Implementer Demand in Program Evaluation

Ashish Shenoy[1*] and Travis J. Lybbert[1]

[1]University of California, Davis

[*]Corresponding author. Email: shenoy@ucdavis.edu.

Abstract

Development policies are commonly piloted by organizations with strong community ties. Reminding beneficiaries a pilot is being evaluated leads them to take costly actions favorable to the implementer. We elicit this behavior in an evaluation of an unsuccessful pilot agricultural extension program that ultimately drove treated farmers away from the target crop. Making the evaluation salient during endline data collection led participants to close the negative gap between treatment and control by altering their cultivation patterns. Participants' desire to support implementing organizations can help explain why promising pilots frequently fail to reproduce measured benefits at scale.

JEL Codes: O22, C90, L31

1 Introduction

Many development programs are initiated as small-scale pilots with the intent of expanding if proven successful. High-quality implementation at the pilot stage is necessary to ensure a program remains faithful to its intended design. As a result, pilots are commonly run by organizations with strong institutional capacity enabled by a history of local expertise and involvement. However, the intensive community engagement and oversight brought to bear at the pilot stage often cannot be replicated as a program expands in scope. There is growing concern that programs shown to have large impacts at small scale can lose effectiveness as expansion brings in new levels of management and administration.

In this paper, we demonstrate how the same features that constrain implementation quality at scale may also lead pilot evaluation to inaccurately overestimate program impacts. This study takes place in the context of an agricultural intervention to promote smallholder cultivation of pulses in Bihar, India. The policy was piloted in a two-year randomized evaluation by four non-governmental organizations (NGOs) selected for their extensive history of local rural development work in the study area. We find evidence that farmers involved in program evaluation take costly actions to make the evaluation result appear more favorable to the implementing organizations.

The primary data for this study come from a revealed-preference elicitation of demand for agricultural inputs at endline. The elicitation had real stakes, as participants purchased pulse seeds in accordance with their stated demand. This exercise was intended to measure farmers' sustained intention to produce pulses after program activities concluded, which was an explicit policy goal. Implementers defined success as an increase in treated farmers' preference for pulse cultivation, resulting in greater input demand relative to control.

To evaluate how implementers' desires affect participant behavior, we experimentally varied the salience of program evaluation during input demand elicitation. Specifically, enumerators introduced the elicitation either as an explicit evaluation of the implementer's efforts or more generally as a study of regional attitudes toward pulse cultivation. After this manipulation, demand elicitation proceeded identically for all participants. Importantly, we ensure the introductory language does not communicate information about product quality by offering a consistent product explicitly sourced and delivered by the local implementer. We interpret differences in participants' willingness to pay for pulse inputs by evaluation salience as a reflection of participants' preferences over the outcome of the evaluation itself.

Increasing the salience of evaluation skews the estimated treatment effect in favor of the implementers. Overall, the two-year intervention actually discouraged pulse cultivation among treated farmers by confirming their belief that growing pulses was not worth the opportunity cost of displacing more lucrative alternatives (see Lybbert et al., 2023, for further details). This belief manifested as 25% lower demand for pulse inputs on average in the incentive-compatible elicitation. However, the negative treatment effect was only observed in elicitations with low evaluation salience, where treated farmers purchased less than half

the input quantity of their control counterparts. By contrast, there was no distinguishable demand difference between treatment and control in elicitations with high evaluation salience. Making the evaluation salient during data collection obscured any evidence of a negative treatment effect.

This shift in input demand represents costly action taken by study participants. Treated farmers spent an average of Rs. 70 ($1.00) more on pulse seeds in high-salience elicitations. More importantly, seed purchases reflected real cultivation choices over the following crop season. We find a strong, positive correlation between seeds purchased and area sewn, with no systematic deviation by salience status. On average, we estimate farmers that were reminded of their participation in a program evaluation devoted 2% of their cropland to influencing the outcome of evaluation in favor of the implementer.

Responsiveness to implementer demand can be thought of as a form of Hawthorne effect in experimental economics. Past work has established that subjects in an experiment may alter their behavior when they know they are being monitored or evaluated (see, e.g., Levitt and List, 2011; Friedman and Gokul, 2014; McCambridge et al., 2014). Most relevant to our study, de Quidt et al. (2018) investigate experimenter demand, whereby participants act in response to their beliefs about researcher objectives, as a possible source of bias. The authors intentionally manipulate experimenter demand and find the resulting distortions to be modest. We extend this type of work to introduce the possibility that the relevant pressure in program evaluation comes not from the experimenter conducting the evaluation, but rather the implementer being evaluated.

Implementer demand bias is particularly concerning for research tools designed exclusively to generate evaluation data. In particular, Becker-DeGroot-Marschack (BDM) elicitation is common in field experiments—including our own study—because it reveals respondents' full demand curve in an incentive-compatible manner (Lusk and Shogren, 2007). This mechanism has come under recent criticism due to concerns about decision fatigue (Brown et al., 2023), misunderstanding of the dominant strategy (Cason and Plott, 2014; Berry et al., 2020), and price anchoring (Berry et al., 2020; Mamadehussene and Sguera, 2023). Nevertheless, it empirically performs well relative to other measures of willingness-to-pay (Berry et al., 2020; Burchardi et al., 2021; Brown et al., 2023). We show that even if BDM elicitation accurately reveals demand, the demand itself may be influenced by participants' preferences over the outcome of evaluation. Such considerations are more likely to arise in activities introduced for the purposes of evaluation—such as BDM elicitation—than in study participants' regular endeavors—such as purchasing inputs at the market.

Our findings more generally contribute to the large literature on external validity in program evaluation (see Banerjee et al., 2017; Al-Ubaydli et al., 2019). Across a range of sectors, programs implemented NGOs have systematically generated greater impact than those run by governments (Vivalt, 2020). Microeconomic evidence highlights community engagement (Usmani et al., 2022) and responsiveness (Björkman Nyqvist et al., 2019) as important contributors to NGO effectiveness, and these features are difficult to replicate at scale (e.g. Dhaliwal and Hanna, 2017; Mitchell et al., 2023). Our study establishes how these factors can undermine accuracy of evaluation independent of their role in implementation quality, presenting a further

threat to external validity.

Implementer demand bias in our setting is closely related to the issue of endogenous participant effort described by Chassang et al. (2012). Many development programs rely on complementary investments from program beneficiaries, and beliefs about implementation quality can alter participants' incentives to invest. For instance, in two field evaluations, Bulte et al. (2014) and Bulte et al. (2023) show how the returns to upgraded seeds in Tanzania depend crucially on farm labor choices. We find that participant behavior responds not only to beliefs about implementation quality, but also preferences over evaluation outcomes. This type of responsiveness to implementer demand may also exacerbate concerns of favorable selection of sites (Allcott, 2015) or indicators (Saccardo et al., 2023) in pilot evaluation, though we find no evidence of such interactions in our setting.

This paper proceeds as follows. In Section 2 we describe the context in which this study takes place. Section 3 discusses the data and methodology, and Section 4 presents results. Finally, we discuss implications in Section 5.

2 Background

This research takes place in the Indian state of Bihar. The state is among the poorest in the country, where over a third of households fall below the national poverty line. The population is also predominantly agrarian with just 12 percent residing in urban centers. As a result, Bihar has been a region of focus for rural development programs by both the Government of India and the NGO sector. There are currently 4,255 registered NGOs and volunteer organizations in the state,[1] the majority of which operate at small scale and rely on heavy engagement with beneficiary communities. We investigate how to translate experience from this type of localized development work into guidance for policy design.

