# Top Down or Bottom Up?

## A Field Experiment on Public Goods Provision and Solid Waste Management

Martina Jakob*

University of Bern

martina.jakob@unibe.ch

Carla Coccia*

University of Bern

carla.coccia@unibe.ch

January 14, 2024

Participatory bottom-up initiatives have become a popular alternative to the traditional top-down provision of local public goods. This study compares the effectiveness of these two approaches. Based on a randomized controlled trial with 120 communities in rural El Salvador, we assess the impact of two interventions addressing solid waste contamination: (i) a top-down intervention where streets were cleaned by an external actor, and (ii) a bottom-up intervention where a facilitator raised awareness and mobilized for collective action. Using an objective measure of pollution based on geotagged photos and deep learning, we find large immediate effects for both interventions, with reductions in waste by 39 percent for the top-down intervention and 28 percent for the bottom-up intervention. Four months after the end of the project, these effects depreciated by 80 percent for the top-down and 60 percent for the bottom-up treatment. Our complementary data from 2,421 surveys and 883 activity records is consistent with a theoretical framework where many individuals are willing to contribute to public goods when others do, but fail to coordinate in the absence of a committed leader.

*Both authors contributed equally to this study. This project was financed through a *doc.CH* grant by the Swiss National Science Foundation (SNSF) awarded to Martina Jakob for the completion of her dissertation projects. A randomized controlled trial registry is available at: `https://www.socialscienceregistry.org/trials/10913`.

# 1 Introduction

Many of the world's most pressing challenges, like curbing greenhouse gas emissions, maintaining global peace, or establishing a functioning health and education infrastructure in low-income countries, are public goods problems. As public goods benefit everyone irrespective of their personal contribution to them, individuals have an incentive to free-ride. To avoid the resulting underprovision, the standard solution calls for a *top-down* intervention by a powerful actor such as the state to provide the public good or enforce rules for its protection (Olson, 1971). Yet, ample empirical evidence documents that groups are often able to act collectively and overcome the social dilemma tied to public goods (e.g., Ostrom, 1990, 1999). This has inspired an alternative line of thinking advocating for *bottom-up* solutions often referred to as community-driven development (CDD). In this study, we compare the effectiveness of these two approaches in the context of solid waste management.

Our paper is based on a randomized controlled trial with 120 communities in rural El Salvador. We study the impact of two programs designed to reduce local solid waste pollution. The first intervention pursued a traditional top-down approach with monthly community visits by an external cleaning team to collect litter from the streets. In the second intervention, a local facilitator was appointed for each community to raise awareness and mobilize for collective action to address the problem in a bottom-up process. Typical activities in this community-driven initiative were educational sessions about waste management, collective monthly cleanups, and community meetings to define common strategies. The two interventions had a duration of four months, were similar in cost, and implemented by the local non-profit organization "Consciente". We randomly assigned communities to three experimental groups: 40 communities received the top-down intervention, 39 participated in the bottom-up initiative, and 41 were assigned to a control group. To track contamination levels in all communities, we took about 200,000 geo-tagged photos along all streets, and evaluated them using a deep learning model. Our model achieves state-of-the-art performance in trash detection, allowing us to establish a reliable and objective measure of contamination. To understand the mechanisms behind potential impacts, these contamination assessments were complemented with survey data from a sample of 2,421 residents and a detailed registry of all the 883 activities conducted in the context of the interventions.

We find large *immediate impacts* for both interventions. The top-down intervention reduced solid waste pollution by 0.7–0.8$\sigma$ or 39 percent ($p < 0.01$). Effects are

significantly smaller ($p < 0.05$), but still substantial for the bottom-up intervention, with a reduction by 0.5–0.6$\sigma$ or 28 percent ($p < 0.01$). Our survey results further show that these improvements did not go unnoticed, as both interventions had significant immediate effects on people's cleanliness perceptions ($\sim$0.15$\sigma$ for both interventions) and self-reported recycling practices ($\sim$10 percentage points for both interventions). For the bottom-up intervention, we also observe a 13 percentage point increase in the share of respondents indicating that they dispose of their waste appropriately, rather than burning, burying or dumping it. *Long-term results* show that four months after the end of the interventions, the impact on observed pollution decreased by 80 percent for the top-down intervention and by 60 percent for the bottom-up intervention. This yields a long-term effect of 0.1$\sigma$ ($p = 0.11$) for the top-down treatment and of 0.2$\sigma$ ($p < 0.05$) for the bottom-up treatment. While this is suggestive evidence for a higher persistence in the bottom-up intervention, the difference between depletion rates is not statistically significant ($p = 0.2$). We observe no depletion in people's cleanliness perceptions, but the immediate changes in self-reported waste management behavior strongly depreciate or disappear for both treatments.

Our rich complementary data offers insights into the mechanisms driving the success and limitations of bottom-up development initiatives. We find limited evidence for *information effects* through increased awareness of the problem or knowledge of others' concern for it. Although the bottom-up intervention had an immediate impact on people's beliefs about the prevalence of littering behavior in their community, this social norms effect was short-lived and not significantly more pronounced than in the top-down intervention. Our results are more consistent with the hypothesis that community-driven development can alleviate *organizational constraints* to collective action. However, much of the success along this dimension appears to be tied to the presence of the facilitator. While the number of cleanup events and participants remained consistently high during the intervention period, we observe a sharp decline in collective efforts – from 0.9 to 0.4 monthly cleanups – after the withdrawal of the NGO, and we find limited evidence for a sustained increase in social capital. Our results are most consistent with a theoretical framework where many individuals are willing to contribute to public goods as long as others do so too, but struggle to coordinate in the absence of a dedicated leader.

This study makes three distinct contributions. First, we add to the debate on the effectiveness of bottom-up development strategies. The rise of bottom-up development strategies represents a major trend in international development cooperation (Mansuri and Rao, 2012; Casey, 2018). Based on an analysis of 250,000 World Bank project

reports, we find that the share of documents mentioning keywords connected with community-driven development increased rapidly from the early 1990s. By 2003, over 40 percent of all documents contained at least one related term. Our pre-survey further shows that practitioners and academics alike tend to be optimistic about bottom-up initiatives, with roughly 80 percent of respondents in both groups believing they would outperform traditional top-down solutions in the long run. Despite the vast importance of the approach, rigorous evaluations of community-driven initiatives remain scarce (Table A13). Most notably, the effectiveness of bottom-up solutions has not yet been compared to that of the traditional to-down alternatives they seek to replace. This study contributes to filling this critical gap in empirical research. Our findings highlight that while bottom-up initiatives can indeed successfully promote the provision of local public goods, they are not always more effective in doing so than top-down interventions.

Second, our study also contributes to the discussion on how to tackle problems related to solid waste management in developing countries. While 96 percent of waste in high-income countries is collected and properly disposed of, only 39 percent of waste in low-income countries is. At the same time, solid waste generation in low-income countries is expected to triple by 2050 (Kaza et al., 2018). Finding effective ways to address the problem and limit the environmental and health repercussions it causes, is thus a critical and timely priority. Our study ties into the nascent literature evaluating different interventions to improve solid waste management (Table A14). We find that raising awareness and empowering communities to address the waste problem can be an important part of the solution, but may not be successful on its own without continued investment. In addition, our results suggest that interventions that focus on changing littering norms alone, without complementary efforts to collect waste that continues to accumulate on the streets, are unlikely to be sustainable.

Finally, our paper advances the burgeoning field of research using machine and deep learning methods to track and understand global development. A rapidly expanding economic literature has shown that important socio-economic outcomes can be accurately predicted from alternative data sources such as satellite imagery (Jean et al., 2016; Yeh et al., 2020), phone records (Blumenstock et al., 2015), or tweets (Jakob and Heinrich, 2023). However, this literature is largely focused on providing proofs of concept, and scientific or practical applications remain scarce. In this study, we use deep learning to derive an objective and reliable measure for our main experimental outcome. By fine-tuning a YOLOv8 object detection model using publicly available trash data and a sample of images from our experiment, we achieve state-of-

the-art performance in trash detection, with an AP50 of 59.5 percent on the popular TACO dataset and of 59.0 percent for our own images. The resulting contamination measure produced more robust results than an alternative approach based on subjective contamination assessments by enumerators. This highlights the potential of deep learning methods in settings where large amounts of data must be processed or human measurements are prone to subjectivity.

# 2 Public Goods and the Rise of Community-Driven Development

The public goods problem models a situation where the benefits of a cooperative outcome accrue to everyone irrespective of people's individual contributions towards it. The dominant strategy for a self-motivated and rational agent is to free-ride by contributing nothing. Standard economic theory considers this a market failure, as it results in a single, Pareto-inefficient equilibrium where the public good is not provided, and individuals fail to realize a mutually beneficial outcome (Olson, 1971; Hardin, 1971, 1982). The conventional approach to addressing market failures associated with public goods calls for a top-down intervention by a powerful entity, such as the state, to either supply the public good or enforce protective regulations.

However, ample research documents that most people do not behave as the standard model of self-interested actors predicts. Zero contributions to public goods are neither the norm in laboratory experiments (e.g., Fischbacher et al., 2001; Willer, 2009; Chaudhuri, 2011) nor in real-world situations. For example, many people volunteer in associations, donate blood, contribute to charities, make environmentally friendly consumption choices, or take part in political protest. Rather than maximizing personal gains, the majority of individuals appear to follow norms of reciprocity and contribute as long as a sufficient number of others do so too (e.g., Keser and Van Winden, 2000; Gächter, 2006; Thöni and Volk, 2018). Under preferences for conditional cooperation, the provision of public goods becomes a coordination problem with multiple possible equilibria. This is consistent with numerous examples showing that groups sometimes succeed and sometimes fail in providing public goods or protecting common resources (e.g., Ostrom, 1990, 1999).

In this context, the idea has gained traction that groups can be empowered to coordinate and guarantee the provision of public goods in a bottom-up process. This approach is variously known as community-driven development (CDD), community-

based development (CBD), community and local development (CLD), or participatory development (Mansuri and Rao, 2012; Casey, 2018).

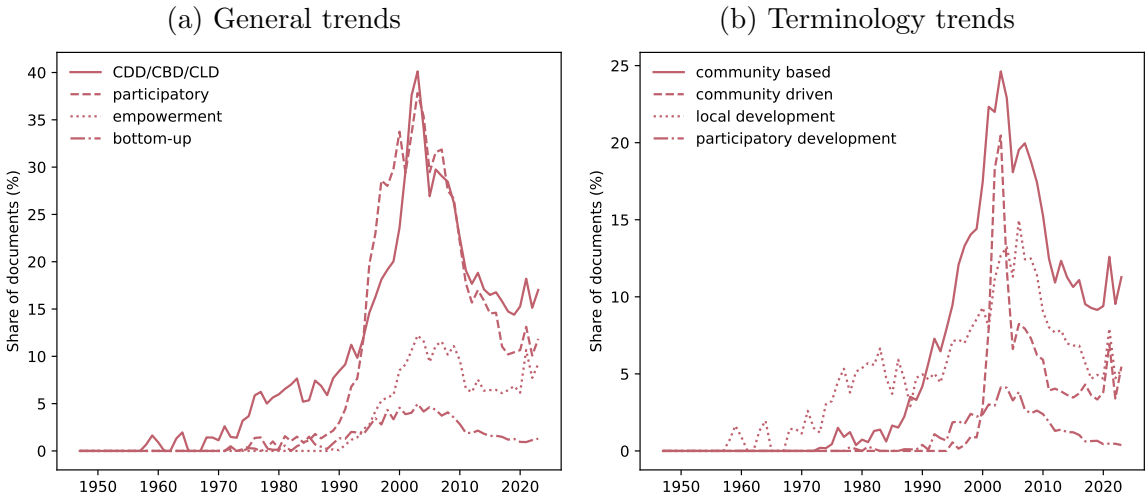



Figure 1: The Rise of Community-Driven Development

Illustration based on 259,668 project documents obtained through the World Bank API. Document types include, among others, procurement plans (23%), implementation reports (18%), project information documents (5%), or environmental assessments (5%). We exclude documents with less than 500 correct English words (10% of all documents), and documents that do not contain the word "development" (15% of the remaining documents). "CDD/CBD/CLD" refers to any of the keywords "community-driven", "community-based", "participatory development", or "local development" (different spellings accounted for).

The rise of community-driven initiatives represents a major strategic shift in international development cooperation. In response to concerns about poorly maintained infrastructure following traditional top-down interventions, governments, NGOs and international organizations have increasingly turned to community-based solutions for public goods provision. This bottom-up approach is often hailed as "more responsive to demands, more inclusive, more sustainable, and more cost-effective than traditional centrally led programs" (Dongier et al., 2003), and believed to sustainably transform and strengthen local institutions. To illustrate this trend, we scraped over 250,000 World Bank project documents, published between 1947 and 2023. We find that the proportion of documents containing keywords directly related to community-driven development, along with more loosely connected keywords such as "participatory" or "empowerment", began to increase rapidly in the early 1990s (Figure 1). At its peak in 2003, more than 40 percent of the documents mentioned at least one CDD keyword. Over the past two decades, this share has declined, but remains high, stabilizing at around 17 percent for the past three years. This strong focus on participatory, bottom-up initiatives is also reflected in funding priorities. In 2022, the World Bank

alone had 373 ongoing community-based initiatives with more than $40 billion in total lending (World Bank, 2022).

Our pre-survey with 100 scientists and local practitioners further substantiates this sense of optimism regarding the potential of bottom-up approaches (Figure A1). Over 90 percent of the practitioners and scientists in our sample expressed confidence that adopting a community-based approach to waste management would lead to a reduction in community pollution in both the short and long term. In addition, about 80 percent of respondents from both groups agreed that a bottom-up approach would outperform a more traditional top-down intervention in the long run. While the majority of academics believed that the relative advantage of bottom-up provision unfolds only in the long term, most practitioners also predicted better short-term outcomes.

The rise of community-driven development has also sparked interest in the academic community, leading to a number of rigorous evaluations to assess the effectiveness of the approach. Table A13 provides a comprehensive overview of this literature. Although the reviewed studies vary in the types of interventions and outcomes they examine, we can draw four general conclusions from this research. First, bottom-up initiatives are indeed often successful in delivering and maintaining public goods and improving the livelihoods of the poor (Avdeenko and Gilligan, 2015; Björkman and Svensson, 2009; Desai and Olofsgård, 2019; Duflo et al., 2015). Second, the evidence is inconclusive on the proposed transformative impact on local institutions. Many evaluations report no lasting effects on collective action capacity (Casey et al., 2012; Casey, 2018; Mansuri and Rao, 2012) or the empowerment of minority groups (Casey et al., 2012; Van der Windt and Mvukiyehe, 2020). Third, existing studies compare bottom-up initiatives with a status quo where no infusion of funds occurs. While this allows to assess whether such initiatives work, it does not tell us if they outperform alternative ways of service delivery, a key limitation noted by recent studies in the field (e.g., Casey, 2018). Fourth, there appears to be little clarity about the precise mechanisms through which participatory bottom-up interventions should affect the provision of public goods, limiting our understanding of where such initiatives may fail and how they can be improved. Our study addresses the limitations raised in the last two points by (i) offering a comparison between two modes of providing the same public good, and (ii) discussing the results within a more general theoretical framework.

# 3 Theoretical Framework

We use a simple theoretical framework to analyze through what channels community-driven development potentially facilitates collective action and the provision of public goods.[1] We assume that individuals are willing to contribute to a public good as long as a certain fraction of the group does, and that they differ in these *thresholds for conditional cooperation*. A threshold of 0 corresponds to people who always cooperate, while a threshold of 1 indicates that someone never cooperates even if everyone else in the group does. Evidence from laboratory studies shows that these extreme types are in the minority, and that most people exhibit behavior consistent with varying degrees of conditional cooperation (Fischbacher et al., 2001). Individual thresholds may be determined by numerous factors, such as the importance the person places on the public good (i.e., preferences), the individual's pro-sociality, or his or her resources. As people usually cannot observe the actual number of contributors, they act based on their beliefs about it. This means that individuals will start contributing as soon as they believe that the proportion of contributors is higher than their personal threshold, and stop doing so if they think that this is no longer the case. For a given distribution of thresholds, multiple equilibria may thus be possible.[2] In a repeated game, we would expect self-reinforcing positive or negative dynamics, as people continually adjust their contributions based on the observed contributions of others until a stable equilibrium is reached (Berger, 2021; Berger et al., 2023).

Even when a socially more desirable equilibrium exists, attaining it often requires coordinated action. Take the example of a group of workers deciding whether to go on strike. If most people are willing to participate as long as most others do so too, the strike can only take place if the group coordinates to act simultaneously. Ample research shows that allowing people to communicate with each other increases the chance of reaching a stable high-level equilibrium (Chaudhuri, 2011). Following Cowen (1992) and Dahlman (1979), we thus assume that coordinating collective action entails *transaction costs*. The magnitude of these costs depends on how well people know and trust each other, and on the institutions they set up to facilitate cooperation. This idea is reflected in the notion of social capital, commonly understood as "the norms

[1] For simplicity, we limit ourselves to the extensive contribution margin (i.e., whether people contribute). Yet, a very similar case can be made for the intensive contribution margin (i.e., how much people contribute).

[2] Consider a community where 40 percent of individuals will contribute as long as at least 30 percent of the population contributes, and 60 percent contribute as long as at least 80 percent contribute. In this case, three stable equilibria could be reached: one where no one contributes, one where 40 percent contribute, and one where everyone contributes.

and networks that facilitate collective action" (Woolcock et al., 2001, p. 9). Thus, at higher levels of social capital, members of a group are more likely to succeed in organizing to collectively provide or protect public goods (e.g., Anderson et al., 2004). A related idea concerns the concept of leadership. In most real-world scenarios, transaction costs are not perfectly divisible, meaning that a single individual (or a small group of individuals) must bear a large portion of these costs. The presence of a committed leader (or leadership team) should thus be critical for a group to overcome organizational constraints to collective action. This is in line with extensive empirical evidence documenting the importance of leadership for collective action and the provision of public goods (e.g., Glowacki and von Rueden, 2015; Sahin et al., 2015).[3]

Finally, the provision of certain public goods requires a significant monetary investment. In a low-income setting, where time is not easily translated into money and people lack access to affordable loans, a group may fail to realize a collectively beneficial outcome due to *credit constraints*. For example, consider a poor community trying to build a paved road that is expected to yield high returns for everyone. Even if individuals are willing to contribute and able to organize themselves, the project will not be realized if the community does not have access to funding. This aligns with numerous studies documenting how financial markets often fail the poor (Banerjee and Duflo, 2007). Therefore, we conclude that collective action succeeds if the distribution of contribution thresholds allows for a high-level equilibrium, if the community is sufficiently organized to coordinate collective action so that this equilibrium can be reached, and if its members have access to sufficient funding to cover potential monetary investments.

