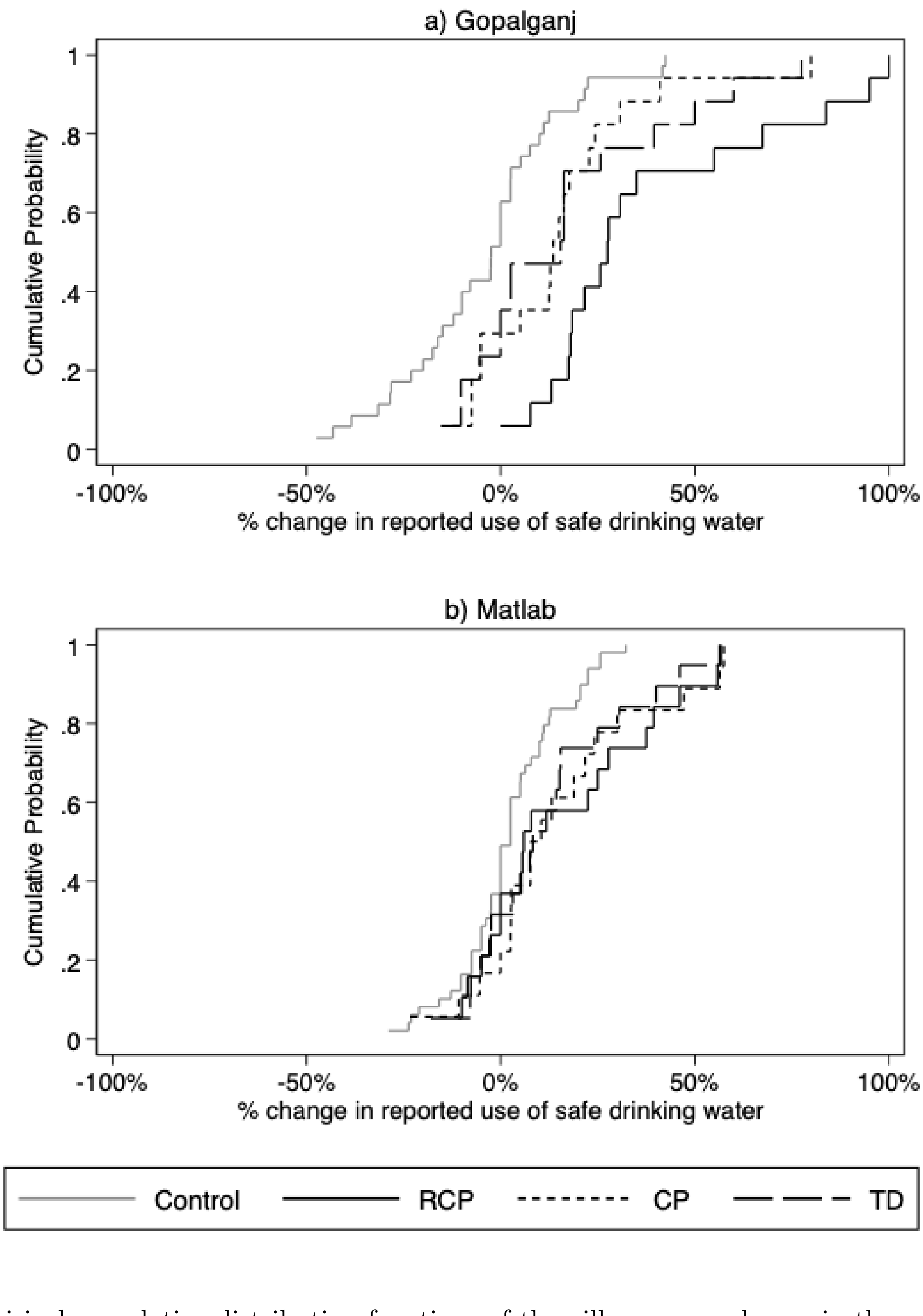
The results are not sensitive to specification choices. The results are very similar if we estimate the same equation using household-level data, weighting these data so that each village counts equally in summary statistics, and clustering standard errors by village, with very small gains in precision (Appendix Table C2). The comparison between decision-making processes is also almost identical if we exclude the control villages, thereby exploiting only the fully experimental variation in the assignment of villages to different decision-making processes (Appendix Table C3). Additionally, as we note above, the results are if anything strengthened by using the full set of control villages instead of the spatially matched control villages (Appendix Table C4).

Table 2: Heterogeneity of program impacts by upazila

	Gopalganj (1)	Matlab (2)		p value
TD	0.21***	0.10**	$\text{TD}_G = \text{TD}_M$	0.238
	(0.08)	(0.05)		
CP	0.19***	0.12**	$\text{CP}_G = \text{CP}_M$	0.433
	(0.07)	(0.06)		
RCP	0.42***	0.14**	$\text{RCP}_G = \text{RCP}_M$	0.005***
	(0.09)	(0.06)		
Average change in control group	-0.04	0.02		
	(0.04)	(0.02)		
RCP vs CP	0.017**	0.886	RCP_G - RCP_M = CP_G - CP_M	0.073*
CP vs TD	0.858	0.748	CP_G - CP_M = TD_G - TD_M	0.732
TD vs RCP	0.037**	0.634	TD_G - TD_M = RCP_G - RCP_M	0.147
TD vs pooled	0.235	0.632	TD_G - pooled_G = TD_M - pooled_M	0.489
CP vs pooled	0.113	0.925	CP_G - pooled_G = CP_M - pooled_M	0.199
RCP vs pooled	0.012**	0.722	RCP_G - pooled_G = RCP_M - pooled_M	0.067*

Note: This table shows the estimated impact on the use of safe drinking water under each of the three decision-making processes in each upazila. Data are collapsed to village-level means, and robust standard errors are in parentheses. Reported values are derived from a fully saturated regression on indicators of treatment under each decision-making process and their interactions with an indicator for Gopalganj. Regressions include 107 villages in which tubewells are feasible and 84 matched control villages. In columns 1 and 2, p values for the following tests are reported: i) the significance of the pairwise difference between the means in each decision-making process pair; and ii) the significance of the difference between means under one decision-making process and for the remainder of the treated villages. *** p <0.01, ** p <0.05, * p <0.1. Source: The authors of this paper.

Figure 5: Change in reported use of safe drinking water



Notes Empirical cumulative distribution functions of the village mean change in the reported use of safe drinking water for villages in which tubewells were feasible and matched control villages in each study region as indicated. Source: The authors of this paper.

6 Which community characteristics best predict differences in impact under different decision-making processes?

The results in the previous section suggest that the relative performance of different decision-making processes varies by context. However, they do not tell us *why* the relative performance varies. Gopalganj and Matlab, as discussed in section 2, differ in many respects. In this section, we make progress on understanding why the relative performance varies by using machine learning to evaluate which community characteristics most strongly correlate with a higher impact under each decision-making process.

6.1 Methodology

We adapt the generic machine learning framework of Chernozhukov et al. (2020) to evaluate which community-level characteristics most strongly correlate with a higher predicted impact. We implement the approach using the following steps, for any two groups, which we designate the reference group (R) and the comparison group (C). Our main objective is to compare impacts across two treatment arms that received the intervention under different decision-making processes, in which case we arbitrarily designate one treatment arm R and the other C. We define D_i so that $D_i = 1$ if $i \in R$ and $D_i = 0$ if $i \in C$. We compare impacts across treatment arms by comparing changes in the use of safe drinking water, denoted by ΔY_v.

Our objective is to understand how impacts vary with a vector of pre-intervention characteristics Z_i, comprising variables that are hypothesized to relate to the success of participatory approaches or collective action. We include measures of the potential impact (baseline access to safe drinking water), group size, heterogeneity, remoteness, education, social cohesion, conflict, leader quality, and poverty. We exclude measures of how well-informed communities are about the arsenic problem, because the information campaign we conducted pre-baseline ensured that all communities were well-informed about the arsenic problem (Madajewicz, Tompsett, and Habib 2021). The full list of all variables in Z_i is given in Appendix Table A. Our objective is to learn about the difference between the conditional average expected outcomes in C and in R, which we designate the Conditional Average Difference in Treatment Effects (CADTE).

We repeat the following steps 100 times:

1. For observations in R and C, we randomly divide the sample (50:50) into training (T) and analysis (A) samples, stratifying by upazila and by treatment status.
2. In T, we use random forests to predict ΔY_v in R and C using the information in Z_i, using the default tuning parameters to reduce the run time.
3. Then, in A, we use the trained models to predict outcomes under R, which we designate by $r(Z_i)$, and under C, which are similarly designated by $c(Z_i)$. The difference between the two predictions is the unit-specific predicted difference in treatment effects between C and R.
4. We then estimate the Best Linear Predictors (BLP) for the CADTE by weighted OLS in A,

i.e.,

$$\begin{aligned}\Delta Y_v = {}_0 &+ \hat{\alpha}_1 r\left(Z_i\right) + \hat{\alpha}_2\left(c\left(Z_i\right) - r\left(Z_i\right)\right) + \hat{\beta}_1\left(D_i - p_i\right) \\ &+ \hat{\beta}_2\left(D_i - p_i\right)\left(\left(c\left(Z_i\right) - r\left(Z_i\right)\right) - \mathbb{E}_{N,A}\left(\left(c\left(Z_i\right) - r\left(Z_i\right)\right)\right)\right) + \hat{\epsilon}_i, i \in A.\end{aligned} \tag{4}$$

The weights $w\left(Z_i\right)$ are set equal to $p_i\left(1-p_i\right)^{-1}$, where p_i is the propensity score for assignment to C, and standard errors are robust. This regression yields an estimate of the average differential treatment effect ($\hat{\beta}_1$) and the best linear prediction for heterogeneity ($\hat{\beta}_2$).

