WHEN LESS IS MORE: EXPERIMENTAL EVIDENCE ON INFORMATION DELIVERY DURING INDIA'S DEMONETIZATION

ABHIJIT BANERJEE[⋆], EMILY BREZA[§], ARUN G. CHANDRASEKHAR[‡], AND BENJAMIN GOLUB[†]

ABSTRACT. How should information be disseminated to large populations? The options include broadcasting (e.g., via mass media) and informing a small number of "seeds" who then spread the message. While it may seem natural to try to reach the maximum number of people from the beginning, we show, theoretically and experimentally, that when incentives to seek information are endogenous, this is not necessarily true. In a field experiment during the 2016 Indian demonetization, we varied how information about the policy was delivered to villages along three dimensions: how many people were initially informed (i.e., broadcasting versus seeding); whether the identities of the initially informed were made common knowledge; and number of facts delivered (2 versus 24). The quality of information aggregation was measured in three ways: the volume of conversations about demonetization, the level of knowledge about demonetization rules, and the likelihood of making the correct choice in a strongly incentivized decision where understanding the rules is key. Under common knowledge, seeding dominates broadcasting. Moreover, common knowledge makes seeding more effective but broadcast *less so*. These comparisons hold for all three outcomes and underscore the importance of the incentive to engage in social learning. Using data on differential behavior across different ability categories, we interpret our results via a model of image concerns, and also consider several alternative explanations.

Date: This Version: October 22, 2021.
We gratefully acknowledge financial support from The Pershing Square Fund for Research on the Foundations of Human Behavior and The National Science Foundation under SES-1658940. We thank Mohammad Akbarpour, George Beknazar-Yuzbashev, Leo Burzstyn, Chris Blattman, Pascaline Dupas, Matt Jackson, Rachel Kranton, Michael Kremer, Ted Miguel, Ben Olken, Evan Sadler and many seminar participants for helpful comments and conversations. We are grateful to Ruben Enikolopov for an insightful discussion. We thank Pooja Suri, Arnesh Chowdhury, Piyush Tank, Tithee Mukhopadhyay, Sumhith Veda Aradhyula, Anushka Chawla, Vasu Chaudhary, Yixi Jiang, Devika Lakhote, Alokik Mishra, Shambhavi Priyam, Aneesha Parvathaneni, Awa Ambra Seck, Vedika Vyas, and Xuan Zhu for excellent research assistance. Stanford IRB-35996.
[⋆]Department of Economics, MIT; NBER; JPAL.
[§]Department of Economics, Harvard University; NBER; JPAL.
[‡]Department of Economics, Stanford University; NBER; JPAL.
[†]Department of Economics, Northwestern University.

1. Introduction

How should new information that is potentially valuable to a large population be delivered? For example, during an epidemic such as zika or COVID-19, there is a useful list of do's and don'ts; how does a government or an NGO get that information to the relevant population? In practice, there are two commonly used strategies: (1) broadcasting information widely to all (e.g., radio, television, newspaper, or a Twitter feed) and (2) delivering information to a select few "seed" individuals and relying on subsequent diffusion (which we see in viral marketing, agricultural extension services, or the introduction of microcredit).[1]

It might seem evident that the dissemination of the information is maximized by delivering the information to the maximum number of people from the beginning. However, the success of dissemination strategies often relies on community members' own engagement in social learning. If the mode of dissemination itself affects engagement, then it is not at all obvious which policy option ultimately generates the most knowledge. This is the question we tackle in this paper.

To fix ideas, suppose people need to talk to others to get help understanding the information they have been given, but they worry about exposing their ignorance or lack of comprehension. Prior surveys and experiments document that image concerns—specifically, the reluctance to reveal compromising information about one's ability—can inhibit engagement in social learning.[2] If information is broadcast widely in a public way, then agents may hesitate to ask for help because everybody will know they too received the information and nevertheless need help. In contrast, other strategies, such as only giving the information to a handful of individuals, may allow people to seek information with less concern about others' inferences about them.

These considerations highlight the potentially vital role played by meta-knowledge—specifically, what people know about the information that has been shared with others—in the effectiveness of information campaigns. In this paper, we examine how meta-knowledge choices affect the value of providing more people with information, and in particular the possibility that meta-knowledge crowds out the benefits of the information itself. This has direct implications for the design of information policies, and also raises new theoretical and empirical questions concerning situational incentives for engagement in learning.

[1] See, e.g., Leskovec et al. (2007); Ryan and Gross (1943); Conley and Udry (2010); Miller and Mobarak (2014); Banerjee et al. (2013); Beaman et al. (2021); Cai et al. (2015).

[2] Chandrasekhar et al. (2018) showed the existence of such a friction in a lab-in-the-field setting in a similar population. This ties into a literature that analyzes image concerns, and leverages them to affect behavior, in settings ranging from retail transactions to educational choices to tax compliance to voter turnout to vaccinations. See, for example, Goldfarb et al. (2015), Bursztyn et al. (2019), Butera et al. (2019), Perez-Truglia and Troiano (2018), Gerber et al. (2008), and Karing (2018). See Bursztyn and Jensen (2017) for a comprehensive survey.

To investigate whether potentially perverse effects of meta-knowledge play an important role in a policy-relevant setting, we conducted a randomized experiment in 200 villages in Odisha, India during the 2016 Indian demonetization, approximately six weeks after Prime Minister Narendra Modi announced the demonetization of all Rs. 500 and Rs. 1000 notes. The policy was unexpected and far-reaching, affecting 86% of India's currency. While there was near-universal awareness of the broad outlines of the policy, its chaotic implementation, involving over 50 rule changes in a seven-week period, led to widespread confusion and misinformation (see Appendix A). For example, in our baseline sample, 15% of respondents thought that the Rs. 10 coin was also being demonetized, though this was never a possibility; 25% did not understand that demonetized currency could only be deposited into a bank account (as opposed to being exchangeable for new bills over the counter). Thus, this is a context where individuals needed help interpreting policy information. At the same time, our survey evidence indicates that image concerns deterred people from asking questions: respondents worried that they would appear ignorant or unintelligent if they asked for clarification; moreover, they also made negative judgments about others in their communities who revealed ignorance about a widely-publicized policy. These facts motivate our experiment.

In our core experimental design, we vary how many people are informed (Seed vs. Broadcast) and whether meta-information is provided (Common Knowledge vs. No Common Knowledge). We focus on comparisons of the four possible dissemination strategies: (1) (Broadcast, Common Knowledge): information is broadcast widely to all households in a village, and this fact about the information policy is itself made evident to all (as in many standard broadcast methods); (2) (Seed, Common Knowledge): information is delivered to a small set of (five) "seed" individuals, and this is again made evident to the community (much like the way extension services publicize the identities of model farmers who were given training on a new technology, etc.); (3) (Seed, No Common Knowledge): information is again seeded with a small set of (five) individuals, but this is not publicized (as in viral marketing); and (4) (Broadcast, No Common Knowledge): information is dispersed widely but this is done in a way that does not generate public awareness of the delivery strategy (for example, through private mailers). We also had a third dimension of variation: the volume of information delivered. In some villages, we delivered two facts (Short), while in others we delivered 24 (Long). This dimension is of obvious practical relevance, since policymakers need to decide how much information to deliver.[3]

We present a simple theoretical framework to analyze our results. An individual decides whether to engage in social learning. The value of not engaging depends on information that the individual has at baseline, as well as information received from an intervention.

[3]Moreover, as we will discuss, it enables important auxiliary tests of our theory.

Engaging yields more valuable information if other people in the community are more informed. To decide whether to seek, the individual assesses the probability that others are informed based on the announcements of the policymaker. Engaging also has physical and opportunity costs. Jointly, these factors determine the net benefits of seeking information. We start with a benchmark frictionless model, where individuals seek exactly when these net benefits are positive, without any distortion. An important testable implication we derive is that (Broadcast, No Common Knowledge) should have less engagement than (Broadcast, Common Knowledge) because the latter treatment makes it clear that information is available. For similar reasons, (Seed, No Common Knowledge) should do worse than (Seed, Common Knowledge). Finally, (Broadcast, Common Knowledge) might involve less seeking than (Seed, Common Knowledge) if, after receiving information, the benefits of seeking clarification are smaller than the costs (we will call this the *high-cost* hypothesis). However, this can happen only if an even larger reduction is seen in (Broadcast, No Common Knowledge) relative to (Seed, Common Knowledge).

Next, we hypothesize a friction that distorts the seeking decision, coming from an image concern—specifically, people caring about how others assess their ability to understand the information they are given.[4] We analyze how the resulting seeking behavior depends on the environment. Public announcements about who has information increase the perceived value of seeking information but also introduce signaling concerns. If it is common knowledge that everyone was informed, those who received information but did not fully comprehend it will worry about what asking questions signals about their comprehension. On the other hand, if it is common knowledge that only a few specific people were informed, or if the breadth of the information delivery is not made public, there is less reason to hesitate about asking questions. This mechanism implies that, in the presence of image concerns, (Broadcast, No Common Knowledge) generates more conversations and more learning than (Broadcast, Common Knowledge), overturning the main prediction of the frictionless model. We also use the model to show when (Seed, Common Knowledge) will outperform (Broadcast, Common Knowledge), despite the fact that the latter delivers more information and offers people more opportunities to seek clarification. The force here is that under (Seed, Common Knowledge),

[4]This ties into the image concerns literature discussed previously, and work in psychology on shame and stereotype threat. The literature on shame experimentally establishes, typically in single-person lab experiments, that feeling negatively judged can lead to general withdrawal (Gruenewald et al., 2007; de Hooge et al., 2010), which guided our hypothesis that it could inhibit engagement in social learning. Chandrasekhar et al. (2018) has a detailed review of distinctions between the shame effects more often examined in psychology and strategic/instrumental decisions to manage image, showing that they are quite different in how they operate and establishing, in lab-in-the-field experiments, that both effects are present. A bit farther afield is a literature on stereotype threat (Spencer et al., 2016; Pennington et al., 2016; Steele and Aronson, 1995), which uses experiments to investigate whether prompting people to think of negative stereotypes of themselves can deter their cognitive performance.

seeking information is normal since those who are not seeds have little information, whereas under (Broadcast, Common Knowledge) seeking is a sign that they did not understand what they were told. The model also provides more detailed predictions concerning how changes in seeking rates depend on an individual's ability, providing a framework to probe the mechanism further and compare this theory to alternatives.

