RUHR
ECONOMIC PAPERS

Ann-Kristin Reitmann
Maximiliane Sievert

Using Framed Field Experiments to Evaluate Real-World Policy Interventions – A Case Study on Changing Environmental Preferences

rwi #1072

Imprint

Ruhr Economic Papers

Published by

RWI – Leibniz-Institut für Wirtschaftsforschung
Hohenzollernstr. 1-3, 45128 Essen, Germany

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstr. 12, 45117 Essen, Germany

Editors

Editorial Office

Ruhr Economic Papers #1072

Responsible Editor: Manuel Frondel

ISSN 1864-4872 (online) – ISBN 978-3-96973-244-1

Ruhr Economic Papers #1072

Ann-Kristin Reitmann and Maximiliane Sievert

Using Framed Field Experiments to Evaluate Real-World Policy Interventions – A Case Study on Changing Environmental Preferences

Bibliografische Informationen der Deutschen Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at **http://dnb.dnb.de**

RWI is funded by the Federal Government and the federal state of North Rhine-Westphalia.

http://dx.doi.org/10.4419/96973244
ISSN 1864-4872 (online)
ISBN 978-3-96973-244-1

Ann-Kristin Reitmann and Maximiliane Sievert*

Using Framed Field Experiments to Evaluate Real-World Policy Interventions – A Case Study on Changing Environmental Preferences

Abstract

Many policy interventions in the environmental sector aim at changing environmental preferences, because these provide the basis for adopting environmental conservation behavior or technologies. To evaluate these changes, standard measurement tools in survey-based impact evaluations have their limitations. We discuss the potential of framed field experiments as an alternative to attain unbiased outcome measures, and present a case study from an impact evaluation in the Colombian coffee sector. While clear advantages exist, we show that indicators from framed field experiments require substantial sample sizes to provide well-powered results. Moreover, preference indicators are highly context specific, which calls for an elaborated framing to attain the intended kind of preferences.

JEL-Codes: C83, C93, O13, O22, Q20

Keywords: Field experiment; environmental preferences; impact evaluation; lessons learned; behavioral economics

March 2024

* *Ann-Kristin Reitmann, Leibniz University Hannover, University of Passau; Maximiliane Sievert, RWI.* *– We are thankful to researchers from the University of Passau, participants of the VfS Annual Conference 2021 and participants of the Leibniz Environment and Development Symposium 2022 for valuable comments on earlier versions of this paper. We also thank Niek de Jong for his support in the study, and Carlos Garcia, Gustavo Ochoa and their team for the great implementation of the survey and the experiment. Funding: The data underlying this research were collected for an impact evaluation commissioned by the Policy and Evaluation Department of the Ministry of Foreign Affairs of the Netherlands (IOB). The IOB had no involvement in any stage of this study. Author contributions: Ann-Kristin Reitmann: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization, Project administration; Maximiliane Sievert: Conceptualization, Methodology, Validation, Investigation, Writing - Review & Editing, Supervision, Project administration, Funding acquisition.* *– All correspondence to: Ann-Kristin Reitmann, Otto-Brenner-Straße 7, 30159 Hannover, Germany; annkristin.reitmann@ihe.uni-hannover.de*

1. Introduction

In impact evaluations of policy interventions, designing measurable outcome indicators is often a challenge. In the environmental sector, typical outcomes are the adoption of environment-friendly technologies and the improvement of natural resource quality, but increasingly also preferences in the form of environmental attitudes, beliefs, and norms. While the former are either objectively measurable or assessed through self-reports, preferences are more difficult to measure. Particularly in developing countries, where the institutional capacity to enforce environmental conservation is often low (Dietz, Ostrom, & Stern, 2003), policies and programs that aim at changing environmental preferences among the population are an important complement to institutional solutions. The assumption is that pro-environmental preferences form the basis for individuals to adopt certain behavior or technologies that lead to environmental conservation (Farrow, Grolleau, & Ibanez, 2017; Frey & Stutzer, 2008).

The difficulty in measuring (environmental) preferences lies in their lack of direct observation and reporting bias in self-reported assessments. Individuals may feel uncomfortable stating their true opinion. As a consequence, they refuse to answer or align their response with what is perceived socially acceptable (De Quidt, Vesterlund, & Wilson, 2019; DeMaio, 1984; Krumpal, 2013; Zizzo, 2010), especially when the response does not imply actions that are binding (Johansson-Stenman & Svedsäter, 2012; Kahneman & Knetsch, 1992). While there are different ways of attaining unbiased responses to sensitive concepts from social psychology and the political science literature[1], researchers in environmental and resource economics mainly revert to field experiments to measure environmental preferences (Harrison & List, 2004; List & Price, 2016). Field experiments use elements of traditional laboratory experiments to observe a subject in a controlled, but natural, real-world setting (Harrison & List, 2004). A large body of research analyzes the moderating role of social preferences and norms in the context of common-pool and natural resource management by means of field experiments (see e.g., Brent, Friesen, Gangadharan, & Leibbrandt, 2017 for a review). Yet, to the best of our knowledge, few studies use evidence from field experiments as the ultimate outcome measure in impact evaluations on environmental conservation interventions.

In this paper, we share our experience with using a framed field experiment for measuring environmental preferences as an outcome of a policy intervention and provide lessons learned. We applied a modified version of a public goods game with actual field context[2] to study the impact of an intervention among Colombian coffee-growing farmers. The

[1] To name a few, such measurement tools involve implicit association tests (see e.g., Greenwald, Nosek, & Banaji, 2003; Greenwald, Poehlman, Uhlmann, & Banaji, 2009; Kim, 2006) and survey experiments such as the endorsement experiment (see e.g., Blair, Fair, Malhotra, & Shapiro, 2013; Lyall, Blair, & Imai, 2013) and the list experiment (see e.g., Blair & Imai, 2012; Corstange, 2009; Glynn, 2013; Kuklinski, Cobb, & Gilens, 1997).

[2] We follow the notion of Voors et al. (2011), who conduct a similar experiment which can be interpreted to mimic a public goods game. Yet, in a strict sense, the applied experiment is rooted in a dictator or donation game (Eckel & Grossman, 1996), although the latter rather involve the measurement of prosocial behavior towards individuals outside the participants' societies and not towards a non-tangible concept such as the environment.

intervention aimed at improving water management in domestic and post-harvest coffee processing activities. An important part of the intervention was the provision of intensified training and extension services to small-scale coffee farmers intended to change their environmental mindset and eventually their behavior. In order to assess this impact, we relied on self-reported (stated) preferences and complemented these measures with results from a framed field experiment to elicit revealed preferences. Farmers were offered to donate an endowment to a local reforestation project.[3] This experimental measure assumes that farmers' willingness to donate correlates not only with their valuation of the particular reforestation project but also with their general valuation of environmental conservation.[4]

We find that that using an experimental measure for revealed preferences as an outcome in impact evaluation settings requires relatively large sample sizes, because of high standard deviations of such measures. This applies particularly if the measured impact is evaluated after a short period of time and, therefore, is expected to be small in size. Moreover, we observe our experimental measure from a framed field experiment to capture a context specific range of environmental preferences. By eliciting donations towards a reforestation project, we most likely measure the valuation of reforestation in particular instead of the valuation of overall environmental conservation. Our findings warrant future research to thoroughly pilot the experimental set-up to inform proper power calculations and test the adequate framing.