5. We divide A into quartiles based on the unit-specific difference in treatment effects, and we estimate the averages of the CADTE in each quartile by weighted OLS in A, which we designate the Group Average Differential Treatment Effects (GADTES), i.e.,

$$\begin{aligned}Y_i = \hat{\alpha}_0 + &\hat{\alpha}_1 r\left(Z_i\right) + \hat{\alpha}_2\left(c\left(Z_i\right) - r\left(Z_i\right)\right) + \\ &\sum_{k=1}^{4} \hat{\gamma}_k\left(D_i - p_i\right)\mathbb{1}\left(c\left(Z_i\right) - r\left(Z_i\right) \in I_k\right) + \hat{\nu}_i, i \in A,\end{aligned} \tag{5}$$

where $I_k = \left[l_{k-1}, l_k\right)$ and l_k is the kth quartile of $\left(c\left(Z_i\right) - r\left(Z_i\right)\right)_{i \in A}$. The weights and standard errors are as above.

6. We conduct a classification analysis (CLAN) in which we describe the average characteristics in the most and least affected quartiles. We estimate the CLAN parameters in A by

$$\hat{\delta_k} = \mathbb{E}_{N,A}\left(Z_i | c\left(Z_i\right) - r\left(Z_i\right) \in I_k\right), \tag{6}$$

where the weights and standard errors are as above.

7. For any coefficients of interest, we record the coefficient, the p value, and confidence interval.

We then take the median value across the 100 iterations of any coefficients of interest. We also construct confidence intervals and p values using the median values, but we double the p value and use 95% confidence intervals to construct 90% confidence intervals to account for the uncertainty generated by splitting the sample. We monotonize the GADTES parameters and their confidence intervals following Chernozhukov, Fernández-Val, and Galichon (2009).

6.2 Results

6.2.1 Heterogeneity in average treatment effect

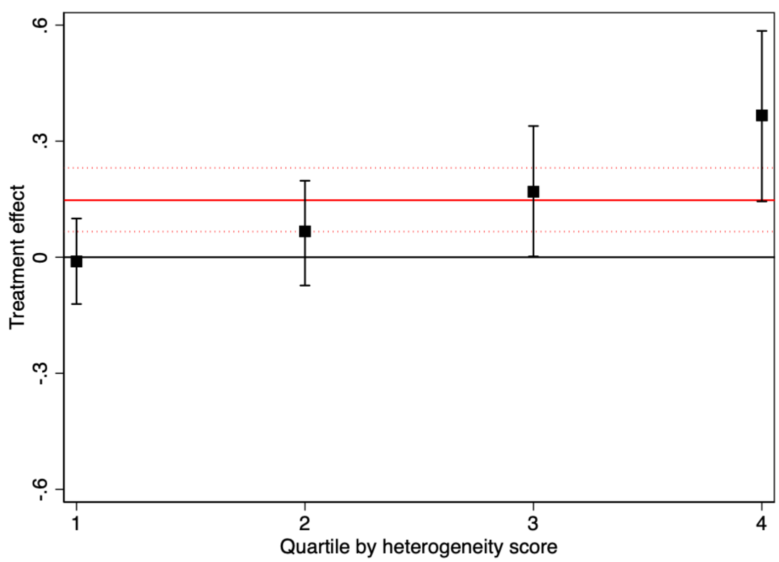
We first apply the framework to learn about heterogeneity in the average treatment effect of the program. We define the reference group as the control group and the comparison group as the treated group, pooled across all three treatment arms. Note that the weighting scheme implicitly assigns zero weight to the villages that were not randomly assigned to the treatment group in Matlab.

The results suggest that there is considerable heterogeneity in the treatment effects of the program when they are averaged across all three decision-making processes. Figure 6 illustrates this. In the least affected quartile, there is essentially no effect of the program, while in the most affected quartile, the treatment effect—a 38 percentage point rise in the use of safe drinking water—is almost 2.5 times as large as the average treatment effect (15 percentage points). Statistical tests confirm that we can strongly reject the null hypothesis of no heterogeneity.[7]

The classification analysis reveals that villages with higher predicted average treatment effects of the program as a whole differ from those with lower predicted average treatment effects in many respects. Among other differences, they have lower baseline access to safe drinking water; they are more remote and poorer; they are more fractionalized in religious terms; they have a higher share of households that are related to one another; and they are less cooperative and more conflict-prone.[8]

These types of difference may explain why treatment effects are larger in Gopalganj than in Matlab for all three treatment arms. These results also illustrate the importance of accounting for overall variation in the treatment effects of a program when the goal is to understand the heterogeneity of the effects of a particular feature of project design. In the remainder of this section, we search for characteristics that differentially predict the impact under different decision-making processes.

Figure 6: Heterogeneity in average treatment effect



Notes The estimated Group Average Treatment Effect for quartiles defined by the heterogeneity score. The x axis shows quartiles defined by the predicted heterogeneity score. The y axis shows the estimated group average treatment effect. The average treatment effect is shown in red. Source: The authors of this paper.

[7]See the full results in Appendix Tables C5 and C6.
[8]See the full results in Appendix Table C7.

6.2.2 Heterogeneity in effects of decision-making processes

We then apply the same framework to learn about heterogeneity in the treatment effects under each decision-making process. For each pair of decision-making approaches, we repeat the same process, treating one (arbitrarily designated) decision-making process as the reference group and the second as the comparison group.

Figure 7 plots the group average differential treatment effects for each pair of decision-making processes. Compared with Figure 6, it is immediately clear that we have less power to detect any heterogeneity in the effects of the decision-making processes: the differences in the impact between decision-making processes are generally smaller than the impacts themselves, and the confidence intervals on the estimates are wider because we have fewer observations. Formally, we are unable to reject the null hypothesis that there is no heterogeneity in the differential treatment effects with respect to the set of baseline characteristics that we can measure for any of the three pairs of decision-making processes (Appendix Tables C8 and C9).[9] There are several potential explanations. The results could reflect limited heterogeneity, although the results in section 5 do not support this interpretation. The results could reflect limited statistical power to detect heterogeneity. The factors that predict heterogeneity may simply be difficult to measure. For instance, if there are few appropriate sites for a community well in a village because suitable land is scarce, the characteristics of the relevant landowners may be very important for shaping the impact of the program. This type of variation would not be reflected in our village-level summary statistics.

The results do show some interesting patterns. Figure 7a suggests that although the RCP process always outperforms the CP process, the difference in performance is negligible for the least differentially affected quartile. The difference in performance between the CP and TD processes is generally small, and the results suggest that the difference may take different signs, ranging from positive to negative. The RCP process always outperforms the TD process and heterogeneity appears to be very limited, at least with respect to the baseline characteristics we consider.

For the differences between the RCP and CP processes, we can identify the characteristics that most strongly differ between the least affected quartile—the group for which outcomes are most similar between the two decision-making processes—and the most affected quartile—the group for which the RCP process results in the greatest increase in impact (Table 7a). The characteristics that predict a greater impact under the RCP process compared to the CP process fall into two main groups. The first group includes characteristics that correlate with having "more at stake," specifically, lower access to safe drinking water at the baseline. The second group includes characteristics that correlate with lower community cohesion. One might be tempted to conclude that this second set of characteristics is important because the RCP restrictions are not necessary to obtain good outcomes when communities are more cohesive. However, less cohesive communities have lower average treatment effects over all decision-making processes, making this explanation less likely (Appendix Table C7). Another possibility is that more cohesive communities had been

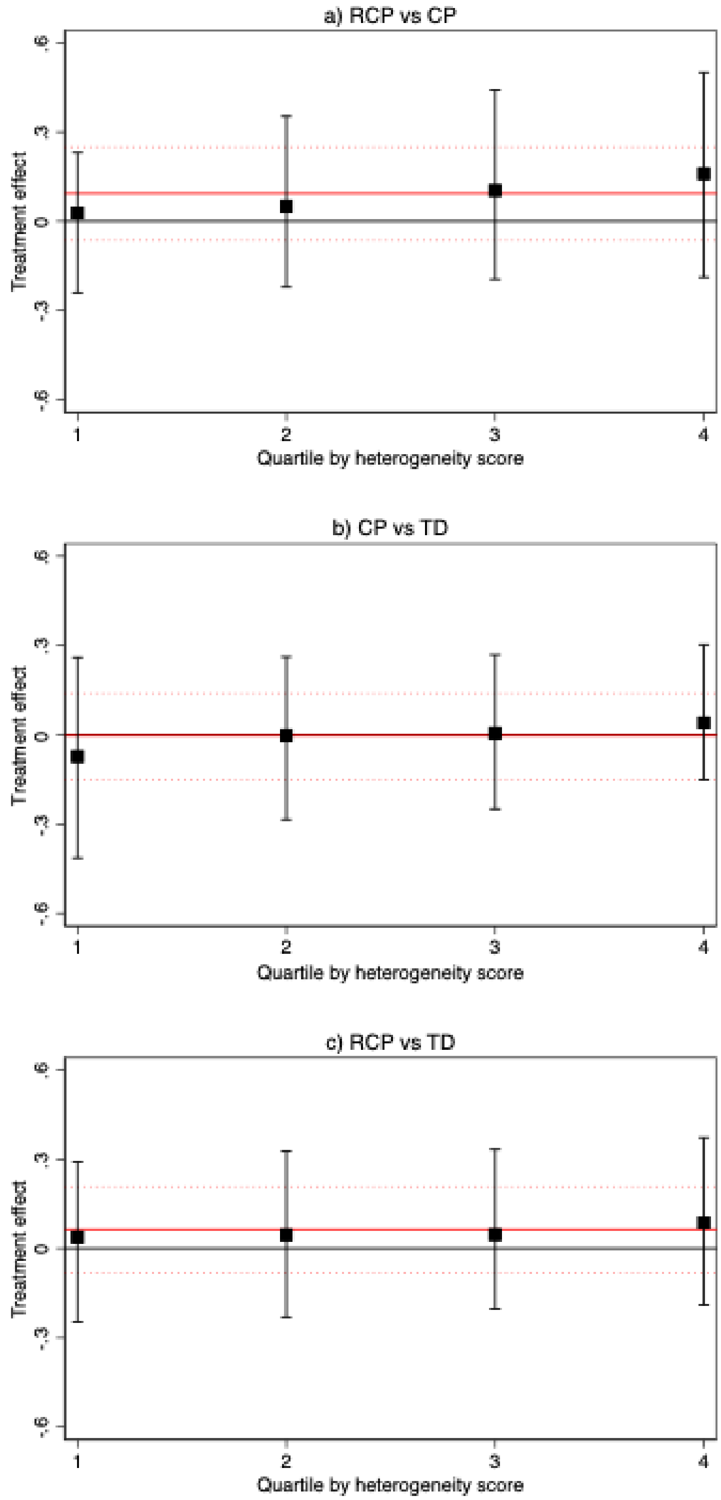
[9]Note that the point estimates and confidence intervals are monotonized in Figure 7 but not in Appendix Table C9. The only affected comparison is the comparison of the RCP and TD processes.

more successful at solving the collective action problem before our intervention, rendering potential impacts smaller in these communities. While not all differences fall neatly into one of those two categories, those that do not tend to have higher p values and may simply arise due to chance.