Based on this framework, we distinguish three basic mechanisms through which community-driven development interventions could facilitate collective action. First, it can help to alleviate *informational constraints*. If individuals underestimate the share of others contributing to a public good, getting people to talk about the problem can eliminate these misconceptions. As interventions often convey information on the topics related to specific public goods and on effective solutions, they may also directly alter the distribution of thresholds, as people begin to care more about the

[3]Note that in some cases, the institutions groups set up to facilitate cooperation can also be seen as modifying the thresholds themselves (rather than lowering transaction costs). For example, if groups devise means of punishing defector, this would change people's thresholds for conditional cooperation. The same can be said if individuals get inspired by a charismatic leader (Jack and Recalde, 2015). For simplicity, we abstract from this alternative conceptualization, and view the organization of the collective action as a second-order public goods problem related to the bearing of transaction costs.

problem or become more confident about their ability to address it. Under certain threshold distributions, this would enable the community to reach a higher equilibrium. A second possible mechanism is related to *organizational constraints* and thus to the transaction costs that effective coordination entails. By bringing people together and encouraging them to set up organizational structures, bottom-up interventions could build social capital and leadership, and thereby facilitate coordination. A final potential channel through which bottom-up initiatives could improve public goods is by mitigating *credit constraints*. People in poor communities may be sufficiently informed and organized to address local public goods problems, but simply lack access to funding to do so. This is the premise of the archetypal community-driven development intervention, which provides block grants to communities to invest in the provision of local public goods.

While the main focus of our study is on comparing the effectiveness of bottom-up provision with a more traditional top-down intervention, we will use this theoretical framework to make sense of patterns in our data, thereby contributing to a better understanding of why community-driven initiatives may work and where they may fail.

# 4 Context and Interventions

We conducted our study in the context of solid waste management in rural communities in El Salvador, a lower-middle-income country in Central America. Inadequate waste management is ubiquitous in developing countries and causes numerous detrimental health and environmental impacts. Over time, waste can spread over large areas, contaminating rivers, oceans, groundwater and soil. In addition to its environmental consequences, contaminated water can pose serious health risks by spreading infectious diseases such as diarrhea or hepatitis (Mohan and Joseph, 2021). While 96 percent of solid waste in high-income countries is collected and properly disposed of, the corresponding figures are only 51 percent for lower-middle-income countries and 39 percent for low-income countries (Kaza et al., 2018). At the same time, solid waste generation in low- and middle-income countries is expected to triple by 2050. Finding out what works to address the problem is thus an important and timely priority (see Table A14 for a review of the emerging literature in this area).

Solid waste contamination poses a typical public goods problem for local communities, as a clean environment benefits everyone regardless of their personal contribution towards it. This is in line with insights from our baseline survey, showing that even

though most people in our sample are bothered by the waste pollution in their communities, the problem remains widespread. Nearly 80 percent of respondents indicate that the waste in their community bothers them much (39%) or very much (38%). Meanwhile, about half the people admit that they dispose of their waste improperly by burning, burying, or dumping it. Similarly, only 32 percent of the communities are visited by a municipal garbage truck collecting household waste at least every two weeks, and 36 percent have no such truck service at all. Almost all communities have considerable amounts of waste in public spaces, and contamination levels are not significantly correlated with the frequency of the garbage truck service.

Communities can adopt two distinct strategies to tackle the problem. People can either stop dumping waste in public spaces, or they can coordinate to collectively remove it and ensure proper disposal. Since both actions involve costs, and gains are shared between all residents, effective solutions are needed to overcome the free-riding problem inherent to the provision of public goods. We partnered with the local NGO Consciente to develop and implement two interventions to address the problem: (i) a traditional top-down intervention and (ii) a community-driven bottom-up intervention. Both initiatives had a duration of four months and were similar in costs.

In the *top-down intervention*, an external team of cleaners employed by the NGO made monthly visits to all communities to collect litter from public areas and gather household waste from residents. The team comprised two cleaners and a garbage truck driver, and each community visit typically lasted half a day. While the intervention was conducted by a non-governmental rather than a governmental institution, it mirrors what a top-down state intervention would look like in this context and corresponds to the typical approach pursued by governments worldwide to tackle solid waste pollution.

In the *bottom-up intervention*, a team of 24 part-time facilitators was hired and trained in topics related to waste pollution and management and community organization strategies. Facilitators were typically young university graduates from the area, but not necessarily from the community, and responsible for one or two communities. Their job consisted in raising awareness for the problem, mobilizing for collective action, and encouraging the creation of local organizational structures to facilitate sustainable solutions. For this purpose, they could draw on extensive teaching materials developed by the NGO, but were instructed to adapt the proposed activities based on local needs. The typical community intervention consisted of an initial meeting, a series of educational sessions and community activities on waste management (such

as input and discussion sessions, hands-on workshops, poster campaigns, or community movie nights), and monthly collective cleanups. The facilitator also assisted the community in organizing the disposal of the waste collected from the cleanups and households. This was typically done by using the private vehicle of a community member or by appealing to the municipality government. At the end of the intervention, each community presented a waste management plan indicating how the problem would be addressed after the withdrawal of the NGO.

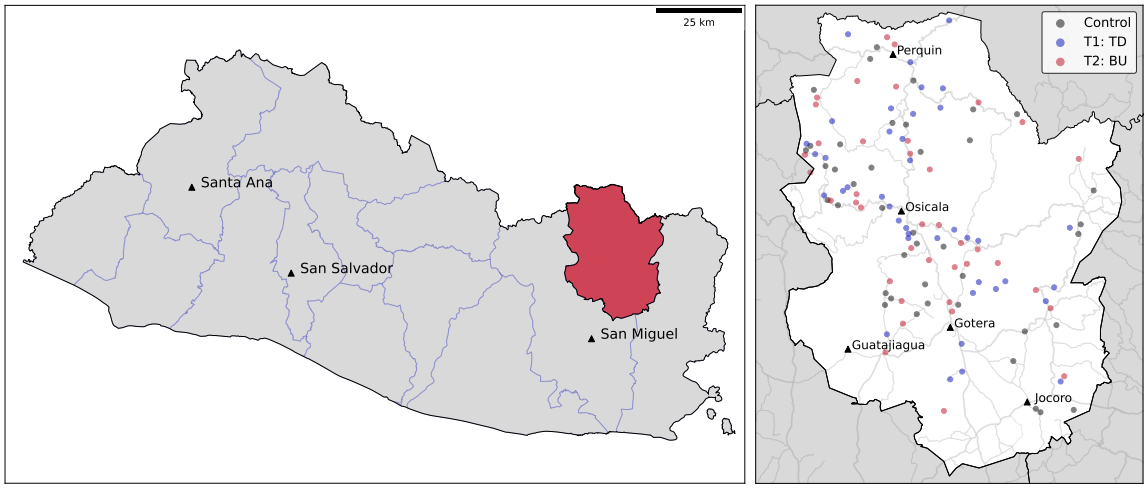



Figure 2: Study Area

# 5 Research Design

To study the impact of these two interventions, we conducted a randomized controlled trial with 120 communities in the rural department of Morazán in El Salvador (see Figure 2). The selection of these communities was undertaken in two steps. First, we compiled a list of medium-sized, non-urban communities (30–300 households) facing waste management problems with the help of municipal governments. In a second step, we conducted a baseline survey in 140 communities and selected 120 communities based on two criteria: contamination levels using our own measurements, and spatial distance to limit spillover effects. The resulting sample is not representative of our study area, but contains a diverse set of communities that have not solved the waste management problem through an endogenous bottom-up process or with the help of local government institutions. We randomly assigned these communities to three experimental conditions: (1) the top-down intervention (40 communities), (2) the bottom-up intervention (39 communities), or (3) a control group that received no

intervention (41 communities).[4] Randomization was stratified by baseline contamination (three bins) and geographic zones (four bins). In all experimental groups, we conducted a measurement wave before the intervention (baseline), toward the end of the intervention (midline, after 3–4 intervention months), and four months after the intervention (endline). This allows us to study both the immediate impact of the two interventions and whether potential effects are sustained after the end of the program.

## 5.1 Data

To track different waste-related outcomes, we collected three types of data: (i) contamination assessments based on images taken along all streets, (ii) survey data on people's perceptions and self-reported behavior, and (iii) monitoring data on all the activities that were conducted in the context of the interventions.

### 5.1.1 Image Data on Contamination

For the main outcome of our experiment, we took *geocoded pictures* along all the streets and public spaces in the 120 communities. For this purpose, enumerators worked in pairs and simultaneously took geotagged photos on both sides of the street every five steps. Enumerators were carefully trained and received a detailed manual explaining how to take the photos. Photos typically show a portion of the street, the roadside, and the background. To ensure spatial consistency across the three measurement waves, we used an application that enabled us to outline the geographic boundaries of each community and display them on an interactive map. Enumerators were instructed to cover all roads, paths and public spaces within this designated area. To account for minor deviations in the covered area, we only include photos with spatial support across all three waves (92% of all images).[5] This procedure results in approximately 500 images per community and wave, ranging from 118 photos in the smallest community to 1,926 photos in the largest community, and a total of 181,393

[4]The number of communities differs between experimental groups because remainders per stratum were assigned with probability 1/3 to each group or group combination (in the case of two remainders). A common alternative is to group remainders over all strata and reassign them randomly. Assigning remainders with probabilities instead of grouping them means equal group sizes cannot be ensured, but assignment balance within the strata is preserved (McKenzie and Bruhn, 2011).

[5]A photo in a given wave is defined as having no spatial support in another wave if the closest photo is more than 8 meters away and the fifth closest photo is more than 25 meters away. This decision rule was found to produce good results, by excluding road segments that were not covered in all waves, but keeping photos in all other segments. We only include photos with spatial support in all waves, e.g., only baseline photos with nearby midline and endline photos.

images across all waves. We then used a deep learning model to predict the amount of trash on each image (see Section 5.2).

A key challenge is to link midline and endline contamination levels in different areas of each community to their baseline contamination values. We use three different approaches of *spatial aggregation*: (i) a kernel approach, (ii) a raster approach, and (iii) raw community averages. The *kernel approach* consists in drawing a circle with a radius of 12.5 meters around each midline or endline image (see Figure 3).[6] We then use a triangular kernel to compute a weighted average of all baseline contamination values within the circle. The average circle contains 7.6 baseline images, and 99 percent of all circles contain at least one baseline image. Our final sample consists of 60,709 observations for the estimation of immediate effects (midline) and 65,673 observations for assessing long-term effects (endline). For the *raster approach*, we lay a fixed 16.5 x 16.5 meter grid over each community and compute the wave-specific average across all photos in each cell (see Figure 4).[7] The average cell contains 4.6 images, and 81 percent of all cells with baseline images also contain midline and endline images. The raster approach results in a final sample of 10,740 cells, with an average of 90 cells per community, ranging from 21 cells in the smallest community to 278 cells in the largest community. Finally, we also compute *raw averages* across all images for each community and wave, resulting in 120 (unclustered) observations.

As a robustness check, enumerators were also told to make a *subjective assessment* of the general cleanliness of the environment every 25 steps or 5 photos. Based on representative example images, they had to classify their environment into four categories, ranging from "very clean" to "very dirty". Our final sample consists of about 100 ratings per community and wave, with 23 assessments in the smallest community and 408 in the largest community.[8] We use a triangular kernel with a radius of

[6]Note that this approach results in different baseline circles for midline and endline measurements respectively. To determine an appropriate radius, we created and examined community maps showing all included observations (with baseline values) and excluded observations (without baseline values) for different circle sizes. With a radius of 12.5 m, almost all dropped observations were at the community boundaries (which were interpreted slightly differently across waves) rather than within communities.

[7]The ideal raster produces enough observations (cells) per community while maintaining a good support across waves, so that few of these observations need to be dropped. A 16.5 x 16.5 meter grid was found to strike a good balance between these competing criteria.

[8]To obtain geocoded ratings, we used a simple low-tech strategy. Enumerators had to take a picture of a placard with the number corresponding to the level of contamination. We then used the weights of a Github model pretrained on the popular Street View House Numbers (SVHN) dataset (Netzer et al., 2011) to predict the number corresponding to each image. To make sure that all predictions were correct, we manually reviewed the few cases where the model predicted low certainties. The number images were integrated with all other photos to determine the spatial support across waves.

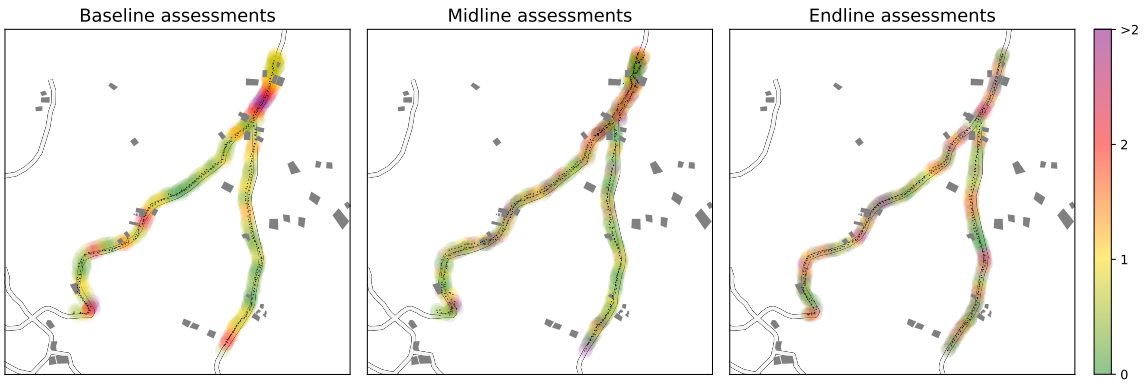



Figure 3: Illustration of Kernel Approach in Example Community

Black dots represent image locations. Circle color corresponds to the number of trash pieces identified on each image. Baseline values are imputed based on circles around each midline and endline assessment respectively. Circle radius is 12.5 m. A triangular kernel is used to give higher weights to closer assessments. Baseline map is shown with respect to the midline assessment. We use OpenStreetMap for all base maps.

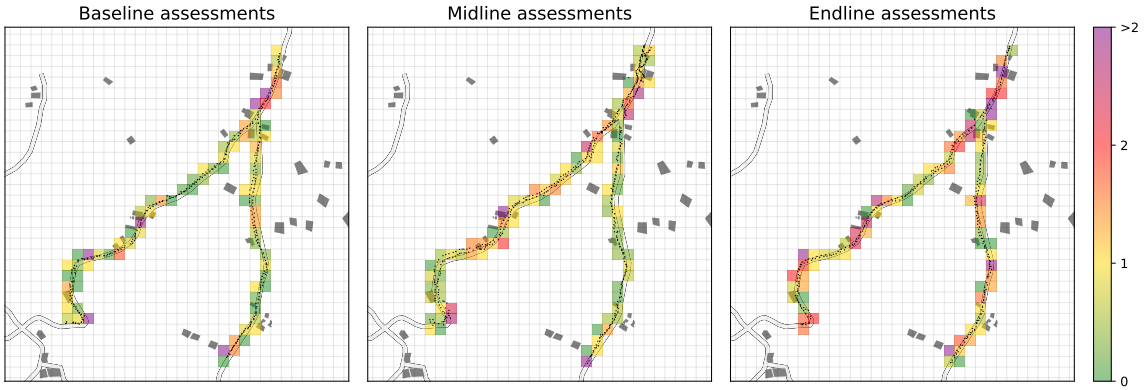



Figure 4: Illustration of Raster Approach in Example Community

Black dots represent image locations. Cell color corresponds to the average number of trash pieces identified on an image in the cell. Resolution of the raster is 0.00015 degrees (approx. 16.5 m).

25 meters and a raster of 33 x 33 meters for spatial aggregation of the enumerator assessment data (see Figures A4 and A5).

### 5.1.2 Survey Data

To better understand the mechanisms behind potential effects, we administered short surveys to 20 residents per community. Our survey includes questions about waste-related activities respondents observed or participated in, the perceived cleanliness of the community, waste disposal and recycling behaviors, littering norms and self-reported littering behaviors, and various measures of social capital. Table A12 provides an overview on all included survey questions. Participants were selected by

enumerators during the community visit for the baseline assessments. Enumerators were instructed to recruit survey participants by randomly knocking on doors until the target of 20 interviews was reached. While the resulting sample is not representative (mainly due to different propensities to be home during the day), it is very diverse and comparable across experimental groups. Our final sample consists of 2,421 individuals.[9] Attrition was 15 percent in the midline and 24 percent in the endline assessment, resulting in 2,066 observations to estimate immediate effects and 1,832 observations for long-term effects. We find no indication of differential attrition by treatment status (see Table A9). Missing values were imputed using the mean of the respective experimental group.[10]

### 5.1.3 Activity Registry

To gain insights into how the program was implemented, a detailed registry of all activities performed under each intervention was compiled. For the top-down intervention, we collected data on every cleaning visit, including the amount of garbage collected and the number of working hours devoted to the task. The activity registry for the community-driven intervention contained information about the type and duration of each activity, the number of participants, facilitator preparation time, and subjective ratings regarding activity success and participant interest. For cleanup campaigns, the log additionally recorded how much litter was collected in how many working hours, and how its removal was organized. All intervention activities were registered by the NGO staff responsible for conducting each activity (i.e., cleaners or facilitators). People were instructed to report honestly on all activities, and neither pay nor promotion was contingent on the successful execution of these activities. In addition, facilitators were required to submit photos of each activity to the project coordination team of the NGO. For the community-driven intervention, we also recorded all activities during the post-intervention period through phone calls to community leaders, allowing us to study to which extent collective action efforts continued after the withdrawal of the NGO. To better understand the challenges communities faced in the post-intervention period, we also conducted interviews with the person who remained in charge in each community after the end of the intervention.