Our study is tied to a joint development initiative between the Government of India and local organizations in Bihar. In 2016, the Government of India partnered with four local NGOs to pilot a policy to increase the production of pulses by farm households. Many households in this region grow small amounts of pulses—primarily pigeon pea—on crop borders or other marginal land for home consumption. The partnership designed and administered a two-year package of input subsidies, agricultural extension services, and marketing support to modernize cropping practices and boost output. This package was piloted in five districts to test whether intensive short-term investment could shift the long-term crop portfolio of participating households.

Implementing partners for this project were selected because of their track record with local development. All four organizations had operated in their respective areas for at least ten years prior, and had been involved in past initiatives ranging from agriculture to health and nutrition to savings and credit. As a result, local implementers had preexisting relationships with study participants before the inception of the

[1]Source: NITI Aayog (`https://ngodarpan.gov.in/index.php/home/statewise`) accessed 10/4/2021

pulses program.

From summer 2017 through spring 2019, implementers procured and distributed modern-variety pulse seeds at subsidized prices, conducted local training and extension sessions to demonstrate best practices, give individualized feedback to program participants through the cropping season, and assisted with the sale of output. The NGO seed distribution network, which brought in higher quality inputs than were previously available in local markets, remained in place after other activities concluded.

Program activities were carried out in a randomized controlled trial, with the intent of using lessons from the pilot to design a statewide policy that could be adopted and run by government agencies. Importantly, both treated and control farmers knew they were part of a pilot evaluation that may scale up if successful. Main program evaluation results are reported by Lybbert et al. (2023). In this paper we quantify how preexisting relationships may affect impact evaluation in pilot experimentation.

3 Data and Methodology

3.1 Data

Data for this study come from laboratory-style elicitations of willingness-to-pay for pulse seeds among farmers participating in the pilot field experiment. A key evaluation outcome for the trial was sustained production of pulses after main program activities had concluded. As part of this evaluation, demand for certified seeds was elicited ahead of the planting season among a sample of study farmers. Elicitations measured demand for pigeon pea (*arhar*) and black gram (*urad*) in the summer (*Kharif*) season and for red lentil (*masoor*) in the winter (*Rabi*) season of 2019. All three crops were initially promoted by the pulses program, but second-year implementation predominantly focused on the latter two.

Each demand elicitation included farmers from the same village who placed orders for certified seed from the NGO supplier through a BDM revelation mechanism. For each variety, participants were given list of possible prices and asked to reveal their quantity demanded at each price.[2] At the end of the session, one price at random was selected from the list, and participants purchased their stated demand at that price. The purchase transaction ensures that each demand decision potentially had real financial stakes.

Participants did not have the option to adjust their quantity demanded after the transaction price was revealed to ensure incentive compatibility during the elicitation. However, there was a delay of several days between demand elicitation and delivery of seeds. At the time of delivery, participants were required to purchase their full stated quantity, but they had the option to purchase a different amount at the supplier price (foregoing the experimental discount) or to purchase no seeds at all. Field reports indicate there were only a few isolated instances where demand fell to zero and no known instances of farmers purchasing a

[2]The maximum price in each elicitation was the prevailing supplier price. Above this level, demand would rationally have been zero because participants always have the option to purchase seeds directly from the supplier. Hence, the elicitation reflects demand when purchasing seeds at a discount relative to their outside option, and we cannot observe hypothetical demand at higher prices.

different amount at the supplier price. Therefore, we interpret experimental responses as farmers' intended demand at the time of the elicitation.

In practice, demand elicitation offered a coupon for certified seeds from the NGO supplier. This detail is important because seeds sourced by the NGO were higher quality than those available in local markets. We offer coupons to ensure that experimental manipulations do not influence participants' beliefs about the source of seeds. Farmers may still have different valuations for the product on offer—this was indeed an intended objective of the two-year period of subsidized learning-by-doing—but coupons enforce uniformity in the product that study participants express demand for.

After the elicitation exercise, participants were asked a short set of questions related to demographics and intended pulse cultivation. Participants in the winter season are also asked about their involvement with the pulses pilot program, their subjective perception of whether the pulse program had been beneficial, and their past involvement with the local implementer's other activities. The exact wording of these questions is provided in Appendix A.

3.2 Study Design

This study leverages two sources of experimental variation. The primary variation comes from the script introducing the demand elicitation, where we experimentally vary the salience of the evaluation and implementing partner. In half the sessions, designated high-salience, we explicitly identified our objective to "evaluate how effective [partner organization]'s efforts to promote pulses are" over the prior two years. In the other half, designated low-salience, we motivated our involvement more generally as a study to "understand more about pulse production in this region."[3] Other than this discrepancy, the demand elicitation proceeded identically for all participants. Salience treatment assignment remained constant within village over the two planting seasons.

[Table 1 about here.]

Table 1 presents descriptive statistics on participants across salience arms. The first four rows describe the participants themselves, and the next three rows provide household demographic information. Only the fraction male differs significantly at the 10% level, and a joint F-test fails to reject balanced at the 10% level.

Evaluation salience cross-cuts the second source of variation inherited from treatment assignment in the pilot evaluation. This study includes farmers from 94 experimentally treated villages that received two years of input subsidies and extension support as well as from 53 experimental control villages that did not. We also elicit demand in 66 non-experimentally selected villages that were treated by the implementing partners alongside pilot evaluation. In Appendix A we discuss this selection process and show there is little

[3]The full introductory text is provided in Appendix A. This manipulation and evaluation were pre-registered as a separate trial with the AEA RCT Registry: AEARCTR-0004405.

difference between those experimentally and non-experimentally selected for treatment, so we pool both to increase power.

Demand elicitations had a fairly high attrition rate with only two thirds of invited households opting to participate. Participation is uncorrelated with salience treatment assignment, and the manipulation only took place after recruitment. Therefore, attrition does not introduce bias into comparisons across salience arms. However, the choice to participate may have been affected by the prior two years of program participation and therefore may introduce selection bias into comparisons across experimental treatment arms. As a result, our findings in this paper can be interpreted as the causal effect of evaluation salience on the demand of those who participate in the elicitation, but selection into the data may differ between treated and control groups in the pilot evaluation.[4]

We evaluate the impact of evaluation salience on demand for pulse seeds using the regression

$$Q_{icp} = \beta_1 Treat_i + \beta_2 Salient_i + \beta_3 Salient_i \times Treat_i + \alpha_c + \gamma_p + \delta_{b(i)} + X_i'\sigma + \epsilon_{icp} \quad (1)$$

where Q_icp denotes the quantity demanded by individual i for seeds of crop c at price p. α_c and γ_p control for crop-specific and price-level demand shifters, respectively. $\delta_{b(i)}$ controls for block-level (sub-district) demand shifters, and $X_i'\sigma$ controls for participant demographic characteristics.

Coefficients β_1 and β_2 in (1) describe how demand differs on average in villages treated in the pilot experiment and those exposed to the high-salience script, respectively. The main coefficient of interest, β_3, reveals how the salience effect differs between treated and untreated villages. A finding of $\beta_3 \neq 0$ would indicate that the estimated treatment effect in the pilot experiment would differ based on whether the fact of evaluation were made salient or not during data collection.

4 Results

Table 2 reports differences in input quantity demanded by treatment assignment in the pilot experiment and by evaluation salience in the demand elicitation. All standard errors are clustered at the village level, and we report p-values from randomization inference over 1,000 iterations reassigning village-level salience status. In Appendix B we discuss heterogeneity by crop and by implementing partner.

[Table 2 about here.]

The first row of Column 1 reproduces the program evaluation result that promotion activities actually lowered input demand among treated farmers. This result arises because two years of experience confirmed treated farmers' prior belief that pulses are no more profitable than the crops they would displace, even with modern cropping practices and high external investment. Relative to the mean, treatment lowered

[4] Attrition is discussed further in Appendix A.

pulse demand by nearly 30%.[5]

4.1 Evaluation Salience and Input Demand

The main finding of this paper is that the effect of evaluation salience varies with treatment assignment in the pilot experiment. The third row of Column 1 reveals making the impact evaluation salient lowers input demand on average. Column 2 breaks this effect up by treatment assignment according to the regression specification in (1). The second row reports a stark difference in response between those in treated and control villages.