For the differences between the CP and TD processes, we repeat the exercise (Table 7b). At the risk of ex post reasoning, all the differences we observe can be rationalized through the lens of an information-elite capture trade-off, in which the TD model reduces the risk of elite capture but results in worse-informed decisions. The TD process may thus be at a disadvantage when many people have already switched to safe wells because this renders local information about outstanding pockets of low access to safe water more important. The CP process may be at a disadvantage when the risk of elite capture is higher, which may be captured here by the presence of fewer women's groups and the presence of a large group of related individuals.

Given that we detect very little heterogeneity for the comparison between the RCP and TD processes, differences between the least and most affected quartiles are hard to interpret and may arise due to chance (Appendix Table C10).

Figure 7: Heterogeneity in impacts of decision-making processes



Notes The estimated Group Average Differential Treatment Effect for quartiles defined by the heterogeneity score. The x axis shows quartiles defined by the predicted heterogeneity score. The y axis shows the estimated differential effect of treatment under the decision-making process listed first compared to the decision-making process listed second. The average differential treatment effect is shown in red. Source: The authors of this paper.

Table 3a: Characteristics that correlate with a greater impact of the RCP process vs the CP process

	25% Least	25% Most	Difference
Baseline reported use of safe drinking water	0.65 (0.42, 0.87)	0.20 (0.06, 0.36)	−0.46 (-0.71, -0.19) [0.002]
Share switching from As-contaminated wells, last 5 years	0.58 (0.36, 0.81)	0.18 (0.04, 0.35)	−0.41 (-0.67, -0.14) [0.007]
Distance to safe drinking water (minutes)	6.99 (3.99, 10.14)	20.32 (12.22, 27.77)	12.92 (5.03, 21.08) [0.006]
Distance to market (hours)	18.22 (11.63, 25.74)	38.67 (21.07, 56.20)	19.44 (0.94, 38.78) [0.079]
Average share of village related	0.05 (0.03, 0.07)	0.14 (0.07, 0.21)	0.09 (0.01, 0.17) [0.050]
Mean no. of community associations participated in	1.17 (1.03, 1.32)	0.91 (0.69, 1.09)	−0.28 (-0.53, -0.03) [0.051]
Mean no. of collective actions, total	1.41 (0.93, 1.83)	0.47 (-0.00, 1.05)	−0.98 (-1.63, -0.30) [0.012]
Mean no. of collective actions participated in	0.69 (0.41, 0.96)	0.23 (-0.01, 0.51)	−0.43 (-0.86, -0.02) [0.040]
Share participating in community decision-making	0.74 (0.62, 0.87)	0.50 (0.35, 0.64)	−0.26 (-0.45, -0.06) [0.018]
Any conflicts	0.22 (-0.08, 0.52)	0.67 (0.33, 1.01)	0.56 (0.14, 0.91) [0.021]
Leader selection is wealth-based	0.89 (0.66, 1.12)	0.44 (0.09, 0.80)	−0.44 (-0.80, -0.02) [0.080]
Mean share of leaders seeking input from many	0.65 (0.59, 0.70)	0.76 (0.69, 0.82)	0.11 (0.01, 0.20) [0.061]
Share poor or very poor (self-reported)	0.31 (0.21, 0.43)	0.44 (0.33, 0.55)	0.14 (-0.01, 0.28) [0.081]

Note: This table shows the mean value of the listed characteristic for the quartiles with i) the smallest (or most negative) differential treatment effect of the decision-making process listed first compared to the decision-making process listed second and ii) the greatest (or most positive) differential treatment effect of the decision-making process listed first compared to the decision-making process listed second, and the difference in mean values between the two quartiles. Only characteristics for which the least and most affected quartiles differ significantly ($p < 0.1$) are shown. Source: The authors of this paper.

Table 3b: Characteristics that correlate with a greater impact of the CP process vs the TD process

	25% Least	25% Most	Difference
Share switching from As-contaminated wells, last 5 years	0.24	0.54	0.31
	(0.06, 0.42)	(0.30, 0.77)	(0.00, 0.59)
			[0.087]
Average share of village related	0.11	0.05	-0.06
	(0.06, 0.17)	(0.03, 0.07)	(-0.12, -0.01)
			[0.058]
No. of women's groups	3.44	8.78	5.00
	(1.54, 5.37)	(4.28, 13.16)	(0.70, 10.03)
			[0.042]

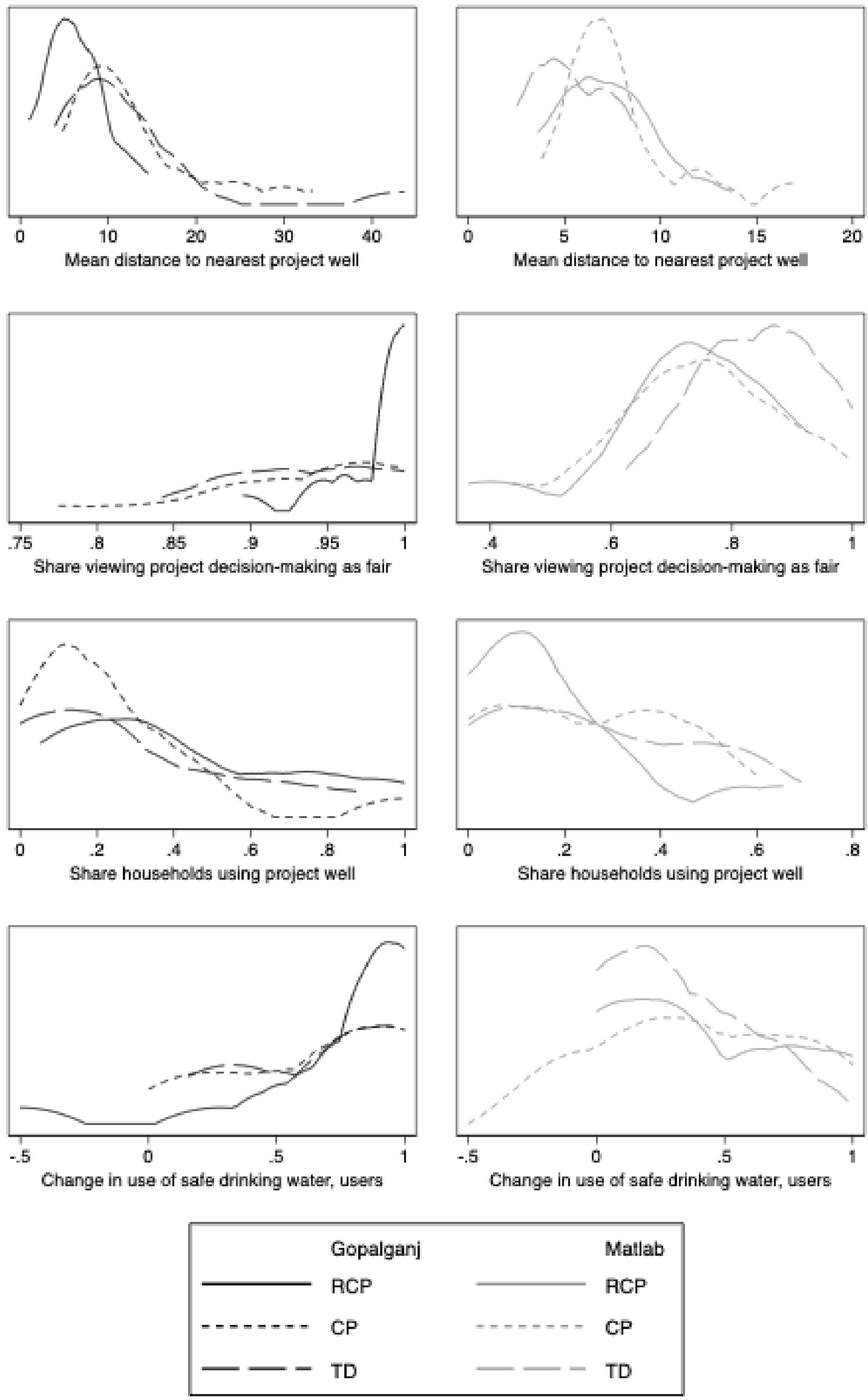
Note: This table shows the mean value of the listed characteristic for the quartiles with i) the smallest (or most negative) differential treatment effect of the decision-making process listed first compared to the decision-making process listed second and ii) the greatest (or most positive) differential treatment effect of the decision-making process listed first compared to the decision-making process listed second, and the difference in mean values between the two quartiles. Only characteristics for which the least and most affected quartiles differ significantly ($p < 0.1$) are shown. Source: The authors of this paper.