[9]This sample is larger than 2,400 because we grouped 8 communities into 4 community clusters at baseline due to geographic proximity and in an effort to avoid spillovers (meaning that our analyses include 124 communities and 120 community clusters). The community clusters received the same treatment, but 40 interviews were conducted instead of 20. Throughout the study, these clusters are treated like communities.

[10]Missing values were rare, with fewer than 1 percent missings in all our main survey variables presented in Table 5.

## 5.2 Deep Learning for Waste Detection

We employ a novel approach that uses deep learning to create an objective measure of contamination based on the approximately 200,000 images included in our analysis. This is achieved by fine-tuning a YOLOv8 object detection model using publicly available trash datasets and manually labeled images from our own study. The YOLOv8 model is the latest addition to the YOLO (You Only Look Once) family, which comprises state-of-the-art object detection systems employed in real-time tasks for robotics, self-driving cars, and video surveillance applications (Terven and Cordova-Esparza, 2023). In contrast to other object detection models, as implied by their name, YOLO models have the ability to simultaneously identify all objects within an image. This is achieved by dividing the image into a grid and making predictions for multiple bounding boxes for each grid section, accompanied by confidence scores and a vector of class probabilities (Redmon et al., 2016). This feature marks a significant improvement in terms of speed while maintaining a high accuracy and is therefore a key factor behind the popularity of the YOLO family. YOLOv8 was released by Ultralytics, the company behind one of the older model versions (YOLOv5), in January 2023. Ultralytics offers five different model sizes, varying in features such as their mean average precision on the popular COCO dataset (200,000 annotated images) and the number of parameters the model has to estimate (ranging from 3.2 million for the smallest and 68.2 million parameters for the largest model). To balance speed, accuracy and necessary computational power, we opted for the median model, YOLOv8m, with an mAP50-95 of 50.2 percent for the COCO dataset and 25.9 million estimated parameters.[11]

To fine-tune the model, we use the publicly available TACO (Trash Annotations in Context) dataset, consisting of 1,500 official images with 4,784 annotated trash bounding boxes (Proença and Simoes, 2020). The TACO data is often used as the benchmark dataset to compare the performance of different trash detection algorithms. In addition to the official images, TACO contains a set of photos with crowd-sourced annotations, which have not yet been subjected to a quality check. We manually reviewed all these unofficial images to exclude instances with incorrect bounding boxes, resulting in 3,432 additional images with 7,511 additional annotations, and a total of roughly 5,000 and 12,000 annotations for the extended TACO dataset (official + unofficial TACO). We also test if the performance is improved by adding a second

[11]The mAP (mean average precision) corresponds to the mean of the average precision (AP) over all classes and IoU (Intersection over Union) thresholds from 0.5 to 0.95 (see below for an explanation).

popular trash detection dataset, the PlastOPol data containing 2,418 images with 5,300 annotations, to the fine-tuning procedure (see Córdova et al., 2022). As the images in this dataset usually center on a single piece of trash in the foreground, they differ markedly from our own images, which depict natural settings potentially containing multiple small pieces of trash, meaning that it is a priori unclear whether adding PlastOPol to our training data would improve or degrade model performance for our task. Finally, we also include 600 manually labeled images with 3,024 annotations from our own images (200 images per wave) and 216 of our own images without any trash.

We trained our model using 70 percent of the data for training and 30 percent for testing, and computed separate performance statistics for each data source. For training, we use 200 epochs and a batch size of 8, mainly determined by computational power limitations. For prediction and evaluation, we set the detection threshold to 50 percent, meaning that objects are only detected if the model is at least 50 percent confident of its prediction. Our principal performance statistic is the AP (Average Precision), a measure that is widely used in the deep learning literature to compare results across different models. This metric is based on the area under the precision-recall curve and thus captures how well the model performs averaging over different certainty thresholds. In line with previous research, we will use AP50, meaning that a predicted bounding box is considered as accurate if the intersection between the true and the predicted box corresponds to at least 50 percent of the union of the two boxes. As additional more intuitive measures, we will also report the precision (the proportion of detected instances that are correct), the recall (the proportion of true instances that are detected), and the F1 score (a combination of precision and recall).

For the TACO dataset, the AP50 reaches 59.5 to 61.2 percent depending on whether we include the PlastOPol dataset for training or not. Table 1 illustrates that these results are similar to the best-performing models reported in the literature, ranging from an AP50 of 57.4 percent (Das et al., 2023) to an AP50 of 63.3 percent (Córdova et al., 2022). Our best model specification performs almost equally well on our own data as on the TACO dataset, achieving an AP50 of 57–59 percent. As including PlastOPol slightly decreases the AP50 for our images (Table 1), we do not use it for the training of our final model. We thus attain an AP50 of 59.0 percent, a precision of 78.6 percent, and a recall of 39.6 percent, suggesting that our model produces few incorrect detections, but misses many true instances. As many pieces of garbage are small, partially hidden, or in the background and thus difficult to detect even for human coders, this is a remarkable performance.

Table 1: Model Performance

| | *Our photos* | | | | *TACO* | | | |
|---|---|---|---|---|---|---|---|---|
| | AP50 | Precision | Recall | F1 | AP50 | Precision | Recall | F1 |
| **Our model** | | | | | | | | |
| With PlastOPol | 56.7 | 75.5 | 38.9 | 51.4 | 61.2 | 83.5 | 37.0 | 51.3 |
| Without PlastOPol | 59.0 | 78.6 | 39.6 | 52.7 | 59.5 | 82.9 | 34.1 | 48.3 |
| **Other models** | | | | | | | | |
| Córdova et al. (2022) | - | - | - | - | 63.3 | 48.4 | 66.4 | 56.0 |
| Das et al. (2023) | - | - | - | - | 57.4 | 82.8 | 49.1 | 61.6 |
| Majchrowska et al. (2022) | - | - | - | - | 62.4 | - | - | - |

Majchrowska et. al (2022) included the extended TACO dataset in their performance evaluation.

The fact that our model is not perfectly accurate at detecting trash has a predictable impact on treatment effect estimates. First, we know that 21.4 percent of all detections are *false positives* due to a tendency of our model to identify other objects, typically stones or leaves, as trash. In our test set, we observe an average of 0.187 false positives per image (46 false positives for 246 images in the test set). Assuming that the number of false positives is unrelated to the treatment status, this implies that the average trash count in all experimental groups is biased upward by 0.187 pieces of trash. This does not, however, affect (absolute) treatment effects, as the bias cancels out when comparing different experimental groups. A second bias is related to *false negatives*. The recall of 0.4 suggests that our model misses a bit more than half of trash on our images (i.e., the false negative rate is 0.6). Assuming that the capacity of the model to detect a given trash piece is unrelated to the treatment status, the average reported trash count for each experimental group thus corresponds to only 40 percent of the true trash count. Consequently, the treatment effect, reported in pieces of trash, is underestimated by the same factor. As the reduced differences between treatment groups are accompanied by a lower variance, this bias disappears when effects are reported in standard deviations. In summary, under plausible assumptions, raw group means and treatment effects can be biased due to the occurrence of false positives and false negatives, while standardized effects are not. When reporting on group means or effects in pieces of trash (or percent), we will thus also present results


Figure 5: Illustration of Deep Learning Model Performance

The image shows model predictions for an example image. The decimal number represents the confidence of the model.

accounting for these two biases. This is done using the following simple correction:

$$Y_g = (\hat{Y}_g - FP) \cdot \frac{1}{recall} \tag{1}$$

where $Y_g$ is the true average trash count in treatment group $g$ after applying the bias correction to the predicted trash count $\hat{Y}_g$, $FP$ is the average number of false positives per image in our test set and thus 0.187, and $recall$ is the overall share of true trash pieces that are correctly detected in our test set and thus 0.396.[12]

[12]To correct the bias in (non-standardized) treatment effects, we only need to multiply the raw treatment effect by $\frac{1}{recall}$, since the first part of the equation cancels out. Note further that the assumptions that the probability of false positives and false negatives is unrelated to the treatment status is likely to be only approximately true. In the case of false positives, one could argue that false detections are more likely in cleaner images (where less space is covered by trash). This would introduce an additional downward bias in treatment effects, as contamination in the (cleaner) treatment group is overstated more strongly compared to the (dirtier) control group. In this case, our corrected treatment effect estimate would represent a lower bound for the true effect. A similar argument holds for false positives. If trash is harder to detect in dirtier environments (where the model may struggle to tell many different trash pieces apart), our corrected estimates would still be too conservative.

## 5.3 Baseline Characteristics

Table 2 shows that contamination levels and survey responses at baseline are well-balanced across experimental groups. Only for two of our main variables, the share of people engaging in voluntary work and the percentage of an endowment people choose to donate in a framed dictator game (altruism), we report significant differences between groups.

Our deep learning model detects approximately one piece of garbage on the average image. Based on Equation 1 in Section 5.2, this implies that an average image contains about 2 real pieces of trash. Considering that the photos were taken randomly along all streets and not specifically in places with garbage, this indicates substantial solid waste contamination. For the average community, this corresponds to roughly 1,000 visible pieces of trash on our images alone. There is considerable variance between communities with 0.19 detections (hardly any real pieces) on the average image in the least polluted community and 3.12 detections ($\approx$ 7.38 real pieces) in the most polluted community. Similarly, enumerators rated the average site across all communities as a 2 ("a bit polluted") on a scale from 1 to 4. Community averages based on these subjective enumerator assessments range from 1.37 in the cleanest community to 2.78 in the dirtiest community.

The average survey respondent is 43 years old, and 75 percent of respondents are female. About two thirds of the individuals in our sample have not completed any educational degree (no schooling: 20%, incomplete primary: 45%), 11 percent have a primary degree, 19 percent have completed high school, and 5 percent possess a tertiary degree. On average, respondents believe that roughly 60 percent of people in their community litter, that 70 percent of people in their community disapprove of littering, and that 55 percent of people in their community would punish litterers with a disapproving gesture. The average community has about 300 residents, corresponding to roughly 90 households. People tend to know each other, with the average person reporting that 70 percent of community members are known and 40 percent are friends or family. Approximately 20 percent of respondents belong to a community organization and 30 percent report having done voluntary work for the community in the last month.

Table 2: Balance at Baseline

| | Control | T1: TD | T2: BU | P-value | N |
|---|---|---|---|---|---|
| **Photo trash count: Contamination** | | | | | |
| Kernel approach wrt. midline (count) | 0.911 | 0.988 | 0.952 | 0.804 | 60709 |
| Kernel approach wrt. endline (count) | 0.899 | 0.958 | 0.957 | 0.831 | 65673 |
| Raster approach (count) | 0.908 | 0.993 | 1.010 | 0.458 | 10740 |
| Raw averages (count) | 0.933 | 0.951 | 0.975 | 0.957 | 120 |
| **Enumerator assessments: Contamination** | | | | | |
| Kernel approach wrt. midline (1-4) | 1.990 | 1.981 | 1.976 | 0.245 | 12216 |
| Kernel approach wrt. endline (1-4) | 1.988 | 1.960 | 1.998 | 0.268 | 13163 |
| Raster approach (1-4) | 2.014 | 1.974 | 1.998 | 0.344 | 4272 |
| Raw averages (1-4) | 2.011 | 1.962 | 2.021 | 0.394 | 120 |
| **Survey: Sociodemographics** | | | | | |
| Female | 0.740 | 0.756 | 0.722 | 0.392 | 2421 |
| Age | 42.881 | 42.186 | 43.417 | 0.433 | 2418 |
| Education | 2.526 | 2.468 | 2.394 | 0.362 | 2421 |
| Poverty (1-5) | 3.162 | 3.046 | 2.979 | 0.178 | 2354 |
| Community size | 305.349 | 300.439 | 294.964 | 0.784 | 2420 |
| **Survey: Contamination and waste disposal** | | | | | |
| Perceived cleanliness (1-5) | 3.146 | 3.103 | 3.118 | 0.587 | 2421 |
| Appropriate disposal (%) | 0.452 | 0.532 | 0.489 | 0.530 | 2421 |
| **Survey: Social norms** | | | | | |
| Littering (%) | 0.599 | 0.603 | 0.571 | 0.131 | 2418 |
| Littering is bad (%) | 0.705 | 0.687 | 0.665 | 0.133 | 2413 |
| Punish littering (%) | 0.565 | 0.562 | 0.532 | 0.108 | 2417 |
| **Survey: Social capital** | | | | | |
| Strong ties (%) | 0.339 | 0.404 | 0.312 | 0.259 | 2419 |
| Weak ties (%) | 0.695 | 0.712 | 0.645 | 0.276 | 2375 |
| Trust (1-5) | 3.485 | 3.499 | 3.629 | 0.249 | 2421 |
| Organizations (%) | 0.174 | 0.191 | 0.207 | 0.445 | 2421 |
| Voluntary work (%) | 0.252 | 0.296 | 0.338 | 0.036 | 2415 |
| Altruism (%) | 0.439 | 0.454 | 0.406 | 0.038 | 2421 |

The last column indicates the p-value of a joint F-test that each treatment dummy coefficient is equal to 0. Education refers to highest completed degree: None = 1, incomplete primary = 2, complete primary = 3, high school degree = 4), technical = 5, and university degree = 6. Standard errors are clustered at the community level.

Table 3: Community Activities Summary Statistics

| Activity | *During intervention* | | | *After intervention* | |
|---|---|---|---|---|---|
| | Completed activities | % with one or more | Participants | Completed activities | % with one or more |
| Sessions | 7.20 | 0.95 | 18.90 | 0.08 | 0.08 |
| Workshops | 3.50 | 0.95 | 18.40 | 0.00 | 0.00 |
| Community activities | 3.60 | 0.95 | 19.10 | 0.03 | 0.03 |
| Cleanup campaigns | 3.40 | 0.92 | 18.70 | 2.05 | 0.67 |
| Meetings | 0.90 | 0.90 | 21.70 | 1.38 | 0.33 |

The period for both during as well as after the intervention spans a total of 4 months. Intervention activities were recorded by facilitators. The post-intervention data was obtained through phone calls to the responsible person at each community. The number of participants is conditional on the activity taking place.

# 6 Empirical Results

This chapter discusses the main findings of our study. We will (i) take a look at how the program was implemented, (ii) present our main findings, and finally (iii) use insights from the survey and the activity registry to discuss potential mechanisms based on the theoretical framework discussed in Section 2.

## 6.1 Program Implementation

Our data suggests that both interventions were successfully implemented. In the top-down intervention, an average of 3.9 cleanups were conducted in each community (i.e., roughly one per month), and no community received fewer than 3 cleanups. The bottom-up intervention was implemented in 95 percent of the communities assigned to this condition, with an average of 0.9 meetings, 14.3 educational activities (two-hour sessions, practical workshops, or community activities such as movie nights), and 3.5 collective cleanups per community (Table 3). This corresponds to a total of 18.6 activities and 4.7 monthly activities per community. About 20 community members, corresponding to 7 percent of the population, participated in a typical activity in this intervention arm.

For the bottom-up intervention, we also collected data during the four months following the intervention to observe whether efforts to keep the community clean continued. In the post-intervention period, around two thirds of the communities re-

port conducting at least one cleanup campaign, with 2.05 campaigns (0.51 per month) in the average community. Similarly, about one third of all communities conducted at least one meeting about solid waste contamination. In line with expectations, educational activities were largely discontinued in the post-intervention period.

Our survey data shows that the sudden increase in activities related to solid waste management activities associated with the interventions did not go unnoticed (Table A1). The bottom-up intervention had a large and significant immediate impact on the percentage of respondents who reported being aware of various waste-related activities – namely community meetings, education sessions, or collection campaigns – in their community within the past four months. In addition, it raised the number of people claiming to have participated in each of these activities. The top-down intervention also increased the number of individuals observing or participating in cleaning efforts, though to a lesser extent than the bottom-up intervention. For the participatory treatment, with the exception of educational sessions, substantial effects on all activities persist into the post-intervention period, while no lasting impacts are observed for the top-down intervention. Overall, our activity registries and survey data consistently indicate proper implementation of both interventions according to the specifications of each experimental group.

## 6.2 Program Effects

To assess the causal effect of the two treatments on contamination levels for each post-treatment $wave \in \{midline, endline\}$, we use

$$Y_{iv}^{wave} = \beta_1 T_1 + \beta_2 T_2 + \delta Y_{iv}^{baseline} + \mu_s + \epsilon_{iv} \tag{2}$$

where $Y_{iv}^{wave}$ are midline or endline outcomes for kernel or raster cell $i$ in village $v$; $T_1$ and $T_2$ are treatment indicators for treatment 1 (top-down) and treatment 2 (bottom-up); $Y_{iv}^{baseline}$ is the baseline kernel or cell contamination level; and $\mu_s$ are strata fixed effects. With the exception of the models analyzing effects on raw community averages, standard errors are clustered at the community level. For survey outcomes, we extend Equation 2 by adding individual-level controls for sex, age, and education.