The positive sign of the coefficient on the interaction term (β_3) indicates treated farmers signal greater input demand when evaluation is salient relative to their control counterparts. This behavior would support implementing partners' desire to demonstrate the success of their intervention. The magnitude of the coefficient—85% as large as the negative effect in the first row—suggests evaluation salience distorts results by enough to nearly erase the decline in demand caused by treatment. A joint test fails to reject that the estimated treatment effect is statistically distinguishable from zero among high-salience participants at the 10% level.

Disappearance of the negative treatment effect is apparent in (inverse) input demand curves, plotted by crop and by experiment status in Figure 1. The figure shows a gap between treated and control farmers who participate in elicitations with low evaluation salience, but this gap closes in elicitations with high salience. The effect is most apparent with black gram and red lentils, the two crops that were the primary focus of second-year program activities. By contrast, for pigeon peas, which farmers were most likely to have already been growing pre-intervention, salience lowers demand overall but has no differential effect between treatment and control.

[Figure 1 about here.]

These results are consistent with farmers adjusting their behavior to satisfy implementers' desire to demonstrate effectiveness when reminded of the program evaluation. Mentioning the evaluation seems to lower demand for inputs as participants are reminded of the unsuccessful intervention over the prior two years, but treated farmers resist this urge in order to deliver a positive evaluation. Alternately, the salience effect may be caused by those in the control group suppressing their underlying desire to experiment with high-quality seeds in order to demonstrate a need for NGO activities. In either case, the behavior change among study participants leads evaluation results to look more favorable to the implementer.

This behavior change reflects a real shift in the agricultural portfolio of evaluation participants. The estimated difference of a half kilogram in seed purchases corresponds to expenditures of around Rs. 70, or

[5] Lybbert et al. (2023) show the demand elicitation result to be consistent with other post-intervention indicators that treated farmers ceased pulse cultivation once subsidies expired, and not attributable to differential attendance at demand elicitations or to buildup of stocks during the intervention period. However, a causal interpretation of the treatment effect is not necessary for the discussion regarding evaluation salience in this study.

$1.00, at market prices. While this expenditure itself may be small, program implementers were active in ensuring farmers planted and cultivated what they purchased. In Appendix A we show a strong positive correlation between seed purchases and area cultivated, with an R-squared of 0.6–0.8, and verify deviations from this relationship are not systematically related to salience status. Records indicate the salience effect led farmers to reallocate around 0.03 acres—or 2% of their total landholdings—to supporting a positive program evaluation.

4.2 Engagement with Implementing NGO

Prior engagement with the local implementer also influences seed demand, but this channel neither explains nor dilutes the relationship between treatment status and evaluation salience. After the second demand elicitation, farmers were asked about their participation in the pulses program, their beliefs about how beneficial the program had been, and their engagement with other NGO development activities in the past. Columns (3) and (4) of Table 2 explore how these variables affect experimental results.

It is worth noting that self-reported beneficiary status may be an endogenous outcome of evaluation salience. Table 3 reports how the three variables relating to program benefits differ with experimental assignment. The first two columns show that while those treated in the pilot experiment are more likely to self-report participation the pulses program, a majority claim to have participated even in control villages. This is likely because all study farmers were initially recruited for a pulses-related evaluation prior to treatment assignment, so those even those who did not directly receive the intervention package may have considered themselves to be part of the pulses program. The final two columns show that those who received the pulses intervention are also more likely to self-report having participated in other NGO activities in the past despite random assignment of treatment. Interestingly, evaluation salience seems to prime participants to recall any NGO engagement, raising the likelihood of self-reporting participation in both the pulses program and in past NGO activities.

[Table 3 about here.]

The middle two columns of Table 3 reveal three quarters of participants claim to believe the pulses intervention was beneficial. There is little variation in this rate with either treatment assignment or evaluation salience. A favorable response to this question is cheap talk and likely less indicative of true beliefs than revealed-preference input demand decisions.

Evaluation salience differentially affects the input demand choices of those who self-report prior NGO engagement. As Column 4 of Table 2 demonstrates, pulse program participants have sightly lower input demand on average, but their demand substantially increases when the evaluation is made salient. The inverse is true of prior engagement with other NGO actives: input demand is slightly greater among past beneficiaries, but evaluation salience substantially lowers it. This latter pattern is consistent with a model in which mentioning the NGO at the start of the evaluation leads participants to anchor expectations around

the organization's typical strategy of providing benefits that are heavily subsidized or free,[6] though other explanations are also possible. In any case, the coefficient estimates on treatment assignment, salience, and their interaction remain equally strong after controlling for self-reported measures of NGO engagement.

5 Discussion

When evaluating a program run by a sympathetic implementer, the salience of the evaluation itself can influence the estimated treatment effect. We establish this fact in the context of a pilot agricultural policy that was negatively received but run by organizations with strong community ties. When the evaluation was made salient, participants altered their behavior in a real-stakes demand elicitation with binding input decisions for the coming crop season. While increasing the salience of evaluation heightened negative responses to the pilot on average, we observe substantial heterogeneity between evaluation arms. Making the evaluation salient effectively closed the gap between treatment and control, masking evidence of the negative impact.

This outcome is consistent with participants adjusting their behavior to satisfy the desired outcome of the implementer. In general, beneficiaries of development programs can reciprocate at the evaluation stage. Motivation for this form of reciprocity may be either backward-looking or forward-looking. That is, participants may seek to reward the implementer's past efforts with a positive evaluation, or they may anticipate that a positive evaluation will increase the likelihood of receiving continued benefits in the future. The exact motivation reciprocity and responsiveness to implementer demand remain open topics for future research.

The behavior described in this study has implications for translating the experience from pilot programs into broader policy lessons. When introducing new development initiatives, it is common practice to partner with established organizations that have strong community ties to leverage their local knowledge and institutional capacity. However, our findings suggest the features that enable successful implementation are precisely those that can undermine accurate evaluation. When faced with the prospect of evaluation, beneficiaries sympathetic to the implementer may take costly actions to help the implementer achieve its goal of demonstrating success. As a consequence, pilot evaluation can uncover misleadingly optimistic results that overstate a policy's true benefits and therefore do not replicate at scale.

[6] In Appendix B we explore potential treatment effects on demand elasticity.

References

Al-Ubaydli, Omar, John A List, and Dana Suskind, "The Science of Using Science: Towards an Understanding of the Threats to Scaling Experiments," Working Paper 25848, National Bureau of Economic Research May 2019.

Allcott, Hunt, "Site Selection Bias in Program Evaluation," *The Quarterly Journal of Economics*, 2015, *130* (3), 1117–1165.

Banerjee, A.V., S. Chassang, and E. Snowberg, "Decision Theoretic Approaches to Experiment Design and External Validity," in Abhijit Vinayak Banerjee and Esther Duflo, eds., *Handbook of Field Experiments*, Vol. 1 of *Handbook of Economic Field Experiments*, North-Holland, 2017, pp. 141–174.

Berry, James, Greg Fischer, and Raymond Guiteras, "Eliciting and Utilizing Willingness to Pay: Evidence from Field Trials in Northern Ghana," *Journal of Political Economy*, 2020, *128* (4), 1436–73.

Brown, Alexander, Jinliang Liu, and Michael Tsoi, "Is There a Better Way to Elicit Valuations than the BDM?," Working Paper 4476764, Social Science Research Network 2023.