The remainder of the paper is organized as follows: We first summarize the opportunities of field experiments for policy evaluation in general in Section 2 and present our case study in the context of environmental and resource economics in Colombia in Section 3. In Section 4, we then discuss the pitfalls of using field experiments to measure outcomes for an impact evaluation and provide lessons learned from our experience. Section 5 concludes.

2. Field experiments for impact evaluation

Harrison and List (2004) distinguish between two types of field experiments: (i) artefactual field experiments and (ii) framed field experiments. The former is essentially the same as a conventional lab experiment, but with field subjects instead of the standard lab subject pool of students. The latter complements the artefactual field experiment with an actual field context in either the commodity, task or information set that subjects can use. Framed field experiments are particularly important in environmental and resource economics, because local (i.e., social and ecological) context is of particular importance for the evaluation of real behavior (Anderies et al., 2011), and hence, the underlying preferences (Brent et al., 2017).

Field experiments have fulfilled various purposes for policy design and evaluation. Werner and Riedl (2019) flag the potential of experiments to serve as 'testbeds' for policy interventions.

[3] This type of experiment is also applied in other contexts to capture more robust measures of environmental conservation preferences and behavior (see e.g., d'Adda, 2011; Karapetyan & d'Adda, 2014; Voors et al., 2011), but not as an outcome of an impact evaluation.

[4] Reforestation has positive multiplier effects on environmental conservation; trees planted around water sources prevent soil erosion and, thus, improve water quantity and quality. These effects were also emphasized during the intervention's training and extension sessions.

They allow for examining decisions of a broader range of subjects (i.e., beyond students, which are usually the subject pool of lab experiments) and to gain insights into systematic deviations from strict rationality that is usually building the foundation of economic decision-making models. These insights can form the basis for improving ex-ante predictions about policy outcomes, and potentially allowing for (re-)designing the respective policies (Gangadharan, Jain, Maitra, & Vecci, 2022; Werner & Riedl, 2019). List and Price (2016) summarize ways of how field experiments have been used to design environmental policies in particular, and they highlight their value in informing cost-benefit analysis and efforts to promote resource conservation.

In terms of policy evaluation, the literature proposes the use of field experiments to collect both explanatory covariates and to conduct them as part of the outcome. The former serves particularly useful for analyzing mechanisms that lead to the observed outcomes (Gneezy & Imas, 2017). Yet, field experiments can also be applied for measuring preferences, social norms, beliefs and actual behavior – both in form of the dependent or independent variable (Gneezy & Imas, 2017; Viceisza, 2016). For both types of variables, impact evaluations require unbiased measures, especially when serving as the main outcome. Experimental measures have the potential to serve as viable measures, because they are less likely to be biased. The concern of biasedness is particularly high for unobservable preference concepts such as attitudes, beliefs and social norms. Self-reports often suffer from various sources of responding bias: respondents tend to overstate their preferences in hypothetical settings (i.e., 'hypothetical bias'), because their response does not imply binding actions (de Corte, Cairns, & Grieve, 2021; Johansson-Stenman & Svedsäter, 2012). The bias is heightened when preferences towards sensitive or 'moral' topics are elicited and respondents align their answer with what they perceive is socially desirable (i.e., 'social desirability bias') (DeMaio, 1984; Krumpal, 2013). Such biases can create a significant mismatch between (self-reported) preferences and actual behavior (Jerolmack & Khan, 2014). This is also a common phenomenon in the environmental sector, where pro-environmental preferences are stated because they are perceived socially desirable, but practical action is costly and/or not publicly visible (Bravo & Farjam, 2022). In fact, various studies report weak to no correlation between self-reported environmental preferences, and actual behavior, questioning the validity of the former (e.g., Bravo & Farjam, 2022; Carrington, Neville, & Whitwell, 2014; Farjam, Nikolaychuk, & Bravo, 2019; Whitmarsh, 2009). Voors et al. (2011) find the same in the context of social preferences and conservation behavior, where they apply different types of experiments. They observe that preferences and behavior as portrayed in a framed field experiment are associated with higher conservation efforts outside the experiment, whereas they fail to establish a strong correlation with stated survey measures (see also Karapetyan & d'Adda, 2014). Considering that 'real stakes' (mostly expressed in monetary terms) are introduced in experiments, the opportunity costs for hypothetical or socially desirable behavior are too high, reducing the possibility of biased preference measures.

Few studies use field experiments to measure outcomes in the impact evaluation literature. Jakiela, Miguel and te Velde (2015) use a modified dictator game to measure changes in social preferences and cultural values as an impact of an education initiative targeting adolescent girls in Kenya. Also within the educational context, Rao (2019) uses lab and field experiments (i.e., a dictator game) to measure social behavior among Indian students. Almas et al. (2018), Brar et al. (2023) and Schaner (2017) use self-designed experiments to measure female empowerment in the context of a conditional cash transfer program in Macedonia, a gender norms intervention targeting adolescents in Somalia and a program that provided bank accounts in Kenya, respectively. However, to the best of our knowledge, no study has applied a (framed) field experiment to measure environmental preferences in the context of environmental policies.

3. Case study: Environmental preferences among Colombian coffee-farmers

As a case study, we use the framed field experiment that we applied in the course of an impact evaluation of an intervention implemented among Colombian small-scale coffee farmers. The intervention had the goal to improve water management in domestic and post-harvest coffee processing activities, because both activities are traditionally very water-intensive and pose a threat to the environment through effluent wastewater that contaminates surface water bodies. Apart from subsidizing improved hardware for a small share of coffee farmers, all beneficiaries received intensified training and extension services, where farmers learned about the general importance of environmental conservation (i.e., in particular, the protection of water sources), and how to implement more environment-friendly practices. By providing crucial information, the latter intervention component had the potential to form pro-environmental preferences. This can create an intrinsic motivation for sustainable behavior that ultimately results in increased adoption of environmental conservation devices and practices. We collected data among coffee farms located in 25 treatment river basins, who benefited from the intervention, before and after the implementation. To evaluate its impacts, we compare the treatment group to a control group, which consisted of coffee farmers in 25 comparable river basins, within a difference-in-differences (DID) design.[5] We rely on two different preference measures: (i) a stated preference measure derived from self-reported statements (potentially suffering from reporting bias)[6], and (ii) a revealed preference measure derived from a framed field experiment.

3.1 Revealed preferences from a framed field experiment

To complement the stated measures on environmental conservation preferences with a more robust, revealed preference measure, we conducted a behavioral experiment after the endline survey with a sub-sample of 661 coffee farmers (equally distributed among the treatment and control group). The sample farmers received 20,000 Colombian Pesos (COP)[7]

[5] For more information on the impact evaluation relating to environmental preferences see Reitmann (2020).

[6] The coffee farmers were asked about how much they identify themselves with seven statements on pro-environmental behavior. For the stated preference measure, responses to all seven statements are summarized into an index of standardized outcomes (see Appendix B for more information).