In all of our experimental treatments, the information we provided consisted of a list of facts in a printed pamphlet, and the same pamphlet was provided to all households who received information in that village. Our experiment was conducted in the ten days starting on December 21, 2016, when banks stopped accepting demonetized notes, and the facts came directly from the Reserve Bank of India's circular (released on December 19th, 2016), containing up-to-date information that the policymakers themselves chose to communicate to the public. Two points are worth making. First, the information was unlikely to cover everything the villagers needed to know about the policy. Even the long lists of 24 facts fell short of a full description of the policy and contained only national information, rather than local implementation details. Second, people were inundated with information, not all of it credible, and had to assess which information to believe and what to do with it. As a result, consulting others was likely beneficial; indeed, we hoped the pamphlets would make the villagers realize that there was hard information to be had and encourage the sharing of information, including topics that were not covered in the pamphlets.

We returned to the study villages three days after the intervention and measured three primary outcomes: engagement in social learning, policy knowledge, and choice in an incentivized decision. To measure engagement, we asked how many people villagers spoke with about demonetization over the prior three days. We refer to this number as the volume of conversations. To measure knowledge, we asked questions about the demonetization rules and calculated an overall knowledge score from the responses. For an incentivized measure of choice quality, we asked the subjects to select one of the following three options: (a) same-day receipt of a Rs. 500 note (worth 2.5 days' wage) in the old currency, which was still legal to deposit in the bank; (b) an IOU for Rs. 200 in Rs. 100 notes (unaffected by demonetization) redeemable 3-5 days later; and (c) an IOU for dal (pigeon peas) worth Rs. 200, again redeemable 3-5 days later. At the time of the choice elicitation, subjects still had time to deposit the Rs. 500 note at the bank, no questions asked, and we show that it was very easy to do so.

From a policy perspective, we are interested in which of the core strategies leads to the greatest social learning. We find, contrary to the frictionless model, that (Broadcast, No CK) outperforms (Broadcast, CK).[5] Next, we observe that (Seed, CK) dominates both (Seed, No

[5]We often abbreviate Common Knowledge by CK.

CK) and (Broadcast, CK). A final striking observation is that (Seed, CK) does as well as (Broadcast, No CK): the social learning that occurs in the former treatment is sufficient to match the value of informing everyone in a way that does not activate signaling concerns.

First, we look at participation in social learning. Adding common knowledge to a seeding strategy makes for more conversations; going from (Seed, No CK) to (Seed, CK) increases the number of conversations by 103% ($p = 0.04$). Among broadcast strategies, however, we find the reverse: (Broadcast, CK) generates 63% fewer conversations ($p = 0.02$) than (Broadcast, No CK). This reversal, as we noted above in discussing theoretical predictions, should not happen in a frictionless model. In addition, going from (Seed, No CK) to (Broadcast, No CK) increases the number of conversations by 113% (p = 0.048), but (Broadcast, CK) leads to 61% *fewer* conversations (p = 0.029) than (Seed, CK). While potentially unintuitive, this alone need not be inconsistent with the frictionless model; recall from the predictions that if people have more information to start with from the broadcast, it may deter further seeking under a high-cost hypothesis. However, the frictionless model cannot account for why this reduction is so dramatic while (Broadcast, No CK) and (Seed, CK) are very similar, since the frictionless model's prediction was that this reduction in the incentive to seek would be, if present, even larger without common knowledge.

Second, we turn to whether the changes in participation correspond to changes in knowledge. Going from (Seed, No CK) to (Seed, CK) increases the knowledge index by 5.6% ($p = 0.0142$). Within broadcast, (Broadcast, CK) has a 3.8% lower knowledge index than (Broadcast, No CK), though the effect is not statistically significant ($p = 0.17$). Finally, while going from (Seeding, No CK) to (Broadcast, No CK) corresponds to a 4.9% increase in the knowledge index ($p = 0.053$), going from (Seed, CK) to (Broadcast, CK) leads to a 3.1% *reduction* in the knowledge index ($p = 0.062$). This is not per se inconsistent with the frictionless model, but the fact that seeding five people generates more knowledge overall than seeding everyone is nevertheless striking. In particular, it implies that engaging in social learning is critical to making the best use of information, which is a key fact.

Third, we look at the incentivized decision—whether subjects choose the Rs. 500 note over an IOU worth Rs. 200. We again see a similar pattern. Going from (Seed, No CK) to (Seed, CK) leads to an 81% increase in the probability of choosing the Rs. 500 note ($p = 0.037$) but going from (Broadcast, No CK) to (Broadcast, CK) leads to a 48% decline in the probability of choosing the Rs. 500 note ($p = 0.041$), which is the reversal that we flagged before. Going from (Seed, CK) to (Broadcast, CK) leads to a 38.5% decline in the probability of choosing the Rs. 500 note ($p = 0.104$). In contrast, there is a 114% increase in the probability of choosing the note when going from (Seed, No CK) to (Broadcast, No CK) ($p = 0.014$). As before, the large magnitude of the reduction in the probability of the right choice between (Seed, CK) and (Broadcast, CK)—without any comparable reduction in the comparison of (Seed, CK) and (Broadcast, No CK)—is evidence in favor of frictions playing an important role.

We then show additional evidence supporting specific predictions of our signaling model. First, in the signaling model the distinction between high and low types is key: the results in the model are driven by the fact that the low types seek more in general and, as a result, the high types cut back on seeking more when reputational concerns are activated. To test this prediction, we use information from our baseline surveys and a random forest approach to construct a mapping from demographic covariates to predicted baseline policy knowledge. We then use this mapping to classify individuals in the endline survey sample into high and low predicted ability, in the language of the model. We find that when Common Knowledge is added to Broadcast, the high types reduce their seeking more than the low types. This is exactly the compositional effect that underlies a signaling explanation. The additional evidence also helps us assess the "high-cost" hypothesis within the frictionless model—that people seek less in (Broadcast, CK) relative to (Seed, CK) simply because they are already endowed with information. We note that this hypothesis cannot explain why the composition of those seeking information is so different across (Broadcast, CK) and (Broadcast, No CK). Thus, the ability results are also evidence of an important role for situational frictions.

Second, we make use of the arm in which we varied the length of information delivered. The main takeaway here is that the perverse effects of meta-information are more strongly observed in the short treatments. Each person has 1.396 fewer conversations in (Broadcast, CK) compared to (Broadcast, No CK) when pamphlets are short ($p = 0.00428$). When the pamphlets are long, the same change results in a smaller reduction in conversations ($p = .0783$), which is not distinguishable from zero ($p = .448$). We see this as additional support for the signaling model. It is more compromising not to understand short, simple messages than long complicated ones, and therefore asking questions is more revealing of one's type when the message is short.

Taken together, we find that in a policy-relevant context, perhaps counter-intuitively, (Seed, CK) is the best of the typically-available strategies, and as good as (Broadcast, No CK) when that is feasible.[6] Our results highlight the central role of meta-information in mediating the success of information campaigns and its perverse interaction with the scope of dissemination activities. Consistent with a model of image concerns, but not the frictionless case, removing common knowledge under broadcasting leads to increased learning. In other words, even if contacting all households is feasible, the policymaker might do just as well by publicly informing a few seeds. Moreover, if broader outreach is carried out, then it may be important not to publicize its breadth.

[6] E.g., when a policymaker can regularly deliver messages to everyone without this practice being widely known.

In the penultimate section of the paper, we discuss several alternative models and narrow the range of theories that are compatible with all the evidence. An interesting class of alternative explanations has people actively sharing what they learned, in a way endogenous to the treatment. For example, we consider explanations based on seeds exerting effort to share information widely when it is commonly known they have it (because they feel responsible for distributing it, for example). However, we find this is inconsistent with their measured active engagement, which is only 0.0853 conversations higher in (Seed, CK) than in (Seed, No CK), relative to non-seed households ($p = 0.891$).[7] Another alternative explanation focuses on a different kind of image concern—where individuals differ in their judgment of what information to share (rather than in their ability to interpret it). Discerning individuals avoid sharing when it is known that everyone got the same information, while less discerning individuals continue to share even when it is not needed. Common knowledge of broadcast then creates a similar signaling issue as in our main model, reducing passing differentially more for the discerning types when it is known that everyone received information. This model can account for several of the important comparisons and may well play some role in the key treatment differences. At the same time, we discuss why this theory cannot by itself account for all of the facts (in particular, the success of seeding with common knowledge), and why endogenous seeking likely plays a substantial role. We also argue that well-known social learning frictions cannot explain our results. Finally, we consider several other more elaborate alternative behavioral models that could be devised to explain our findings. While we cannot rule out all combinations of alternative explanations, we argue that a mechanism based on image concerns has substantial advantages in explaining the data parsimoniously.

By emphasizing the role of choosing to engage in conversation and the image concerns involved in doing so, this paper highlights the importance of a force relevant for the large and growing literature on social learning, but not typically studied in social learning models.