7 What characterizes successful projects under each decision-making process?

The last stage of our analysis examines potential mechanisms through which the different decision-making processes affect the program's impact. Madajewicz, Tompsett, and Habib (2021) report three possible explanations for why the RCP process outperforms the other two decision-making processes on average. First, the RCP process on average results in decisions that place the wells closer to more households. Second, communities may negotiate broader access to new and existing safe water sources under the RCP process. A third possibility is that being involved in decision-making later induces more households to switch to safe wells, although this last hypothesis is difficult to test empirically. The differences in performance are not driven by different numbers of wells being installed: communities install similar numbers of wells under all three decision-making processes in both regions. The results suggest a trade-off between information and elite capture: the participatory approaches make better use of information than the TD approach, meaning that the TP provider has to resort to inefficient strategies—like placing wells on public land even when the public land is inconvenient to access—because they lack the local information that would allow them to more efficiently avoid elite capture. However, the CP process is more vulnerable to elite capture.

Figure 8 shows that these average differences are not consistently observed in the two regions. The RCP process yields well locations that are closer to a greater number of households only in Gopalganj. In the smaller communities in Matlab, this advantage disappears. The RCP process also only has a clear advantage in terms of producing decisions that most perceive as fair in Gopalganj. In Matlab, both participatory approaches result in decisions that are perceived as less fair than the top-down approach. The RCP process also only results in higher use rates and better targeting of households who previously lacked access to safe water in Gopalganj. In Matlab, the RCP process results in the lowest use rates and the most limited targeting. The reason that the RCP process still performs the best in Matlab is that communities apparently renegotiate access to existing safe or other new water sources, resulting in increases in the use of safe drinking water among those who do not use project wells.

Figure 8: Mechanisms



Notes Empirical densities of key variables that highlight heterogeneous mechanisms in Gopalganj and Matlab. Source: The authors of this paper.

7.1 Methodology

To further investigate the role of different mechanisms under the different decision-making processes, we adapt the methodology we describe in subsection 6.1 to identify whether variables that reflect different potential mechanisms correlate with changes in the use of safe drinking water, and if so, to determine which mechanisms most strongly predict changes in the use of safe drinking water. We cannot simply apply the methodology from subsection 6.1 directly, because the mechanisms we study are themselves secondary outcomes. We cannot assume that communities with the same secondary outcomes—for example, the same share of households reporting decisions as fair or the same mean distance to a project well—are comparable across treatment arms. Figure 8 suggests that they almost certainly are not. We thus study heterogeneity separately within each treatment arm in this section.

We compare how changes in the use of safe drinking water, ΔY_v, correlate with M_i, where M_i comprises variables that measure different potential mechanisms in five categories: the number of wells installed, location of wells, contributions, self-reported project evaluation, and use of wells. The variables in M_i are those that Madajewicz, Tompsett, and Habib (2021) consider when investigating the mechanisms through which the differences in average treatment effects arise. Appendix Table B provides a full list of these variables. Our objective then is to learn about the Conditional Average Value (CAV) of ΔY_v in the treatment arm of interest.

We repeat steps similar to steps 1–7 from subsection 6.1 100 times, with the following differences. We only train one random forest model in T, to predict outcomes that we designate $y\left(M_i\right)$. We then use the model to predict the outcomes in the analysis sample A and estimate a Best Linear Predictor (BLP) for the CAV by weighted OLS, also in A:

$$Y_i = \hat{\alpha} + \hat{\beta}\left(y\left(M_i\right) - \mathbb{E}_{N,A}\left(y\left(M_i\right)\right)\right) + \hat{\epsilon}_i, i \in A. \tag{7}$$

We estimate the sorted Group Average Values (GAVs), which are averages of the CAV by quartile ranked by $y\left(M\right)$, using weighted OLS in A, i.e.,

$$Y_i = \sum_{k=1}^{4} \hat{\gamma}_k \mathbb{1}\left(y\left(M_i\right) \in I_k\right) + \hat{\nu}_i, i \in A, \tag{8}$$

where $I_k = [l_{k-1}, l_k)$ and l_k is the kth quartile of $\left(y\left(M_i\right)\right)_{i \in A}$.

We conduct a classification analysis (CLAN) in which we describe the average characteristics in the quartiles with the smallest and largest changes in the use of safe drinking water. We estimate the CLAN parameters in A by

$$\hat{\delta}_k = \mathbb{E}_{N,A}\left(M_i | y\left(M_i\right) \in I_k\right). \tag{9}$$

We implicitly interpret changes in the use of safe drinking water as a proxy for the program's impact. Changes in the use of safe drinking water are a valid proxy for the program's impact if changes in the use of safe drinking water in the absence of the program have an expected value equal to zero, which seems to be a reasonable assumption in both Gopalganj and Matlab (Figure 5).

7.2 Results

Figure 9 shows the predicted changes in the use of safe drinking water in quartiles defined by the heterogeneity score. We estimate significant heterogeneity under the RCP process, a modest amount of heterogeneity under the TD process, and very little heterogeneity under the CP process. Statistically, we reject the null hypothesis that there is no heterogeneity with respect to the mechanism variables we consider only for the RCP process. Again, detecting little heterogeneity can mean that there is limited heterogeneity but it can also mean that the variables we use to describe mechanisms are not good predictors of impact or that we have limited statistical power to detect heterogeneity.[10]

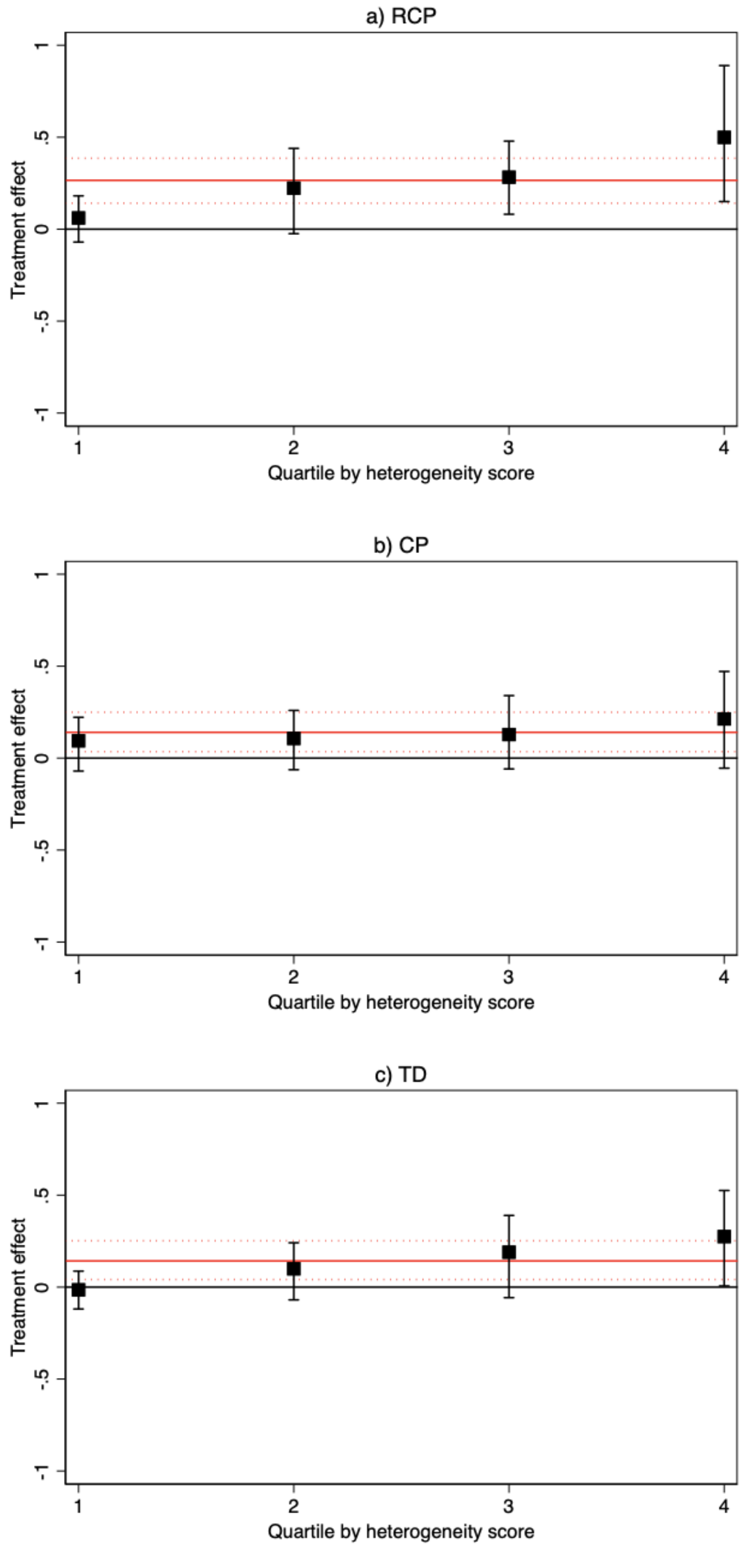
For the RCP process, the use of safe drinking water in the most impacted quartile increases by 50 percentage points but by only 6 percentage points in the least impacted quartile. The mechanisms that are most strongly associated with the magnitude of the change in the use of safe drinking water fall into two categories (Table 4a). First, larger changes in use are associated with greater reductions in the distance to the nearest safe well and greater changes in the use of safe drinking water among well adopters. One potential interpretation of these differences is that the RCP process results in better targeting when it is most effective. An alternative interpretation is that these differences reflect the differences we observed in the previous section with respect to lower baseline access to safe drinking water ("more at stake"). Second, larger changes in the use of safe drinking water are associated with better decisions and lower levels of elite capture: households report that they agree with decisions and that decision-making was fair at higher rates in higher-impact quartiles, and they are less likely to report that decisions favored influential people or that wells on private land were hard to access. This latter group of differences suggests that, when the RCP process works well, it yields decisions that are fair and equitable, and vice versa.