For our main outcome based on *trash counts*, we find large immediate effects for both interventions (Figure 6, Table 4, and Table A2). The top-down intervention reduced solid waste contamination by 0.7–0.8$\sigma$ or roughly 0.5–0.6 detected trash pieces on the average image ($p < 0.01$). Applying our bias correction, this corresponds

to a decrease of about 1.35 trash pieces or 39 percent. The bottom-up intervention had a significantly smaller ($p < 0.05$), but still substantial impact of 0.5–0.6$\sigma$ or approximately 0.4 trash detections ($p < 0.01$). This translates into an effect of 1 piece of garbage or 28 percent. Estimates for long-term impacts reveal a stark depletion of effects for both treatments four months after the end of the intervention. Communities in the top-down intervention outperform the control group by only 0.1$\sigma$, an effect that is statistically indistinguishable from zero at conventional levels (p $\approx$ 0.2). This corresponds to a depreciation by about 0.6$\sigma$ or 80 percent compared to immediate effects. For the community-driven intervention, we document a slightly larger and statistically significant long-term effect of 0.2$\sigma$ or 0.25 trash pieces ($p < 0.05$). The depletion of immediate impacts corresponds to about 0.3$\sigma$ or 60 percent. While the absolute depreciation (i.e., in standard deviations) is significantly lower in the bottom-up intervention than in the top-down intervention ($p < 0.01$), the difference in relative depletion rates is not statistically significant ($p = 0.2$, see Table A4).[13]

As a robustness check, we compare these results with effects based on *subjective enumerator assessments* (lower panel in Table 4 and Table A2). In line with our

[13]Whether absolute or relative depreciation is more appropriate depends on the assumptions about counterfactual trends in the two groups. Under a parallel trends assumption, absolute depreciation would be the correct measure. On the other hand, if we assume convergence back to the level of the control group, we should use a relative measure. Since the second scenario seems more plausible, we use relative depreciation as our main measure.

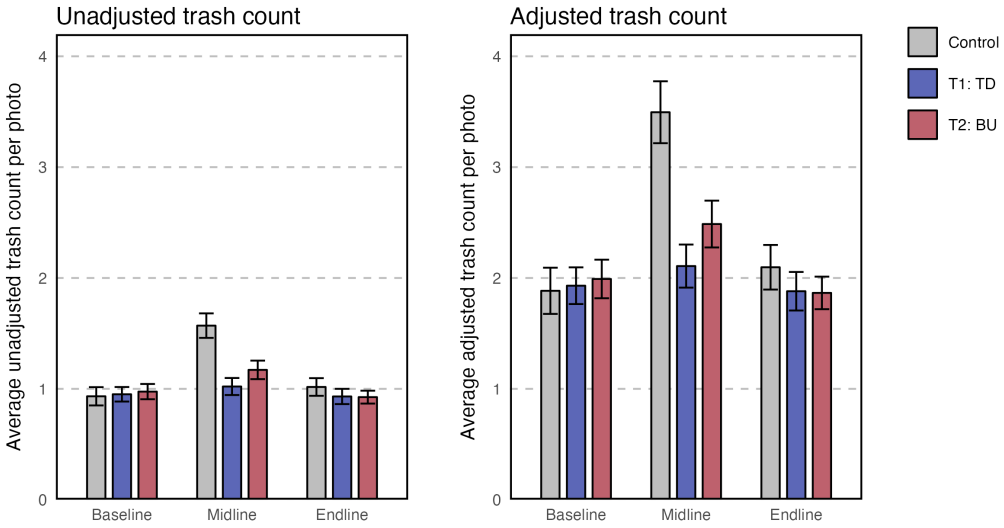



Figure 6: Average Trash Count per Image by Wave and Treatment

The baseline measurement was conducted in September and October 2022, the midline in March 2023, and the endline in July 2023. The increase in the amount of litter during the midline assessment is likely to be a seasonal effect, as this was the only measurement conducted during the dry season, when waste is less likely to be washed away or covered by vegetation.

Table 4: Main Results Based on Trash Detection and Enumerator Assessments

| | *Immediate effects* | | | | *Long-term effects* | | | |
|---|---|---|---|---|---|---|---|---|
| | T1: TD | T2: BU | T2 - T1 | N | T1: TD | T2: BU | T2 - T1 | N |
| **Photo trash detection** | | | | | | | | |
| Kernel approach | -0.755*** | -0.540*** | 0.215** | 60709 | -0.129 | -0.199** | -0.070 | 65673 |
| | (0.129) | (0.130) | (0.107) | | (0.107) | (0.100) | (0.096) | |
| Raster approach | -0.727*** | -0.471*** | 0.256** | 10740 | -0.134 | -0.200* | -0.066 | 10740 |
| | (0.142) | (0.158) | (0.128) | | (0.116) | (0.109) | (0.106) | |
| Raw averages | -0.792*** | -0.604*** | 0.188 | 120 | -0.175 | -0.248** | -0.073 | 120 |
| | (0.119) | (0.120) | (0.121) | | (0.113) | (0.114) | (0.115) | |
| **Enumerator assessments** | | | | | | | | |
| Kernel approach | -0.932*** | -0.771*** | 0.161 | 12216 | -0.047 | -0.057 | -0.009 | 13163 |
| | (0.184) | (0.178) | (0.193) | | (0.210) | (0.173) | (0.230) | |
| Raster approach | -0.803*** | -0.616*** | 0.187 | 4272 | 0.007 | -0.006 | -0.012 | 4272 |
| | (0.185) | (0.176) | (0.185) | | (0.218) | (0.171) | (0.234) | |
| Raw averages | -0.999*** | -0.853*** | 0.146 | 120 | -0.089 | -0.088 | 0.000 | 120 |
| | (0.204) | (0.204) | (0.206) | | (0.195) | (0.195) | (0.197) | |

Results reported in standard deviations at the community level. Controls include contamination at baseline and strata fixed effects. Standard errors are clustered at the community level for the kernel and the raster approach. $*p < 0.1$; $**p < 0.05$; $***p < 0.01$

main outcome based on trash detections, we observe large immediate effects for both interventions. However, the difference between the two treatments is no longer significant and the long-term effects disappear. A likely explanation for these deviations is that the subjectivity of the ratings introduced considerable noise into the assessment measure. While the resulting measurement errors should be uncorrelated with the treatment, they are clustered at the community level (because enumerators always covered an entire community), which considerably reduces the precision of the estimates. Indeed, if we include enumerator fixed effects, estimates for long-term effects based on enumerator assessments change markedly, indicating significant long-term effects for both treatments (Table A3). Our main results based on trash detection are less sensitive to the inclusion of these fixed effects. This underscores the advantages of the objective contamination measure that we derive using deep learning.

Our *survey results* show that the changes in solid waste pollution did not go unnoticed (Table 5, panel "Contamination and waste disposal"). In line with our findings from trash detections and contamination ratings, perceived cleanliness improved significantly by about $0.16\sigma$ immediately after both treatments. In addition, both interventions had a significant short-term impact on people's recycling practices, with a 10 percentage point increase in the share of people recycling at least one type of solid waste. For the community-driven intervention, we further report a significant immediate improvement in self-reported waste disposal practices. The share of people indicating that they use an official deposit or a garbage truck to dispose of their waste, as opposed to burning, burying, or dumping it, increased by about 10 percentage points. Estimates for long-term effects show that impacts on perceived cleanliness persist, with effects of $0.17$–$0.18\sigma$ for both interventions. For the bottom-up intervention, we further report a sustained increase in the share of people indicating appropriate waste disposal by roughly 7 percentage points (50% depreciation compared to immediate effects). Recycling effects disappear in the long run for both interventions.

## 6.3 Discussion

Our complementary data from surveys, activity records, and interviews allow us to explore the mechanisms behind the observed effect patterns using the theoretical framework we propose in Section 2. In this chapter, we shed light on two key policy questions arising from our project. We will (i) explore the extent to which the community-driven intervention may have addressed informational, organizational, and credit constraints,

Table 5: Survey Regression Results

| | *Immediate effects* | | | *Long-term effects* | | |
|---|---|---|---|---|---|---|
| | T1: TD | T2: BU | T2 - T1 | T1: TD | T2: BU | T2 - T1 |
| **Contamination and waste disposal** | | | | | | |
| Perceived cleanliness (sd) | 0.158* | 0.163** | 0.005 | 0.173** | 0.182** | 0.009 |
| | (0.085) | (0.083) | (0.076) | (0.079) | (0.082) | (0.078) |
| Appropriate disposal (%) | 0.040 | 0.137*** | 0.097** | -0.005 | 0.067* | 0.072* |
| | (0.044) | (0.041) | (0.046) | (0.042) | (0.039) | (0.040) |
| Recycling (%) | 0.083** | 0.112*** | 0.029 | -0.026 | -0.001 | 0.025 |
| | (0.041) | (0.037) | (0.029) | (0.040) | (0.040) | (0.040) |
| **Social norms** | | | | | | |
| Littering (%) | -0.065** | -0.104*** | -0.039 | 0.011 | -0.019 | -0.030 |
| | (0.028) | (0.026) | (0.026) | (0.019) | (0.019) | (0.021) |
| Littering is bad (%) | -0.015 | -0.037** | -0.022 | -0.001 | 0.004 | 0.005 |
| | (0.020) | (0.017) | (0.020) | (0.018) | (0.018) | (0.016) |
| Punish littering (%) | 0.027 | 0.003 | -0.024 | 0.017 | 0.036* | 0.019 |
| | (0.028) | (0.025) | (0.027) | (0.022) | (0.020) | (0.022) |
| **Social capital** | | | | | | |
| Strong ties (%) | 0.059 | 0.059* | 0.001 | 0.024 | 0.006 | -0.018 |
| | (0.038) | (0.034) | (0.038) | (0.031) | (0.029) | (0.032) |
| Weak ties (%) | -0.000 | 0.007 | 0.007 | 0.006 | 0.023** | 0.017** |
| | (0.017) | (0.014) | (0.016) | (0.010) | (0.009) | (0.008) |
| Trust (sd) | 0.113 | 0.031 | -0.082 | 0.013 | 0.054 | 0.041 |
| | (0.095) | (0.090) | (0.096) | (0.086) | (0.080) | (0.081) |
| Organizations (%) | -0.007 | 0.042 | 0.049* | -0.009 | 0.011 | 0.020 |
| | (0.023) | (0.026) | (0.028) | (0.023) | (0.025) | (0.027) |
| Voluntary work (%) | 0.011 | 0.159*** | 0.149*** | 0.035 | 0.129*** | 0.094*** |
| | (0.031) | (0.034) | (0.037) | (0.029) | (0.031) | (0.031) |
| Altruism (%) | -0.014 | -0.001 | 0.013 | 0.012 | 0.017 | 0.005 |
| | (0.019) | (0.020) | (0.020) | (0.017) | (0.015) | (0.015) |

Sample sizes are n = 2066 for the estimation of immediate effects and n = 1832 for long-term effects. Social norm variables refer to beliefs about other people's behavior. Controls include strata fixed effects, sex, age, education (dummies), and the baseline value for the respective outcome. Standard errors are clustered at the community level. $*p < 0.1$; $**p < 0.05$; $***p < 0.01$

and (ii) discuss if impacts were mainly driven by cleaning efforts or by changes in littering behavior.

### 6.3.1 How Can Bottom-Up Development Alleviate Constraints to Collective Action?

Community-based initiatives can mitigate *information constraints* in two ways. Residents could become more aware of the problem and of effective means to address it, inducing them to lower their thresholds for cooperative behavior, or they could correct their (potentially biased) beliefs about the number of others who are contributing. In either case, a successful intervention would induce a gradual shift toward a stable higher equilibrium, as more and more individuals join the camp of cooperators. Thus, we should observe increasing participation in cleanups over time, and a gradual and sustained reduction in littering. We find limited evidence for either of these patterns. The number of participants in the average cleanup was stable throughout the intervention and the post-intervention period, suggesting that when campaigns were organized, similar numbers of residents continued to participate.[14] Results for

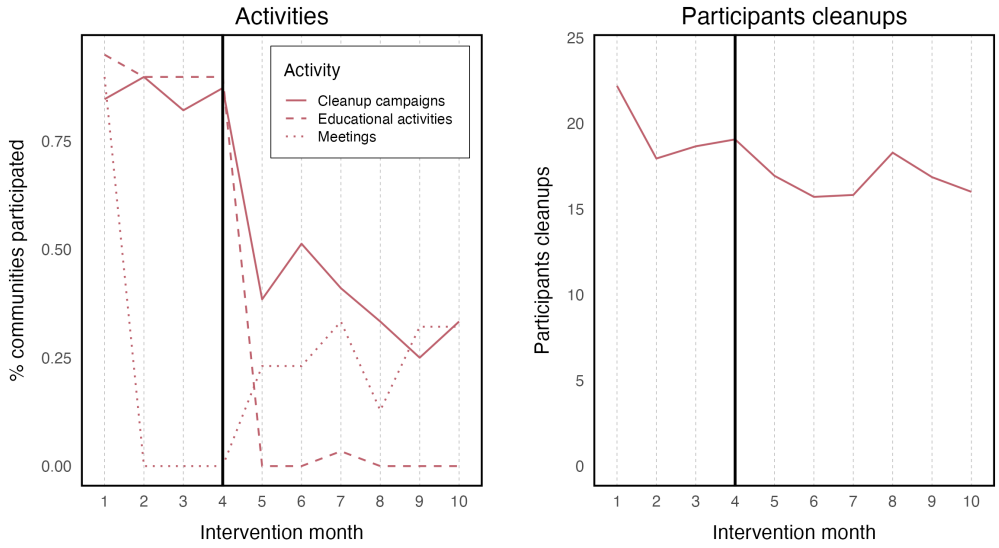



Figure 7: Collective Action in the Bottom-Up Intervention Over Time

The left figure shows the number of participants in the average cleanup per month. The right figure documents what share of communities realized different types of activities in each month. The black line corresponds to the end of the intervention.

littering behavior are inconclusive as well. In line with potential information effects,

[14]We also inspected separate trends for all communities to see if averages mask diverging trends toward high levels of collective action in some communities and low levels in others, and find no support for this hypothesis.

survey respondents tend to be much more positive about their own littering behavior than that of their neighbors, and the community-driven intervention narrowed this gap: The bottom-up treatment reduced the proportion of residents who respondents believed to engage in littering by about 10 percentage points ($p < 0.01$, Table 5, panel "Social norms").[15] However, a similar change in descriptive norms, namely a reduction by 7 percentage points, occurred in the top-down intervention, and both effects disappear in the long run. In addition, no clear effects are found for all other outcomes related to littering norms and behaviors (Table A7). This is consistent with our registry data on the amount of trash collected during cleanups (see Figure 8). While the average number of garbage bags collected in bottom-up cleanups decreases by about 50 percent from the first to the last intervention month – a potential indication for reduced littering – we observe a very similar and statistically indistinguishable decline for the traditional top-down intervention. Similarly, we do not find a steeper reduction in the total number of working hours required for cleaning in the bottom-up intervention than in the top-down intervention.[16]

A key argument for community-driven interventions is that they alleviate *organizational constraints* to collective action, thereby enabling communities to coordinate the provision of public goods. If groups manage to get organized and agree on joint actions, this would lead to an immediate shift towards a higher equilibrium. This is consistent with the observation that participation in cleanups was high from the first month and remained stable throughout the intervention. This suggests that a sufficiently large number of community members were willing to (conditionally) commit time to a cleaner environment from the outset, and that the intervention succeeded in bringing them together to do so. Organizational effects can either be limited to the intervention period, where paid facilitators take the lead in mobilizing for collective action, or, ideally, be enduring if communities succeed in strengthening local institutions. Our survey data provides only limited support for the latter type of effects (Table A12, panel "Social Capital"). We present clear evidence that the bottom-up intervention increased engagement in voluntary work (likely through participation in

[15] While only 15 percent of people say they have littered in the past month, the average person believes that 60 percent of others have done so (see Table A5). Note, however, that this does not necessarily indicate that people's perceptions are biased, as responses about self-reported behavior may be driven by a social desirability bias.

[16] Note that the reduction in collected waste (or the time used to do so) over the course of the intervention is not only driven by social norms, but also by the fact that waste in the early months may have accumulated over longer periods of time. For work hours, we may also observe changes in the efficiency of the group, as a large group may be more productive at collecting large amounts of waste compared to smaller amounts.

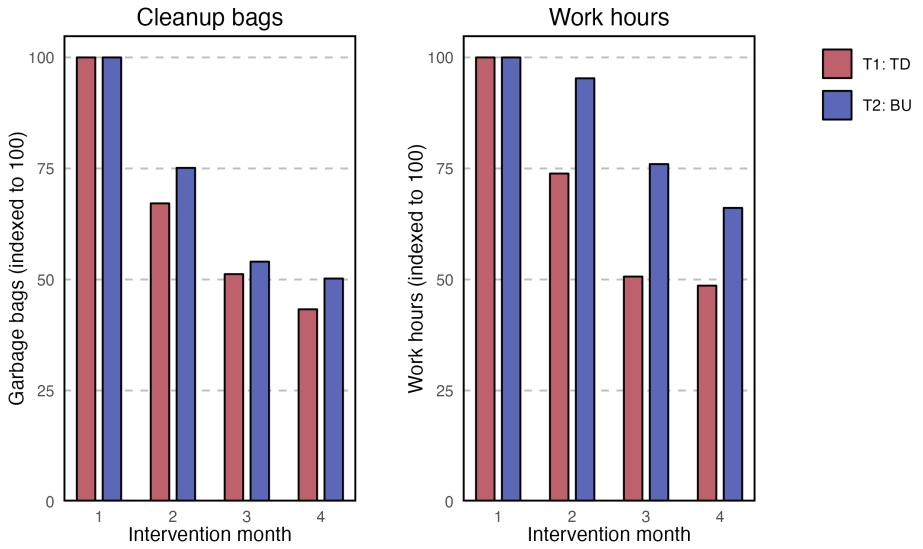



Figure 8: Cleanup Statistics Over Time by Treatment

The left figure shows the number of garbage bags filled in the average cleanup over time. The right figure documents the average number of working hours (added over all contributors) needed for the task. To make trends comparable, results are expressed as percentages of the treatment-specific average of bags (T1: 4.7, T2: 9.5) and work hours (T1: 11.6, T2: 15.1) in the first month.

the cleanups) and suggestive evidence that it improved social ties (strong ties in the short run and weak ties in the long run), but we find no immediate or lasting effects on trust, membership in organizations, and altruism. Together with the steep decline in collective action immediately after the end of the intervention (Figure 7), these findings suggest that much of the success in the organizational dimension was tied to the presence of the facilitator. Mobilizing for collective action is time-consuming and demands a disproportionate contribution from the person (or persons) taking the lead in the endeavor. If people are willing to contribute about as much as others do, no such leader will emerge to take over from the facilitator. This aligns with qualitative evidence from interviews with the community members who assumed responsibility after the departure of the facilitator, where "time constraints to mobilize people and organize campaigns" emerged as the most frequently mentioned challenge to project continuation. It is also consistent with our heterogeneity analyses, which suggest that the bottom-up intervention had a higher short-term impact at lower initial levels of social capital ($p = 0.07$), where organizational constraints addressed by the temporary leadership of the facilitator might have been more binding (Figure A2).