The remainder of the paper is organized as follows. Section 2 describes the context and setting, as well as motivating evidence for the importance of image concerns. Section 3 describes the experimental design and its implementation. Section 4.1 presents theoretical predictions first in a frictionless benchmark model, then in a model with image concerns. We present our empirical results in Section 5 and a discussion of alternative models in Section 6. Section 7 provides a discussion.

[7] Niehaus (2011) emphasizes a different aspect of endogenous participation. In his model, the informed party decides whether or not to reveal what they have learned.

2. Context and setting

2.1. **Demonetization.** On November 8, 2016, Indian Prime Minister Narendra Modi announced a large-scale demonetization. At midnight after the announcement, all outstanding Rs. 500 and Rs. 1000 notes (the "specified bank notes" or SBNs) ceased to be legal tender. Demonetization affected 86% of circulating currency (in terms of value), and individuals holding SBNs had until December 30, 2016 to deposit them in a bank or post office account. Modi intended for the surprise policy to curb "black money" and, more broadly, to accelerate the digitization of the Indian economy. The policy affected almost every household in the country, either directly because they held the SBNs or indirectly through the cash shortages that resulted from problems in printing and distributing enough new bills fast enough.[8]

The implementation was chaotic. The initial rollout revealed a number of ambiguities, loopholes, and unintended outcomes. As a result, the government changed the rules over 50 times in the seven weeks following the announcement. The changes concerned issues such as the time frame for over-the-counter exchange of SBNs, the cash withdrawal limit, the SBN deposit limit, and various exemptions—e.g., for weddings, which tend to be paid for in cash. See Appendix A for a timeline of these rule changes.

2.2. **Setting.** Our study took place in 225 villages across 9 sub-districts in the state of Odisha, India. The baseline was conducted starting December 21, 2016, the intervention on December 23, 2016, and the endline ran from December 26 to 30, 2016. All survey activities were completed before the December 30 bank deposit deadline.

Our study villages have two or more hamlets, each dominated by a different caste group. Typically one hamlet consists of scheduled caste and/or scheduled tribe individuals (SCST), commonly referred to as lower caste. The other hamlet consists of general or otherwise-backwards caste (GMOBC) individuals, commonly referred to as upper caste. The hamlets are typically 1/2 to 1 km apart. Given the hamlet structure of the study area, all of our treatments and outcomes were focused on only one randomly-chosen hamlet in each village.

Basic sample statistics are provided in Table 1. 89% of individual respondents in the sample had some kind of formal bank account, 80% of respondents were literate, and major occupations included casual laborer (21%), domestic worker (16%), landed farmer (16%) and share-cropper (9%).

2.3. **Baseline knowledge of demonetization rules.** Using responses from our baseline survey, we first explore the beliefs of villagers about the rules prior to our intervention. While villagers almost universally understood that the Rs. 500 and Rs. 1000 notes were being taken out of circulation, Panel A of Table 2 documents that many households had

[8] Chodorow-Reich et al. (2020) estimate an aggregate decline in employment and night lights from the policy.

inaccurate beliefs about other aspects of the policy. For example, approximately 15% of the population thought (inaccurately) that the Rs. 10 coin was also being taken out of circulation;[9] 25% of villagers believed (falsely) that, at the time of our baseline survey, they could still exchange notes at the bank without first depositing them into an account. Moreover, only a small fraction of respondents could accurately tell us the deadline for being able to exchange the demonetized notes; only 50% of respondents could tell us that the notes could be deposited at post offices, RBI offices, or village government offices. Our subjects were particularly uninformed about some of the economically important details, such as the weekly withdrawal limits from banks. 33% of respondents reported that they did not know the limit, and, in total, only 22% of respondents could tell us the correct answer (Rs. 24,000). Respondents also had very poor knowledge about limits on ATM withdrawals (10% accuracy) and withdrawal limits on the low documentation *Jan Dhan* accounts used by the poor (13% accuracy). Given the widespread penetration of bank accounts noted above, the low levels of knowledge are not due to limits to financial inclusion in the study setting.

Panel B of Table 2 shows the incidence of the respondent reporting to us that they "don't know" the answer to the question.[10] While almost all respondents believed they knew which notes were being demonetized, more than 30% of respondents reported that they did not know about the withdrawal limits or how to deposit the demonetized notes anywhere besides a bank branch. So a large fraction of individuals were willing to acknowledge to us (and thus, to themselves) that they were uninformed about important aspects of the policy.

One might ask whether it was important for relatively poor households with limited formal savings to understand various details of the policy. One major implementation problem associated with demonetization was that there simply were not enough notes to meet demand, which ended up affecting the lives of most people. For example, microfinance borrowers were not able to service their loans, and demand for cash purchases at small shops fell. Even for individuals without bank accounts, properly understanding the rules would have been useful for a variety of decisions: e.g., whether to accept an IOU from an employer or customer, or how much inventory to order for a small business.

More importantly, the policy took place during Odisha's primary agricultural harvest, when labor demand is high and when rural households receive a large share of annual labor and agricultural self-employment income. Many employers reported not having enough cash to pay workers. This would have affected the majority of households in our sample.

[9]This specific rumor spread across much of the country and was reported in the Indian press (e.g., `http://www.thehindu.com/news/national/tamil-nadu/Rs.10-coins-pile-up-as-rumours-take-toll/article16966261.ece`).

[10]If a respondent answered "don't know" to any of the questions, they were then asked to make their best guess. These guesses are included in our measures of errors in Panel A.

2.4. **Motivating evidence for image concerns.** Our motivating hypothesis—that people's desire to seek out clarification, even when it is needed, may conflict with their desire to signal desirable attributes—came out of conversations about demonetization during the field-scoping phase of the project, and was also motivated by prior work (Chandrasekhar et al., 2018). That paper develops a theory of image concerns and the decision to seek information that we build on here. It also provides supporting evidence from both an experiment and a field survey conducted in Indian villages. The survey asked villagers how they seek information on several topics: farming, health, and household finance. 88% of respondents reported feeling constrained in seeking advice from others; of these, 64% felt the reason they were constrained was that they did not want to appear "weak" or uninformed. In the field experiment, when signaling concerns are randomly switched on, there is a 55% decline in the probability of a low-ability subject seeking out information that has a high monetary return.

Our field work suggests that the types of seeking frictions documented in (Chandrasekhar et al., 2018) were also relevant during demonetization. In 2018, we conducted interviews and surveys with 102 randomly-selected subjects from 4 villages in rural Karnataka, India. We first include some representative quotes and then summarize the survey results.

Consistent with the chaotic policy implementation and the low levels of baseline knowledge we document above, respondents recall feeling confused or knowing others were confused during the period of the demonetization. They also report that, because information was abundant, asking for clarification was potentially compromising.

> "There was confusion about where to deposit money, how much to deposit, where to withdraw from, where all money could be deposited and last date. People hesitate to ask because they may think, 'even after showing so much on TV, if I ask, what will they think of me. They will think I don't understand.' " – Respondent 1

> "People with more money hesitate to ask because they will worry what others will think about them [. . .] Others will think, 'Don't they know anything? People with money should know more. But if they are still asking, they must be less intelligent.' " – Respondent 2

Relatedly, individuals who understood the key points of the policy report judging others for not understanding them.

> "If someone didn't exchange money till December, they must definitely be the biggest *bewakoof* (fool) in the world." – Respondent 3

> "Not everyone knew the deadline and application process. In December if someone comes and asks even after showing on TV, I will think they are dumb. They can't understand so they must be unintelligent. Fearing that others will think like [me], some people who were confused didn't ask." – Respondent 4

And this of course reinforced the hesitancy to ask in the first place. That is, people are indeed cognizant of such judgments.

> I came to know a little later that I had 2 old notes with me. I didn't exchange because I didn't know when the last date was. If I ask someone, I was worried what they will say about me. What will people think? They will say, 'Were you lazy? Were you sleeping till now? Everything was shown on TV.' " – Respondent 5

Figure 4 displays the survey results. We find that 80% of respondents said they felt confused, and 79% felt that even at the end of the demonetization period they did not understand the note-ban's policy-relevant implications completely. 94% reported that others in the village were confused as well. At the same time, 96% of the individuals felt that people were responsible for understanding the policy. If someone in the village asked about the policy in December (after extensive public information campaigns), 80% of respondents said that the individual would seem unintelligent, while 85% said the individual would appear irresponsible. Finally, 85% said that even if they were confused, they held back from asking questions of acquaintances for fear of being judged.

In short, this is a setting in which: the policy implementation made it hard to learn; individuals felt confused; they felt that confusion was associated with being unintelligent or irresponsible; they worried that seeking out information would therefore look bad; and they therefore reduced their information-seeking. Though a large fraction of people were somewhat confused themselves, they readily admitted they were willing to pass judgment on others who did not understand how to behave. This motivates our interest in a model with image concerns.

3. Experiment: Design and implementation

Motivated by the evidence on image concerns, we designed our experiment to explore how to convey policy information to entire communities in the context of a real-world, high-stakes setting. Central to the design is the observation that meta-information might have a perverse impact on social learning if it activates image concerns and changes the willingness of individuals to participate. The ultimate goal was to influence the mode of delivery of information—to understand, for example, whether the common policy of delivering information to everyone by loudspeaker might actually do worse than other simple alternatives.