For the TD process, greater changes in the use of safe drinking water among well adopters are also associated with greater changes in the use of safe drinking water in the community more generally (Table 4b). As for the RCP process, this pattern has several potential interpretations. One interpretation that is consistent with other results is that this pattern reflects the problems with information under the TD process. When the TD process works poorly, it does not apparently successfully target those most in need, possibly as a consequence of the inefficient strategies to which field staff resort in the absence of other effective ways to reduce elite capture, such as locating the well on public land even when this results in the well being inconvenient to access or distant from those most in need of access to safe drinking water.

For the CP process, for which we detected little heterogeneity, the classification analysis does not identify any mechanisms that are associated with predicted changes in use.

[10] Appendix Tables C11 and C12 show the full results.

Figure 9: Heterogeneity in impact by decision-making process



Notes Group average change in the use of safe drinking water as predicted by mechanism variables for quartiles defined by the heterogeneity score. The x axis shows quartiles defined by the predicted heterogeneity score. The y axis shows the estimated group average change. The average change in the treatment arm is shown in red. Source: The authors of this paper.

Table 4a: Mechanism characteristics that correlate with a greater impact of the RCP process

	25% Least	25% Most	Difference
Mean change in distance to nearest safe well	-1.06 (-2.75, 0.57)	-22.05 (-36.14, -6.21)	-21.06 (-35.79, -5.04) [0.025]
Share agreeing with project decisions	0.66 (0.50, 0.80)	0.88 (0.71, 1.01)	0.22 (0.00, 0.44) [0.093]
Share viewing project decision-making as fair	0.74 (0.52, 0.93)	0.98 (0.95, 1.01)	0.21 (0.03, 0.46) [0.051]
Share reporting decisions favored influential people	0.22 (0.06, 0.41)	0.03 (-0.01, 0.08)	-0.19 (-0.39, -0.02) [0.059]
Share reporting wells hard to access, as they are on private land	0.13 (0.05, 0.21)	0.03 (-0.00, 0.07)	-0.09 (-0.19, -0.01) [0.072]
Change in use of safe drinking water, users	0.25 (-0.03, 0.59)	0.91 (0.67, 1.03)	0.66 (0.30, 1.00) [0.002]

Table 4b: Mechanism characteristics that correlate with a greater impact of the TD process

	25% Least	25% Most	Difference
Change in use of safe drinking water, users	0.27 (-0.03, 0.59)	0.80 (0.55, 1.02)	0.54 (0.13, 0.91) [0.029]

Note: These tables show the mean value of the listed characteristic for the quartiles with i) the smallest (or most negative) predicted change in the use of safe drinking water and ii) the greatest (or most positive) predicted change in the use of safe drinking water, and the difference in the mean values between the two quartiles. Only characteristics for which the least and most affected quartiles differ significantly ($p < 0.1$) are shown. Source: The authors of this paper.

8 Conclusion

This paper investigates heterogeneity in the effects of community participation in decision-making in projects to provide safe drinking water in two regions of rural Bangladesh. We randomly assigned villages to receive otherwise identical interventions under three different decision-making processes. A deliberative, consensus-based approach to community decision-making increased the average project impact compared to the two other decision-making processes we study—top-down decision-making and community decision-making without restrictions—but the effects of the consensus-based approach are strongly heterogeneous. In one study region, the consensus-based process doubled the project's impact relative to either of the other processes, while in the second region, all three processes yielded very similar impacts. Formally, we can reject the null hypothesis that the difference in performance between the consensus-based process and the other two processes is the same in the two regions. The same field staff implemented the same intervention in each region under all three decision-making processes, implying that the differences in impact are driven by contextual differences between the two regions.

We then use machine learning to better understand the potential drivers of these heterogeneous impacts. We consider a wide range of contextual factors that previous studies have hypothesized might affect the success of participatory decision-making or, more generally, collective action. We also investigate the characteristics of successful projects under each decision-making process. Our results suggest that the consensus-based approach yields better outcomes when there is more at stake: specifically, when access to safe drinking water is lower before the intervention, implying that the potential gains from the intervention are larger. When the consensus-based process works best, it does so by yielding better, fairer decisions that are less prone to elite capture. When the consensus-based process does not result in large changes in the use of safe drinking water, households are more likely to report that decisions are unfair and that elites capture the process and benefits. We interpret these results as suggesting that it is costly for communities to achieve better decisions and outcomes under the consensus-based process, an insight we formalize in a simple model. When less is at stake, communities may be unwilling to incur these costs, and decision-making may default to local norms, which are typically elite-controlled and non-inclusive.

Our results shed light on a central puzzle in the literature on participation in decision-making: why are the effects of participation so heterogeneous? Reflecting the results in the literature more generally, our study nests both a case in which the consensus-based approach was highly successful and one in which it made little difference to the impact. Our study design holds the intervention and the implementers constant, allowing us to attribute these differences in impact to contextual differences between the two regions. Our results thus suggest that context may be an important explanation for heterogeneity in the impacts of participatory decision-making.

The central limitation of our study is its relatively small sample size. The machine learning approaches that we use are most effective with larger samples: in our context, we detect heterogeneity in the average impacts of the intervention, pooled across all treatment arms and both study regions, with much greater precision than heterogeneity in the differences between impacts under different treatment arms. A potential avenue for future research is to homogenize and pool data across many studies to investigate the same questions in much larger samples, although this approach is not without its difficulties, not the least of which is resolving how to account for differences in approaches to decision-making and interventions across different studies, which this study holds constant by design. Our study also focuses exclusively on heterogeneity as a function of community characteristics. Future research might adapt our approach to evaluate the heterogeneity of impacts at the household level, for example, to understand how different decision-making processes affect outcomes for the poor or other potentially marginalized groups of particular interest to policy-makers.

The potential benefits of inclusive participatory approaches to decision-making must be compared to the costs they impose on implementing organizations and beneficiary communities. The costs to the implementing organization of the consensus-based process and the top-down process are similar, while the unregulated community decision-making process saves at most one day of staff time, equivalent to a 20% decrease in overhead costs (Madajewicz, Tompsett, and Habib 2021). Time costs for communities are somewhat higher for the consensus-based approach. Project staff recorded that on average, 33 households attended project meetings under the consensus-based approach to decision-making, with meetings taking approximately 90 minutes. The top-down and unregulated processes also involved community meetings, which were to provide information only and were thus somewhat shorter and attended by slightly fewer people. Communities may also incur less tangible social costs under the consensus-based process in terms of stress and strain on community relationships. Our model suggests, however, that communities only incur these social costs if they actively engage with the consensus-based decision-making process and that they only do this when enough is at stake to make incurring these costs worthwhile. Communities choose the consensus-based approach over the other two approaches when offered a systematic choice between them, suggesting that the costs in general do not outweigh the benefits for communities (Cocciolo, Habib, and Tompsett 2019b).

The consensus-based decision-making process always appears to perform better than or equivalently to the other decision-making processes in the context we study, and our evidence suggests that its costs to communities and implementing organizations are limited. Our results do not, however, provide unequivocal support for inclusive participatory approaches to decision-making. Our analysis implies that there will be contexts in which any kind of participatory decision-making performs poorly and that one characteristic of such contexts is that less is at stake. Our results suggest that implementing organizations might consider limiting the use of inclusive participatory approaches to decision-making to problems that are especially urgent or salient to beneficiary communities.

References

Ahuja, Amrita, Michael Kremer, and Alix Peterson Zwane. 2010. Providing Safe Water: Evidence from Randomized Evaluations. *Annual Review of Resource Economics* 2.

Alatas, Vivi et al. 2012. Targeting the Poor: Evidence from a Field Experiment in Indonesia. *American Economic Review* 102(4).

Alatas, Vivi et al. 2019. Does Elite Capture Matter? Local Elites and Targeted Welfare Programs in Indonesia. *AEA Papers and Proceedings* 109.

Alesina, Alberto, Reza Baqir, and William Easterly. 1999. Public Goods and Ethnic Divisions. *The Quarterly Journal of Economics* 114(4).

Allakulov, Umrbek et al. 2023. Transparency, Governance, and Water and Sanitation: Experimental Evidence from Schools in Rural Bangladesh. *Journal of Development Economics* 163.

Arkedis, Jean et al. 2021. Can Transparency and Accountability Programs Improve Health? Experimental Evidence from Indonesia and Tanzania. *World Development* 142.

Baland, Jean-Marie and Jean-Philippe Platteau. 2018. Collective Action on the Commons: The role of inequality. *Inequality, Cooperation, and Environmental Sustainability*. Ed. by Jean-Marie Baland, Pranab Bardhan, and Samuel Bowles. Princeton University Press. Chap. 2.

Banerjee, Abhijit, L. Iyer, and Rohini Somanathan. 2008. Public Action for Public Goods. *Handbook of Development Economics* 4(49).

Banerjee, Abhijit and Rohini Somanathan. 2007. The Political Economy of Public Goods: Some Evidence from India. *Journal of Development Economics* 82.

Banerjee, Abhijit et al. 2010. Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India. *American Economic Journal: Economic Policy* 2(1).

Barr, Abigail et al. 2012. Information and Collective Action in Community-Based Monitoring of Schools: Field and Lab Experimental Evidence from Uganda. Working Paper.

BenYishay, Ariel et al. 2022. Testing Multi-Stakeholder Dialogue for Better Local Governance in Niger: An Experiment. Can We Talk Our Way Out of Development Problems? *Development Policy Review* 40.

Björkman, Martina and Jakob Svensson. 2009. Power to the People: Evidence from a Randomized Field Experiment on Community-Based Monitoring in Uganda. *Quarterly Journal of Economics* 124(2).