Bulte, Erwin, Gonne Beekman, Salvatore Di Falco, Joseph Hella, and Pan Lei, "Behavioral Responses and the Impact of New Agricultural Technologies: Evidence from a Double-blind Field Experiment in Tanzania," *American Journal of Agricultural Economics*, 2014, *96* (3), 813–830.

_ , Salvatore Di Falco, Menale Kassie, and Xavier Vollenweider, "Low-Quality Seeds, Labor Supply and Economic Returns: Experimental Evidence from Tanzania," *The Review of Economics and Statistics*, 2023, *forthcoming*.

Burchardi, Konrad B., Jonathan de Quidt, Selim Gulesci, Benedetta Lerva, and Stefano Tripodi, "Testing willingness to pay elicitation mechanisms in the field: Evidence from Uganda," *Journal of Development Economics*, 2021, *152* (Sep), 102701.

Cason, Timothy N. and Charles R. Plott, "Misconceptions and Game Form Recognition: Challenges to Theories of Revealed Preference and Framing," *Journal of Political Economy*, 2014, *122* (6), 1235–70.

Chassang, Sylvain, Gerard Padró I Miquel, and Erik Snowberg, "Selective Trials: A Principal-Agent Approach to Randomized Controlled Experiments," *American Economic Review*, 2012, *102* (4), 1279–1309.

de Quidt, Jonathan, Johannes Haushofer, and Christopher Roth, "Measuring and Bounding Experimenter Demand," *American Economic Review*, November 2018, *108* (11), 3266–3302.

Dhaliwal, Iqbal and Rema Hanna, "The devil is in the details: The successes and limitations of bureaucratic reform in India," *Journal of Development Economics*, 2017, *124*, 1–21.

Friedman, Jed and Brinda Gokul, "Quantifying the Hawthorne Effect," Technical Report October 16, World Bank Blogs 2014.

Levitt, Steven D. and John A. List, "Was There Really a Hawthorne Effect at the Hawthorne Plant? An Analysis of the Original Illumination Experiments," *American Economic Journal: Applied Economics*, 2011, *3* (1), 224–38.

Lusk, Jayson L. and Jason F. Shogren, *Experimental Auctions: Methods and Applications in Economic and Marketing Research*, Cambridge, UK: Cambridge University Press, 2007.

Lybbert, Travis, Ashish Shenoy, Tomoe Bourdier, and Caitlin Kieran, "Striving to Revive Pulses in India with Extension, Input Subsidies, and Output Price Supports," *American Journal of Agricultural Economics*, 2023, *forthcoming*.

Mamadehussene, Samir and Francesco Sguera, "On the Reliability of the BDM Mechanism," *Management Science*, 2023, *69* (2), 1166—-79.

McCambridge, Jim, John Witton, and Diana R. Elbourne, "Hawthorne effect: new concepts are needed to study research participation effects," *Journal of Clinical Epidemiology*, 2014, *67* (3), 267–77.

Mitchell, Harrison, A. Mushfiq Mobarak, Karim Naguib, Maira Rei ao, and Ashish Shenoy, "Delegation Risk and Implementation at Scale: Evidence from a Migration Loan Program in Bangladesh," Unpublished manuscript 2023.

Nyqvist, Martina Björkman, Andrea Guariso, Jakob Svensson, and David Yanagizawa-Drott, "Reducing Child Mortality in the Last Mile: Experimental Evidence on Community Health Promoters in Uganda," *American Economic Journal: Applied Economics*, 2019, *11* (3), 155–92.

Saccardo, Silvia, Hengchen Dai, Maria Han, Naveen Raja, Sitaram Vangala, and Daniel Croymans, "Assessing Nudge Scalability," Working Paper 3971192, Social Science Research Network 2023.

Usmani, Faraz, Marc Jeuland, and Subhrendu K. Pattanayak, "NGOs and the Effectiveness of Interventions," *Review of Economics and Statistics*, 2022, *forthcoming*.

Vivalt, Eva, "How Much Can We Generalize From Impact Evaluations?," *Journal of the European Economic Association*, 2020, *18* (6), 3045–3089.

Table 1: Participant Characteristics by Treatment Assignment

	Evaluation Salience		
	Low	High	Difference
Male	0.87	0.81	−0.06*
	(0.34)	(0.39)	
Age	46.79	47.73	0.94
	(16.53)	(16.70)	
Primary School	0.66	0.65	−0.01
	(0.47)	(0.48)	
Secondary School	0.49	0.49	0.00
	(0.50)	(0.50)	
HH Size	7.40	7.53	0.13
	(3.76)	(4.04)	
SC/ST	0.14	0.11	−0.03
	(0.35)	(0.31)	
Acres Owned	1.74	1.66	−0.08
	(1.96)	(1.80)	
Joint Significance			0.18
Participation Rate	0.67	0.65	−0.02
Participants	420	476	
Villages	100	113	

Group averages with standard deviations in parentheses. Rows correspond to fraction male, participant age, primary school completion, secondary school completion, household size, fraction belonging to a schedule caste or scheduled tribe, and land area owned by household at pilot baseline. Participation rate reflects fraction of those invited who appeared at either demand elicitation.

Table 2: Effect of Evaluation Salience on Seed Quantity Demanded

	(1)	(2)	(3)	(4)
Treated	-0.289	-0.534	-0.229	-0.500
	(0.11)	(0.18)	(0.12)	(0.20)
Salient × Treated		0.448		0.463
		(0.21)		(0.23)
		[0.04]		[0.08]
Salient	-0.171	-0.493	-0.186	-0.563
	(0.08)	(0.19)	(0.10)	(0.23)
	[0.05]	[0.01]	[0.09]	[0.04]
Salient × Pulse Program				0.670
				(0.20)
				[0.01]
Pulse Program Participant			0.085	-0.248
			(0.13)	(0.18)
Salient × Prior NGO				-0.529
				(0.19)
				[0.03]
Prior NGO Beneficiary			0.067	0.339
			(0.12)	(0.17)
Mean Demand	1.09	1.09	1.19	1.19
Salient Treat Effect (p-val.)		0.46		0.74
R-Squared	0.16	0.16	0.17	0.18
Observations	9390	9390	6305	6305

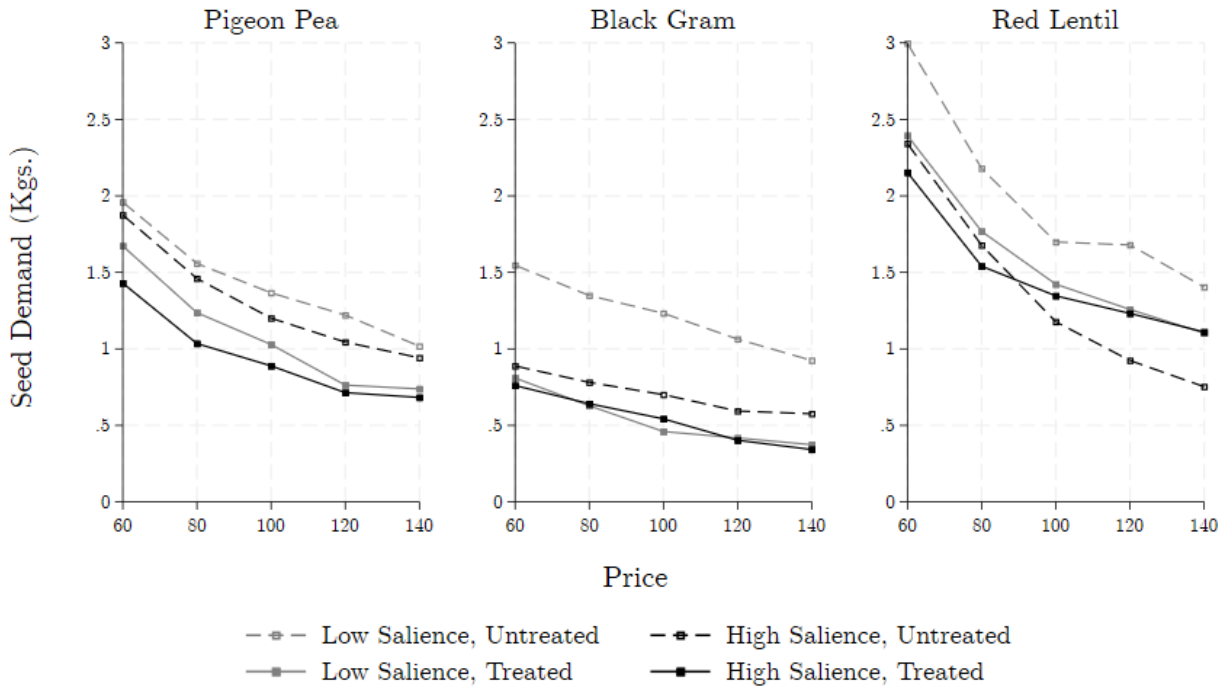
Notes: Outcome is pulse seed quantity demanded. Treated: Village received pulse program treatment in prior two years. Salient: High-salience script in elicitation. Pulse Program: Participant self-identifies as beneficiary of pulse program. Prior NGO: Participant self-identifies as beneficiary of previous NGO programs. Salient Treat Effect: p-value from test of sum of coefficients on Treated and Salient × Treated. All regressions include crop and price fixed effects, block (sub-district) fixed effects, and participant demographic controls. Columns (3) and (4) restrict to those that self-reported program participation during the red lentil (winter) elicitation. Standard errors clustered by village in parentheses; p-values from randomization inference over 1,000 re-draws of village-level salience treatment status in square brackets.