3.1. **Treatments.** All of our experimental treatment arms involved distributing pamphlets with information about demonetization to the study villages. Our goal was to spread the official policy rules, and thus all information came from the RBI circulars released up until December 19th, 2016. We took this official information, published by the central bank, and subdivided it into 30 distinct policy rules. As we implemented our experiment over the last week before the December 30 deadline, the rules that our pamphlet touched on did not change over the course of our experiment. Through informal conversations in pilot villages, we identified the 10 most useful rules for a typical villager in the study area.[11] Our experimental protocol involved giving a randomly-selected set of facts to each village—below we describe how the selection was done. All individuals receiving lists of facts in a village received the same list.

Our core design is a 2 × 2 that varies how many people got information and the extent of common knowledge. Because another important dimension for information policies concerns which facts to disclose, we added an arm varying whether villages received long or short lists of facts. Prior work has shown that more information can overwhelm individuals and harm learning (Carvalho and Silverman (2017), Beshears, Choi, Laibson, and Madrian (2013), Abaluck and Gruber (2011)), so we wanted to examine whether similar effects would be present in our social learning setting. The variation in the length of information also provides a useful test of our signaling model. Figure 1 summarizes the design.

Thus, the treatments are as follows:[12]

(1) Information dissemination:
- *Broadcast:* information pamphlets were provided to all households in the hamlet.
- *Seed:* information pamphlets were provided to 5 seed households in the hamlet, chosen as the individuals best situated to spread information in the village according to the "gossip survey" methodology of Banerjee et al. (2016).[13,14]

(2) Common knowledge:
- *No Common Knowledge (No CK):* we did not tell any subject that we were providing information to anyone else in the community.
- *Common Knowledge (CK):* we provided common knowledge of the information dissemination protocol. In "Broadcast" treatments in arm (1), every pamphlet contained a note that all other households received the same pamphlet. (Thus, if subjects understood and believed us, then they had common knowledge of the

[11]For example, one rule that we do not classify as "useful" explained how foreigners could exchange SBNs.
[12]We also attempted to get data from 30 villages where we did not intervene whatsoever and instead only collected endline data. We call these "status quo" villages. Unfortunately, these villages are not entirely comparable to our core set due to implementation failures that led to violations of randomization. We detail this in Online Appendix M.
[13]Seed households were not told that they were chosen for any particular reason.
[14]We asked each individual "If we want to spread information about the money change policy put in place by the government recently, whom do you suggest we talk to? This person should be quick to understand and follow, spread the information widely, and explain it well to other people in the village. Who do you think are the best people to do this for your hamlet?" and we allowed them to nominate anywhere from 0 to 4 individuals. The results reported in Banerjee et al. (2016) show that this methodology identified the best people in the village to spread information—informing gossips led to three times as many people being reached as informing random people or informing prominent people.

pamphlet's distribution.) In the "Seed" treatments, every household received a notification that five individuals in their community (who were identified) were provided information about demonetization by us, and that the seeds were informed that we would identify them to everyone.

(3) Information volume:
- *Long:* 24 facts were provided. The "Long" lists of facts were drawn uniformly from the overall list.
- *Short:* 2 facts were provided.
  The "Short" lists of facts contained one of the 10 most "useful" facts, drawn uniformly at random, and a second fact drawn uniformly at random from the remaining 20.[15]

Appendix B provides the total list of facts from which we selected the list for each pamphlet, and Appendix C provides examples of the pamphlets we handed out.[16] We simplified official facts from the RBI circular into ordinary language.

3.2. **Sample.** Our data collection was constrained by the fact that the project went from conception to completion in less than a month. Demonetization was announced on November 8, 2016 and banks stopped accepting the demonetized notes after December 30, 2016. We saw that there was a need to provide information that also offered a research opportunity and came up with an implementable research design as quickly as possible. However, by the time we were ready to launch the intervention it was already early December and the study had to be completed by the end of the year. This imposed constraints on what we could do and led to some implementation errors.

We started with a list of 276 villages in an area where one of us had previously worked.[17] We required that all villages in the study have multiple hamlets (the predominant village organization in the study area) and that each hamlet have at least 20 households. One hamlet in each village was supposed to be in our study; in half of the villages, chosen at random, this was the Scheduled Caste or Scheduled Tribe (SCST) hamlet while in the other half, it was the non-SCST hamlet. To facilitate planning, we randomized villages to treatments before we verified that each village met our criteria. As a result, only a set of 221 villages

[15]Thus, on average, in the Long treatment, 8 facts were useful. In the Short treatment, at least one fact was always useful, and the additional fact was useful with probability 1/3.

[16]Appendix H contains a version of our main analysis, looking separately at the endline knowledge of useful facts, facts that were reported in that particular village, and facts that were omitted from that village's pamphlets.

[17]Our list included some places where the research team had been before over the course of work on Breza et al. (2017), Breza et al. (2021), Breza et al. (2020), and Kaur et al. (2019) though the presence of researchers in these villages had ended many months before the baseline survey was conducted for this study.

were eventually treated. Sixteen villages in a new subdistrict were then added to increase the sample to 237.[18]

We collected a repeated cross-section (rather than a panel, due to the time-cost of tracking each respondent multiple times) in each village. A baseline survey was administered for 5 randomly chosen individuals in each study hamlet. We also implemented an endline survey, after treatment, with a total sample of 1248 households. Given the rush of implementing 200+ interventions in a matter of days, some additional field errors were made. Endline data was not collected in 6 villages and the intervention did not happen in 5 villages (we also did not collect endline data there). In two villages, the elders refused entry to our surveyors. In Appendix N, we show that the village-level attrition caused by these issues was not differential by treatment status. Ultimately, we have a sample of 225 villages that were treated and received endline surveys.[19,20] Figure 2 presents a timeline of the roll-out.

In each survey round, the enumerators selected households using standard circular random sampling. We asked to speak with any adult permanent resident of the household. Almost all of the survey refusals were from households in which no adult permanent resident was home at the time of the enumerator's visit.[21] In the endline surveys, we also attempted to over-sample seeds and potential seeds. Because the gossip survey was administered in the baseline, we can identify seeds and counterfactual seeds in all treatment cells.

We present a test for baseline covariate balance across our four main experimental treatment arms in Table 3. Columns 1-4 present means by covariate in each treatment cell, while columns 5-10 present p-values of pairwise comparisons of differences in means across cells. Of the 54 pairwise comparisons, only 11% have a p-value below 0.1 and only 5.5% have a p-value below 0.05, indicating balance.

3.3. **Outcomes.** We have three main outcomes of interest at endline: engagement in social learning; general knowledge about facts surrounding demonetization; and whether the respondent selected the demonetized Rs. 500 note as opposed to an IOU payable in 3-5 days for either Rs. 200 in non-demonetized notes or Rs. 200 in *dal*, a staple food.

[18]Online Appendix L repeats our main analysis dropping these new villages and shows that our conclusions remain the same.

[19]Unfortunately, also due to the intense time pressure, in 16 of the villages our field team administered the intervention and endline to the wrong hamlet. While this should be idiosyncratic and orthogonal to treatment, we collected outcome data in the intended hamlet and we redo our estimation using treatment assignment as instruments for treatment in Online Appendix K. All our results look nearly identical.

[20]We use clustered standard errors as described below. Simonsohn (2021) notes that the appropriate HAC estimators perform as well as randomization inference, but given the complex nature of implementation and attrition, coding clustered standard errors is considerably simpler.

[21]In these cases the enumerators made at least two additional attempts to conduct surveys on the day of the visit. The biggest reason for locked doors was time of day—it was much easier to find respondents early in the morning or in the evening. Because surveyors were dispatched to villages in randomized order, we control for time of entry in the village in all of our main regression specifications.

First, we collected data on the volume of conversations about demonetization, measured as the number of people each respondent spoke to about demonetization in the prior three days. This allows us to see whether engagement in social learning increased or decreased based on the dissemination strategy.

Second, we assessed knowledge of facts surrounding demonetization. We surveyed the respondent on 34 facts and calculate the fraction of correct responses.[22]

Third, at endline, we offered subjects an unanticipated choice between: (a) a demonetized Rs. 500 note; (b) an IOU to be filled in 3-5 days for Rs. 200 in two Rs. 100 notes; (c) an IOU to be filled in 3-5 days for Rs. 200 worth of dal. With a probability of 1/6, subjects actually received the item they chose. To implement the payment, we returned to each household in the sample before exiting the village, rolled the die, and provided either the Rs. 500 or the IOU notice.[23] The reason for using the IOU, which obviously relied on the villagers trusting us, was to make sure that the villagers did not go for the lower amount because they could get it right away, rather than after going to the bank. We nevertheless worried about the cost of going to the bank and depositing the 500 rupee note into an account. As noted already, 89% of respondents had bank accounts. We also collected data about the actual cost of going to the bank (see Table 4); based on the data we collected, the median wait time at banks was 10 minutes in the area and the median village in our sample was about 20 minutes from a bank by foot.[24] At the time of our experiment, depositing the bill required no documentation of the source of the cash. Thus, selecting Rs. 200 or the equivalent was giving up more than one day's wages, even accounting for the time spent traveling and at the bank. We argue that this is evidence of confusion and measures a willingness to pay to avoid holding on to the demonetized note in a period where it was both legal and easy to convert.[25] Further, we asked respondents who did not choose the Rs. 500 to provide an open-ended justification for their choice at the end of the survey module. Figure 3 shows that most individuals who did not choose the Rs. 500 note believed, mistakenly, that the

[22]It is certainly the case that some of the facts in our index are more relevant for decision-making than others. Thus, our knowledge score should be viewed as a noisy measure of decision-relevant information.

[23]In practice, we surprised the respondents by giving them the value in non-demonetized notes (Rs. 100 notes) even when they chose the Rs. 500 bill, saving them the cost of going to the bank. Note that this was our last action before we exited the village; it occurred after each subject had already locked in their response.