— 2010. When is Community-Based Monitoring Effective? Evidence from a Randomized Experiment in Primary Health in Uganda. *Journal of the European Economic Association* 8(2-3).

Casey, Katherine, Rachel Glennerster, and Edward Miguel. 2012. Reshaping Institutions: Evidence on Aid Impacts Using a Preanalysis Plan. *Quarterly Journal of Economics* 127(4).

Chernozhukov, Victor, Ivan Fernández-Val, and Alfred Galichon. 2009. Improving Point and Interval Estimates of Monotone Functions by Rearrangement. *Biometrika* 36(3).

Chernozhukov, Victor et al. 2020. Generic Machine Learning Inference on Heterogenous Treatment Effects in Randomized Experiments. arXiv preprint 1712.04802.

Christensen, D. et al. 2021. Building Resilient Health systems: Experimental Evidence from Sierra Leone and the 2014 Ebola Outbreak. *Quarterly Journal of Economics* 136(2).

Cocciolo, Serena, Ahasan Habib, and Anna Tompsett. 2019a. Group Size and Collective Action: Evidence from Bangladesh. Working Paper.

Cocciolo, Serena, Md. Ahasan Habib, and Anna Tompsett. 2019b. Understanding Institutional Persistence: Exposure to Democratic Decision-Making and the Value of Autonomy and Democracy. Pre-registered Report and Pre-analysis Plan.

Cocciolo, Serena et al. 2021. Do Community Water Sources Provide Safe Drinking Water? Evidence from a Randomized Experiment in Rural Bangladesh. *World Bank Economic Review* 35(4).

Cocciolo, Serena et al. 2023. Time Is Not Money: An Experiment With Community Contribution Requirements in Cash and Labour. Working Paper.

Fearon, James D., Macartan Humphreys, and Jeremy M. Weinstein. 2015. How Does Development Assistance Affect Collective Action Capacity? Results from a Field Experiment in Post-Conflict Liberia. *American Political Science Review* 109(3).

Galasso, Emanuela and Martin Ravallion. 2005. Decentralized Targeting of an Antipoverty Program. *Journal of Public Economics* 89(4), pp. 705–727.

Humphreys, Macartan, Raul Sanchez de la Sierra, and Peter van der Windt. 2019. Exporting Democratic Practices: Evidence from a Village Governance Intervention in Eastern Congo. *Journal of Development Economics* 140.

Johansson, Per-Olov. 1992. Altruism in Cost-Benefit Analysis. *Environmental and Resource Economics* 2.

Khemani, S. 2007. Can Information Campaigns Overcome Political Obstacles to Serving the Poor? *The Politics of Service Delivery in Democracies: Better Access for the Poor*. Ed. by Shantayanan Devarajan and Ingrid Widlund. EGDI Secretariat. Chap. 5.

Khwaja, Asim Ijaz. 2009. Can Good Projects Succeed in Bad Communities? *Journal of Public Economics* 93.

Long, J. Scott and Laurie H. Ervin. 2000. Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model. *The American Statistician* 54(3).

Madajewicz, Malgosia, Anna Tompsett, and Ahasan Habib. 2021. How Does Delegating Decisions to Communities Affect the Provision and Use of a Public Service? Evidence From a Field Experiment in Bangladesh. *Journal of Development Economics* 150.

Mansuri, G. and V. Rao. 2013. Localizing Development: Does Participation Work? Washington, DC: World Bank.

Miguel, Edward and Mary Kay Gugerty. 2005. Ethnic Diversity, Social Sanctions and Public Goods in Kenya. *Journal of Public Economics* 89(11-12).

Mohanan, Manoj et al. 2020. Information and Facilitation Interventions for Accountability in Health and Nutrition: Evidence from a Randomized Evaluation in India. Economic Research Initiatives at Duke (ERID) Working Paper 295.

Nyqvist, Martina Björkman, Damien de Walque, and Jakob Svensson. 2017. Experimental Evidence on the Long-Run Impact of Community-Based Monitoring. *American Economic Journal: Applied Economics* 9(1).

Olken, Ben A. 2007. Monitoring Corruption: Evidence from a Field Experiment in Indonesia. *Journal of Political Economy* 115(2).

— 2010. Direct Democracy and Local Public Goods: Evidence from a Field Experiment in Indonesia. *American Political Science Review* 104(2).

Olson, M. 1971. The Logic of Collective Action. Harvard University Press.

Pradhan, Menno et al. 2014. Improving Educational Quality through Enhancing Community Participation: Results from a Randomized Field Experiment in Indonesia. *American Economic Journal: Applied Economics* 6(2).

Raffler, Pia, Daniel N. Posner, and Doug Parkerson. 2020. Can Citizen Pressure be Induced to Improve Public Service Provision? Working Paper.

Appendices
For Online Publication

A List of variables included in heterogeneity analysis

- Safe water access:
  - Baseline reported use of safe drinking water
  - Share of wells As-contaminated (BAMWSP)
  - Share knowing someone who became seriously ill / died from As
  - Share switching from As-contaminated wells, last 5 years
  - Depth to safe groundwater unknown
  - Depth to safe groundwater (feet)
  - Distance to safe drinking water (minutes)
  - Distance to collect drinking water (minutes)
  - Share owning own private well
  - No. of users per private well
- Village demographics:
  - No. of households in village
  - No. of in-migrants
  - No. of out-migrants
- Remoteness:
  - Distance to road (minutes)
  - Distance to market (minutes)
  - Distance to school (minutes)
- Heterogeneity/fractionalization:
  - Fraction Muslim
  - Religious fractionalization
  - Minority groups poor
- Education:
  - Fraction literate
- Social cohesion:
  - Average share of village related
  - Mean out-network size
  - Mean no. of community associations, total
  - Mean no. of community associations participated in
  - Mean no. of collective actions, total
  - Mean no. of collective actions participated in
  - No. of women's groups

  - Share participating in community decision-making
  - Community involved in decision-making (FG)
  - Share reporting high community cooperation

- Conflict:

  - Religious conflicts
  - Ethnic conflicts
  - Any conflicts
  - Conflicts frequent
  - Conflicts violent

- Leader quality:

  - Leaders consult community
  - Leader selection is hereditary
  - Leader selection is wealth-based
  - Leader selection is democratic
  - Leaders include women
  - Mean share of leaders serving community very well
  - Mean share of leaders seeking input from many
  - Mean share of leaders very trustworthy
  - Mean leader tenure

- Poverty:

  - Mean value of household assets (BDT)
  - Inequality in assets (GINI)
  - Share poor or very poor (self-reported)
  - Share poor or very poor (enumerator assessment)
  - Share high-income (self-reported)
  - Share high-income (enumerator assessment)
  - Share in lowest asset quintile
  - Share in second-lowest asset quintile
  - Share in middle asset quintile
  - Share in second-highest asset quintile
  - Share in highest asset quintile

B List of variables included in mechanism correlation analysis

- Wells installed:
  - No. of wells installed
  - Proportion of offered wells installed
- Location of wells:
  - Fraction of wells in public places
  - Mean distance to nearest project well
  - Mean change in distance to nearest safe well
  - Mean change in distance to main source
- Contributions:
  - No. of contributing households
  - One contributing household/well
  - Residualized assets (contributors)
  - Residualized leader-connectedness (contributors)
  - Residualized leader kin-connectedness (contributors)
- Self-reported project evaluation:
  - Share agreeing with project decisions
  - Share viewing project decision-making as fair
  - Share reporting decisions favored influential people
  - Share reporting wells hard to access, as they are on private land
  - Share reporting access to wells restricted by landowners
  - Share reporting wells too far away
- Use of wells:
  - Share of households using project well
  - Change in use of safe drinking water, users
  - Change in use of safe drinking water, non-users
  - Change in number of users, other wells
  - Change in number of users, other safe wells

C Additional figures and tables

Figure C1a: Impact for each decision-making process and cycle: Gopalganj

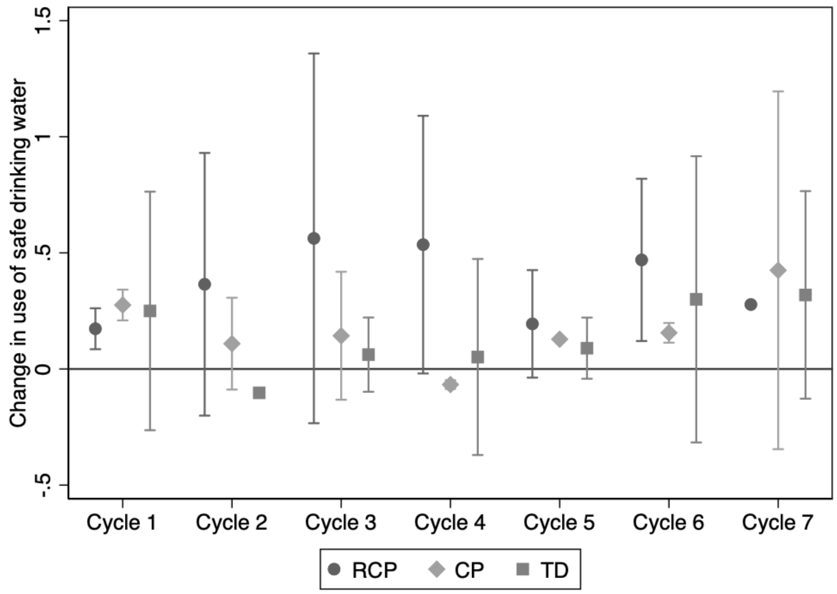
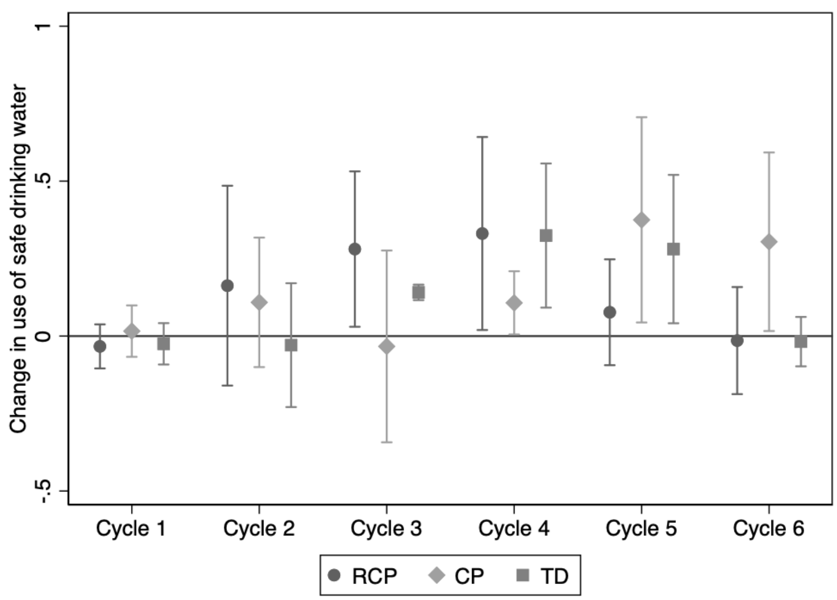