[7] 20,000 COP is equivalent to around 5.50 EUR (date of conversion: 19.02.2020).

and were asked how much they would be willing to donate to a reforestation project. Money not donated to the reforestation project could be kept by the farmers. The experiment was introduced as an opportunity for donating to a reforestation project, not as an experiment. No concrete project was named, the farmer was only informed that 20,000 COP suffice for planting 25 trees (including operation costs) and that they would be planted in their own district.[8] To ensure anonymity, the coffee farmer received an envelope to make the donation in privacy.[9] The donations were kept private to measure the intrinsic motivation of the individual farmer (d'Adda, 2011). 20,000 COP were slightly below the daily minimum wage of 24,600 COP at the time of the survey and, therefore, a substantial amount of money to the rural coffee farmers.

The experiment requires farmers to weigh the individual monetary short-term benefits from keeping the money against longer-term environmental benefits for the whole community through reforestation. Although the experiment is framed in the context of reforestation, it aims to capture not only the farmer's valuation of reforestation in particular, but the valuation of environmental conservation in general. The importance of reforestation and the implications for the quality and quantity of local water resources (e.g., prevention of erosion) was emphasized throughout the intervention. Hence, the underlying hypothesis is that farmers who value (local) environmental quality, namely those who have high environmental conservation preferences, donate more. If the intervention has a positive impact on environmental conservation preferences, farmers in the treatment area should ultimately donate more than control group farmers. The revealed preferences can be expected to be less biased than the stated preferences for environmental conservation, because the 'real stakes' introduce opportunity costs for socially desirable behavior.

Given that the donation in our experiment is directed towards a local reforestation project, it can be interpreted to mimic a public goods game (Voors et al., 2011). The donation experiment has been implemented in varying settings to elicit pro-environmental preferences. d'Adda (2011), for instance, applies the experiment in Bolivia to examine how non-monetary and non-regulatory incentives affect pro-social behavior for environmental conservation. In another study, Karapetyan and d'Adda (2014) use donations to an environmental NGO in Sierra Leone to capture individual preferences for environmental conservation. However, to the best of our knowledge, this is the first to use the donation experiment for evaluating an actual intervention.

3.2 Evaluation strategy

To analyze the impact of the intervention on revealed preferences, which were only collected at endline, we use a simple difference approach (i.e., the difference between treatment and control group at endline only):

[8] After project closure, the collected donations were given to local organizations in the river basins with a signed agreement to direct them towards reforestation in the district.

[9] The endowment was given out in ten 2,000 COP bills. Hence, donations could only be made in 2,000 COP steps.

$$D_{frd} = \beta_{Treat}Treat_r + \beta_{Cov}X_{frd} + \alpha_d + u_{frd}, \quad (1)$$

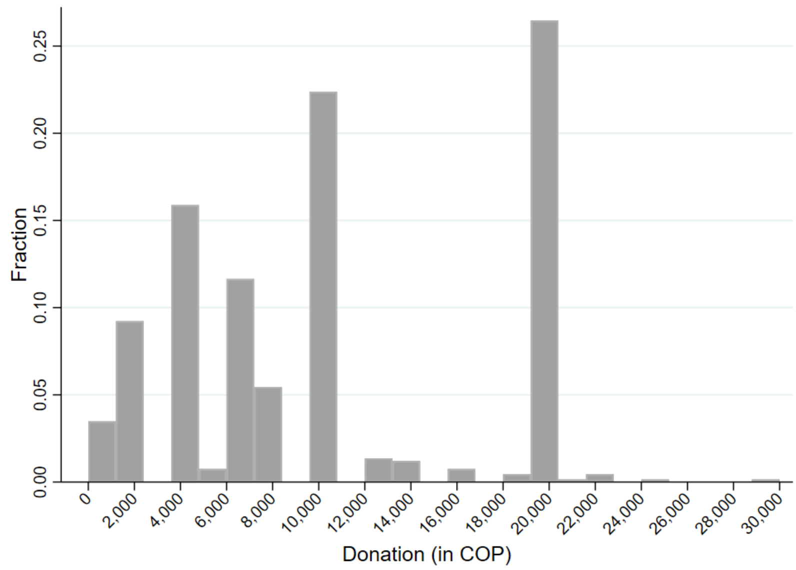
where D_{frd} denotes the contribution to the reforestation project from farm (or coffee farmer) f in river basin r in department d, which serves as a proxy fir revealed environmental conservation (EC) preferences. $Treat_r$ is a binary variable equal to one when the intervention was implemented in a treatment river basin. β_{Treat} measures the impact of the intervention on the outcome. α_d is the department fixed effect that controls for time-invariant departmental specific characteristics.[10] We include further covariates such as household and farm characteristics in the vector X_{frd}. The vector also contains other variables that might affect donation behavior for a reforestation project. The residual is allowed to be correlated within the river basin (i.e., the cluster), so all standard errors and tests are robust to intra-cluster correlation. Considering that the treatment and control group farmers did not differ significantly at the baseline for a wide range of observable variables (see Reitmann, 2020), we assume that donation behavior would have been similar as well, which justifies the use of the simple difference approach. The training was offered to all coffee farmers in the specified 'area of influence' in the treatment river basins. Participation, however, was not mandatory and invited farmers (and their families) could decide how many training sessions they wanted to attend. Therefore, the estimates need to be interpreted as intention-to-treat (ITT) effects. That is, the average impact of giving access to the intervention's intensified training and extension services in the treatment river basins, regardless of actual uptake.

3.3 The impact of the intervention on revealed preferences

On average, farmers donated 10,277 COP (with a standard deviation of 6,814 COP), which is around half the endowment they received. Figure 1 illustrates that 23% of the farmers donated exactly half of their endowment. 25% of the farmers donated the whole amount and some farmers even donated more than their initial endowment of 20,000 COP by adding up to 10,000 COP from their own resources. By contrast, only a very small fraction of less than 5% kept the whole amount of 20,000 COP to themselves.

[10] The 25 treatment river basins where equally distributed among five departments in the Colombian coffee-farming region. They differ with respect to climatic conditions and, moreover, environmental laws or other regulations are often implemented at the department level.

Figure 1: Donations to reforestation project



Note: Data from endline (2017) survey. The number of observations is 661 (i.e., treatment and control group farms are pooled).

Table 1: Impact of the intervention on revealed preferences (ITT effect)

	Donation (in COP)				
	(1)	(2)	(3)	(4)	(5)
Treatment (=1)	655.4	805.6	785.4	820.2	783.2
	(1,075.4)	(914.8)	(899.3)	(902.3)	(890.6)
Constant	12,480.6***	7,420.2***	5,053.0**	7,364.3***	5,063.6**
	(858.4)	(1,788.9)	(1,986.1)	(1,753.3)	(2,019.3)
Observations	661	661	629	661	629
R-squared	0.036	0.099	0.114	0.099	0.114
Baseline controls	YES	YES	YES	YES	YES
Endline controls	NO	YES	YES	YES	YES
Trust controls	NO	NO	YES	NO	YES
Stated preferences control	NO	NO	NO	YES	YES
Department FE	YES	YES	YES	YES	YES
Endline mean in control group			9,960.5		

Note: Data from endline (2017) survey. Standard errors (in parentheses) and tests are robust to intra-river basin correlation (there are 30 clusters/river basins). All regressions also include controls for unbalanced baseline characteristics (i.e., the perception of erosion) and department fixed effects. The coefficients in Columns 1–5 are OLS estimates of β_{Treat} in Equation 1. The dependent variable is the donation in COP. The endline means reported at the bottom are calculated for control river basins. *** indicates significance at 1%, ** at 5%, and * at 10%.