Table 3: Effect of Evaluation Salience on Self-Reported NGO Benefits

	Pulse Program				Other NGO	
	Participation		Beneficial		Participation	
	(1)	(2)	(3)	(4)	(5)	(6)
Treated	0.124	0.123	0.036	0.091	0.154	0.272
	(0.06)	(0.09)	(0.05)	(0.08)	(0.07)	(0.10)
Salient × Treated		-0.005		-0.106		-0.231
		(0.12)		(0.11)		(0.13)
		[0.97]		[0.39]		[0.13]
Salient		0.093		0.078		0.225
		(0.10)		(0.09)		(0.11)
		[0.42]		[0.44]		[0.08]
Variable Mean	0.59	0.59	0.73	0.73	0.66	0.66
Salient Treat Effect (p-val.)		0.15		0.84		0.62
R-Squared	0.41	0.42	0.06	0.06	0.17	0.19
Observations	563	563	563	563	563	563

Notes: Outcomes are self-reported status in NGO programs. Columns (1) and (2) report participation in pulses program; (3) and (4) report stated belief that pulse program was beneficial; (5) and (6) report participation in prior NGO initiatives. Treated: Village received pulse program treatment in prior two years. Salient: High-salience script in elicitation. Salient Treat Effect: p-value from test of sum of coefficients on Treated and Salient × Treated. All regressions include block (sub-district) fixed effects and participant demographic controls. Standard errors clustered by village in parentheses; p-values from randomization inference over 1,000 re-draws of village-level salience treatment status in square brackets.

Figure 1: Inverse Demand Curves by Crop and by Experimental Status



Notes: Average seed demand among farmers by pilot treatment status and evaluation salience at each price level for each crop.

Supplementary Appendix for "Implementer Demand in Program Evaluation" For Online Publication Only

A Experiment Details

A.1 Village Selection

This study takes place in villages that participated in a pilot evaluation of an initiative to promote pulse production. The evaluation initially comprised 158 villages, of which 99 were experimentally assigned to receive the intervention package and the other 59 were assigned to control. The input demand elicitation took place after two years of intervention. At the time of elicitation, we included farmers from an additional 70 villages selected non-experimentally for treatment by the implementing partners. Farmers from 94 of the 99 experimentally treated villages, 53 of the 59 control villages, and 66 of the 70 non-experimentally treated villages consented to participate in demand elicitations.

We observe little difference in both input demand and responsiveness to evaluation salience between experimentally and non-experimentally treated farmers. Table S1 reports differences in quantity demanded and self-reported measures of NGO engagement between these groups. Experimentally treated farmers are more likely to report believing the pulse program to be beneficial, but this difference disappears when the evaluation is made salient. Other than this discrepancy, differences between farmers in experimentally and non-experimentally selected villages are quantitatively small relative to the variable mean and statistically indistinguishable from zero.

[Table S1 about here.]

This finding is surprising in light of evidence on favorable site selection for pilot evaluations (e.g. Allcott, 2015). We hypothesize two contributing factors: first, the intervention took place in a geographically compact area. Activities were administered at the block (sub-district) level by organizations with strong local ties and long histories of local involvement. As a result, nearly all experimental and non-experimental villages had NGO involvement in the past and there may not have been much scope for favorable site selection. Second, non-experimentally treated villages were enrolled throughout the two-year intervention period. These include villages reserved for treatment before experimental randomization—which we would expect to be the most favorable sites—as well as those enrolled after experimental activities were underway—which we would expect to be less favorable than those identified for randomization. Unfortunately the administrative record-keeping lacks sufficient detail to explore heterogeneity in the timing of non-experimental village enrollment.

A.2 Participant Recruitment

Study participants from the pilot experiment—both treated and control—were selected based on interest in pulses prior to randomization, two years before the input demand elicitation. Before the pilot intervention began, implementing partners held kickoff meetings in each experimental village to advertise the program and identify interested farmers. Seven–eight attendees from each kickoff meeting were selected at random for surveying during the intervention period,

and they constitute the set of invited participants in seed demand elicitation from experimental villages. In non-experimentally treated villages, we attempt to recreate this selection procedure as closely as possible by sampling at random out of the initial set of farmers engaged by the NGO when they first began program activities in the village.

In practice, only two thirds of invited farmers actually participated in either demand elicitation, with little difference between control, experimental treatment, and non-experimental treatment. Among those invited, 50% participated in the summer (Kharif) session and 42% in the subsequent winter (Rabi) session. Roughly a third of invitees for the summer session and half in the winter session declined due to lack of interest or other local engagements. The rest either could not be reached or were unavailable on the day of the elicitation for reasons such as travel outside the village. Those who were invited and participated constitute the sample used in this study.

[Table S2 about here.]

[Table S3 about here.]

The top panels of Tables S2 and S3 report the baseline characteristics of invited households by participation status in each demand elicitations. Participants and non-participants are nearly identical on baseline household characteristics, though the obviously differ in their desire to purchase pulse seeds. Participation was determined before reading the evaluation salience script, so attrition does not affect the causal interpretation of the effect of evaluation salience. However, attrition should be taken into account when drawing conclusions about who responds to evaluation salience in their input demand.

To reach the target of 8–10 participants in each demand elicitation session, we requested village leaders to invite additional farmers that were interested in purchasing pulse seeds. The bottom panels of Tables S2 and S3 compare this supplemental sample to the main sample on the limited set of characteristics we collected data on during demand elicitation. Supplemental participants are typically older and more male than invited participants, and the difference is statistically significant at the 1% level.

More worryingly, field reports indicate many of these supplemental participants came from neighboring villages so we cannot accurately determine their treatment assignment or even verify their participation in the pilot experiment. This fact is also reflected in their substantially lower rate of self-identifying as a participant in the pulses program, shown in Table S3. As a result, we exclude them from the main analysis.

[Table S4 about here.]

Table S4 presents regression results from this supplemental sample as the counterpart to Table 2. Estimated effects are in the same direction as in the main sample, but not as strong. This result is consistent with variable participation in the pilot experiment, which would both weaken the effect of evaluation salience and introduce attenuation bias into regressors based on pilot treatment status.