Figure C1b: Impact for each decision-making process and cycle: Matlab



Note: The estimated change in the use of safe drinking water for each decision-making process, cycle, and region. Cycles correspond to groups of geographically clustered villages in which field staff implemented the program sequentially, and thus they share similar characteristics and have similar amounts of time between implementation and endline data collection. 90% confidence intervals are shown.

Table C1: Average program impacts

	(1)
TD	0.15***
	(0.04)
CP	0.16***
	(0.04)
RCP	0.27***
	(0.05)
Average value in control group	−0.01
	(0.02)
RCP vs CP	0.069*
CP vs TD	0.920
TD vs RCP	0.057*
TD vs pooled	0.226
CP vs pooled	0.293
RCP vs pooled	0.034**
N	191

Note: This table shows the estimated change in the use of safe drinking water under each of the three decision-making processes, averaged across upazilas. Data are collapsed to village-level means, and robust standard errors are in parentheses. The reported values are derived from a regression on indicators of treatment under each decision-making process and their interactions with a demeaned indicator for Gopalganj. Regressions include 107 villages in which tubewells are feasible and 84 matched control villages. Asterisks reflect the significance of effects under each decision-making process compared to the control group. p values for the following tests are reported: i) the significance of the pairwise difference between the means in each decision-making process pair; and ii) the significance of the difference between means under one decision-making process and for the remainder of the treated villages. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Source: The authors of this paper.

Table C2: Heterogeneity of program impacts by upazila: robustness test using household-level data

	Gopalganj (1)	Matlab (2)		p value
TD	0.21***	0.10**	$\text{TD}_G = \text{TD}_M$	0.213
	(0.07)	(0.05)		
CP	0.19***	0.13**	$\text{CP}_G = \text{CP}_M$	0.443
	(0.06)	(0.05)		
RCP	0.42***	0.14**	$\text{RCP}_G = \text{RCP}_M$	0.004***
	(0.08)	(0.05)		
Average change in control group	−0.04	0.02		
	(0.04)	(0.02)		
RCP vs CP	0.012**	0.875	RCP_G - RCP_M = CP_G - CP_M	0.061*
CP vs TD	0.807	0.724	CP_G - CP_M = TD_G - TD_M	0.680
TD vs RCP	0.032**	0.601	TD_G - TD_M = RCP_G - RCP_M	0.145
TD vs pooled	0.249	0.603	TD_G - pooled_G = TD_M - pooled_M	0.522
CP vs pooled	0.096*	0.918	CP_G - pooled_G = CP_M - pooled_M	0.176
RCP vs pooled	0.009***	0.698	RCP_G - pooled_G = RCP_M - pooled_M	0.061*

Note: This table shows the estimated change in the use of safe drinking water under each of the three decision-making processes in each upazila. Household-level data are weighted so that each village counts equally in summary statistics. The values are derived from a fully saturated regression on indicators of treatment under each decision-making process and their interactions with an indicator for Gopalganj. Regressions include data from 7,182 households in 107 villages in which tubewells are feasible and 84 matched control villages. In columns 1 and 2, p values for the following tests are reported: i) the significance of the pairwise difference between the means in each decision-making process pair; and ii) the significance of the difference between means under one decision-making process and for the remainder of the treated villages. *** p <0.01, ** p <0.05, * p <0.1. Source: The authors of this paper.

Table C3: Heterogeneity of program impacts by upazila: robustness test using treated villages only

	Gopalganj (1)	Matlab (2)		p value
TD	0.16**	0.12***	$\text{TD}_G = \text{TD}_M$	0.609
	(0.07)	(0.05)		
CP	0.15***	0.15***	$\text{CP}_G = \text{CP}_M$	0.969
	(0.06)	(0.05)		
RCP	0.38***	0.16***	$\text{RCP}_G = \text{RCP}_M$	0.020**
	(0.08)	(0.05)		
RCP vs CP	0.018**	0.886	RCP_G - RCP_M = CP_G - CP_M	0.075*
CP vs TD	0.858	0.748	CP_G - CP_M = TD_G - TD_M	0.732
TD vs RCP	0.038**	0.634	TD_G - TD_M = RCP_G - RCP_M	0.148
TD vs pooled	0.235	0.632	TD_G - pooled_G = TD_M - pooled_M	0.489
CP vs pooled	0.113	0.925	CP_G - pooled_G = CP_M - pooled_M	0.199
RCP vs pooled	0.012**	0.722	RCP_G - pooled_G = RCP_M - pooled_M	0.067*

Note: This table shows the estimated change in the use of safe drinking water under each of the three decision-making processes in each upazila. Data are collapsed to village-level means, and robust standard errors are in parentheses. The reported values are derived from a fully saturated regression on indicators of treatment under each decision-making process and their interactions with an indicator for Gopalganj. Regressions include 107 villages in which tubewells are feasible. In columns 1 and 2, p values for the following tests are reported: i) the significance of the pairwise difference between the means in each decision-making process pair; and ii) the significance of the difference between means under one decision-making process and for the remainder of the treated villages. *** p <0.01, ** p <0.05, * p <0.1. Source: The authors of this paper.

Table C4: Heterogeneity of program impacts by upazila: robustness test using full control group

	Gopalganj (1)	Matlab (2)		p value
TD	0.21***	0.10**	$\text{TD}_G = \text{TD}_M$	0.238
	(0.08)	(0.05)		
CP	0.19***	0.12**	$\text{CP}_G = \text{CP}_M$	0.433
	(0.07)	(0.06)		
RCP	0.42***	0.14**	$\text{RCP}_G = \text{RCP}_M$	0.005***
	(0.09)	(0.06)		
Average change in control group	−0.04	0.02		
	(0.04)	(0.02)		
RCP vs CP	0.017**	0.886	RCP_G - RCP_M = CP_G - CP_M	0.073*
CP vs TD	0.858	0.748	CP_G - CP_M = TD_G - TD_M	0.732
TD vs RCP	0.037**	0.634	TD_G - TD_M = RCP_G - RCP_M	0.147
TD vs pooled	0.235	0.632	TD_G - pooled_G = TD_M - pooled_M	0.489
CP vs pooled	0.113	0.925	CP_G - pooled_G = CP_M - pooled_M	0.199
RCP vs pooled	0.012**	0.722	RCP_G - pooled_G = RCP_M - pooled_M	0.067*

Note: This table shows the estimated change in the use of safe drinking water under each of the three decision-making processes in each upazila. Data are collapsed to village-level means, and robust standard errors are in parentheses. The reported values are derived from a fully saturated regression on indicators of treatment under each decision-making process and their interactions with an indicator for Gopalganj. Regressions include 107 villages in which tubewells are feasible and 99 control villages. In columns 1 and 2, p values for the following tests are reported: i) the significance of the pairwise difference between the means in each decision-making process pair; and ii) the significance of the difference between means under one decision-making process and for the remainder of the treated villages. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Source: The authors of this paper.

Table C5: Heterogeneity in average treatment effects: best linear prediction

Treated vs control	
ATE	HET
0.15	1.29
(0.07, 0.23)	(0.54, 2.07)
[0.001]	[0.002]

Note: This table shows the Best Linear Predictor of the Conditional Average Treatment Effect. Medians over 100 splits are given. 90% confidence intervals are given in parentheses. p values for the hypothesis that the parameter is equal to zero are given in square brackets. Source: The authors of this paper.

Table C6: Heterogeneity in average treatment effects: group average treatment effects

25% Least	25% Most	Difference
Treated vs control		
-0.01	0.37	0.38
(-0.12, 0.10)	(0.14, 0.58)	(0.12, 0.62)
[1.000]	[0.003]	[0.008]

Note: This table shows quartiles defined by Conditional Average Treatment Effects. Medians over 100 splits are given. 90% confidence intervals are given in parentheses. p values for the hypothesis that the parameter is equal to zero are given in square brackets. Source: The authors of this paper.