A final channel through which community-driven interventions could facilitate the provision of public goods is by easing *credit constraints*. This mechanism is less relevant for the particular public good we study, because a clean environment can

be maintained with a minimal financial investment. Removing litter from the streets can be accomplished with voluntary work and a few plastic bags, and communities typically took advantage of the municipal garbage truck or a resident's journey to transport the collected waste to an official depot.[17] Accordingly, no financial transfer was made to communities in the context of the project. A notable exception is the provision of snacks to volunteers during the collective cleanups. However, few interviewees mentioned the lack of such provisions as a key constraint to the continuation of collective action activities in the post-intervention phase. Hence, it appears unlikely that the short-term and partial long-term success of the bottom-up treatment was driven by mechanism related to credit constraints.

Overall, our results are most consistent with a theoretical framework where many individuals are willing to contribute to public goods as long as others do so too, but struggle to coordinate in the absence of a dedicated leader. Community-based interventions have the potential to build leadership and strengthen the local institutions needed to coordinate collective action. However, achieving transformations that outlive the presence of a paid facilitator may often be beyond the scope of a four-month intervention.

### 6.3.2 Should Solid Waste Interventions Aim for Cleanups or Changes in Littering Norms?

Communities can pursue two interrelated strategies to provide the public good of a clean environment. They can either mobilize for regular collective cleanups, or establish informal institutions to discourage littering in the first place. The complementary data discussed in the previous section can also be used to gauge the importance of each of these channels. The cleanups clearly played an important role, as the typical community in the bottom-up group conducted about one monthly cleanup with roughly 20 participants (Table 3), and the intervention had a large effect on the proportion of survey respondents who reported observing or participating in such campaigns (Table A1). As discussed above, the results for littering behavior are more mixed. The shift in beliefs about other people's littering practices induced by the bottom-up intervention was mirrored by a similar change for the top-down intervention, and did not persist

[17] Our post-intervention interviews with community leaders reveal that removing the collected trash from the community was an major challenge in a few communities where the municipality charged a (usually substantial) fee to send the garbage truck or a private vehicle had to be hired. However, no financial support was provided for the removal of the collected waste during the intervention, and facilitators successfully devised solutions in coordination with community members. This underscores that waste transportation was primarily an organizational challenge rather than a financial one.

after the end of the program (Table 5). Similarly, while we find that the amount of trash collected in monthly campaigns decreased substantially over time, this decline was not greater for the bottom-up than for the top-down intervention (Figure 8). A plausible explanation is that individuals form their beliefs about other people's littering behavior based on the amount of waste they observe on the streets, and modify their own practices in response to this inferred social norm. This is in line with ample research documenting that people are substantially less likely to litter in clean than in dirty environments (Cialdini et al., 1990; Ramos and Torgler, 2012; Bateson et al., 2013; Sagebiel et al., 2020). As the two interventions lead to similar reductions in solid waste pollution due to the cleaning efforts, individuals in both treatment groups may have concluded that fewer people are littering and, potentially, adapted their own behaviors accordingly.

Overall, our data points to the cleanups as the main driver of the success of both interventions. While a shift in littering norms may also have played a role, our data does not provide much support for the hypothesis that the community-driven intervention was more effective in inducing this change. Viewed through the theoretical framework developed in Section 2, our findings suggest that interventions focusing on changes in littering behavior alone are unlikely to be sustainable. Maintaining a clean environment without any cleaning requires perfect adherence to a non-littering norm by all community members and visitors. If a small minority litters regardless of what others do, waste will accumulate, inducing conditional cooperators to start littering as well. As a result, communities will revert to a low equilibrium where everyone litters except those who are willing to cooperate irrespective of what others do. In contrast, reaching a stable high equilibrium through collective cleanups requires the cooperation of only a small group of committed residents, which may be much easier to achieve. The positive dynamics induced by the cleanups may then be reinforced by changes in littering behavior, as people are less likely to dump waste into clean environments.

# 7 Conclusion

Participatory bottom-up initiatives have become a popular alternative to the conventional top-down approach to the provision of public goods. While several recent studies have evaluated such programs, their effectiveness has not yet been compared to the more traditional top-down strategies they often replace. In this study, we present the results of a randomized controlled trial comparing the effectiveness of bottom-

up and top-down strategies to address local waste pollution in rural El Salvador. Immediate effects on contamination level are substantial for both interventions, but significantly larger in the top-down intervention. Four months after the end of the intervention, we observe a strong diminution of these effects, which is only slightly less pronounced in the bottom-up intervention. Our complementary data suggests that the presence of the facilitator may have helped the communities overcome organizational constraints to collective action, but many communities were unable to sustain these efforts independently.

Our findings have important implications for the policy debates around *bottom-up development strategies*. We find that while community-driven initiatives can indeed successfully promote the provision of local public goods, they are not always more effective in doing so than top-down interventions. More specifically, our findings highlight that many individuals are willing to voluntarily contribute to public goods, and involving them in the development of their communities may indeed produce more sustainable outcomes. However, sustaining the high levels of collective action needed to provide public goods at optimal levels requires strong informal institutions and local leadership. Building such capabilities may be beyond the scope of a short-term intervention, and entail considerable costs, including facilitation expenses for the implementing organization, and opportunity costs for participants. A combined approach that strengthens government institutions alongside communities may thus be a promising long-term strategy. How much and what kind of bottom-up participation produces the most sustainable and cost-effective solutions is an important question for future research. In this context, two important limitations of our study should not go unmentioned. First, our study is based on the provision of a specific public good in a particular context, meaning that more research is needed to draw confident conclusions about the relative effectiveness of bottom-up development initiatives. Second, the top-down intervention in our study was implemented by a committed NGO rather than a governmental institution and its effectiveness may thus be an upper bound for what a state-led arrangement in developing countries could achieve. Nevertheless, by providing a first rigorous comparison between a top-down and a bottom-up provision strategy, our study constitutes a critical starting point for the necessary discussion on the relative effectiveness of different approaches to local public good provision.

The findings presented in this study are also relevant to policy makers seeking to devise effective *solid waste management* strategies. Based on our findings and theoretical considerations, we draw two cautious conclusions. First, raising awareness and empowering communities to tackle the waste problem can be an important part of the

solution, but the assumption that a one-time investment in facilitation will effectively solve the problem forever is clearly unrealistic. Second, picking up waste may be more critical to the success of waste management interventions than inducing changes in littering behavior. Although shifts in social norms may reinforce the positive trend induced by cleaning efforts, interventions focusing exclusively on littering behavior are unlikely to lead to a stable high-level equilibrium. In light of the rapid increase in solid waste production in developing countries and the scarcity of research on how best to address the problem, these are crucial and timely insights.

Finally, our study also advances the use of *deep learning methods* to understand, track, and improve outcomes related to global development. A rapidly growing body of research has shown that a variety of outcomes, including poverty, education or agricultural yields, can be predicted from alternative data sources such as satellite imagery (Kuwata and Shibasaki, 2015; Jean et al., 2016; Yeh et al., 2020), phone records (Blumenstock et al., 2015), social media posts (Jakob and Heinrich, 2023), or Google Street View images (Suel et al., 2019). However, the main focus of this literature is on proof-of-concept, and applications that bridge real gaps in data availability remain scarce. By using image data and deep learning to derive an objective measure of contamination, our study provides such an application. We illustrate how predicted measures can be used in an experimental setup, and how potential biases can be accounted for. As deep learning methods continue to penetrate the social sciences, such applications and discussions of the biases they may introduce, are likely to become increasingly important.

# 3 References


Anderson, L. R., Mellor, J. M., and Milyo, J. (2004). Social capital and contributions in a public-goods experiment. *American Economic Review*, 94(2):373–376.

Arcand, J.-L. (2008). Does community driven development work? Evidence from Senegal. Available at SSRN: `http://dx.doi.org/10.2139/ssrn.1265231`.

Avdeenko, A. and Gilligan, M. J. (2015). International interventions to build social capital: Evidence from a field experiment in Sudan. *American Political Science Review*, 109(3):427–449.

Banerjee, A. V., Banerji, R., Duflo, E., Glennerster, R., and Khemani, S. (2010). Pitfalls of participatory programs: Evidence from a randomized evaluation in education in India. *American Economic Journal: Economic Policy*, 2(1):1–30.

Banerjee, A. V. and Duflo, E. (2007). The economic lives of the poor. *Journal of Economic Perspectives*, 21(1):141–167.

Bateson, M., Callow, L., Holmes, J. R., Redmond Roche, M. L., and Nettle, D. (2013). Do images of "watching eyes" induce behaviour that is more pro-social or more normative? A field experiment on littering. *PloS One*, 8(12):e82055.

Beath, A., Christia, F., and Enikolopov, R. (2013). Empowering women through development aid: Evidence from a field experiment in Afghanistan. *American Political Science Review*, 107(3):540–557.

Berger, J. (2021). Social tipping interventions can promote the diffusion or decay of sustainable consumption norms in the field. Evidence from a quasi-experimental intervention study. *Sustainability*, 13(6):3529.

Berger, J., Efferson, C., and Vogt, S. (2023). Tipping pro-environmental norm diffusion at scale: Opportunities and limitations. *Behavioural Public Policy*, 7(3):581–606.

Björkman, M., de Walque, D., and Svensson, J. (2017). Experimental evidence on the long-run impact of community-based monitoring. *American Economic Journal: Applied Economics*, 9(1):33–69.

Björkman, M. and Svensson, J. (2009). Power to the people: Evidence from a randomized field experiment on community-based monitoring in Uganda. *The Quarterly Journal of Economics*, 124(2):735–769.

Blumenstock, J., Cadamuro, G., and On, R. (2015). Predicting poverty and wealth from mobile phone metadata. *Science*, 350(6264):1073–1076.

Casey, K. (2018). Radical decentralization: Does community-driven development work? *Annual Review of Economics*, 10:139–163.

Casey, K., Glennerster, R., and Miguel, E. (2012). Reshaping institutions: Evidence on aid impacts using a preanalysis plan. *The Quarterly Journal of Economics*, 127(4):1755–1812.

Castaldi, G., Cecere, G., and Zoli, M. (2021). Smoke on the beach: On the use of economic vs behavioral policies to reduce environmental pollution by cigarette littering. *Economia Politica*, 38:1025–1048.

Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: A selective survey of the literature. *Experimental Economics*, 14:47–83.

Chitotombe, J. W. (2014). Interrogating factors associated with littering along road servitudes on Zimbabwean highways. *Environmental Management and Sustainable Development*, 3(1):181.

Cialdini, R. B., Reno, R. R., and Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, 58(6):1015.

Córdova, M., Pinto, A., Hellevik, C. C., Alaliyat, S. A.-A., Hameed, I. A., Pedrini, H., and Torres, R. d. S. (2022). Litter detection with deep learning: A comparative study. *Sensors*, 22(2):548.

Cowen, T. (1992). *Public goods and market failures: A critical examination*. Transaction Publishers.

Dahlman, C. J. (1979). The problem of externality. *The Journal of Law and Economics*, 22(1):141–162.

Das, D., Deb, K., Sayeed, T., Dhar, P. K., and Shimamura, T. (2023). Outdoor trash detection in natural environment using a deep learning model. *IEEE Access*.

Desai, R. M. and Olofsgård, A. (2019). Can the poor organize? Public goods and self-help groups in rural India. *World Development*, 121:33–52.

Dongier, P., Van Domelen, J., Ostrom, E., Ryan, A., Wakeman, W., Bebbington, A., Alkire, S., Esmail, T., and Polski, M. (2003). Community driven development. *World Bank Poverty Reduction Strategy Paper*, 1:303–327.

Duflo, E., Dupas, P., and Kremer, M. (2015). School governance, teacher incentives, and pupil–teacher ratios: Experimental evidence from Kenyan primary schools. *Journal of Public Economics*, 123:92–110.

Dur, R. and Vollaard, B. (2015). The power of a bad example: A field experiment in household garbage disposal. *Environment and Behavior*, 47(9):970–1000.

Fearon, J. D., Humphreys, M., and Weinstein, J. M. (2009). Can development aid contribute to social cohesion after civil war? Evidence from a field experiment in post-conflict Liberia. *American Economic Review*, 99(2):287–91.

Fischbacher, U., Gächter, S., and Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, 71(3):397–404.

Gächter, S. (2006). Conditional cooperation: Behavioral regularities from the lab and the field and their policy implications. In *CeDEx Discussion Paper No. 2006–03 Available at: http://hdl.handle.net/10419/67977*.

Glowacki, L. and von Rueden, C. (2015). Leadership solves collective action problems in small-scale societies. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1683):20150010.

Hardin, R. (1971). Collective action as an agreeable n-prisoners' dilemma. *Behavioral Science*, 16(5):472–481.

Hardin, R. (1982). *Collective action*. Johns Hopkins University Press., Baltimore, MD.

Humphreys, M., Sánchez de la Sierra, R., and Van der Windt, P. (2019). Exporting democratic practices: Evidence from a village governance intervention in Eastern Congo. *Journal of Development Economics*, 140:279–301.

Jack, B. K. and Recalde, M. P. (2015). Leadership and the voluntary provision of public goods: Field evidence from Bolivia. *Journal of Public Economics*, 122:80–93.

Jakob, M. S. and Heinrich, S. (2023). Measuring human capital with social media data and machine learning. *University of Bern Social Sciences Working Papers No. 46.* Available at: `https://ideas.repec.org/p/bss/wpaper/46.html`.

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., and Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*, 353(6301):790–794.

Kaza, S., Yao, L., Bhada-Tata, P., and Van Woerden, F. (2018). *What a waste 2.0: a global snapshot of solid waste management to 2050*. World Bank Publications.

Keser, C. and Van Winden, F. (2000). Conditional cooperation and voluntary contributions to public goods. *Scandinavian Journal of Economics*, 102(1):23–39.

Kuwata, K. and Shibasaki, R. (2015). Estimating crop yields with deep learning and remotely sensed data. In *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 858–861.

Labonne, J. and Chase, R. S. (2011). Do community-driven development projects enhance social capital? Evidence from the Philippines. *Journal of Development Economics*, 96(2):348–358.

Lewis, A., Turton, P., and Sweetman, T. (2009). *Litterbugs: How to deal with the problem of littering*. Policy Exchange.

Liu, J. H. and Sibley, C. G. (2004). Attitudes and behavior in social space: Public good interventions based on shared representations and environmental influences. *Journal of Environmental Psychology*, 24(3):373–384.

Majchrowska, S., Mikołajczyk, A., Ferlin, M., Klawikowska, Z., Plantykow, M. A., Kwasigroch, A., and Majek, K. (2022). Deep learning-based waste detection in natural and urban environments. *Waste Management*, 138:274–284.

Mansuri, G. and Rao, V. (2012). *Localizing development: Does participation work?* World Bank Publications.

McKenzie, D. and Bruhn, M. (2011). Tools of the trade: Doing stratified randomization with uneven numbers in some strata. Available at: https://blogs.worldbank.org/impactevaluations/tools-of-the-trade-doing-stratified-randomization-with-uneven-numbers-in-some-strata. Last accessed: 2023-10-29.

Mohan, S. and Joseph, C. P. (2021). Potential hazards due to municipal solid waste open dumping in India. *Journal of the Indian Institute of Science*, 101(4):523–536.

Nepal, M., Karki Nepal, A., Khadayat, M. S., Rai, R. K., Shyamsundar, P., and Somanathan, E. (2023). Low-cost strategies to improve municipal solid waste management in developing countries: Experimental evidence from Nepal. *Environmental and Resource Economics*, 84(3):729–752.

Netzer, Y., Wang, T., Coates, A., Bissacco, A., Wu, B., and Ng, A. Y. (2011). Reading digits in natural images with unsupervised feature learning. In *NIPS Workshop on Deep Learning and Unsupervised Feature Learning 2011*.

Nguyen, T. C. and Rieger, M. (2017). Community-driven development and social capital: Evidence from Morocco. *World Development*, 91:28–52.

Nkwocha, E. E. and Okeoma, I. O. (2009). Street littering in Nigerian towns: Towards framework for sustainable urban cleanliness. *African Research Review*, 3(5).

Olken, B. A. (2007). Monitoring corruption: Evidence from a field experiment in Indonesia. *Journal of Political Economy*, 115(2):200–249.

Olson, M. (1971). *The logic of collective action: Public goods and the theory of groups, with a new preface and appendix*. Harvard University Press.

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge University Press.

Ostrom, E. (1999). Coping with tragedies of the commons. *Annual Review of Political Science*, 2(1):493–535.

Proença, P. F. and Simoes, P. (2020). Taco: Trash annotations in context for litter detection. *arXiv preprint arXiv:2003.06975*.

Raffler, P., Posner, D. N., and Parkerson, D. (2019). The weakness of bottom-up accountability: Experimental evidence from the Ugandan health sector. *Innovations for Poverty Action Working Paper*.

Ramos, J. and Torgler, B. (2012). Are academics messy? Testing the broken windows theory with a field experiment in the work environment. *Review of Law and Economics*, 8(3):563–577.

Rangoni, R. and Jager, A. (2017). Social dynamics of littering and adaptive cleaning strategies explored using agent-based modelling. *The Journal of Artificial Societies and Social Simulation*, 20(2):1.

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 779–788.

Sagebiel, J., Karok, L., Grund, J., and Rommel, J. (2020). Clean environments as a social norm: A field experiment on cigarette littering. *Environmental Research Communications*, 2(9):091002.

Saguin, K. (2018). Why the poor do not benefit from community-driven development: Lessons from participatory budgeting. *World Development*, 112:220–232.

Sahin, S. G., Eckel, C., and Komai, M. (2015). An experimental study of leadership institutions in collective action games. *Journal of the Economic Science Association*, 1:100–113.