[24]At this time, there were still news reports of very long queues at banks and ATMs in other, more urban parts of the country. In our study area, the waits had become much more manageable compared to the weeks following the policy announcement. Nevertheless, we were concerned that the villagers' perceived wait time could be very large. Our survey data showed that this was not the case—the median perceived wait time was 15 minutes, which was consistent with reality.

[25]One issue is that given the tight time constraints, some households may not have been able to get to the bank on time. Thus, choosing the Rs. 500 option is likely an underestimate of the decision payoff that we would have seen for a less time-constrained decision.

deposit deadline had already passed. The choice between 200 rupees and the equivalent in *dal* was intended to capture general trust in paper currency and confusion about whether the 100 rupee bills had also become demonetized. Taking the money offered more flexibility, since *dal* was easy to buy in village stores.[26]

4. Model

We present a simple framework to organize our analysis of how the treatments affect endogenous communication. The model plays two roles. First, it allows us to precisely articulate predictions based on the image concern frictions that motivated our study. Second, it provides a vocabulary on which we rely after presenting our results to consider a number of alternative stories of communication, both with and without other frictions, and assess how they line up with our empirical findings.

4.1. **Basic framework.** This model is designed to study a situation where deliberate engagement in learning to acquire information is the first-order driver of differences across treatments.[27] An individual's information comes from three sources: pre-existing knowledge, information delivered by the experiment, and information acquired from social learning. We will define notation to keep track of these variables and analyze the engagement decision, both with and without frictions.

Let us focus on one decision-maker, called D, and denote by $d \in \{0, 1\}$ his decision of whether or not to engage in conversations. Let I_{D} be the indicator variable of whether this individual directly receives information in our experiment. Finally, let I_{S} be an indicator variable (potentially unknown to the decision-maker) of whether social information is available.

Let the random variable $V^{(I_{\mathrm{D}})}(0)$ denote the privately-known instrumental value to D if he chooses $d = 0$ (does *not* engage in social learning); let $V^{(I_{\mathrm{D}},I_{\mathrm{S}})}(1)$ be the privately-known value to D of information if he chooses $d = 1$ (*does* engage in social learning). Both random variables depend on I_{D}, where D directly received information, but only the latter depends on I_{S}, the presence of social information. These values include all technological features of engaging in conversation—for example, the opportunity costs of time.

What is relevant to the individual's engagement decision is the the instrumental payoff of engagement. This is a random variable

$$\Delta^{(I_{\mathrm{D}},I_{\mathrm{S}})} := V^{(I_{\mathrm{D}},I_{\mathrm{S}})}(1) - V^{(I_{\mathrm{D}})}(0) \tag{4.1}$$

[26] We explore this further in Online Appendix H.

[27] For the reasons behind this choice, as opposed to some alternative models where differences in information sharing drive the effects, see Section 6.

This random variable has a c.d.f. $F_a^{(I_\mathrm{D},I_\mathrm{S})}$, which depends on the agent's *ability*, $a \in \{\mathrm{H}, \mathrm{L}\}$, in addition to $(I_\mathrm{D}, I_\mathrm{S})$.

The instrumental payoff of seeking depends on both I_D and I_S; D's beliefs about these are determined by what he knows in the given treatment. The individual always knows whether he directly received information, I_D, but he may be uncertain about whether social information is available, I_S.

The timing is:

(1) A dissemination treatment $\mathfrak{t}$ is exogenously selected. The treatment has two dimensions: its breadth (Broadcast or Seed) and its publicness (Common Knowledge, meaning that everyone is informed of the breadth, or No Common Knowledge, where no information about breadth is delivered). The breadth determines the direct delivery I_D.
(2) Individuals learn whether they are informed and form beliefs about the presence of social information, I_S.
(3) Individuals learn their idiosyncratic values, and hence their draw of $\Delta^{(I_\mathrm{D},I_\mathrm{S})}$. They may then engage in social learning.

We will take the agent's payoff from selecting $d = 1$ as opposed to $d = 0$ to be

$$U = \Delta^{(I_\mathrm{D},I_\mathrm{S})} - f_\mathfrak{t},$$

where $f_\mathfrak{t}$ is a treatment-dependent friction distorting the privately optimal engagement decision, such as an image cost of discussing the topic. For now we leave this abstract, but our main hypothesis proposes a concrete form for this friction.

We begin by assuming that the distributions $F_a^{(I_\mathrm{D},I_\mathrm{S})}$ are exogenously given for each $(I_\mathrm{D}, I_\mathrm{S})$ and are not dependent on others' behavior. Section 6 discusses richer models where others' engagement affects seeking behavior.

In our analysis throughout this section, we focus on the decision-making of a non-seed D, who receives information only in the Broadcast treatments. Such individuals were the large majority of people in any village. Appendix D works out the details of seed D's as well.

4.2. **Frictionless model.** We first consider, as a benchmark, a frictionless model of engagement in learning. In this model, $f_\mathfrak{t} = 0$ for all $\mathfrak{t}$, and individuals engage in social learning if and only if $\Delta^{(I_\mathrm{D},I_\mathrm{S})} \geq 0$, without any wedge distorting the decision.

We give two conditions, satisfied in our setting, under which (Broadcast, CK) dominates (Broadcast, No CK) in terms of volume of conversation as well as our other outcomes.

The first condition is that $\Delta^{(I_\mathrm{D},I_\mathrm{S})}$ is increasing in I_S: the incremental value of engaging in social learning is increasing in the availability of social information. The second condition is that D's subjective probability of the event that $I_\mathrm{S} = 1$ is higher in (Broadcast, CK) than under (Broadcast, No CK). This is natural, since in the former case the fact that everyone else has information is publicized along with the pamphlets themselves; the inequality is strict for our non-seed D. We maintain the assumption from now on that these conditions hold.

Under these two conditions, *there is strictly more engagement in social learning in (Broadcast, CK) than in (Broadcast, No CK)*: D places a higher probability on $I_{\mathrm{S}} = 1$ and the distribution of values is shifted up in that case. This is because the monotonicity assumptions mean that D places more probability on the events where Δ is large, and thus finds it worthwhile to engage in more states of the world. Moreover, since in (Broadcast, CK) agents could have had a weakly greater payoff than in (Broadcast, No CK) had they chosen not to seek, they must receive a higher $\Delta^{(I_{\mathrm{D}},I_{\mathrm{S}})}$ by engaging in learning. This, in turn, implies that agents should be receiving greater informational benefits by engaging—i.e., greater knowledge and better choice outcomes. For similar reasons, (Seed, CK) and (Broadcast, CK) should both dominate (Seed, No CK).

The comparison of (Seed, CK) and (Broadcast, CK) under the frictionless model is, in general, a bit more delicate. While the latter treatment makes clarification more available ($V^{(I_{\mathrm{D}},I_{\mathrm{S}})}(1)$ increases), it also increases the endowment $V^{(I_{\mathrm{D}},I_{\mathrm{S}})}(0)$. The latter effect may dominate, reducing the net value of engagement and reducing overall engagement rates. That is, people may end up with less information in (Broadcast, CK), precisely because they started with more and therefore have a weaker incentive to seek. Of course, this requires costs of seeking to be sufficiently high that the benefits of getting clarification are not worth it—in the introduction, we called this the high-cost hypothesis. However, if we observe that (Broadcast, No CK) has equal or greater volume than (Seed, CK), then the hypothesis is ruled out. In that case, since people become more willing to seek given the mere possibility of information, even when endowed with some, we know that $\Delta^{(I_{\mathrm{D}}=0,I_{\mathrm{S}}=1)}$ stochastically dominates $\Delta^{(I_{\mathrm{D}}=1,I_{\mathrm{S}}=0)}$. It is clear that then the frictionless model predicts that (Broadcast, CK) should dominate (Seed, CK) as well.

Adding Common Knowledge to Broadcast thus constitutes a key test of the frictionless model, and if this *hurts* outcomes, the frictionless hypothesis is rejected.

4.3. **Seeking frictions from image concerns.** Now we posit that the friction term, f_{t}, comes from an image concern. We model image as being assessed by an observer, called O—a random person in the village who observes the seeking decision d that D makes and forms beliefs about D's ability. In turn, D cares about these beliefs. In particular, D values being perceived as more likely to be the high-ability type. Let $q_{\mathrm{O}}(d)$ be the observer's subjective probability that $a = \mathrm{H}$ (i.e., person D has high ability). Then the benefit of engaging in social learning is $\Delta^{(I_{\mathrm{D}},I_{\mathrm{S}})}$, while a potential cost is changing O's belief from $q_{\mathrm{O}}(0)$ to $q_{\mathrm{O}}(1)$, if the latter is lower. To incorporate both considerations, we posit that the net payoff of engaging in social learning is

$$U = \Delta^{(I_{\mathrm{D}},I_{\mathrm{S}})} - \underbrace{\lambda \mathbb{E}^{\mathrm{D}}[q_{\mathrm{O}}(0) - q_{\mathrm{O}}(1)]}_{\text{friction } f_{\mathrm{t}}},$$

where and $\lambda > 0$ is a weight and the expectation is taken from D's perspective. D engages if and only if $U \geq 0$, i.e. if $\Delta^{(I_{\mathrm{D}},I_{\mathrm{S}})} \geq f_{\mathrm{t}}$.

Solving this model involves solving for an equilibrium: D's seeking decision depends on the expected image payoff. In turn, the observer's $q_{\mathrm{O}}(d)$ is calculated using Bayes' rule, taking into account the different engagement rates of high and low types.