A.3 Salience Script and Demand Elicitation

Enumerators introduced themselves before the demand elicitation exercise using one of two scripts. In the high-salience version, enumerators read the following paragraph:

> "Your participation in this auction and the survey is part of an evaluation project. We are here because of our partnership with [PARTNER ORGANIZATION] to help evaluate how effective their efforts to promote pulses are. To do so, we want to understand how beneficial you think pulse production would be to you as a farmer after their work in this region."

In the low-salience version, enumerators introduced themselves with the following paragraph:

> "Your participation in this auction and the survey is part of a research project. As such, we are here as a research team, not a sales team. We are not here to promote any kind of pulses. We simply want to understand how beneficial you think pulse production would be to you as a farmer and understand more about pulse production in this region."

If participants asked specifically about the partner NGO or the pulses program during low-salience sessions, enumerators were instructed to provide the following response:

> "We are aware of their activities, but this exercise is designed to learn about attitudes to pulses in this region overall. We are visiting several villages in this area, including many where [PARTNER ORGANIZATION] is not operating."

Other than this difference, demand elicitation proceeded identically in all villages.

In each elicitation, participants were given a list of possible prices for certified varieties of pigeon pea and black gram in the summer (Kharif) session, and of red lentil in the winter (Rabi) session. They reported quantity demanded at each possible price, and then one price was selected at random for actual purchase. To ensure incentive-compatibility, participants could not adjust their quantity demanded after the price was announced.

Prior to demand elicitation, participants played two practice rounds to build familiarity with the mechanism. In the first practice round, they were given a participation fee and could opt to purchase sweets from the enumerators. In the second practice round, they stated hypothetical seed demand, drew a hypothetical transaction price, and were told the quantity of seed they would have received were it the real elicitation. In field testing, we found this second practice round was necessary to ensure participants understood they would not be able to adjust their demand after observing the real transaction price.

A.4 Seed Purchases and Planting

Demand elicitation participants ultimately received a coupon to purchase exactly their stated amount at the randomly selected elicitation price, which was typically below the price set by the NGO supplier. Coupon amounts were also shared directly with the supplier so that farmers could receive their target quantity even if they lost their coupon. Suppliers did not keep detailed records of coupon redemption, but field reports suggest the vast majority of coupons were redeemed for the stated amount.

After delivering seeds, implementing NGOs were active in ensuring farmers planted what they purchased. Seed quantity and area cultivated were documented in administrative records for all farmers from villages that received the pulses intervention. Unfortunately these records cannot be linked to participants in the demand elicitation, and no comparable records exist for farmers from control villages in the pilot evaluation.

[Figure S1 about here.]

[Table S5 about here.]

Figure S1 reveals a strong, positive relationship between these two variables for each crop in the demand elicitation. In Table S5 we confirm seed purchase quantity strongly predicts acreage, with a regression R-squared of between 0.6 and 0.8. For black gram and red lentil—the two crops where evaluation salience has the greatest impact—one kilogram of seeds purchased corresponds to roughly 0.07 acres cultivated.

Importantly, there is no systematic pattern in deviations from this relationship by evaluation salience. Figure S1 shows that the acreage of farmers in both high-salience and low-salience villages are evenly distributed around the trend line, and Table S5 confirms that after controlling for seed quantity, the effect of evaluation salience on acreage is quantitatively small and statistically indistinguishable from zero. Given this evidence, we conclude that differences in elicited seed demand correspond to real differences in area planted, and are not merely performative purchases made at low cost and subsequently discarded.

Table S1: Outcomes for Experimentally and Non-Experimentally Treated Farmers

	Input	Pulse Program		Other NGO
	Demand	Participation	Beneficial	Participation
	(1)	(2)	(3)	(4)
Experimental	-0.041	-0.073	0.222	-0.001
	(0.15)	(0.08)	(0.09)	(0.09)
Salient × Experimental	0.172	0.097	-0.222	-0.077
	(0.18)	(0.12)	(0.13)	(0.13)
Salient	-0.168	0.007	0.125	0.041
	(0.13)	(0.09)	(0.10)	(0.10)
Variable Mean	1.00	0.64	0.73	0.71
R-Squared	0.17	0.42	0.10	0.16
Observations	6690	409	409	409

Notes: Regressions restrict to study farmers in treated villages. Outcomes are input quantity demanded (Column 1) and self-reported status in NGO programs (Columns 2–4). Experimental: Village selected for treatment through experimental randomization. Salient: High-salience script in elicitation. All regressions include block (sub-district) fixed effects and participant demographic controls, and Column 1 includes crop and price-level fixed effects. Standard errors clustered by village in parentheses.

Table S2: Participation in the Summer (Kharif) Demand Elicitation

Household Characteristics	Main Sample	Declined	Unavailable	
Farmer Male	0.86	0.84	0.85	
	(0.35)	(0.37)	(0.36)	
Farmer Age	48.92	48.20	48.73	
	(16.34)	(15.68)	(15.24)	
Farmer Primary Sch.	0.63	0.59	0.66	
	(0.48)	(0.49)	(0.47)	
Farmer Secondary Sch.	0.45	0.44	0.51	
	(0.50)	(0.50)	(0.50)	
HH Size	7.42	7.20	7.27	
	(3.74)	(3.70)	(4.58)	
SC/ST	0.12	0.14	0.11	
	(0.33)	(0.35)	(0.31)	
Acres Owned	1.61	1.57	1.66	
	(1.74)	(1.96)	(1.75)	
Joint Difference		0.92	0.97	
Participants	681	453	226	
Participant Characteristics	Main Sample			Supplemental
Male	0.84			0.92***
	(0.36)			(0.28)
Age	46.42			46.17
	(16.77)			(46.69)
Primary School	0.66			0.64
	(0.47)			(0.48)
Secondary School	0.49			0.46
	(0.50)			(0.50)
HH Size	7.42			7.90**
	(3.74)			(4.00)
SC/ST	0.12			0.13
	(0.33)			(0.34)
Saved Seeds	1.92			1.86
	(4.78)			(4.36)
Joint Difference				0.00
Participants	681			1125

Notes: Group averages with standard deviations in parentheses. Top panel describes household characteristics by participation among invited farmers. Rows correspond to fraction male, farmer age, primary school completion, secondary school completion, household size, fraction belonging to a schedule caste or scheduled tribe, and land area owned by household at pilot baseline. Bottom panel describes participant characteristics among main and supplemental sample. Rows correspond to fraction male, participant age, primary school completion, secondary school completion, household size, fraction belonging to a schedule caste or scheduled tribe, and self-reported stock of saved pulse seeds for planting.

Table S3: Participation in the Summer Winter (Rabi) Demand Elicitation

Household Characteristics	Main Sample	Declined	Unavailable	
Farmer Male	0.85	0.84	0.88	
	(0.35)	(0.37)	(0.33)	
Farmer Age	49.93	47.91**	45.00**	
	(16.07)	(15.83)	(14.80)	
Farmer Primary Sch.	0.63	0.60	0.70	
	(0.48)	(0.49)	(0.46)	
Farmer Secondary Sch.	0.46	0.45	0.48	
	(0.50)	(0.50)	(0.51)	
HH Size	7.55	7.17*	6.81	
	(3.91)	(3.86)	(3.19)	
SC/ST	0.13	0.12	0.08	
	(0.34)	(0.33)	(0.28)	
Acres Owned	1.73	1.48**	1.72	
	(1.93)	(1.73)	(1.84)	
Joint Difference		0.04	0.43	
Participants	573	727	61	
Participant Characteristics	Main Sample			Supplemental
Male	0.82			0.91***
	(0.39)			(0.29)
Age	47.81			47.55
	(16.32)			(15.28)
Primary School	0.64			0.62
	(0.48)			(0.48)
Secondary School	0.47			0.45
	(0.50)			(0.50)
HH Size	7.55			7.94*
	(3.91)			(3.89)
SC/ST	0.13			0.11
	(0.34)			(0.32)
Saved Seeds	3.05			3.01
	(6.69)			(6.93)
Pulse Program Participant	0.59			0.39***
	(0.49)			(0.49)
Pulse Program Beneficial	0.73			0.69
	(0.45)			(0.46)
Prior NGO Beneficiary	0.66			0.45***
	(0.48)			(0.50)
Joint Difference				0.00
Participants	573			1080

Notes: Group averages with standard deviations in parentheses. Rows defined as in Table S2 with the addition of self-reported stock of saved pulse seeds for planting, self-reported participation in the pulses program, subjective belief about whether the pulse program was beneficial, and self-reported participation in prior NGO initiatives.