Schultz, P. W. (1999). Changing behavior with normative feedback interventions: A field experiment on curbside recycling. *Basic and Applied Social Psychology*, 21(1):25–36.

Sheely, R. (2013). Maintaining local public goods: Evidence from rural Kenya. In *CID Working Papers 273, Center for International Development at Harvard University.*

Suel, E., Polak, J. W., Bennett, J. E., and Ezzati, M. (2019). Measuring social, environmental and health inequalities using deep learning and street imagery. *Scientific Reports*, 9(1):6229.

Tanyanyiwa, V. I. (2015). Motivational factors influencing littering in Harare's Central Business District (CBD), Zimbabwe. *IOSR Journal of Human and Social Sciences*, 20(2):58–65.

Terven, J. and Cordova-Esparza, D. (2023). A comprehensive review of YOLO: From YOLOv1 and beyond. *arXiv preprint arXiv:2304.00501.*

Thöni, C. and Volk, S. (2018). Conditional cooperation: Review and refinement. *Economics Letters*, 171:37–40.

Torgler, B., Frey, B. S., and Wilson, C. (2009). Environmental and pro-social norms: Evidence on littering. *The BE Journal of Economic Analysis and Policy*, 9(1).

Van der Windt, P. and Mvukiyehe, E. (2020). Assessing the longer term impact of community-driven development programs: Evidence from a field experiment in

the Democratic Republic of Congo. *World Bank Policy Research Working Paper*, (9140).

Willer, R. (2009). Groups reward individual sacrifice: The status solution to the collective action problem. *American Sociological Review*, 74(1):23–43.

Woolcock, M. et al. (2001). The place of social capital in understanding social and economic outcomes. *Canadian Journal of Policy Research*, 2(1):11–17.

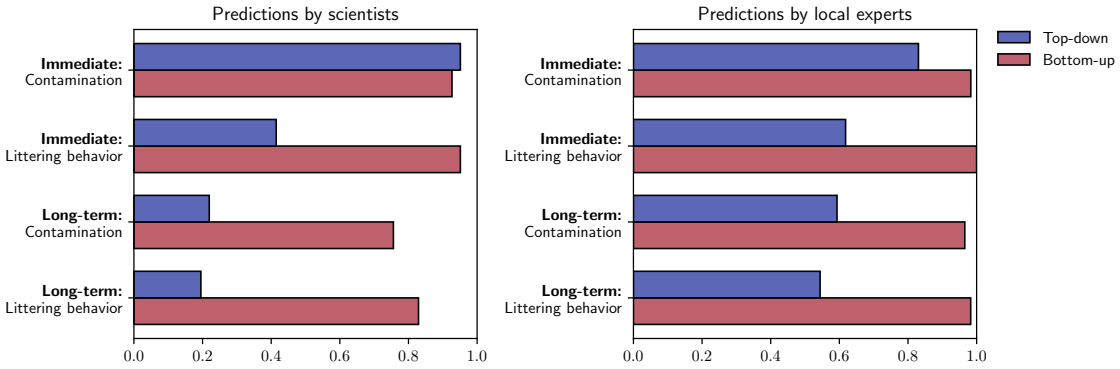
World Bank (2022). Community and local development. Available at: `https://www.worldbank.org/en/topic/communitydrivendevelopment`. Last accessed: 2023-10-26.

Yeh, C., Perez, A., Driscoll, A., Azzari, G., Tang, Z., Lobell, D., Ermon, S., and Burke, M. (2020). Using publicly available satellite imagery and deep learning to understand economic well-being in Africa. *Nature Communications*, 11(1):2583.

# A Appendix

## A1 Additional Results

(a) Does the Intervention Work?




(b) Which Intervention Works Best?

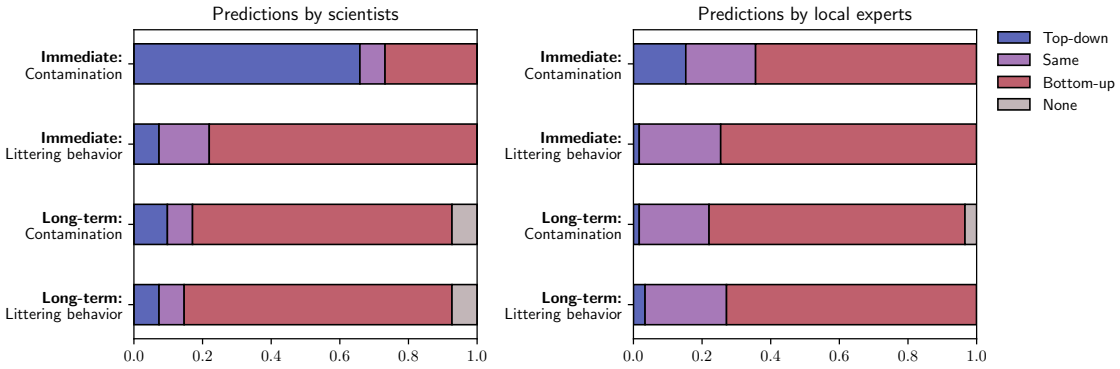



Figure A1: Pre-Survey Results

Illustration based on a prediction survey with 41 social scientists and 59 local experts. The upper figure shows the percentage of respondents who expect each intervention to have a positive effect. The lower figure shows the share of respondents indicating that a particular intervention worked best.

Table A1: Community Activities: Survey Answers

| | *Immediate effects* | | | *Long-term effects* | | |
|---|---|---|---|---|---|---|
| | Control | T1: TD | T2: BU | Control | T1: TD | T2: BU |
| **Activities observed** | | | | | | |
| Community meeting | 0.29 | 0.35 | 0.70*** | 0.24 | 0.21 | 0.40*** |
| Session or workshop | 0.06 | 0.13** | 0.48*** | 0.01 | 0.02 | 0.05* |
| Cleaning | 0.26 | 0.55*** | 0.74*** | 0.36 | 0.45 | 0.59*** |
| None | 0.55 | 0.38*** | 0.15*** | 0.57 | 0.50 | 0.31*** |
| **Activities participated** | | | | | | |
| Community meeting | 0.16 | 0.15 | 0.42*** | 0.18 | 0.16 | 0.27** |
| Session or workshop | 0.03 | 0.06 | 0.33*** | 0.01 | 0.01 | 0.04 |
| Cleaning | 0.21 | 0.34*** | 0.51*** | 0.30 | 0.36 | 0.47*** |
| None | 0.69 | 0.62 | 0.43*** | 0.64 | 0.61 | 0.47*** |
| **Perception** | | | | | | |
| Level of activities (sd) | 0.00 | 0.16 | 0.81*** | 0.00 | -0.06 | 0.20** |
| Waste management organization (sd) | 0.00 | 0.25*** | 0.59*** | 0.00 | 0.00 | 0.22*** |
| Frequency waste truck | 1.86 | 2.44 | 2.26 | 1.79 | 2.19 | 2.20 |
| Frequency waste truck usage | 1.64 | 2.20 | 2.07 | 1.66 | 2.04 | 1.95 |
| Frequency community cleaning | 0.96 | 1.18 | 1.40 | 0.82 | 0.71 | 0.89 |

Sample sizes are n=2066 for the estimation of immediate effects and n=1832 for long-term effects. Missings are imputed using the mean value per treatment group. The displayed values are sample means per group. The stars indicate the p-value of a joint F-test that each treatment dummy coefficient is equal to 0. Standard errors were clustered at the community level, controls are baseline education, sex, age and strata fixed effects. $*p < 0.1$; $**p < 0.05$; $***p < 0.01$

Table A2: Raw Contamination Results

| | *Immediate effects* | | | | *Long-term effects* | | | |
|---|---|---|---|---|---|---|---|---|
| | T1: TD | T2: BU | T2 - T1 | N | T1: TD | T2: BU | T2 - T1 | N |
| **Photo trash detection** | | | | | | | | |
| Kernel approach | -0.536*** | -0.384*** | 0.153** | 60709 | -0.067 | -0.103** | -0.037 | 65673 |
| | (0.092) | (0.092) | (0.076) | | (0.055) | (0.052) | (0.050) | |
| Raster approach | -0.511*** | -0.331*** | 0.180** | 10740 | -0.071 | -0.106* | -0.035 | 10740 |
| | (0.100) | (0.112) | (0.090) | | (0.062) | (0.058) | (0.056) | |
| Raw averages | -0.561*** | -0.427*** | 0.133 | 120 | -0.089 | -0.126** | -0.037 | 120 |
| | (0.084) | (0.085) | (0.086) | | (0.058) | (0.058) | (0.059) | |
| **Enumerator assessments** | | | | | | | | |
| Kernel approach | -0.283*** | -0.234*** | 0.049 | 12216 | -0.016 | -0.019 | -0.003 | 13163 |
| | (0.056) | (0.054) | (0.059) | | (0.071) | (0.058) | (0.077) | |
| Raster approach | -0.263*** | -0.202*** | 0.061 | 4272 | 0.002 | -0.002 | -0.004 | 4272 |
| | (0.060) | (0.058) | (0.061) | | (0.071) | (0.055) | (0.076) | |
| Raw averages | -0.311*** | -0.266*** | 0.046 | 120 | -0.030 | -0.030 | 0.000 | 120 |
| | (0.063) | (0.063) | (0.064) | | (0.066) | (0.066) | (0.066) | |

Outcomes refer to the number of detected trash items per image in the upper panel and to enumerator assessment scores (1-4) in the lower panel. Controls include contamination at baseline and strata fixed effects. Standard errors are clustered at the community level for the kernel and the raster approach. $*p < 0.1$; $**p < 0.05$; $***p < 0.01$

Table A3: Contamination Results with Coder Fixed Effects

| | *Immediate effects* | | | | *Long-term effects* | | | |
|---|---|---|---|---|---|---|---|---|
| | T1: TD | T2: BU | T2 - T1 | N | T1: TD | T2: BU | T2 - T1 | N |
| **Photo trash detection** | | | | | | | | |
| Kernel approach | -0.893*** | -0.689*** | 0.204** | 60709 | -0.193** | -0.232** | -0.039 | 65673 |
| | (0.135) | (0.138) | (0.103) | | (0.096) | (0.099) | (0.097) | |
| Raster approach | -0.968*** | -0.737*** | 0.231** | 10740 | -0.238** | -0.243** | -0.005 | 10740 |
| | (0.154) | (0.140) | (0.115) | | (0.117) | (0.113) | (0.114) | |
| Raw averages | -0.923*** | -0.641*** | 0.282* | 120 | -0.250* | -0.327** | -0.076 | 120 |
| | (0.157) | (0.158) | (0.159) | | (0.147) | (0.151) | (0.152) | |
| **Enumerator assessments** | | | | | | | | |
| Kernel approach | -1.220*** | -0.949*** | 0.272 | 12216 | -0.403*** | -0.391*** | 0.013 | 13163 |
| | (0.155) | (0.185) | (0.181) | | (0.150) | (0.149) | (0.159) | |
| Raster approach | -1.124*** | -0.884*** | 0.240 | 4272 | -0.375** | -0.378** | -0.003 | 4272 |
| | (0.160) | (0.189) | (0.177) | | (0.154) | (0.165) | (0.182) | |
| Raw averages | -1.157*** | -1.004*** | 0.153 | 120 | -0.364** | -0.452*** | -0.088 | 120 |
| | (0.209) | (0.208) | (0.211) | | (0.165) | (0.170) | (0.170) | |

Results reported in standard deviations at the community level. Controls include contamination at baseline, coder fixed effects, and strata fixed effects. Standard errors are clustered at the community level for the kernel and the raster approach. $*p < 0.1$; $**p < 0.05$; $***p < 0.01$

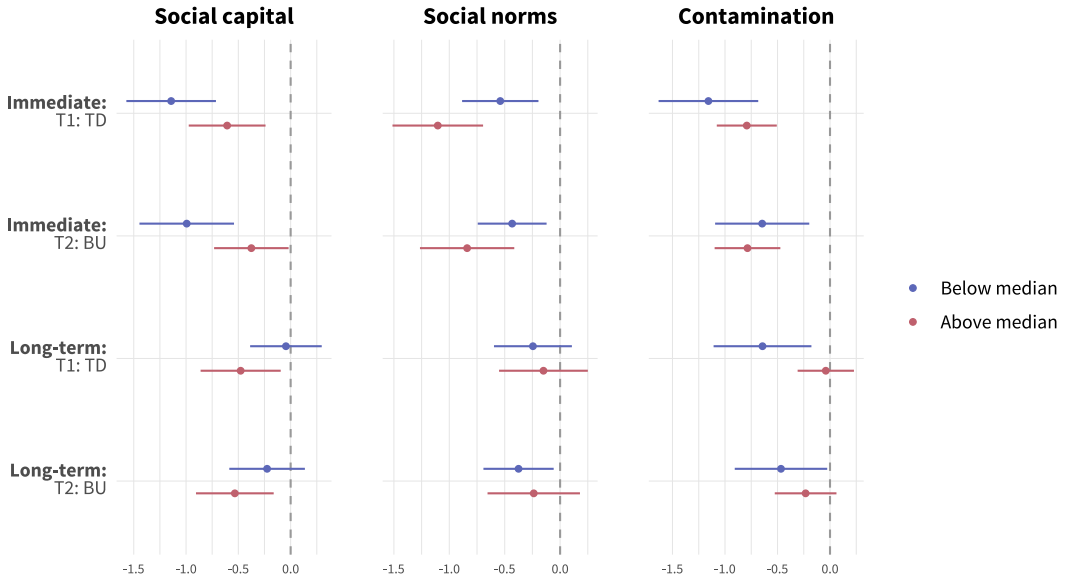



Figure A2: Effect Heterogeneity by Social Capital, Social Norms, and Contamination

The outcome variable is standardized trash counts per image. Social capital refers to an index of networks (strong ties), trust, organizations, and voluntary work (sum of standardized variables); social norms is an index of the share of villagers believed to egage in littering, believed to disapprove of littering, and believed to punish littering (sum of standardized variables); and contamination is the baseline contamination level, measured as standardized trash counts. Heterogeneity analyses are conducted at the community level.

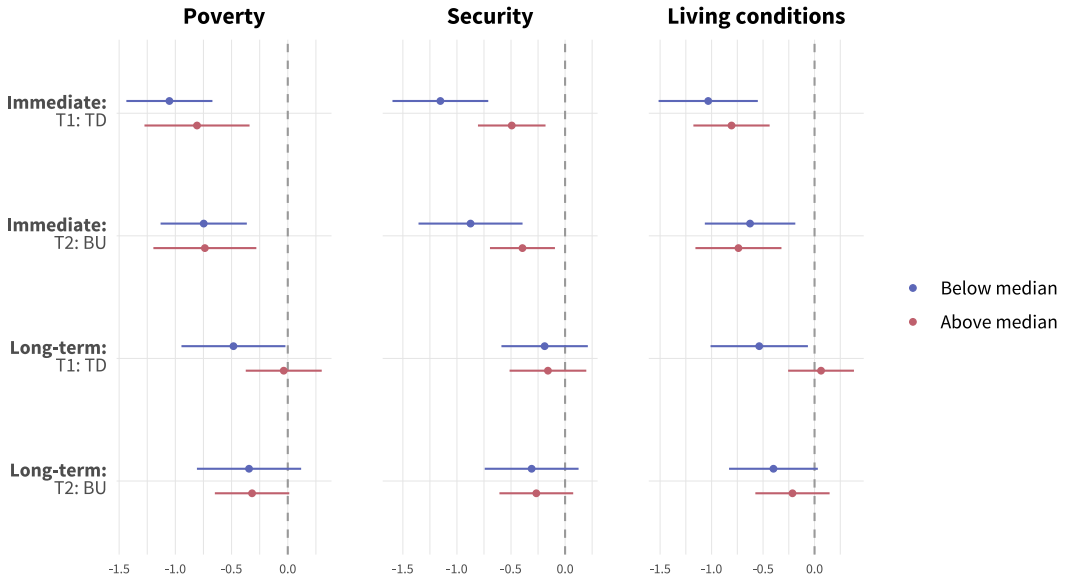



Figure A3: Effect Heterogeneity by Poverty, Security, and Living Conditions

The outcome variable is standardized trash counts per image. Hetereogeneity analyses are conducted at the community level.

Table A4: Coefficient Depletion Rates: Models without Fixed Effects

| | *Absolute depletion* | | | *Relative depletion* | | |
|---|---|---|---|---|---|---|
| | T1: TD | T2: BU | T1 - T2 | T1: TD. | T2: BU | T1 - T2 |
| **Photo trash detection** | | | | | | |
| Raster approach | 0.593 | 0.271 | 0.322 *** | 0.816 | 0.575 | 0.241 |
| Kernel approach | 0.627 | 0.341 | 0.285 *** | 0.829 | 0.631 | 0.198 |
| Raw averages | 0.617 | 0.356 | 0.261 ** | 0.779 | 0.590 | 0.189 |
| **Enumerator assessments** | | | | | | |
| Raster approach | 0.810 | 0.610 | 0.200 | 1.008 | 0.991 | 0.018 |
| Kernel approach | 0.885 | 0.714 | 0.171 | 0.949 | 0.926 | 0.023 |
| Raw averages | 0.911 | 0.765 | 0.146 | 0.911 | 0.896 | 0.015 |

Absolute depletion indicates the difference between short-term and long-term effects in standard deviations. Relative depletion indicates the difference between short-term and long-term effects as a percentage value of short-term effect. For linear differences, the p-values were obtained with a t-test. For nonlinear differences, the p-values were obtained with the delta method. $*p < 0.1$; $**p < 0.05$; $***p < 0.01$

Table A5: Balance at Baseline for Additional Survey Variables

| | Control | T1: TD | T2: BU | P-value | N |
|---|---|---|---|---|---|
| Pay for cleaning, me (log) | 0.352 | 0.374 | 0.315 | 0.889 | 2420 |
| Pay for cleaning, others (log) | 0.454 | 0.344 | 0.341 | 0.385 | 2417 |
| Bothered by litter (1-5) | 3.894 | 3.849 | 3.982 | 0.511 | 2420 |
| Littering, me (%) | 0.142 | 0.163 | 0.141 | 0.532 | 2421 |
| Littering is bad, me (1-5) | 4.574 | 4.548 | 4.635 | 0.307 | 2420 |
| Punish littering, me (%) | 0.394 | 0.343 | 0.324 | 0.035 | 2421 |
| Living conditions (1-5) | 3.267 | 3.215 | 3.165 | 0.081 | 2420 |
| Security (1-5) | 4.133 | 4.096 | 4.141 | 0.441 | 2421 |
| Trust comm. leaders (1-5) | 3.318 | 3.150 | 3.304 | 0.151 | 2416 |
| Trust municipal gov. (1-5) | 2.739 | 2.718 | 2.759 | 0.912 | 2415 |
| Trust central gov. (1-5) | 3.386 | 3.215 | 3.276 | 0.196 | 2413 |

The last row indicates the p-value of a joint F-test that each treatment dummy coefficient is equal to 0. Standard errors are clustered at the community level.