Table 1 shows the results of estimating Equation 1. For now, we focus on the impact of the intervention, while only controlling for unbalanced baseline characteristics and department fixed effects (see Column 1).[11] Farmers in the treatment group donate on average more than

[11] When including further controls that potentially affect donations, the difference in donations between treatment and control group slightly increases, but the difference remains statistically insignificant (see Table 1, Columns 2-5).

the farmers in the control group. However, the difference in donations is not statistically significant at standard levels. The high average donations indicate a relatively high level of environmental valuation among the overall experimental sample. However, there is no evidence of stronger pro-environmental preferences among those that benefited from the intervention compared to those that did not.

Factoring in the results from the stated and revealed preferences (the former are not discussed in this paper), we do not find significant impacts of the intervention on stated or revealed preferences towards environmental conservation among treated relative to control farmers – even though around three-quarter of the sampled coffee farmers in the treatment group took up the training. The ITT effects for both preference measures are positive, but the null hypothesis of no impact cannot be rejected at standard levels. In the original impact evaluation paper, we discuss various explanations for these null findings (see Reitmann, 2020). Those related to the framed field experiment are discussed in the following and will serve as our lessons learned.

4. Lessons learned

4.1 Experimental measures require large sample sizes

Due to budget and logistical constraints, the donation experiment was implemented only (i) at follow-up and (ii) with a sub-sample of 661 farmers in 30 river basins. These circumstances contribute to important limitations of the experiment related to statistical power.

Looking at the minimum detectable effect (MDE) size ex-post, we can see that our study is substantially underpowered for detecting small changes in our outcome variable. The MDE is the smallest effect that one can distinguish from a null hypothesis of no effect with the given power. Assuming a power of 80%, we calculate an MDE of 3,012 COP for a 5% significance level and 2,668 COP for a 10% significance level, which is both far above the estimated effect of 655 COP (see Table 1, Column 1). The effect of 655 COP represents a fairly small effect of 0.1 standard deviations,[12] which requires a larger overall sample size or – as in our case – a higher number of clusters. In fact, with an average number of 22 farmers in each river basin, we would need 328 control and 328 treatment cluster to detect a significant effect of 655 COP. This would amount to a total sample size of 14,432 farmers.

If we compare these power calculations with those for our other impact indicators in the full impact evaluation study (i.e., the adoption of water and soil conservation devices and practices in domestic and productive coffee farming activities), the experimental indicator fares particularly bad in terms of power. The reason is that adoption is often measured in binary units, while preference measures are more complex and can involve measurement error, which, while not causing estimation bias under certain conditions, will increase the

[12] The mean donation in the control group is 9,960 COP with a standard deviation of 6,868 COP. The intra-cluster correlation is 0.2.

standard errors of the estimates (see also Tarozzi, Desai, & Johnson, 2015). Preferences on environmental conservation might be guided by prevailing social norms, which means that intra-cluster correlation is high, increasing standard errors even further. This means that detecting effects on experimental measures for preferences and achieving more precise estimates requires a higher sample size than for less variable outcomes from non-experimental measures.[13,14] This calls for separate power calculations for all relevant types of outcome measures and consequently, if two or more data collections are planned, the consideration of the experimental sample size before baseline collection.

4.2 Preferences are very context-specific

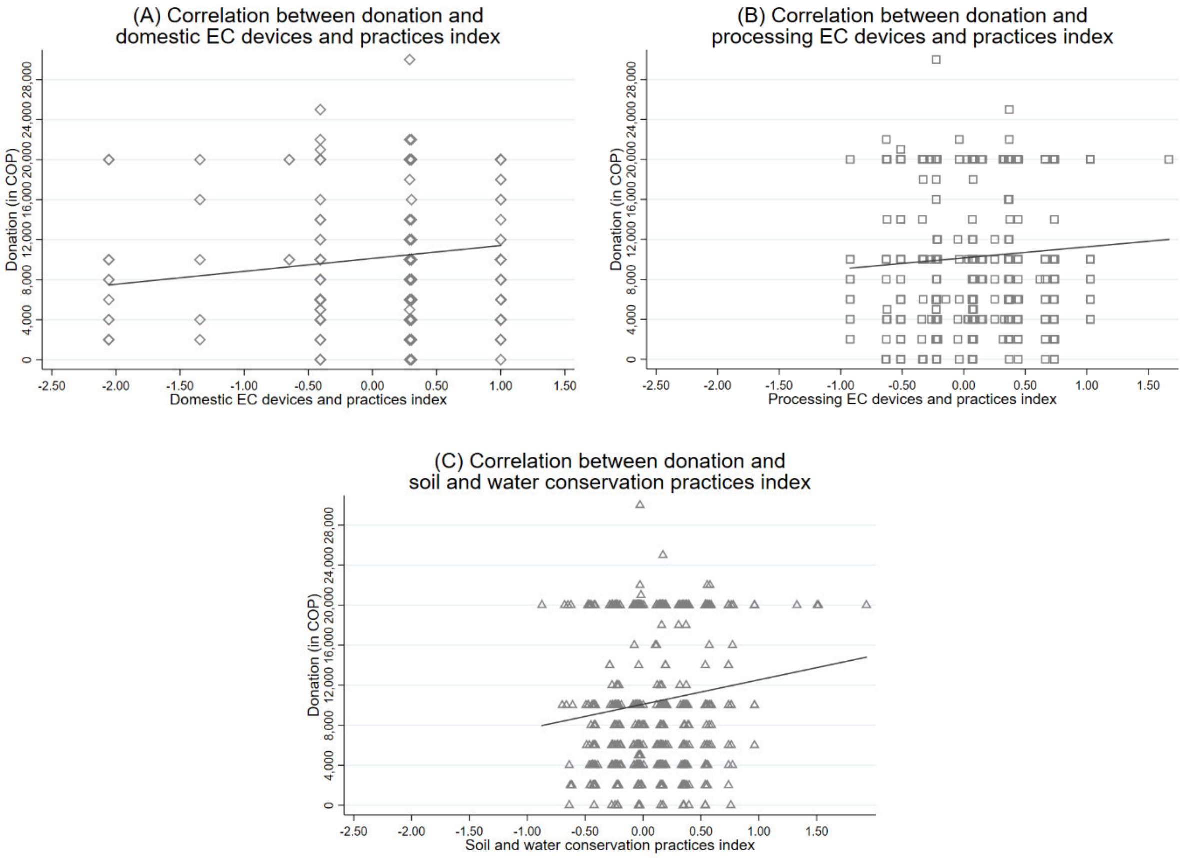
In this section, we critically reflect on using our specific framed field experiment for measuring unbiased preferences on environmental conservation. We initially argued that the reforestation frame does not just capture the farmer's valuation of reforestation in particular, but the valuation of environmental conservation in general. This argumentation is based on the premise that proper reforestation has positive multiplier effects that contribute to environmental and water resource conservation; that is trees planted around water sources prevent soil erosion and, thus, improve water quantity and quality (see e.g., d'Adda, 2011). These effects were also emphasized throughout the intervention and particularly in the training focusing on environmental conservation. Yet, we cannot exclude the possibility that we are indeed just capturing the valuation of reforestation in particular.