Table S4: Effect of Evaluation Salience among Supplemental Farmer Sample

	(1)	(2)	(3)	(4)
Treated	-0.190	-0.257	-0.262	-0.333
	(0.10)	(0.16)	(0.12)	(0.19)
Salient × Treated		0.130		0.158
		(0.19)		(0.22)
		[0.52]		[0.48]
Salient	-0.066	-0.167	-0.040	-0.145
	(0.07)	(0.17)	(0.08)	(0.19)
	[0.35]	[0.38]	[0.63]	[0.47]
Salient × Pulse Program				0.415
				(0.15)
				[0.01]
Pulse Program Participant			0.111	-0.101
			(0.09)	(0.12)
Salient × Prior NGO				-0.401
				(0.15)
				[0.01]
Prior NGO Beneficiary			-0.015	0.192
			(0.10)	(0.14)
Mean Demand	1.14	1.14	1.19	1.19
Salient Treat Effect (p-val.)		0.28		0.15
R-Squared	0.15	0.15	0.16	0.17
Observations	16650	16650	13800	13800

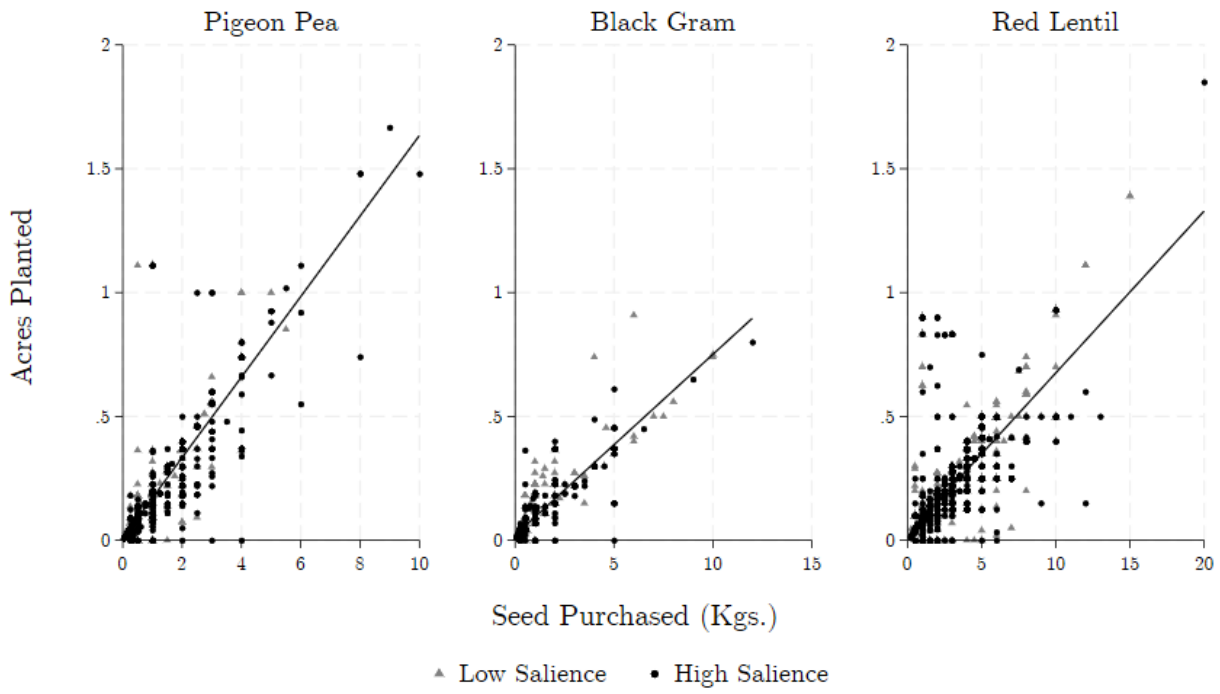
Notes: Outcome is pulse seed quantity demanded. Sample is the set of farmers invited by village leaders to reach participation target in demand elicitation. Treated: Village received pulse program treatment in prior two years. Salient: High-salience script in elicitation. Pulse Program: Participant self-identifies as beneficiary of pulse program. Prior NGO: Participant self-identifies as beneficiary of previous NGO programs. Salient Treat Effect: p-value from test of sum of coefficients on Treated and Salient × Treated. All regressions include crop and price fixed effects, block (sub-district) fixed effects, and participant demographic controls. Columns (3) and (4) restrict to those that self-reported program participation during the red lentil (winter) elicitation. Standard errors clustered by village in parentheses; p-values from randomization inference over 1,000 re-draws of village-level salience treatment status in square brackets.

Table S5: Seed Purchases and Cultivated Area (Treated Villages)

	All	Pigeon Pea	Black Gram	Red Lentil
	(1)	(2)	(3)	(4)
Input Quantity	0.078	0.163	0.073	0.065
	(0.00)	(0.01)	(0.00)	(0.00)
Salience Treatment	-0.004	-0.010	-0.011	-0.006
	(0.01)	(0.01)	(0.01)	(0.01)
Mean Seed (kg.)	1.61	0.97	0.97	1.92
Mean Area (acre)	0.15	0.17	0.09	0.15
R-Squared	0.60	0.79	0.75	0.61
Observations	11087	2402	1138	7547

Notes: Outcome is acres devoted to pulse crop. Input quantity is measured in kgs. of seed purchased from NGO supplier. Column (1) uses data from all crops and includes crop fixed effects; Columns (2)–(4) present results by crop. Data only include farmers from villages treated in pilot evaluation because records were not kept in control villages. Standard errors clustered by village in parentheses.

Figure S1: Seed Purchases and Cultivated Area (Treated Villages)



Notes: Data come from implementer administrative records, which include farmers that did not participate in experimental demand elicitation. Records were only kept in treated villages from the pilot experiment. Hollow triangles represent villages with low evaluation salience at demand elicitation, and solid squares represent high evaluation salience.

B Supplemental Results

B.1 Treatment Effect Heterogeneity

In Figures S2 and S3 we explore heterogeneity in treatment effects by crop and by implementing NGO. Figure S2 confirms the demand patterns observable in Figure 1: The effect of evaluation salience is strongest for black gram and red lentil. These are the two crops the pulses program focused most on in the second implementation year. By contrast, pigeon pea was included in initial program activities but subsequently de-emphasized by implementing partners. The effect of evaluation salience is correspondingly weaker for this variety.

We also observe heterogeneity in the effect of evaluation salience between implementing partners. The sign of the effect is the same for all four partners, as shown in Figure S3. However, the magnitude is greater for two—Kaushalya Foundation and SSEVS— and smaller for the other two—AKRSP and Nav Jagrati. We observe no systematic differences in participants' self-reported participation in the pulses program, subjective belief that the program was beneficial, or self-reported participation in prior NGO initiatives that correspond to this pattern of treatment effect. With only four implementers in the study, we are hesitant to speculate on possible causes of heterogeneity and leave it to future research.

[Figure S2 about here.]