Table A6: Survey Results for Contamination and Waste Disposal

| | *Immediate effects* | | | *Long-term effects* | | |
|---|---|---|---|---|---|---|
| | T1: TD | T2: BU | T2 - T1 | T1: TD | T2: BU | T2 - T1 |
| Perceived cleanliness (sd) | 0.158* | 0.163** | 0.005 | 0.173** | 0.182** | 0.009 |
| | (0.085) | (0.083) | (0.076) | (0.079) | (0.082) | (0.078) |
| Bothered by litter (sd) | -0.069 | -0.008 | 0.061 | -0.023 | 0.073 | 0.096 |
| | (0.093) | (0.092) | (0.087) | (0.086) | (0.078) | (0.082) |
| Appropriate disposal (%) | 0.040 | 0.137*** | 0.097** | -0.005 | 0.067* | 0.072* |
| | (0.044) | (0.041) | (0.046) | (0.042) | (0.039) | (0.040) |
| Recycling (%) | 0.083** | 0.112*** | 0.029 | -0.026 | -0.001 | 0.025 |
| | (0.041) | (0.037) | (0.029) | (0.040) | (0.040) | (0.040) |
| Recycling items (nr) | 0.093 | 0.268** | 0.175 | -0.162 | 0.055 | 0.216** |
| | (0.113) | (0.125) | (0.111) | (0.098) | (0.102) | (0.095) |
| Pay for cleaning, me (log) | -0.109** | -0.109* | 0.001 | 0.060 | 0.028 | -0.031 |
| | (0.056) | (0.061) | (0.050) | (0.068) | (0.061) | (0.062) |
| Pay for cleaning, others (log) | -0.085 | -0.081 | 0.004 | 0.024 | -0.026 | -0.050 |
| | (0.062) | (0.069) | (0.053) | (0.058) | (0.053) | (0.047) |

Sample sizes are n = 2066 for the estimation of immediate effects and n = 1832 for long-term effects. Controls include strata fixed effects, sex, age, education (dummies), and the baseline value for the respective outcome. Note that no baseline values are available for recycling outcomes. Standard errors are clustered at the community level. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$

Table A7: Survey Results for Self-Reported Behaviors and Social Norms

| | *Immediate effects* | | | *Long-term effects* | | |
|---|---|---|---|---|---|---|
| | T1: TD | T2: BU | T2 - T1 | T1: TD | T2: BU | T2 - T1 |
| Littering, me (%) | -0.030** | -0.016 | 0.014 | -0.011 | -0.006 | 0.005 |
| | (0.015) | (0.015) | (0.013) | (0.016) | (0.014) | (0.016) |
| Littering is bad, me (sd) | 0.041 | 0.043 | 0.002 | -0.011 | 0.022 | 0.034 |
| | (0.041) | (0.044) | (0.039) | (0.054) | (0.044) | (0.052) |
| Punish littering, me (%) | -0.027 | -0.002 | 0.024 | 0.005 | -0.014 | -0.020 |
| | (0.026) | (0.026) | (0.026) | (0.031) | (0.029) | (0.030) |
| Littering, others (%) | -0.065** | -0.104*** | -0.039 | 0.011 | -0.019 | -0.030 |
| | (0.028) | (0.026) | (0.026) | (0.019) | (0.019) | (0.021) |
| Littering is bad, others (%) | -0.015 | -0.037** | -0.022 | -0.001 | 0.004 | 0.005 |
| | (0.020) | (0.017) | (0.020) | (0.018) | (0.018) | (0.016) |
| Punish littering, others (%) | 0.027 | 0.003 | -0.024 | 0.017 | 0.036* | 0.019 |
| | (0.028) | (0.025) | (0.027) | (0.022) | (0.020) | (0.022) |

Sample sizes are n = 2066 for the estimation of immediate effects and n = 1832 for long-term effects. Controls include strata fixed effects, sex, age, education (dummies), and the baseline value for the respective outcome. Standard errors are clustered at the community level. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$

Table A8: Survey Results for Other Outcomes

| | *Immediate effects* | | | *Long-term effects* | | |
|---|---|---|---|---|---|---|
| | T1: TD | T2: BU | T2 - T1 | T1: TD | T2: BU | T2 - T1 |
| Living conditions (sd) | -0.141** | -0.094 | 0.046 | -0.075 | 0.029 | 0.103 |
| | (0.059) | (0.066) | (0.068) | (0.059) | (0.053) | (0.063) |
| Security (sd) | -0.014 | 0.053 | 0.068 | -0.035 | -0.074 | -0.040 |
| | (0.081) | (0.086) | (0.085) | (0.097) | (0.088) | (0.109) |
| Trust comm. leaders (sd) | 0.035 | 0.104 | 0.069 | -0.096 | 0.056 | 0.152** |
| | (0.103) | (0.093) | (0.106) | (0.076) | (0.065) | (0.064) |
| Trust municipal gov. (sd) | 0.024 | 0.097 | 0.072 | 0.012 | -0.061 | -0.074 |
| | (0.085) | (0.084) | (0.081) | (0.088) | (0.071) | (0.082) |
| Trust central gov. (sd) | 0.045 | 0.130 | 0.085 | -0.060 | -0.044 | 0.016 |
| | (0.114) | (0.110) | (0.107) | (0.073) | (0.064) | (0.070) |

Sample sizes are n = 2066 for the estimation of immediate effects and n = 1832 for long-term effects. Controls include strata fixed effects, sex, age, education (dummies), and the baseline value for the respective outcome. Standard errors are clustered at the community level. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$

Table A9: Attrition by Treatment Group

| | Control | T1: TD | T2: BU | P-value | N |
|---|---|---|---|---|---|
| Attrition Midline | 0.160 | 0.151 | 0.128 | 0.423 | 2421 |
| Attrition Endline | 0.228 | 0.275 | 0.228 | 0.131 | 2421 |

The last row indicates the p-value of a joint F-test that each treatment dummy coefficient is equal to 0. Standard errors are clustered at the community level.

Table A10: Attriter Characteristics at Midline by Treatment Group

| | Control | T1: TD | T2: BU | P-value | N |
|---|---|---|---|---|---|
| **Sociodemographics** | | | | | |
| Female | 0.706 | 0.664 | 0.670 | 0.936 | 355 |
| Age | 47.360 | 40.899 | 41.810 | 0.002 | 355 |
| Education | 2.147 | 2.361 | 2.360 | 0.464 | 355 |
| Poverty (1-5) | 3.000 | 3.104 | 3.010 | 0.819 | 342 |
| Community size | 305.463 | 279.706 | 269.530 | 0.645 | 355 |
| **Contamination and waste disposal** | | | | | |
| Perceived cleanliness (1-5) | 3.154 | 3.008 | 3.290 | 0.102 | 355 |
| Appropriate disposal (%) | 0.353 | 0.445 | 0.460 | 0.164 | 355 |
| **Social norms** | | | | | |
| Littering (%) | 0.606 | 0.605 | 0.514 | 0.083 | 355 |
| Littering is bad (%) | 0.707 | 0.733 | 0.662 | 0.141 | 354 |
| Punish littering (%) | 0.571 | 0.582 | 0.493 | 0.016 | 355 |
| **Social capital** | | | | | |
| Strong ties (%) | 0.310 | 0.461 | 0.311 | 0.110 | 354 |
| Weak ties (%) | 0.687 | 0.756 | 0.654 | 0.088 | 345 |
| Trust (1-5) | 3.463 | 3.370 | 3.540 | 0.578 | 355 |
| Organizations (%) | 0.110 | 0.109 | 0.110 | 0.911 | 355 |
| Voluntary work (%) | 0.154 | 0.254 | 0.300 | 0.071 | 354 |
| Altruism (%) | 0.405 | 0.505 | 0.413 | 0.066 | 355 |

The first three columns represent attriter group means for each treatment. The last column indicates the p-value of a joint F-test that each treatment dummy coefficient is equal to 0. Education refers to highest completed degree: None = 1, incomplete primary = 2, complete primary = 3, high school degree = 4), technical = 5, and university degree = 6. Standard errors are clustered at the community level.

Table A11: Attriter Characteristics at Endline by Treatment Group

| | Control | T1: TD | T2: BU | P-value | N |
|---|---|---|---|---|---|
| **Sociodemographics** | | | | | |
| Female | 0.660 | 0.724 | 0.607 | 0.065 | 589 |
| Age | 40.984 | 39.654 | 41.421 | 0.440 | 588 |
| Education | 2.546 | 2.599 | 2.522 | 0.687 | 589 |
| Poverty (1-5) | 3.247 | 3.118 | 3.034 | 0.469 | 572 |
| Community size | 300.289 | 288.507 | 282.854 | 0.620 | 589 |
| **Contamination and waste disposal** | | | | | |
| Perceived cleanliness (1-5) | 3.093 | 3.115 | 3.129 | 0.999 | 589 |
| Appropriate disposal (%) | 0.330 | 0.498 | 0.478 | 0.112 | 589 |
| **Social norms** | | | | | |
| Littering (%) | 0.618 | 0.612 | 0.548 | 0.013 | 589 |
| Littering is bad (%) | 0.730 | 0.715 | 0.678 | 0.163 | 589 |
| Punish littering (%) | 0.584 | 0.582 | 0.534 | 0.100 | 588 |
| **Social capital** | | | | | |
| Strong ties (%) | 0.333 | 0.408 | 0.307 | 0.423 | 589 |
| Weak ties (%) | 0.697 | 0.711 | 0.638 | 0.530 | 578 |
| Trust (1-5) | 3.392 | 3.461 | 3.567 | 0.601 | 589 |
| Organizations (%) | 0.144 | 0.143 | 0.135 | 0.923 | 589 |
| Voluntary work (%) | 0.201 | 0.194 | 0.316 | 0.060 | 587 |
| Altruism (%) | 0.467 | 0.456 | 0.430 | 0.628 | 589 |

The first three columns represent attriter group means for each treatment. The last column indicates the p-value of a joint F-test that each treatment dummy coefficient is equal to 0. Education refers to highest completed degree: None = 1, incomplete primary = 2, complete primary = 3, high school degree = 4), technical = 5, and university degree = 6. Standard errors are clustered at the community level.

# A2 Supplementary Information on Data and Measurement Instruments

Table A12: Coding of the Survey Questions

| *Variable* | *Survey question* | *Possible answers* | *Computation* |
|---|---|---|---|
| **Contamination and waste disposal** | | | |
| Perceived cleanliness | How do you evaluate the garbage contamination situation in your community? | Scale from 1 (Very clean) to 5 (Very dirty) | Standardized |
| Bothered by litter | Personally, how bothered are you by the trash in your community? | Scale from 1 (Not at all) to 5 (Very much) | Standardized |
| Appropriate disposal | In the past month, how has your household gotten rid of trash? | 1: Trash truck; 2: Deposit; 3: Bury it; 4: Burn it; 5: Informal deposit, street | Percentage that used the trash truck or formal deposits |
| Recycling | In the past month, has your household separated any trash for recycling? | 0: None of the below; 1: At least one of the below | Percentage |
| Recycling items | What types of garbage have been recycled? | 0: None; 1: Plastic; 2: Glass; 3: Paper; 4: Organic waste | Number of different items |
| Pay for cleaning, me | Imagine if a service was hired in your community to clean the streets. How much would you be willing to contribute per month? | Decimal | Log |
| Pay for cleaning, others | On average, how much do you think a person in your community would be willing to contribute? | Decimal | Log |
| **Self-Reported behaviors and social norms** | | | |

Table A12: Coding of the Survey Questions

| *Variable* | *Survey question* | *Possible answers* | *Computation* |
|---|---|---|---|
| Littering, me | Being very, very honest, in the last month, have you ever thrown trash in the street? | 0: No; 1: Yes | Percentage |
| Littering is bad, me | In your personal opinion, is it bad to litter on the street? | Scale from 1 (Not at all bad) to 5 (Very bad) | Standardized |
| Punish littering, me | If you observed someone in your community throwing trash in the street, what would you do? | 0: Nothing, I do not want to get involved / it does not seem serious to me; 1: React with disapproval | Percentage of people reacting with disapproval |
| Littering, others | Out of every 10 people in your community, how many do you think have thrown trash in the street in the last month? | Integer | Percentage |
| Littering is bad, others | Out of every 10 people in your community, how many do you think believe it is wrong to litter in the street? | Integer | Percentage |
| Punish littering, others | Out of every 10 people in your community, how many do you think would react with a gesture of disapproval to a person throwing trash in the street? | Integer | Percentage |
| **Social capital** | | | |
| Strong ties | Q1: Approximately how many people live in your community?; Q2: Of these people, how many persons are close acquaintances (family, friends)? | Integers | Percentage (Q2/Q1) |

Table A12: Coding of the Survey Questions

| *Variable* | *Survey question* | *Possible answers* | *Computation* |
|---|---|---|---|
| Weak ties | Q1: Approximately how many people live in your community?; Q2: Of these people, how many persons are close acquaintances (family, friends)?; Q3: Of these people, how many persons are casual acquaintances? | Integers | Percentage ((Q2+Q3)/Q1) |
| Trust | Compared to other people, do you trust people in your community more or less? | Scale from 1 (Much less) to 5 (Much more) | Standardized |
| Organizations | Do you belong to one or more community organizations or groups (e.g. ADESCO, youth organization, collectives, etc.)? | 0: No; 1: Yes | Percentage |
| Voluntary work | In the past month, have you participated in any type of volunteer work for the community? | 0: No; 1: Yes | Percentage |
| **Other outcomes** | | | |
| Living conditions | Compared to other communities, how do you evaluate the living conditions in your community? | Scale from 1 ( Very bad) to 5 (Very good) | Standardized |
| Security | Compared to other communities, how do you evaluate the security situation in your community? | Scale from 1 (Very safe) to 5 (Very dangerous) | Standardized |
| Trust community leaders | How much do you trust your community leaders? | Scale from 1 (Not at all) to 5 (Very much) | Standardized |
| Trust municipal government | How much do you trust municipal officials? | Scale from 1 (Not at all) to 5 (Very much) | Standardized |

Table A12: Coding of the Survey Questions

| *Variable* | *Survey question* | *Possible answers* | *Computation* |
|---|---|---|---|
| Trust central government | How much do you trust central government officials? | Scale from 1 (Not at all) to 5 (Very much) | Standardized |
| Altruism | There will be a lottery for 100 USD among the participants. The winner will have to decide how much of this money to keep and how much to donate to a family in need in the department (photos of the delivery would be sent). | Integers | Percentage of how much was donated |
| **Socio-demographic variables** | | | |
| Female | Gender | 0: Male; 1: Female | |
| Age | Age | Integer | |
| Education | Highest level of education | 1: None; 2: Incomplete primary; 3: Complete primary; 4: High school degree; 5: Technical; 6: University degree | Dummies for each level |
| Poverty | What is the family's economic situation like? The family's poverty level was recorded by the enumerators based on pictures of potential housing conditions. | Scale of 1 (Not poor) to 5 (Very poor) | Standardized |
| Community size | Approximately how many people live in your community? | Integer | |
| **Waste management activities** | | | |

Table A12: Coding of the Survey Questions

| *Variable* | *Survey question* | *Possible answers* | *Computation* |
|---|---|---|---|
| Activities observed | In the last 4 months, have you heard of any activity related to the issue of garbage in your community? | Community meeting; Session or workshop; Cleaning; None | Dummies for each activity |
| Activities participated | In the last 4 months, have you participated in any activity related to the issue of garbage in your community? | Community meeting; Session or workshop; Cleaning; None | Dummies for each activity |
| Level of activities | In the last 4 months, do you think there were more or fewer activities than before regarding the issue of garbage in your community? | Scale from 1 (Much less) to 5 (Much more) | Standardized |
| Waste management organization | In your opinion, how organized is your community in relation to garbage management? | Scale from 1 (Not at all organized) to 5 (Perfectly organized) | Standardized |
| Frequency waste truck | In the last 4 months, how often has a toilet train arrived in your community? | 1: Never; 2: Every 2 months; 3: Every month; 4: Every 2 weeks; 5: Every week; 6: Twice a week; 7: Every day | Frequency per month |
| Frequency waste truck usage | In the last 4 months, how often have you used the garbage train to dispose of your garbage? | 1: Never; 2: Every 2 months; 3: Every month; 4: Every 2 weeks; 5: Every week; 6: Twice a week; 7: Every day | Frequency per month |
| Frequency community cleaning | In the last 4 months, how often has your community been cleaned? | 1: Never; 2: Every 2 months; 3: Every month; 4: Every 2 weeks; 5: Every week; 6: Twice a week; 7: Every day | Frequency per month |

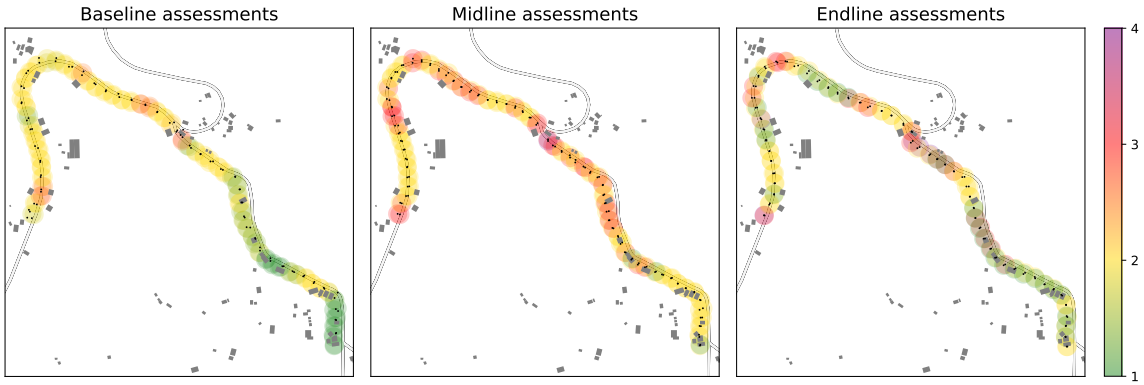



Figure A4: Illustration of Kernel Approach for Subjective Enumerator Assessments

Black dots represent assessment locations. Circle color corresponds to contamination level: 1 = very clean, 4 = very dirty. Baseline values are imputed based on circles around each midline and endline assessment respectively. A triangular kernel is used to give higher weights to closer assessments. Circle radius is 25m. Baseline map is shown with respect to the midline assessment and would be slightly different for the endline assessment.