Table C7: Characteristics that correlate with a greater average treatment effect

	25% Least	25% Most	Difference
Baseline reported use of safe drinking water	0.84 (0.77, 0.91)	0.14 (0.07, 0.21)	−0.70 (-0.80, -0.59) [<0.001]
Share knowing someone who became seriously ill / died from As	0.07 (0.03, 0.10)	0.01 (0.00, 0.03)	−0.05 (-0.09, -0.01) [0.014]
Share switching from As-contaminated wells, last 5 years	0.79 (0.71, 0.86)	0.10 (0.04, 0.16)	−0.69 (-0.79, -0.59) [<0.001]
Depth to safe groundwater unknown	0.00 (-0.04, 0.12)	0.33 (0.14, 0.53)	0.33 (0.14, 0.53) [0.002]
Distance to safe drinking water (minutes)	5.19 (4.16, 6.29)	20.18 (15.93, 24.57)	15.06 (10.63, 19.38) [<0.001]
Distance to collect drinking water (minutes)	3.83 (3.21, 4.48)	2.27 (1.72, 2.84)	−1.48 (-2.36, -0.63) [0.002]
Share owning own private well	0.12 (0.09, 0.16)	0.30 (0.23, 0.38)	0.18 (0.11, 0.26) [<0.001]
No. of users per private well	12.11 (7.29, 17.12)	4.95 (3.91, 5.98)	−7.21 (-12.55, -2.39) [0.009]
Distance to road (hours)	13.35 (9.25, 17.55)	29.92 (19.57, 39.90)	16.64 (4.99, 27.65) [0.009]
Distance to market (hours)	17.56 (14.15, 21.06)	32.69 (23.53, 42.07)	14.94 (4.97, 25.12) [0.007]
Fraction Muslim	0.92 (0.84, 0.99)	0.50 (0.32, 0.69)	−0.42 (-0.61, -0.22) [<0.001]
Average share of village related	0.05 (0.03, 0.06)	0.17 (0.10, 0.24)	0.13 (0.05, 0.19) [0.002]
Mean out-network size	4.52 (4.25, 4.79)	3.33 (2.99, 3.68)	−1.20 (-1.63, -0.75) [<0.001]

Continued on next page

Table C7: Characteristics that correlate with a greater average treatment effect

	25% Least	25% Most	Difference
Mean no. of community associations participated in	1.24 (1.15, 1.32)	0.89 (0.76, 1.00)	−0.36 (-0.51, -0.22) [<0.001]
Mean no. of collective actions, total	1.62 (1.39, 1.82)	0.31 (0.07, 0.53)	−1.30 (-1.62, -0.97) [<0.001]
Mean no. of collective actions participated in	0.82 (0.70, 0.95)	0.16 (0.01, 0.31)	−0.66 (-0.84, -0.46) [<0.001]
No. of women's groups	8.40 (5.53, 11.31)	4.85 (3.66, 6.02)	−3.38 (-6.60, -0.47) [0.052]
Share participating in community decision-making	0.80 (0.75, 0.86)	0.47 (0.39, 0.54)	−0.34 (-0.43, -0.25) [<0.001]
Ethnic conflicts	0.00 (-0.04, 0.12)	0.33 (0.14, 0.53)	0.29 (0.10, 0.52) [0.006]
Any conflicts	0.08 (-0.03, 0.20)	0.75 (0.57, 0.93)	0.67 (0.46, 0.88) [<0.001]
Conflicts frequent	0.04 (-0.03, 0.20)	0.42 (0.21, 0.62)	0.38 (0.17, 0.60) [0.001]
Conflicts violent	0.04 (-0.04, 0.12)	0.29 (0.10, 0.48)	0.25 (0.04, 0.43) [0.035]
Leaders consult community	0.54 (0.34, 0.75)	0.96 (0.88, 1.03)	0.38 (0.15, 0.62) [0.002]
Leader selection is hereditary	0.17 (0.01, 0.32)	0.67 (0.47, 0.86)	0.50 (0.26, 0.75) [<0.001]
Leader selection is wealth-based	0.92 (0.80, 1.03)	0.42 (0.21, 0.62)	−0.50 (-0.73, -0.27) [<0.001]
Leader selection is democratic	0.62 (0.42, 0.83)	0.33 (0.14, 0.53)	−0.29 (-0.57, -0.01) [0.079]
Mean share of leaders seeking input from many	0.63 (0.60, 0.66)	0.72 (0.67, 0.78)	0.09 (0.03, 0.15) [0.011]

Continued on next page

Table C7: Characteristics that correlate with a greater average treatment effect

	25% Least	25% Most	Difference
Mean leader tenure	14.72 (13.80, 15.68)	16.49 (15.16, 17.79)	1.79 (0.21, 3.34) [0.055]
Mean value of household assets (BDT)	684.16 (551.66, 816.55)	410.95 (345.46, 476.71)	−271.56 (-424.80, -124.86) [<0.001]
Share poor or very poor (self-reported)	0.25 (0.20, 0.29)	0.46 (0.40, 0.53)	0.22 (0.15, 0.30) [<0.001]
Share poor or very poor (enumerator assessment)	0.24 (0.20, 0.29)	0.38 (0.33, 0.43)	0.14 (0.07, 0.21) [<0.001]
Share high-income (self-reported)	0.01 (0.01, 0.02)	0.00 (-0.00, 0.01)	−0.01 (-0.02, -0.00) [0.056]
Share high-income (enumerator assessment)	0.04 (0.03, 0.05)	0.01 (0.00, 0.02)	−0.02 (-0.04, -0.01) [0.009]
Share in lowest asset quintile	0.13 (0.10, 0.17)	0.25 (0.18, 0.30)	0.11 (0.04, 0.17) [0.007]
Share in second-highest asset quintile	0.23 (0.20, 0.26)	0.17 (0.14, 0.20)	−0.06 (-0.10, -0.01) [0.022]
Share in highest asset quintile	0.26 (0.21, 0.31)	0.13 (0.10, 0.16)	−0.13 (-0.19, -0.07) [<0.001]

Note: This table shows the mean value of the listed characteristic for the quartiles with i) the smallest (or most negative) treatment effect and ii) the greatest (or most positive) treatment effect, and the difference in mean values between the two quartiles. Only characteristics for which the least and most affected quartiles differ significantly ($p < 0.1$) are shown. Source: The authors of this paper.

Table C8: Heterogeneity of differences between models: best linear prediction

RCP vs CP		CP vs TD		RCP vs TD	
ATE	HET	ATE	HET	ATE	HET
0.09 (-0.06, 0.25) [0.460]	0.44 (-1.19, 2.13) [1.000]	0.00 (-0.15, 0.14) [1.000]	0.48 (-1.34, 2.18) [0.956]	0.06 (-0.08, 0.21) [0.697]	−0.34 (-2.04, 1.52) [0.845]

Note: This table shows the Best Linear Predictor of the conditional average differential treatment effect. Medians over 100 splits are given. 90% confidence intervals are given in parentheses. p values for the hypothesis that the parameter is equal to zero are given in brackets. Source: The authors of this paper.

Table C9: Heterogeneity of differences between decision-making processes: group average differential treatment effects

25% Least	25% Most	Difference
a) RCP vs CP		
0.03 (-0.19, 0.23) [0.813]	0.16 (-0.20, 0.50) [0.618]	0.13 (-0.31, 0.55) [1.000]
b) CP vs TD		
−0.07 (-0.41, 0.26) [1.000]	0.04 (-0.15, 0.26) [0.975]	0.11 (-0.29, 0.51) [1.000]
c) RCP vs TD		
0.09 (-0.20, 0.37) [1.000]	0.04 (-0.19, 0.29) [0.975]	−0.06 (-0.45, 0.36) [1.000]

Note: This table shows quartiles defined by Conditional Average Differential Treatment Effects. Medians over 100 splits are given. 90% confidence intervals are given in parentheses. p values for the hypothesis that the parameter is equal to zero are given in square brackets. Source: The authors of this paper.

Table C10: Characteristics that correlate with a greater impact of the RCP process vs the TD process

	25% Least	25% Most	Difference
Mean out-network size	4.02	3.38	-0.69
	(3.65, 4.39)	(2.79, 3.98)	(-1.37, 0.01)
			[0.085]
Mean no. of collective actions participated in	0.68	0.26	-0.40
	(0.32, 1.03)	(-0.01, 0.58)	(-0.89, 0.07)
			[0.067]

Note: This table shows the mean value of the listed characteristic for the quartiles with i) the smallest (or most negative) differential treatment effect of the decision-making process listed first compared to the decision-making process listed second and ii) the greatest (or most positive) differential treatment effect of the decision-making process listed first compared to the decision-making process listed second, and the difference in mean values between the two quartiles. Only characteristics for which the least and most affected quartiles differ significantly ($p < 0.1$) are shown. Source: The authors of this paper.

Table C11: Heterogeneity in predicted changes in use with respect to observable mechanisms: best linear prediction

RCP		CP		TD	
ATE	HET	ATE	HET	ATE	HET
0.27	1.63	0.14	0.73	0.14	1.30
(0.14, 0.39)	(0.46, 2.86)	(0.04, 0.25)	(-0.80, 2.42)	(0.04, 0.25)	(-0.17, 2.74)
[<0.001]	[0.015]	[0.024]	[0.545]	[0.018]	[0.164]

Note: This table shows the Best Linear Predictor of the Conditional Average Value. Medians over 100 splits are given. 90% confidence intervals are given in parentheses. p values for the hypothesis that the parameter is equal to zero are given in brackets. Source: The authors of this paper.

Table C12: Heterogeneity in predicted changes in use with respect to observable mechanisms: group average values

25% Least	25% Most	Difference
a) RCP		
0.06	0.50	0.44
(-0.07, 0.18)	(0.15, 0.89)	(0.05, 0.83)
[0.595]	[0.017]	[0.063]
b) CP		
0.09	0.04	0.11
(-0.41, 0.26)	(-0.15, 0.26)	(-0.29, 0.51)
[1.000]	[0.975]	[1.000]
c) TD		
−0.01	0.27	0.27
(-0.12, 0.09)	(-0.06, 0.53)	(-0.07, 0.57)
[1.000]	[0.975]	[1.000]

Note: This table shows quartiles defined by Conditional Average Values. Medians over 100 splits are given. 90% confidence intervals are given in parentheses. p values for the hypothesis that the parameter is equal to zero are given in square brackets. Source: The authors of this paper.