We assume that $\Delta^{(I_{\mathrm{D}},I_{\mathrm{S}})}$ is stochastically higher for low-ability types,[28] so that a low-ability D always has at least as much to gain from seeking as a high-ability one, all else equal.[29] Then seeking is always weakly a signal of low ability, and thus $q_{\mathrm{O}}(1) \leq q_{\mathrm{O}}(0)$, and strictly so when low-ability types have strictly more to gain. This implies that engagement in learning is distorted relative to the frictionless model.

The treatment affects both the informational benefits of engaging in social learning and the scope for image concerns to play a role. We do not develop the details here formally, but state the key ideas in a simple special case. (Full details can be found in Appendix D, where we consider weaker assumptions on payoffs.) Suppose that when they do not have information, both high- and low-ability types have equal distributions of $\Delta^{(I_{\mathrm{D}}=0,I_{\mathrm{S}}=1)}$: they stand to learn equally from others. But when they are informed, a low type needs information more than a high type. Then signaling concerns are strong when it is known that D has information, weaker when it is uncertain whether D has information, and weakest—indeed, in this special case, absent—when it is known that D does not have information.

4.3.1. *Aggregate engagement rates.* We now derive predictions about the seeking rates in all four treatments.

(1) *(Broadcast, No CK) dominates (Broadcast, CK).* The expected informational benefits under (Broadcast, CK) are weakly greater. But under CK, it is known that the seeker received information and thus signaling concerns are fully active; in contrast, in No CK, it is not known whether D is informed, and therefore less inference is drawn from his behavior.

[28]I.e. that the distribution $F_{\mathrm{L}}^{(I_{\mathrm{D}},I_{\mathrm{S}})}$ first-order stochastically dominates $F_{\mathrm{H}}^{(I_{\mathrm{D}},I_{\mathrm{S}})}$.

[29]Because the only thing that O observes is the decision d, while a and idiosyncratic value draws are private, the expected image penalty from D's perspective depends on d only.

(2) *(Seed, CK) dominates (Broadcast, CK).* In both cases, D assesses the same informational benefits from seeking but has less information under Seed.[30] Moreover, Broadcast turns on signaling concerns (since it is known D got information) whereas Seed makes it plain that D is uninformed, eliminating them.

(3) *(Seed, CK) dominates (Seed, No CK).* In the latter case, there is no reason to expect information to be available, whereas in the former, it is known that it can easily be found. Signaling concerns are small in either case, because others either know D is not informed or have no reason to believe that he is.

(4) *(Broadcast, No CK) dominates (Seed, No CK).* In the former treatment, all individuals are in the same situation as seeds in the latter treatment: they have received information and do not know anything about who else has received it. Thus, to the extent that they engage in social learning, there should be much more of it happening in (Broadcast, No CK), since there are at least 10 times more seeds than non-seeds.

In other words, in our model more is not always more. Under No CK, the Broadcast arm *increases* engagement by alerting people to at least the existence of information, without differentially activating signaling concerns, as explained above in (4). But under CK, a Broacast policy *decreases* engagement: with CK, people know about the social availability of information regardless, but Broadcasting makes it clear that D is informed and activates ability-signaling concerns.

So far we have focused on volume of conversation, rather than knowledge or choice outcomes. We consider these other outcomes in the appendix. These comparisons are more delicate, because in some cases the information endowment is increased even as engagement in learning is decreased, making a comparison potentially ambiguous. For cases (1) and (3), the results noted above for volume of conversations extend straightforwardly, as endowments do not change in the comparisons. Nevertheless, we show that under assumptions that are reasonable in our setting, the other comparisons also extend. For changes in endowments not to reverse our effects, we need that social learning is sufficiently important for enough of the population, relative to private processing of information. We believe that this is reasonable in our empirical context. The modeling details are in Appendix D.

The image concerns model has implications beyond the aggregate treatment effects discussed above. We next present predictions of the model over two dimensions of heterogeneity. First, we consider how low versus high ability agents respond to the mode of information delivery. Second, we look at how the complexity of the information itself affects the differences between treatment arms.

[30] If the information is easier to find in Broadcast, then this could in principle make the instrumental seeking benefits $V^{(I_{\mathrm{D}},I_{\mathrm{S}})}(1)$ greater in (Broadcast, CK). In that case, the prediction that we make here holds under sufficiently strong image concerns.

4.3.2. *Differential predictions for high vs. low ability agents.* In our model, ability is the key dimension of heterogeneity across agents that drives the image concern. It is low ability—in the specific sense of low facility with understanding demonetization facts—that agents are reluctant to reveal. Of course, whether engagement reveals low ability is endogenous.

We now elaborate on this in a bit more detail. Letting π_{H} be O's prior probability that $a = \mathrm{H}$, we have that, for either decision $d' \in \{0, 1\}$,

$$\frac{1 - q_{\mathrm{O}}(d')}{q_{\mathrm{O}}(d')} = \frac{1 - \pi_{\mathrm{H}}}{\pi_{\mathrm{H}}} \frac{\mathbb{P}(d = d' \mid a = \mathrm{L})}{\mathbb{P}(d = d' \mid a = \mathrm{H})}.$$

Thus, seeking decisions are most informative about type (i.e., they change beliefs from the prior) when the rates of seeking are most different across types. The probabilities $q_{\mathrm{O}}(d')$ determine the seeking friction $f_{\mathrm{t}} = \lambda[q_{\mathrm{O}}(1) - q_{\mathrm{O}}(d)]$, and D seeks if and only if $\Delta^{(I_{\mathrm{D}}, I_{\mathrm{S}})} \geq f_{\mathrm{t}}$.

Thus, in the treatments where we predict that image concerns would drive down engagement, it should be that seeking is indeed done mostly by those of low ability. Thus, our model predicts that low aggregate engagement rates and over-representation of low-ability individuals among those engaging go hand in hand. To restate these predictions:

(1) In (Broadcast, CK), seeking should be less than in (Broadcast, No CK) and (Seed, CK) for both types.
(2) The reduction in seeking should be greater for high-ability types than for low-ability types.[31]

4.3.3. *The difference between the Long and Short treatments.* We now turn to the comparison of long and short pamphlets. Whether signaling concerns will be activated depends, per the previous subsection, on the extent to which the instrumental payoffs to engaging differ between high- and low-ability types. That is, does the probability that $\Delta^{(I_{\mathrm{D}}, I_{\mathrm{S}})} \geq f_{\mathrm{t}}$ differ significantly between H and L types? We will focus on the case (Broadcast, CK), which, as we have argued above, creates the most potential for this difference.

Let us compare the Short and Long treatments (a two-pamphlet information delivery versus a lengthy pamphlet of 24 facts) through the lens of the model. One possibility is that long pamphlets confuse everyone, and nobody sees much value in speaking. In this case we would expect engagement to go down in all treatments relative to Short, with no specific prediction for the relative seeking rates.

We now turn to the case where individuals do see value in discussing the pamphlets. In this case, a natural hypothesis is that high-ability people actually have a greater value of conversation in the Long than in the Short treatment. While a high-ability person is likely to be able to read and comprehend two facts, it is less likely that the person can make use of all 24 facts correctly without discussion. If this is the case, then relative to Short,

[31]See Corollary 1 in Appendix D.

the random variable $V^{(I_{\mathrm{D}}=1, I_{\mathrm{S}}=1)}$ places more mass on high realizations even for high-ability types. Suppose this happens in such a way that for all relevant values of the friction f_{t}

$$\frac{\mathbb{P}(d=1 \mid a=\mathrm{L})}{\mathbb{P}(d=1 \mid a=\mathrm{H})} = \frac{1-F_{\mathrm{L}}(f_{\mathrm{t}})}{1-F_{\mathrm{H}}(f_{\mathrm{t}})}$$

is smaller in Long than in Short. This decreases the scope for signaling and increases seeking relative to Short.[32] Under this hypothesis, going from (Broadcast, No CK, Long) to (Broadcast, CK, Long) should generate less of a reduction in endogenous participation in social learning than going from (Broadcast, No CK, Short) to (Broadcast, CK, Short).

5. Results

We now report the main empirical results and assess them in view of the model-based predictions (with and without frictions) that we have presented.

We begin with a visual inspection of the raw data for each of the primary outcomes in Figure 5. Panel A presents the number of people with whom the respondent had a conversation, Panel B presents the knowledge index, and Panel C presents the share choosing the Rs. 500 note. We can see the main results of the paper in the raw data itself. (Seed, CK) dominates (Seed, No CK) in all three measures. Similarly, (Broadcast, No CK) dominates (Broadcast, CK) in all outcomes, with most power in the raw data for volume of conversations. Finally, strikingly, (Seed, CK) and (Broadcast, No CK) deliver similar, indistinguishable results. Below we detail the results with more structured regression analysis and demonstrate that these findings carry through.

5.1. **Endogenous participation in social learning.**

5.1.1. *Volume of conversations.* We begin by looking at which delivery mechanisms led to more or less engagement in social learning, measured by the number of people with whom the subject had spoken (henceforth "volume of conversations") over the prior three days about demonetization.[33] Results are from regressions of the following form:

$$y_{ivd} = \alpha_d + \beta_1 \mathrm{CK}_v + \beta_2 \mathrm{Broadcast}_v + \beta_3 \mathrm{Broadcast}_v \times \mathrm{CK}_v + \gamma X_v + \lambda X_i + \varepsilon_{ivd} \tag{5.1}$$

where i indexes the individual respondent, v indexes village, and d indexes the subdistrict, which was our unit of stratification. (Seed, No CK) is the omitted treatment arm. Village-level controls X_v include date and time of entry into the village, the caste category of the hamlet treated (and surveyed) in the village, and distance from the village to an urban

[32]Because the length of the pamphlet is common knowledge, both the decision-maker and observer know that the friction is lower, which makes things simple. See Appendix D for details.