We probe into this by comparing the experimental revealed preference measure with actual environmental conservation behavior. We constructed three standardized indices for different types of behavior: domestic activities, coffee processing activities, soil and water conservation activities (see Appendix A for more information). Figure 2 maps the level of donations against the three standardized indices for environmental conservation behavior (i.e., Panel A: domestic activities, Panel B: coffee processing activities, Panel C: soil and water conservation activities). All three measures of environmental conservation behavior are positively correlated. Yet, only for the latter two the pairwise correlation coefficients are significant and range around 0.12, indicating a small correlation. Conclusions from these pairwise correlations have to be interpreted with caution, as other factors can play a role for the connection between the preference measures and behavior. However, the weak correlation between revealed preferences and environmental conservation behavior might indicate that the conducted donation experiment does not necessarily fully capture preferences towards environmental conservation.

---

[13] This is easily illustrated with the following formula: $SE = \frac{\sigma}{\sqrt{n}}$; if the variance of a variable is high, it increases σ. Hence, in order to get more precise estimates (i.e., a smaller standard error SE), a larger sample size n is required.
[14] The variance of the experimental measure also depends on the form of measurement. In our particular experiment, for instance, a lower number of groups would have reduced the variance. Hence, instead of donation in 2,000 COP steps, we could have chosen 4,000 COP steps.

Figure 2: Correlation between revealed preferences and actual behavior



Note: Data from endline (2017) survey. The number of observations is 661 (i.e., treatment and control group farms are pooled). Panel A: the pairwise correlation coefficient is 0.076. Panel B: the pairwise correlation coefficient is 0.121***. Panel C: the pairwise correlation coefficient is 0.125***. *** indicates significance at 1%, ** at 5%, and * at 10%.

Moreover, we look more closely into determinants of higher revealed EC preferences in our experiment. Table 2 shows the full set of results regressing the amount donated on the treatment dummy and additional control variables (measured at endline)[15]. While the coefficient for the treatment dummy remains statistically insignificant, three other variables are significantly affecting donations.

First, if the farmer perceived deforestation as a problem in the district, donations are, on average, 2,000 COP higher. This finding suggests that the desire to contribute to conservation is fostered by awareness of environmental problems, in this case particularly perceived deforestation. This finding is in line with theoretical considerations from Forsyth et al. (2004); Gould, Saupe, and Klemme (1989); and Ostrom (1999), and further supported by empirical studies (see e.g., Bardhan, 2000; Jumbe & Angelsen, 2007; Karapetyan & d'Adda, 2014; Traoré, Landry, & Amara, 1998; Voors et al., 2011). However, we observe that perception of other

[15] We use the endline value of these variables, because we are interested in the correlation between characteristics of the respondent that might affect donations at the time of the experiment (the experiment was conducted subsequent to the endline survey).

environmental problems (i.e., water shortage and erosion) have no significant influence. This indicates that farmers do not necessarily associate reforestation with environmental conservation beyond the positive effects for reforestation in particular.[16] Hence, support for the reforestation might not be a proxy for valuing environmental conservation in general, but only for valuing reforestation.

Second, wealthier farmers (measured through an asset index) donate, on average, higher amounts. The endowment is close to the daily minimum wage and poorer farmers might be constrained to support the reforestation project due to financial hardship.

Third, farmers with higher levels of trust towards neighbors (particularly neighboring coffee farmers) donate, on average, 1,600 COP more.[17] This finding is in line with studies that show that a high level of trust within the community is necessary to ensure successful common pool resource management and minimize free riding (see e.g., Aida, 2019; Bouma, van Soest, & Bulte, 2008). For reforestation projects, a high level of trust towards neighboring coffee farmers is also of particular relevance.

Although the donation experiment was intended to capture overall preferences on environmental conservation, it seems to capture solely reforestation-specific preferences. Researchers that plan to use a framed field experiment to measure environmental preferences therefore need to make sure to carefully select the framing of their setting and to thoroughly test it.

[16] As mentioned earlier, reforestation reduces erosion with multiplier effects on water quantity and quality. Knowing that, the perception of water shortage and erosion should influence the support for a reforestation project as well. In fact, the survey posed the following knowledge question to the farmers: "What practice protects water sources?". Among three possible answers, the correct one is "plant native shrubs and trees". A control for whether they answered the question correctly is included (see coefficient "correct knowledge question"), but suggests no influence.

[17] A small share of farmers refused to answer the trust questions, which is why the number of observations drops to 629 when controlling for trust. When running the regressions in Columns 1,2 and 4 with the steady sample of 629 observations, most effect sizes only change to a small extent and statistical significance remains largely unchanged.

Table 2: Impact of the program on revealed preferences (ITT effect and covariates)

	Donation (in COP)				
	(1)	(2)	(3)	(4)	(5)
Treatment (=1)	655.4	805.6	785.4	820.2	783.2
	(1,075.4)	(914.8)	(899.3)	(902.3)	(890.6)
Correct knowledge quest. (=1)		-255.6	879.2	-229.5	875.0
		(1,453.4)	(1,285.5)	(1,412.0)	(1,226.6)
Conducted reforestation (=1)		-280.9	-184.3	-275.1	-185.2
		(542.4)	(493.6)	(551.6)	(496.9)
Perceived water shortage (=1)		415.4	377.0	407.3	378.1
		(704.0)	(691.3)	(691.9)	(680.9)
Perceived erosion (=1)		2,094.1***	2,040.5***	2,095.1***	2,040.4***
		(502.0)	(535.2)	(505.5)	(536.0)
Perceived deforestation (=1)		779.2	481.3	790.5	479.4
		(641.2)	(607.8)	(642.9)	(609.6)
Group member (=1)		749.6	467.9	784	476.7
		(668.2)	(637.6)	(664.9)	(628.2)
Trust progr. implementer (=1)			737.7		736.4
			(808.9)		(818.2)
Trust neighbors (=1)			1,654.6***		1,654.5***
			(552.8)		(553.9)
Asset index; 1st quartile		Ref.	Ref.	Ref.	Ref.
2nd quartile		1,642.1***	1,395.7**	1,647.1***	1,394.8**
		(542.5)	(572.2)	(553.2)	(583.4)
3rd quartile		2,667.1***	2,650.9***	2,678.7***	2,648.9***
		(685.3)	(719.0)	(711.3)	(738.7)
4th quartile		2,211.8***	1,976.1**	2,219.0***	1,974.7**
		(667.0)	(742.7)	(667.8)	(743.3)
Secure land ownership (=1)		908.7	875.7	916.4	874.2
		(694.8)	(693.8)	(706.5)	(709.1)
Total farm area (ha)		34.0	32.2	34.1	32.1
		(26.7)	(26.7)	(26.6)	(26.7)
Coffee production (@cps)		1.2	1.2	1.2	1.2
		(1.3)	(1.2)	(1.3)	(1.2)
Hired workers (=1)		449.0	532.5	457.8	531.3
		(935.1)	(901.4)	(940.5)	(910.0)
EC stated preferences index				-88.5	15.0
				(600.9)	(600.4)
Constant	12,480.6***	7,420.2***	5,053.0**	7,364.3***	5,063.6**
	(858.4)	(1,788.9)	(1,986.1)	(1,753.3)	(2,019.3)
Observations	661	661	629	661	629
R-squared	0.036	0.099	0.114	0.099	0.114
Baseline controls	YES	YES	YES	YES	YES
Endline controls	NO	YES	YES	YES	YES
Department FE	YES	YES	YES	YES	YES
Endline mean in control group			9,960.5		