[Figure S3 about here.]

B.2 Effects on Input Demand Elasticity

In Table S6 we report estimated treatment effects on the own-price elasticity of seed demand. The outcome is defined at the farmer–crop level by running a regression of log quantity demanded on log price separately for each farmer and each crop. There are no statistically significant effects of either pilot treatment or evaluation salience on demand elasticity.

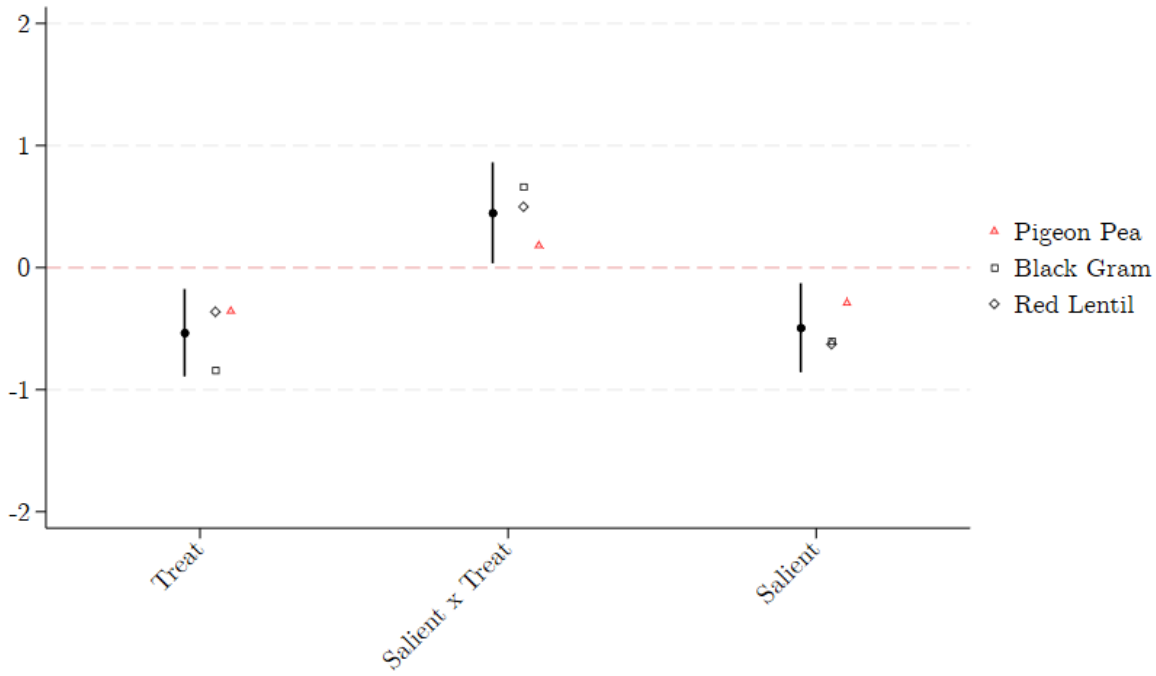
[Table S6 about here.]

Table S6: Effect of Evaluation Salience on Input Demand Elasticity

	(1)	(2)	(3)	(4)
Treated	-0.074	-0.143	-0.010	-0.060
	(0.08)	(0.12)	(0.09)	(0.14)
Salient × Treated		0.131		0.114
		(0.15)		(0.17)
		[0.44]		[0.58]
Salient	0.017	-0.075	0.010	-0.326
	(0.08)	(0.13)	(0.09)	(0.17)
	[0.83]	[0.59]	[0.91]	[0.11]
Salient × Pulse Program				0.430
				(0.16)
				[0.01]
Pulse Program Participant			0.111	-0.113
			(0.12)	(0.16)
Salient × Prior NGO				0.015
				(0.15)
				[0.93]
Prior NGO Beneficiary			0.100	0.089
			(0.09)	(0.11)
Mean Elasticity	-0.83	-0.83	-0.88	-0.88
Salient Treat Effect (p-val.)		0.91		0.61
R-Squared	0.07	0.07	0.10	0.11
Observations	1459	1459	1046	1046

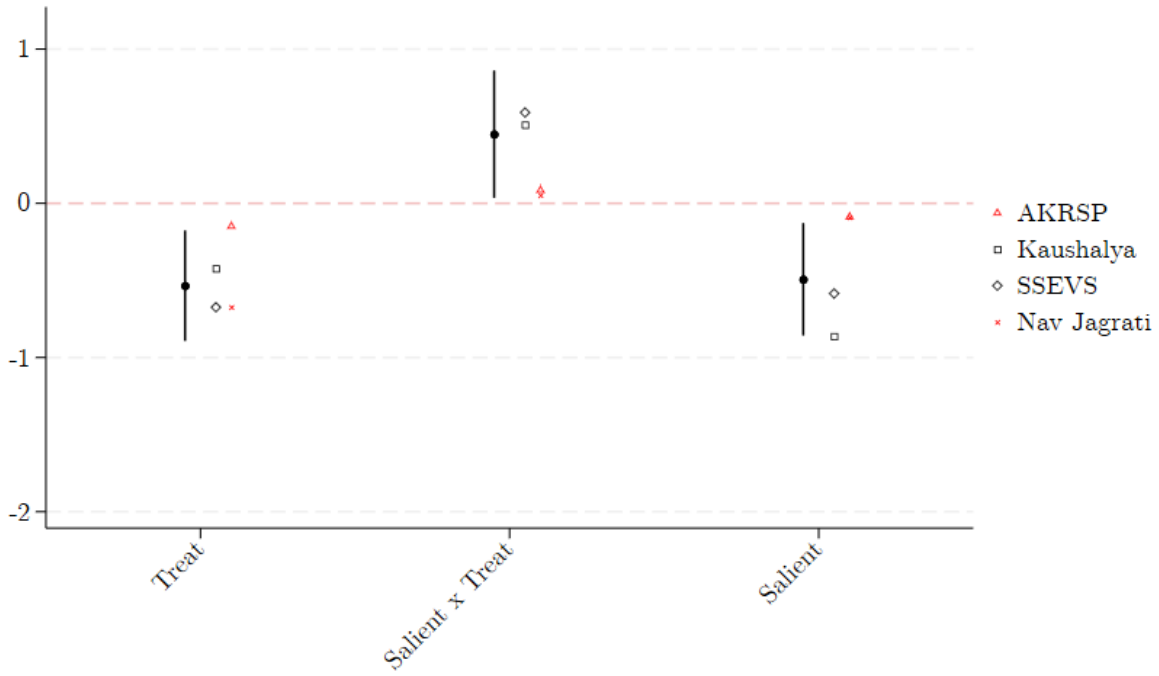
Notes: Outcome is own-price elasticity of pulse demand. Treated: Village received pulse program treatment in prior two years. Salient: High-salience script in elicitation. Pulse Program: Participant self-identifies as beneficiary of pulse program. Prior NGO: Participant self-identifies as beneficiary of previous NGO programs. Salient Treat Effect: p-value from test of sum of coefficients on Treated and Salient × Treated. All regressions include crop fixed effects, block (sub-district) fixed effects, and participant demographic controls. Columns (3) and (4) restrict to those that self-reported program participation during the red lentil (winter) elicitation. Standard errors clustered by village in parentheses; p-values from randomization inference over 1,000 re-draws of village-level salience treatment status in square brackets.

Figure S2: Treatment Effect Heterogeneity by Crop



Notes: Solid markers reproduce estimates from Column 2 of Table 2 following regression equation in (1). Hollow markers each present point estimates for a single crop.

Figure S3: Treatment Effect Heterogeneity by Implementing NGO



Notes: Solid markers reproduce estimates from Column 2 of Table 2 following regression equation in (1). Hollow markers each present point estimates for a single implementing NGO.