[33]Table P.1 of Appendix P presents results from the same regression where we look at whether an individual had any conversation as compared to the count of the number of conversation partners.

center. The respondent-level controls X_i include age, gender, literacy and potential seed status. In all regressions we use post double-selection LASSO (see Belloni, Chernozhukov, and Hansen (2014)) in order to efficiently select controls. Standard errors are clustered at the village level.

Table 5, columns 1-3, presents OLS regressions of the volume of conversations on the various treatments.[34] The coefficients are additive, so to compare (Broadcast, CK) to the omitted category, it is necessary to add the coefficients: CK, Broadcast, and Broadcast $\times$ CK. In each regression specification, we present the p-values throughout, with standard errors clustered at the village level, and three additional key comparisons. The test (CK + Broadcast $\times$ CK = 0) allows us to compare (Broadcast, CK) to (Broadcast, No CK), which, as argued above, represents a direct test of the frictionless model. The test "Broadcast + Broadcast $\times$ CK = 0" allows us to compare (Broadcast, CK) with (Seed, CK), while the test "Broadcast = CK" allows us to compare (Broadcast, no CK) with (Seed, CK).

The outcome variable in column 1 is the volume of conversations about the demonetization in which the respondent took part over the prior three days. Going from (Seed, No CK) to (Seed, CK) increases the number of conversations per capita by 103% (0.64 more conversations, $p = 0.04$). This result is consistent with the frictionless model detailed above—adding information about the identity of the seeds makes it easier to find someone to consult. It is clear in principle that this particular result could come from the fact that seeds have a stronger motivation to spread information under (Seed, CK). However, we do not think this kind of supply response fully drives our results for two reasons. First, in Online Appendix I, we show the same regression split by whether the household was a seed or not and demonstrate that the increase in conversations by seeds in (Seed, CK) cannot account for all the increase in conversations that non-seed households must have had in (Seed, CK). That is, conversations between non-seed households must have increased.[35] To see why, we can do a conservative back-of-the-envelope calculation. Table I.1 shows that a seed individual in (Seed, CK) participates in 1.8 extra conversations relative to (Seed, No CK). There are five such individuals and so there are 9 more conversations generated. Looking at the non-seeded individuals, we see an increase of 0.6 conversations per respondent. In a village of 50 households, say with two eligible respondents per household (household head and spouse), this leaves 90 potential respondents out of whom we have an average increase of 0.6 conversations

[34]For all of our main results, we focus on our core 2 $\times$ 2 treatment design, pooling across the Long and Short lists of facts. Appendix G provides the analysis separately for Long and Short information and also discusses how one might interpret the length of the fact list through the lens of the model.

[35]Recall that every village had "seed" households selected by the same process ex ante, but in Broadcast treatments all households were treated. In Online Appendix I, Table I.1, shows that all our main results hold for the households that are not seeds.

per respondent. That means that there are 27 conversations that involve at least one non-seeded household; from before we know at most 9 of these can be with seeds. So a minimum of 18, or 67%, of the conversations must be among non-seeds.[36] Second, we collected data about the nature of the conversations—whether they were the result of a directed question or statement about demonetization (what we call primary conversations) or merely something that came up in a broader conversation (secondary conversations). These results are reported in Section 5.1.2, below. They make it clear that most of the increase came from secondary conversations, not from people going to ask questions of seeds or seeds coming to deliver a message.

Next, we compare the Common Knowledge treatments. Going from (Seed, CK) to (Broadcast, CK)—which typically corresponds to a tenfold increase in the number of households informed (from 5 households to all households)—leads to a 61% *decline* in the volume of conversations (0.78 fewer conversations, $p = 0.025$). This is a main prediction of the signaling model. It also could be consistent with the frictionless model if receiving the broadcast information substantially lowers the net value of seeking. But then, as discussed in Section 4.2, we would see at least as much of a decline between (Seed, CK) vs. (Broadcast, No CK); in fact the two point estimates are similar and not statistically distinguishable ($p = 0.861$). (The signaling theory does not predict a clear ranking between these two, but does imply that seeking in both should be high if it is high in any treatment—i.e., if we are outside the high-cost regime.)

When we go to (Broadcast, No CK) from (Seed, No CK), then, in sharp contrast to the previous result, we find an increase in the volume of conversations by 113% (0.709 more conversations, $p = 0.042$). This makes intuitive sense and is a prediction of any version of the model: essentially with (Seed, No CK) a typical household doesn't even know that there is something to converse about, whereas that is not true with (Broadcast, No CK).

The move from (Broadcast, No CK) to (Broadcast, CK) leads to a 63% *decline* in the volume of conversations (0.84 fewer conversations, $p = 0.017$). Since it is easier to find people to consult when there is common knowledge of who is informed, this is inconsistent, as we argued above, with the frictionless version of the model.

All of these observations are consistent with the version of the model with signaling frictions.[37]

[36]We can be even more conservative and imagine for some reason that every household has only one individual. Even in that case the same calculation yields 13.5 new conversations involving at least one non-seed, and at the maximum 9 of these could be between seed and non-seed, still leaving 33% purely among non-seeds.

[37]Note that the same patterns emerge when using a binary indicator for having any conversations as the dependent variable. See Appendix Table P.1.

5.1.2. *Impacts on the types of conversations.* We collected information on the number of conversations by type: primary (initiated with the purpose of talking about demonetization) and secondary (the meeting was initiated for some other purpose but then one of the parties brought up demonetization). Columns 2 and 3 of Table 5 break up the number of conversations that the subject participated in by whether they were secondary (column 2) or primary (column 3). Secondary conversations comprise the vast majority, 78%, of reported conversations. As columns 2 and 3 make clear, our core results broadly go through for each type of conversation, but significantly more of the impact of the interventions comes from the secondary conversations.[38] Consistent with that, column 3 of Appendix Table I.1 shows that the gap between the number of conversations in (Seed, CK) relative to (Seed, No CK) does not appear to be driven by the seed actively going out to explain the information to others, nor by others actively seeking out the seeds. The primary driver of the increase in conversations here is conversations among non-seeds, and we see no evidence of any effort by seeds to coordinate conversations about the topic.

5.2. **Information aggregation and choice.** Columns 4-5 of Table 5 show how knowledge of the demonetization rules and incentivized choice behavior depend on the (randomized) information environment. Recall that the quality of the respondents' choices depended on their understanding of the demonetization rules.

In column 4 we look at knowledge. It should be evident that more conversations need not imply greater knowledge—for example, even though there are fewer conversations happening in (Broadcast, CK) as compared to (Seed, CK), ten times more households received information under broadcast treatments and it is entirely possible that they knew more.

Our main measurement of the knowledge outcome after our intereventions is based on the answers to 34 questions about the demonetization policy asked at the endline.[39] The mean in the (Seed, No CK) group is 0.566. Going from seeding to broadcast under common knowledge leads to a 3.8% *reduction* in the knowledge index ($p = 0.057$). This shows that though 100% of households receive information instead of 10%, the amount of aggregated information that a random household has at the end of the day is actually less, not more. As we have discussed, while striking, this could happen in a frictionless world if endowments deter seeking enough (which happens under the high-cost hypothesis); still, by the same argument as before, the very similar performance of (Broadcast, No CK) and (Seed, CK) refutes this theory.

Within broadcast strategies, adding common knowledge leads to a 3.2% *reduction* in knowledge, though the effect is not quite statistically significant ($p = 0.154$). In addition, going

[38]On the other hand, the *relative* increase in conversations is larger for the primary variety.

[39]Recall that our treatment only gave information on a small subset of these 34 facts. We explore whether knowledge improvements are driven by the facts that were actually on the pamphlets in Appendix H.

from (Seed, No CK) to (Seed, CK) increases the score on the knowledge index by 5.5% ($p = 0.0129$), and going from (Seed, No CK) to (Broadcast, No CK) actually makes people better informed and improves knowledge by 4.9% ($p = 0.046$). Reassuringly, reductions in knowledge happen exactly where we see conversations declining, suggesting that conversations are an important source of information. In column 5, we turn to the impact of our experimental treatments on incentivized choice. We look at whether subjects choose the Rs. 500 note on the spot, which they could still deposit in their accounts, or an IOU worth Rs. 200 to be paid in 3-5 days, taking a loss of about 1.5 days' wages. The probability of selecting the Rs. 500 note in the omitted category (Seed, No CK) is only 5.92%. Within Broadcast strategies, adding Common Knowledge leads to a 48% *decline* in the probability of choosing the Rs. 500 note ($p = 0.0385$). This reversal is prima facie inconsistent with the frictionless model, as already argued. Going from Seed to Broadcast, conditional on Common Knowledge, leads to a 38.6% or 4.14pp *decline* in the probability of choosing the Rs. 500 note ($p = 0.104$), which is also striking. In addition, going from (Seed, No CK) to (Seed, CK) leads to a 4.8pp or an 81% increase in the probability of choosing the Rs. 500 note ($p = 0.03$), but going from (Seed, No CK) to (Broadcast, No CK) corresponds to a 6.76 pp or 114% increase in the probability of choosing the Rs. 500 note ($p = 0.011$). These results are fully consistent with the results on conversations and knowledge. More conversations led to better knowledge, which in turn, allowed for improved decision-making.

In a world without common knowledge, the conventional wisdom holds: increasing the number informed encourages more conversations and better decision making. However, under common knowledge, broadcasting information actually backfires, leading to worse outcomes across the board. One bottom-line result is that seeding just five households combined with common knowledge makes the outcomes indistinguishable from (Broadcast, No CK), where ten times as many people were seeded. Perhaps more strikingly, either holding Common Knowledge fixed and moving from Seed to Broadcast or holding Broadcast fixed and moving from No Common Knowledge to Common Knowledge actually reduces conversation volume, knowledge, and quality of choice. Less is more.