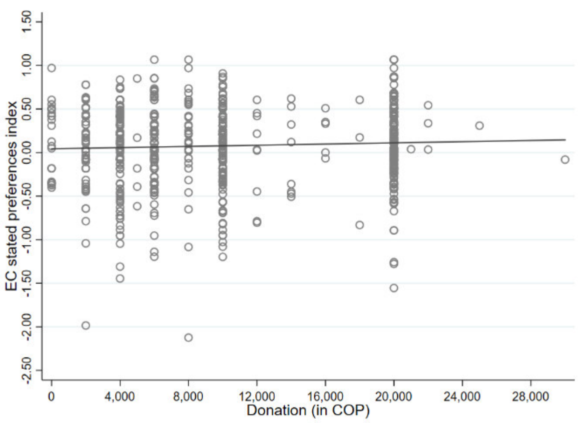
Note: Data from endline (2017) survey. Robust standard errors in parentheses. All regressions include controls for unbalanced baseline characteristics (i.e., the perception of erosion) and department fixed effects. The coefficients in Columns 1–5 are OLS estimates of β_{Treat} in Equation 1. The dependent variable is the donation in COP. The knowledge questions was 'What practice protects water sources?' with the correct answer being 'plant native shrubs and trees'. Group membership indicates that the farmer or his/her spouse participates in a local community organization or association. The trust variable was transformed into a binary variable with 1 = medium to strong trust. The asset index is calculated using factor analysis based on housing characteristics and ownership of various durables (first factor divided into quartiles). Secure land ownership is defined as holding the deed or any other relevant document to the land. Coffee production is measured in "arrobas de café pergamino seco" (@cps) with 1 @cps equaling 12.5 kg of dry parchment coffee. The EC preferences index is described in detail in Appendix B. The endline means reported at the bottom are calculated for control river basins. *** indicates significance at 1%, ** at 5%, and * at 10%.

4.3 Experimental measures can be used to validate non-experimental measures

Conditional on having a well-framed field experiment, the experimental measure can serve as an unbiased outcome for pro-environmental preferences. Though, as mentioned in Section 4.1, this strategy might involve high sample sizes and particularly a large number of clusters in clustered samples. If the necessary sample size is not realizable due to budget constraints or limitations in possible clusters, the experimental measure can be used to validate non-experimental measures (see e.g., Schaner, 2017).

The validation exercise allows to assess whether the non-experimental measures suffer from reporting bias. One could think of eliciting several non-experimental measures using the full sample size and assessing their biasedness in a subsample with an experimental measure. This requires a much smaller sample size for the experimental measure (as opposed to using the experimental measure directly as an outcome), as long as the experimental sample can be considered representative for the overall sample.

Figure 3: Correlation between revealed and stated preferences



Note: Data from endline (2017) survey. The number of observations is 661 (i.e., treatment and control group farms are pooled). The pairwise correlation coefficient is 0.051.

In our study, coffee farmers were directly asked about how much they identify themselves with seven statements on pro-environmental behavior. To construct a measure for stated preferences on environmental conservation, the responses to all seven statements is combined in an index of standardized outcomes (see Appendix B for more information). Figure 3 maps the values of both stated and revealed preference measures against each other. The line represents the linear fit. The almost flat linear line reveals a very weak correlation between the two preference measures.[18] The pairwise correlation coefficient is very small at

[18] The stated EC preferences index also enters the regression of the results presented in Table 2 in Column 4 and 5. Similar to results in Figure 3, the stated EC preferences do not significantly relate to revealed EC preferences.

0.05 and statistically insignificant at standard levels. Yet, in combination with what we discussed in Section 4.2, the lack of correlation may also stem from the fact that our experimental measure captures something different than overall environmental valuation. More specifically, none of the seven statements is directly related to reforestation behavior, which may explain the low level of correlation.

5. Conclusion

This paper discusses the application of framed field experiments for measuring outcomes in impact evaluation studies. Particularly in the field of environmental conservation, unbiased estimates of preferences are important but challenging to elicit due to potentially biased reporting behavior in survey questions and stated preference settings. We therefore discuss our experience with a modified version of a public goods game with Columbian coffee growers as a revealed preference measure for changes in pro-environmental preferences.

Our results show that substantial sample sizes are required in impact evaluation settings. Effect sizes that can realistically be expected from policy interventions, often measured after a short timeframe, are modest and the experimental indicators often have a high standard deviation. Accordingly, statistical power to detect such changes is often low unless big sample sizes are available. Furthermore, careful design decisions have to be taken: our experimental preferences indicator turns out to be rather context specific and cannot be interpreted as an overall measure of pro-environmental preferences.

For future studies that intend to use framed field experiments as an outcome measure, we therefore recommend thorough piloting of the experimental set-up. First, realistic insights into the measure's variance should be obtained, because they are important for proper power calculations. Often, such well-informed power calculations will illustrate the need for big sample sizes that are possibly even more challenging than in purely survey-based settings if real-stakes experiments are implemented requiring meaningful payouts. Second, the specific context and the target population's interpretation of the experiment need to be understood. The context dependence of lab-experiments has widely been acknowledged and particularly in framed field experiments, a realistic framing is crucial.

References

Aida, T. (2019). Social capital as an instrument for common pool resource management: A case study of irrigation management in Sri Lanka. *Oxford Economic Papers*, *71*(4), 952–978. https://doi.org/10.1093/oep/gpy058

Almas, I., Armand, A., Attanasio, O., & Carneiro, P. (2018). Measuring and changing control: Women's empowerment and targeted transfers. *Economic Journal*, *128*, F609–F639. https://doi.org/10.1111/ecoj.12517

Anderies, J. M., Janssen, M. A., Bousquet, F., Cardenas, J. C., Castillo, D., Lopez, M. C., … Wutich, A. (2011). The challenge of understanding decisions in experimental studies of common pool resource governance. *Ecological Economics*, *70*(9), 1571–1579. https://doi.org/10.1016/j.ecolecon.2011.01.011

Bardhan, P. (2000). Irrigation and Cooperation: An Empirical Analysis of 48 Irrigation Communities in South India. *Economic Development and Cultural Change*, *48*(4), 847–865. https://doi.org/10.1086/452480

Blair, G., Fair, C. C., Malhotra, N., & Shapiro, J. N. (2013). Poverty and Support for Militant Politics: Evidence from Pakistan. *American Journal of Political Science*, *57*(1), 30–48. https://doi.org/10.1111/j.1540-5907.2012.00604.x

Blair, G., & Imai, K. (2012). Statistical Analysis of List Experiments. *Political Analysis*, *20*(1), 47–77. https://doi.org/10.1093/pan/mpr048

Bouma, J., van Soest, D., & Bulte, E. (2008). Trust, Trustworthiness and Cooperation: Social Capital and Community Resource Management. *Journal of Environmental Economics and Management*, *56*(2), 155–166. https://doi.org/10.1016/j.jeem.2008.03.004

Brar, R., Buehren, N., Papineni, S., & Sulaiman, M. (2023). *Rebel with a Cause: Effects of a Gender Norms Intervention for Adolescents in Somalia*. Policy Research Working Paper (No. 10567). Washington, D.C.: World Bank.