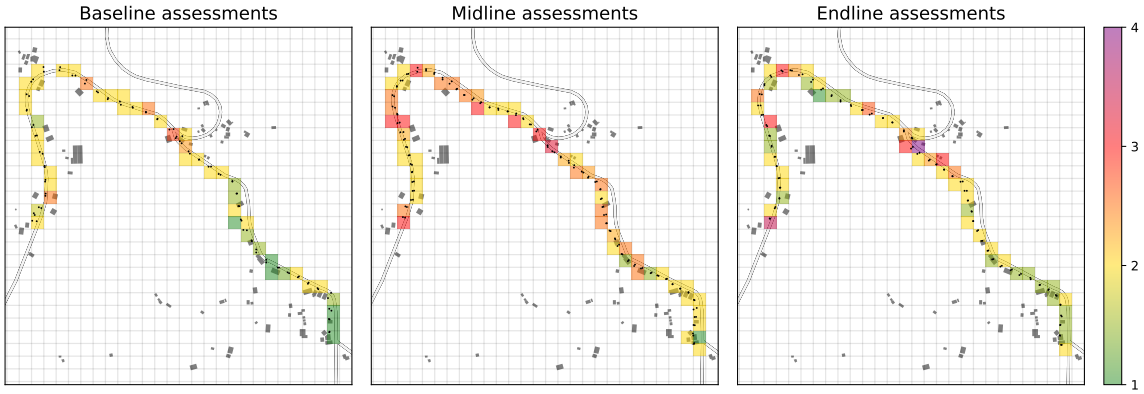



Figure A5: Illustration of Raster Approach for Subjective Enumerator Assessments

Black dots represent assessment locations. Cell color corresponds to contamination level: 1 = very clean, 4 = very dirty. Resolution of the raster is 0.0003 degrees (approx. 33m).

## A3 Literature Review

Table A13: Literature Review on Community-Driven Development

| *Study* | *Study type* | *Description and results* |
|---|---|---|
| **Meta studies of community-driven development programs** | | |
| Casey 2018 | Meta study on evolution of CDD and CDD RCTs. | A synthesis of seven CDD RCTs shows that CDD effectively delivers public goods and some economic benefits at a low cost in challenging environments. However, it does not seem to lead to lasting transformations in local decision-making or empowerment of the poor. This raises the question of how much participation is necessary to preserve the benefits of decentralization while minimizing the time costs imposed on impoverished communities. |
| Mansuri and Rao 2012 | Meta study proposing general concept of CDD based on literature from different fields. | The report discusses the history of participatory development and presents a framework for understanding participatory development, emphasizing the concept of "civil society failure" and its interaction with government and market failures. It is based on literature from anthropology, economics, political science and sociology. Evidence on key development outcomes, public service delivery and quality, but also on issues related to CDD is reviewed. The report also discusses World Bank-funded projects, emphasizing the importance of local context, as well as effective monitoring and evaluation for successful outcomes. |
| **Evaluations of community-driven development programs** | | |
| Arcand 2008 | IV study with panel data on 71 villages with 756 households in Senegal. | This paper investigates the impact of a national CDD program on access to basic services, household expenditures, and child wellbeing. The program had a positive effect on villagers' access to clean water and health services, as well as on child malnutrition. Completed income-generating agricultural infrastructure projects and improved primary education significantly increased household expenditures per capita, while health and hydraulic projects did not. |

Table A13: Literature Review on Community-Driven Development

| *Study* | *Study type* | *Description and results* |
|---|---|---|
| Avdeenko and Gilligan 2015 | RCT with 576 households in 24 communities, and 475 lab-in-the-field subjects. | The intervention had no impact on networks and social norms, but it increased people's involvement in civic activities and local governance. Therefore, the authors attribute the increase in citizen participation not to the growth of social capital, but to the greater openness of the local governments. |
| Beath et al. 2013 | RCT with 500 villages in Afghanistan | The RCT examines the impact of a CDD program that requires female participation on several outcomes related to women's empowerment. Positive effects on women's participation in economic, social, and political activities are reported. However, no impacts on gender roles or family decision-making are found. |
| Casey et al. 2012 | RCT with 2,832 households in 236 villages in the Republic of Sierra Leone. | The study evaluates a CDD program aiming to make local institutions more democratic and egalitarian by imposing participation requirements for marginalized groups. The program had positive short-term effects on local public services and economic outcomes. However, it did not result in sustained impacts on collective action, decision-making, or the involvement of marginalized groups, indicating that the intervention did not durably reshape local institutions. |
| Desai and Olofsgård 2019 | RCT combined with behavioral experiment with 80 villages in India. | The "self-help" groups established in treatment villages significantly improved people's access to and the quality of certain public goods, especially water, due to better information through the groups, stronger community engagement and reduced coordination costs. The behavioral experiment 4 years after the RCT revealed that cooperative norms are stronger in villages that had self-help groups. |
| Fearon et al. 2009 | RCT with 83 communities in Liberia | The study evaluates the impact of a community-driven (post-war) reconstruction project on social cohesion, as measured by an anonymous public goods game. Contributions were significantly higher in the treated communities, with a 9 percent increase in funds raised for a community-selected public good. |

Table A13: Literature Review on Community-Driven Development

| *Study* | *Study type* | *Description and results* |
|---|---|---|
| Humphreys et al. 2019 | RCT with 1,250 communities in the Congo | The study evaluates the impact of a community-driven reconstruction program on democratic governance. Behavior in an unconditional cash transfer program is used to assess whether the intervention had an impact on elite capture. No effects are found. |
| Labonne and Chase 2011 | DiD with 2,100 households in 135 communities in the Philippines. | Using difference-in-differences (DiD) and propensity score matching, the study evaluates a CDD program where communities competed for grants for infrastructure investments. The program increased the participation in village meetings and the frequency of interactions between local officials and village officials, but had a negative impact on collective action. |
| Nguyen and Rieger 2017 | RDD with 1,300 communes in Morocco | The study assesses the impact of a CDD initiative on social capital, employing a regression discontinuity design (RDD) based on the program's poverty selection threshold. The program increased contributions in a public goods game, but had no effect on altruism and a negative effect on trust. |
| Saguin 2018 | DiD based on surveys in 16 municipalities in the Philippines | The "KALAHI-CIDSS" CDD program was found to increase the incomes of poor households. However, it did not improve outcomes such as solidarity and trust. In addition, poor households are underrepresented in village assemblies, with declining participation over time. |
| Van der Windt and Mvukiyehe 2020 | RCT with 1,250 villages in the Republic of Congo. | The study assesses the long-term impact of a CDD initiative 8 years after its launch. The program had a lasting impact on infrastructure quality (e.g., of schools or hospitals), but no effects on other dimensions of service delivery, on economic welfare, and on local institutions (e.g., governance, social cohesion, or female empowerment) were found. |
| **Related studies** | | |

Table A13: Literature Review on Community-Driven Development

| *Study* | *Study type* | *Description and results* |
|---|---|---|
| Banerjee et al. 2010 | RCT with three interventions in India. | This paper examines if citizen involvement can shape public service provision in education. Three interventions were evaluated: (i) providing information on public school organization, (ii) introducing citizens to a simple monitoring tool for their local school, and (iii) training volunteers to hold reading camps in order to improve literacy knowledge. Information and monitoring did not improve outcomes, but the volunteer-led reading camps did. |
| Björkman and Svensson 2009 | RCT with 50 public dispensaries in Uganda. | The intervention aimed at encouraging community engagement in monitoring health services and holding local health providers accountable for their performance. To this end, community members developed village action plans together with the health care providers. One year after the intervention, treatment communities exhibited greater involvement in monitoring providers, resulting in increased effort from health workers to serve the community as well as significant improvements in healthcare utilization and health outcomes. |
| Björkman et al. 2017 | Follow-up of RCT in Björkman and Svensson (2009). | The authors evaluate the long-run impact (4 years) of the experiment in Björkman and Svensson (2009). Even with minimal follow-up, short-term enhancements in healthcare delivery and health outcomes were sustained over the long run. The results indicate that a lower-cost version of the treatment, which primarily aimed to boost participation without information on staff performance, did not influence the quality of care or health outcomes both in the short and in the in the longer run. |

Table A13: Literature Review on Community-Driven Development

| *Study* | *Study type* | *Description and results* |
|---|---|---|
| Duflo et al. 2015 | RCT with 70 schools in Kenya. | The study evaluates a program in Kenya where parents and the school committees of randomly selected schools received (i) funding or (ii) funding and a short School-Based Management (SBM) empowerment training to hire an additional teacher, outside normal Ministry of Education civil-service channels. Centrally hired civil-service teachers in schools receiving only funding endogenously reduced their effort and captured rents for their families by getting relatives the contract teacher positions. The SBM program cut by half both the reduction in the regular teacher effort in response to the program and the fraction of contract teachers who were relatives of regular teachers. |
| Olken 2007 | RCT in 608 villages in Indonesia. | The paper evaluates different interventions aiming at reducing corruption, measured by missing expenditures, in village road projects in Indonesia. Results show that increased government audits significantly reduce missing expenditures. In contrast, enhancing grassroots monitoring had limited impacts on corrupution. |
| Raffler et al. 2019 | RCT with 376 health care centers and 14,609 households in rural Uganda. | The authors evaluate a large-scale information intervention aiming to improve bottom-up monitoring of health service delivery. The study finds only modest positive effects of citizen monitoring on service quality and patient satisfaction, and no effects on utilization and health outcomes such as child mortality. |

Table A14: Literature Review on Waste Management

| *Study* | *Study type* | *Description and results* |
|---|---|---|
| **Evaluations of waste interventions** | | |
| Bateson et al. 2013 | Field experiment with 620 bicycle riders on Newcastle university campus, UK. | This study tests if displaying images of "watching eyes" causes people to litter less and if a potential effect depends on the cleanliness of the environment. People were more likely to litter in dirty environments, but the watching eyes only had an effect when many people were around, and this effect does not depend on the amount of litter in the environment. |
| Castaldi et al. 2021 | RCT in 8 beach resorts in Italy. | The resorts were randomly assigned to 3 groups: (i) free portable ashtrays, (ii) free portable ashtrays and anti-littering message, and (iii) control. Results show a reduction in daily litter (cigarette butts in sand on day/costumers): -10% to -12% for the ashtray group; -7% to -10% for the ashtray + message group. |
| Cialdini et al. 1990 | 5 field experiments in different public spaces with 127–484 observations. | The authors argue that injunctive and descriptive norms must be separated to understand littering behavior since behavior changes only in accordance with the more salient type. In their experiments, they find that littering increases in littered environments, and even more so when someone is observed littering. Conversely, littering decreases when someone is observed littering into a very clean environment. Men are more likely to litter than women across different settings. |
| Dur and Vollaard 2015 | Field experiment with 4,000 households in the Netherlands. | This paper studies littering behavior and free-riding mechanisms related to public services. In a randomly assigned part of residential area, the frequency of cleaning around the garbage containers is drastically reduced from daily cleanups to 2-3 times a week during a 3-month period. Removing the morning cleanup increased the presence of litter in the early afternoon (11% to 27%). Litter accumulation around the garbage disposal increased (from 20% to 75%). Telephone appointments for retrieval of large trash increased, meaning that some people started to clean up more by themselves. The effects persisted at least one month after the treatment ended. |

Table A14: Literature Review on Waste Management

| *Study* | *Study type* | *Description and results* |
|---|---|---|
| Lewis et al. 2009 | Nationwide survey on littering attitude and field study in cinemas in the UK. | The survey revealed personal differences in acceptance and justification of littering depending on the age group, rural/urban living environment, smoker/non-smoker, feeling connected/unconnected to community. Also, missing infrastructure was identified as a cause for littering. In the cinema field study, leaflets with a (i) control message unrelated to littering, (ii) polite anti-littering message, or (iii) direct anti-littering message were distributed, and it was observed how much litter was left behind. People in the control group littered more than people that were politely or directly asked not to. |
| Liu and Sibley 2004 | Field study in a public space in New Zealand with over 3,000 observations. | In a first sub-study, littering attitudes were observed during 3 weeks and the people who disposed of waste (correctly and incorrectly) were interviewed. In the second week, a banner with an anti-littering message was added. People were found to litter less in crowded public spaces compared to less-crowded public places. The banner did not change littering behavior. In a second sub-study, bins and ashtrays were installed, and found to reduce littering by 64% without changing attitudes towards littering. |
| Nepal et al. 2023 | RCT with 75 treatment and 75 control communities in Nepal. | The study evaluates a low-cost treatment to improve municipal solid waste management: Providing information to households and installing waste bins on the streets. Perceived cleanliness in treatment communities increased by 25% at midline (3 months after installation) and 43% at endline (9 months after installation). Giving household waste to collectors increased by 13% at midline and 9% at endline while there was no statistically significant change in at-source waste segregation. |

Table A14: Literature Review on Waste Management

| *Study* | *Study type* | *Description and results* |
|---|---|---|
| Ramos and Torgler 2012 | Field study with 98 observations in Australia. | The authors test the broken-window-theory in a field study, which was conducted over 6 days in university common rooms, alternating between an orderly and a disorderly environment. 59% of participants littered in the disorderly room, compared to 18% in the clean room. Multivariate analysis shows that the disorder variable is always large and statistically significant. Older individuals and senior staff were more likely to litter. |
| Rangoni and Jager 2017 | Simulation in an agent-based model with 100 simulated pedestrians. | The goal of the simulation is to evaluate how social influence may cause a transition from a clean to a littered environment in 3 situations: (i) no trash bins; (ii) trash bins which can get full, and (iii) adding cleaners who can pick up litter and empty bins. For the parameterization of the model, data from a field study is used. The simulations suggest that litter does not grow linearly. Furthermore, a dynamic cleaning regime is cheaper and more effective than pre-determined regimes. |
| Sagebiel et al. 2020 | Field experiment with 200 observations on university benches in Germany. | To test the broken-window theory in the context of littering cigarette butts, two types of environment were prepared: (i) clean environment in which all cigarette butts were removed around the benches; (ii) dirty environment in which 25 cigarette butts were placed around each bench. The authors conclude that increased cleaning effort reduces littering a little, but the effect might be too small to justify additional cleaning costs. |
| Schultz 1999 | Field experiment with 605 residents of single-family dwellings in the US. | The study aims to find out if a plea alone or accompanied by (i) information, (ii) neighbor feedback, or (iii) household feedback increases proper waste disposal. Results show that feedback targeting personal or social norms increased the proportion of people recycling and the amount of recycled materials while not changing the level of contamination through littering. The author argues that a link between norm activation and behavior change exists. |

Table A14: Literature Review on Waste Management

| *Study* | *Study type* | *Description and results* |
|---|---|---|
| Sheely 2013 | RCT with 36 communities in Kenya (based on qualitative study). | This study aims to explain variation in maintaining a clean environment through interactions between government and community institutions. Communities are randomly assigned to 4 experimental groups: (i) collective action to organize cleanups promoted by a local NGO, (ii) collective action and punishment for littering by government chiefs, (iii) collective action and punishment for littering by traditional elders, and (iv) a control group. The author finds that communities with no formal punishment for littering experienced a sustained reduction in littering behavior and an increase in the frequency of public cleanups. Communities in which government administrators or traditional leaders punished littering experienced short-term reductions in littering that were not sustained. |
| **Other waste studies** | | |
| Chitotombe 2014 | Interviews in Zimbabwe and literature review. | The unavailability of bins, socio-cultural consumption styles in particular related to fast foods, illegal display of posters in the streets, and abandoned motor vehicles are mentioned as problems in Zimbabwe. Anti-littering campaigns have shown little success in the past. The interviews revealed that bins were not used even if available and that communities are reluctant to participate in cleanups. Also, the study shows that language barriers and political inefficiencies impede proper waste management. |
| Nkwocha and Okeoma 2009 | Interviews of 6,000 individuals in 6 geo-political zones in Nigeria. | Littering is very common in Nigeria. Reasons given by respondents included lack of bins or long distances to dumpsites, inefficiency of local authorities in keeping public spaces clean, missing legislation against littering, convenience, and ignorance of the environmental and health consequences of littering. Low levels of education were highly correlated with littering. |

Table A14: Literature Review on Waste Management

| *Study* | *Study type* | *Description and results* |
|---|---|---|
| Tanyanyiwa 2015 | Interviews with residents and street workers in Zimbabwe. | Reasons for the high littering levels are identified as: missing sense of ownership of public areas, the belief that someone else will clean up, and that littering is tolerated. Suggested ways to reduce litter include the provision of dedicated recycling bins, a volunteer environmental police force, and the establishment of a coordinated waste management system. |
| Torgler et al. 2009 | Analysis of over 30,000 respondents of the European Value Survey (EVS). | Using EVS data on basic values and beliefs of people in Europe, the authors find a positive, albeit small, relationship between how people perceive environmental cooperation (public littering) and their voluntary environmental morale. |