Bravo, G., & Farjam, M. (2022). Actions speak louder than words: Attitudes, behaviour, and partisan identity in a polarised environmental domain. *Energy Research and Social Science*, *90*, 102547. https://doi.org/10.1016/j.erss.2022.102547

Brent, D. A., Friesen, L., Gangadharan, L., & Leibbrandt, A. (2017). Behavioral Insights from Field Experiments in Environmental Economics. *International Review of Environmental and Resource Economics*, *10*, 95–143. https://doi.org/10.1561/101.00000084

Carrington, M. J., Neville, B. A., & Whitwell, G. J. (2014). Lost in translation: Exploring the ethical consumer intention-behavior gap. *Journal of Business Research*, *67*(1), 2759–2767. https://doi.org/10.1016/j.jbusres.2012.09.022

Corstange, D. (2009). Sensitive Questions, Truthful Answers? Modeling the List Experiment with LISTIT. *Political Analysis*, *17*(1), 45–63. https://doi.org/10.1093/pan/mpn013

d'Adda, G. (2011). Motivation crowding in environmental protection: Evidence from an artefactual field experiment. *Ecological Economics*, *70*, 2083–2097. https://doi.org/10.1016/j.ecolecon.2011.06.006

de Corte, K., Cairns, J., & Grieve, R. (2021). Stated versus revealed preferences: An approach to reduce bias. *Health Economics*, *30*(5), 1095–1123. https://doi.org/10.1002/hec.4246

De Quidt, J., Vesterlund, L., & Wilson, A. J. (2019). Experimenter demand effects. In A.

Schram & A. Ule (Eds.), *Handbook of Research Methods and Applications in Experimental Economics* (pp. 384–400). https://doi.org/10.4337/9781788110563

DeMaio, T. J. (1984). Social desirability and survey measurement: A review. In C. F. Turner & E. Martin (Eds.), *Surveying Subjective Phenomena* (Vol. 2, pp. 257–282). New York, NY: Russell Sage Foundation.

Dietz, T., Ostrom, E., & Stern, P. C. (2003). The Struggle to Govern the Commons. *Science*, *302*(5652), 1907–1912. https://doi.org/10.1126/science.1091015

Eckel, C. C., & Grossman, P. J. (1996). Altruism in Anonymous Dictator Games. *Games and Economic Behavior*, *16*(2), 181–191. https://doi.org/10.1006/game.1996.0081

Farjam, M., Nikolaychuk, O., & Bravo, G. (2019). Experimental evidence of an environmental attitude-behavior gap in high-cost situations. *Ecological Economics*, *166*, 106434. https://doi.org/10.1016/j.ecolecon.2019.106434

Farrow, K., Grolleau, G., & Ibanez, L. (2017). Social Norms and Pro-environmental Behavior: A Review of the Evidence. *Ecological Economics*, *140*, 1–13. https://doi.org/10.1016/j.ecolecon.2017.04.017

Forsyth, D. R., Garcia, M., Zyzniewski, L. E., Story, P. A., & Kerr, N. A. (2004). Watershed Pollution and Preservation: The Awareness-Appraisal Model of Environmentally Positive Intentions and Behaviors. *Analyses of Social Issues and Public Policy*, *4*(1), 115–128. https://doi.org/10.1111/j.1530-2415.2004.00037.x

Frey, B. S., & Stutzer, A. (2008). Environmental morale and motivation. In A. Lewis (Ed.), *Cambridge Handbook of Psychology and Economic Behaviour* (pp. 406–428). Cambridge, UK: Cambridge University Press.

Gangadharan, L., Jain, T., Maitra, P., & Vecci, J. (2022). Lab-in-the-field experiments: perspectives from research on gender. *Japanese Economic Review*, *73*(1), 31–59. https://doi.org/10.1007/s42973-021-00088-6

Glynn, A. N. (2013). What can we learn with statistical truth serum? *Public Opinion Quarterly*, *77*(S1), 159–172. https://doi.org/10.1093/poq/nfs070

Gneezy, U., & Imas, A. (2017). Lab in the Field: Measuring Preferences in the Wild. In *Handbook of Economic Field Experiments* (Vol. 1, pp. 439–464). Elsevier Ltd. https://doi.org/10.1016/bs.hefe.2016.08.003

Gould, B. W., Saupe, W. E., & Klemme, R. M. (1989). Conservation Tillage: The Role of Farm and Operator Characteristics and the Perception of Soil Erosion. *Land Economics*, *65*(2), 167–182. https://doi.org/10.2307/3146791

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and Using the Implicit Association Test: I. An Improved Scoring Algorithm. *Journal of Personality and Social Psychology*, *85*(2), 197–216. https://doi.org/10.1037/0022-3514.85.2.197

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and Using the Implicit Association Test: III. Meta-Analysis of Predictive Validity. *Journal of Personality and Social Psychology*, *97*(1), 17–41. https://doi.org/10.1037/a0015575

Harrison, G. W., & List, J. A. (2004). Field experiments. *Journal of Economic Literature*, *42*(4), 1009–1055. https://doi.org/10.1257/0022051043004577

Jakiela, P., Miguel, E., & te Velde, V. L. (2015). You've earned it: Combining field and lab experiments to estimate the impact of human capital on social preferences. *Experimental Economics*, *18*, 385–407. https://doi.org/10.1007/s10683-014-9409-9

Jerolmack, C., & Khan, S. (2014). Talk Is Cheap: Ethnography and the Attitudinal Fallacy. *Sociological Methods and Research*, *43*(2), 178–209. https://doi.org/10.1177/0049124114523396

Johansson-Stenman, O., & Svedsäter, H. (2012). Self-image and valuation of moral goods: Stated versus actual willingness to pay. *Journal of Economic Behavior and Organization*, *84*(3), 879–891. https://doi.org/10.1016/j.jebo.2012.10.006

Jumbe, C. B. L., & Angelsen, A. (2007). Forest dependence and participation in CPR management: Empirical evidence from forest co-management in Malawi. *Ecological Economics*, *62*(3–4), 661–672. https://doi.org/10.1016/j.ecolecon.2006.08.008

Kahneman, D., & Knetsch, J. L. (1992). Valuing public goods: The purchase of moral satisfaction. *Journal of Environmental Economics and Management*, *22*(1), 57–70. https://doi.org/10.1016/0095-0696(92)90019-S

Karapetyan, D., & d'Adda, G. (2014). Determinants of conservation among the rural poor: A charitable contribution experiment. *Ecological Economics*, *99*, 74–87. https://doi.org/10.1016/j.ecolecon.2014.01.009

Kim, D.-Y. (2006). Voluntary Controllability of the Implicit Association Test (IAT). *Social Psychology Quarterly*, *66*(1), 83–96. https://doi.org/10.2307/3090143

Kling, J. R., Liebman, J. B., & Katz, L. F. (2007). Experimental analysis of neighborhood effects. *Econometrica*, *75*(1), 83–119. https://doi.org/10.1111/j.1468-0262.2007.00733.x

Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: A literature review. *Quality and Quantity*, *47*(4), 2025–2047. https://doi.org/10.1007/s11135-011-9640-9

Kuklinski, J. H., Cobb, M. D., & Gilens, M. (1997). Racial Attitudes and the "New South." *The Journal of Politics*, *59*(2), 323–349.

List, J. A., & Price, M. K. (2016). The Use of Field Experiments in Environmental and Resource Economics. *Review of Environmental Economics and Policy*, *10*(2), 206–225. https://doi.org/10.1093/reep/rew008

Lyall, J., Blair, G., & Imai, K. (2013). Explaining Support for Combatants during Wartime: A Survey Experiment in Afghanistan. *American Political Science Review*, *107*(4), 679–705. https://doi.org/10.1017/S0003055413000403

Ostrom, E. (1999). *Self-Governance and Forest Resources*. Occasional Paper (No. 20). Jakarta: Center for International Forestry Research.

Rao, G. (2019). Familiarity Does Not Breed Contempt: Generosity, Discrimination, and Diversity in Delhi Schools. *American Economic Review*, *109*(3), 774–809. https://doi.org/10.1257/aer.20180044

Reitmann, A.-K. (2020). *Changing environmental conservation attitudes: Evidence from a framed field experiment among small-scale coffee farmers in Colombia*. Passauer Diskussionspapiere - Volkswirtschaftliche Reihe (No. V-82-29). Passau: Universität Passau.

Schaner, S. (2017). The Cost of Convenience? Transaction Costs, Bargaining Power, and

Savings Account Use in Kenya. *Journal of Human Resources*, *52*(4), 919–945. https://doi.org/10.3368/jhr.52.4.0815-7350R1

Tarozzi, A., Desai, J., & Johnson, K. (2015). The impacts of microcredit: Evidence from Ethiopia. *American Economic Journal: Applied Economics*, *7*(1), 54–89. https://doi.org/10.1257/app.20130475

Traoré, N., Landry, É., & Amara, N. (1998). On-Farm Adoption of Conservation Practices: The Role of Farm and Farmer Characteristics, Perceptions, and Health Hazards. *Land Economics*, *74*(1), 114–127. https://doi.org/10.2307/3147217

Viceisza, A. C. G. (2016). Creating a Lab in the Field: Economics Experiments for Policymaking. *Journal of Economic Surveys*, *30*(5), 835–854. https://doi.org/10.1111/joes.12118

Voors, M., Bulte, E., Kontoleon, A., List, J. A., & Turley, T. (2011). Using Artefactual Field Experiments to Learn about the Incentives for Sustainable Forest Use in Developing Economies. *American Economic Review: Papers & Proceedings*, *101*(3), 329–333. https://doi.org/10.1257/aer.101.3.329

Werner, P., & Riedl, A. (2019). The role of experiments for policy design. In A. Schram & A. Ule (Eds.), *Handbook of Research Methods and Applications in Experimental Economics* (pp. 295–311). Edward Elgar Publishing Ltd. https://doi.org/10.4337/9781788110563.00025

Whitmarsh, L. (2009). Behavioural responses to climate change: Asymmetry of intentions and impacts. *Journal of Environmental Psychology*, *29*(1), 13–23. https://doi.org/10.1016/j.jenvp.2008.05.003

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

(Online) Appendix

Appendix A: Indices of standardized outcomes

We use four farm-level indices of standardized outcomes in Section 4.2: one index for environmental conservation (EC) preferences and three indices for actual EC behavior. They are constructed in the spirit of Kling et al. (2007) and following Tarozzi et al. (2015). In these indices, a 'family' of outcomes is combined, namely EC preferences, domestic EC devices and practices, coffee processing EC devices and practices, and soil and water conservation practices (beyond coffee processing).

We illustrate the procedure using the EC preferences index. First, each of the seven statements (see Appendix B) is standardized by subtracting the mean and dividing by the standard deviation of the variable calculated for the control group at endline. Second, the index is calculated as the simple average of the seven standardized variables. All four indices are calculated in a similar way and constructed in a way that larger, positive values indicate higher EC preferences and behavior.

For the three behavioral indices, we included the following practices and devices. The survey asked for actual usage of the devices and not just the possession. Domestic environmental conservation: application of water savers, usage of domestic wastewater treatment system, separation of organic and inorganic waste. Environmental conservation during coffee processing: usage of dry hopper, hydraulic separator and ecological coffee washing device (e.g., tub tank), pulping and transporting pulp without water, usage of processing wastewater treatment system, usage of a (proper) pit. Soil and water conservation practices (beyond coffee processing): reforestation, no burnings, plan contouring, soil coverage, designation of protection areas, living fences and barriers, noble weeds, drainage channels, contouring lines and other conservation practices.

Appendix B: Stated preferences

The coffee farmers were asked about how much they identify themselves with seven statements on pro-environmental behavior before and after the implementation of the intervention.[19] The statements are the following:

(1) I reuse my water for several tasks.
(2) I am inspecting that none of the faucets, pipes and toilet are leaking.
(3) In my household, we save water.
(4) I conserve water even if my neighbors don't.
(5) Water conservation is not only a governmental obligation.

[19] The coffee farmers had to respond on a 5-point Likert scale: 1 – no identification, 2 – weak identification, 3 – indifferent, 4 – medium identification, 5 – complete identification. Due to the various meanings of 'level of identification', we introduce the term 'level of agreement' from this point on. That is, on which level do the coffee farmers agree that the behavior described in each statement applies to their own behavior and mindset.

(6) Coffee is also well washed if not a lot of water is used.
(7) Good practices in coffee processing do not only include those that ensure good coffee quality.[20]

The different training modules aim to transmit values and information to the participant, which should ultimately be reflected in high agreement with all of these statements.

Stated EC preference measures can potentially suffer from social desirability bias, namely respondents might provide answers that do not necessarily reflect their true opinion but rather align with what is viewed favorably by others (DeMaio, 1984). Although the survey team was independent of the institution implementing the intervention, respondents might perceive a pressure for 'environmental correctness' and hence overstate their pro-environmental preferences during the interview.

We summarize all seven statements into an index of standardized outcomes in the spirit of Kling, Liebman, and Katz (2007). The index is constructed as the simple average of the outcomes within the family of activities, standardized using the mean and the standard deviation of the outcome estimated from the control group at endline (see also Tarozzi et al., 2015). For more info, see Appendix A.

[20] Statements 5, 6 and 7 were originally asked the other way around. In order to have high values reflecting pro-environmental preferences, the statements were reversed for the analysis. While statements 1 and 2 indicate preferences towards domestic and productive water conservation behavior, the last two statements directly refer to post-harvest coffee processing. Statements 3, 4 and 5 indicate the farmer's own preferences towards water conservation and also in comparison to their neighbors and the government.