We also note that we find stark impacts of providing meta-knowledge despite the fact that our Common Knowledge treatment was most likely only partial. Some individuals in (Broadcast, No CK) almost surely observed neighbors receiving pamphlets, and some in (Broadcast, CK) may have never read the common knowledge information that they were provided. This points to the power of meta-knowledge in our setting.

We now turn to some additional cuts of the data. First, we consider how low versus high ability agents respond to the mode of information delivery. Second, we look at how the complexity of the information itself affects the differences between treatment arms.

5.3. **Differential outcomes for high vs. low ability agents.** In order to test the ability-based predictions laid out in Section 4.3.2, we need a pre-determined measure of agent ability pertaining to demonetization, such as pre-intervention policy knowledge. We use information from our baseline surveys to construct a mapping from demographic covariates to predicted baseline policy knowledge. We then use this mapping to classify individuals in the endline survey sample into either high or low predicted ability categories.

Specifically, we first construct a random forest (RF) algorithm applied to our baseline sample to predict their baseline knowledge (raw score in our baseline knowledge survey). We allow as predictors all the demographic variables that were collected in both the baseline and endline surveys. This set of predictors includes age, gender, a coarse occupation category, subdistrict, distance to city, subcaste, and caste category. Our random forest model is calibrated to minimize the root mean squared error of the predictions relative to the true knowledge at baseline (which we measured), and achieves a performance of 0.12 on out-of-sample data, which accounts for 17.9 percent of the variation in the data. Once we have the mapping from baseline characteristics to the knowledge score, we next apply it to the endline sample to generate predicted baseline knowledge for each household. We then classify endline respondents based on whether they are above- or below-mean within their village in this ability measure. This helps to ensure balance in ability across treatment assignment.

In Figure 6, we show that the model's ability-based predictions hold. Specifically, we construct and plot likelihood ratios of low-type to high-type seeking in each core treatment cell.[40] First, we see clearly that in (Broadcast, CK), the likelihood ratio exceeds 1 ($p < 0.1$)—that is, more Low types seek relative to High. Second, we find that when common knowledge is added to broadcast, the relative seeking rate for low versus high ability individuals increases substantially ($p = 0.034$). That is, it is the high types whose conversations disproportionately decrease when moving from (Broadcast, no CK) to (Broadcast, CK).[41]

In Appendix Table F.1 we show the results of expanding Equation (5.1) to incorporate heterogeneous treatment effects by predicted ability. Both High and Low ability types decrease conversations in (Broadcast, CK) relative to (Broadcast, No CK) and (Seed, CK), although the reductions are not statistically significant for the Low ability types. Moreover, moving from (Broadcast, No CK) to (Broadcast, CK) leads to a greater differential reduction in communication for the High types ($p = 0.012$).

[40]To do this, we regress an indicator for having any conversations on an indicator for high ability, treatment indicators, and their interactions, controlling only for subdistrict fixed effects. From the estimated coefficients, we construct the treatment-specific, relative seeking rates of Low versus High ability agents. We use the delta method to estimate standard errors for the ratios and test the differences across treatments.

[41]We would expect to see effects on both predicted types even if one is unaffected, since ability predictions are noisy and the separation between the types is imperfect.

5.4. **Differential outcomes for complex (*Long*) vs. simple (*Short*) signals.** In Appendix G, we test for the differential predictions in the Long versus Short treatments, laid out in Section 4.3.3. Specifically, we look at how going from two to 24 facts differentially impacts the effects of interest. Focusing on volume first, in Appendix Table G.1, we find that going from (Broadcast, No CK) to (Broadcast, CK) is less of a deterrent to primary conversations when there are many facts ($p = 0.078$).[42] This finding is consistent with our signaling story—there is much less image cost involved in asking when it is known that they received a lengthy booklet of facts than there is for someone to go ask about information when it is known that they have received two facts. We find qualitatively similar patterns for our knowledge and choice outcomes, but, unsurprisingly, the results are noisier there.

For completeness, in Appendix Table G.1, we show the treatment effects from distributing Long versus Short pamphlets, pooling the data across the other two treatment cells. More information per pamphlet does not lead to more conversations or better outcomes. Providing a 12-fold increase in the number of facts leads to no statistically significant benefits in any of our primary outcomes. In all three cases, the coefficients are negative.

6. Assessing alternative mechanisms

We have presented evidence that our results are consistent with the image concerns mechanism that motivated our experimental design. We now turn to whether other mechanisms could be consistent with our findings. An important class of alternative models considers endogenous decisions of whether to *share* information. Indeed, sharing behavior shaped by image concerns or other incentives can explain some of our findings. In this section, we examine both the parallels and the contrasts between models of endogenous seeking and sharing.

We also discuss a number of other alternatives based on some well-studied social learning frictions to argue that the image frictions we emphasize are quite distinct in their predictions.

Our overall takeaway is that seeking with image concerns is a particularly parsimonious explanation of the facts. We do not insist that there cannot be other equally parsimonious stories, or that the signaling mechanism is the only one operating. However, we do give reasons why a number of natural alternatives cannot by themselves explain all the patterns we observe. In the process, we sketch out how to combine the image concerns model with other important mechanisms relevant to endogenous engagement.

6.1. **Active information sharing.** In our basic model, the main endogenous decision is whether to engage in social learning in order to acquire information, and the theoretical

[42]We find a clearly negative effect of going from (Broadcast, No CK) to (Broadcast, CK) for short pamphlets, corresponding to the test on CK + BC×CK ($p = 0.0043$). The same effect is less pronounced when making the same move for long pamphlets.

counterpart of conversation volume is $\overline{d}$, average seeking rate. Learning comes from active seeking. In this subsection, we introduce a distinct type of conversation, in which people spontaneously bring up the topic and actively share what they know. This would be a different contribution to volume, which we would also measure in our outcomes, but which is driven by different behavior. Our goal is to examine which of our results can be explained by incorporating active information sharing that is endogenous to the treatments.

6.1.1. *A simple active sharing model.* The simplest model is that people share information when they have information and believe that others may not have it. To tie this into the model, we can think of the Observer O as the active sharer, and posit that she is likelier to share when her counterparty is more likely to be uninformed; i.e., her sharing increases in $\mathbb{P}^{\mathrm{O}}(I_{\mathrm{D}} = 0)$. For now, we study this as a mechanical rule: we simply posit that people like to offer helpful or interesting information for whatever reason. Can this type of theory by itself explain our results? To sharpen this question, consider a model with *only* the active sharing. Each individual participates in a number of active sharing conversations; since there is no seeking effort, these conversations are accessed homogeneously throughout the village.

This hypothesis predicts the least active sharing in the (Broadcast, CK) treatment: everyone is sure that others have heard the information. In (Broadcast, No CK), everyone has information, but if they think it sufficiently likely that others might not have it, then we expect to see much more sharing. In (Seed, CK), there are also more conversations than under (Broadcast, CK), because the seeds are aware that they know but others do not. Thus, this theory can explain at least some of the quantitative patterns.

This simple model has a harder time accounting for heterogeneity across ability types. Recall from Section 5.3 that in (Broadcast, No CK) and (Seed, CK) people of high and low predicted ability type (knowledge about demonetization issues) report similar engagement in conversations about demonetization. This is consistent with the active sharing story, where everyone is exposed to information. However, in (Broadcast, CK), the agents of low predicted ability have a much smaller reduction in conversation volume than those of high ability. This is inconsistent with the hypothesis that the overall reduction in this treatment is driven by a reduction in active sharing to which everyone is exposed.[43]

To account for these facts, one could layer on top of the active sharing behavior a seeking decision—the frictionless seeking model from Section 4.2. Under this more elaborate alternative, while there is little active sharing in (Broadcast, CK) for the reasons discussed above, there is substantial active seeking, by low types only. This theory's explanation of the high types' not engaging is that their endowment strongly deters seeking in (Broadcast,

[43]One could hypothesize that conversations are targeted by the active sharers to reach those in need, but this would not explain this happening only in the (Broadcast, CK) treatment.

CK). The difficulty with this story is that we know that high types' informational benefits from conversation are actually considerable, based on knowledge and choice outcomes in (Broadcast, No CK). So it would have to be that the *costs* of receiving information via active sharing are much lower than of seeking information, so that high types are willing to do only the former—the high-cost hypothesis we have mentioned above. These observations identify the conditions under which this theory could work. In examining this explanation, it is worth recalling that in our data a large majority of reported conversations are secondary to some other conversation (rather than sought out mainly for discussing demonetization), so a large difference in costs based on physical or time costs seems unlikely. If the cost difference is based on some social discomfort of asking *per se* (rather than simply talking about the topic), we return to the realm of frictions in seeking.

6.1.2. *Active sharing with image concerns.* To account for heterogeneity across types, it is worth noting that many of the concerns that encourage or deter seeking may also encourage or deter speaking. For example, suppose people are judged positively for being discerning about what information is interesting to others. Discerning types share novel information, whereas boring types share redundant information. Such image concerns reproduce the behavior of the simple active sharing model above (at least for the discerning types). Moreover, the heterogeneity by ability type that we observe in the data could be driven by this variant of a signaling story: the people we predict to have high ability are discerning, and they are the ones who refrain from sharing in (Broadcast, CK), which explains why they are underrepresented among conversations there (even if they still benefit from and report hearing some active sharing).

This model is much closer to our main hypothesis, and so it could explain many of our comparisons. We thus certainly cannot reject that it plays a role. We will, however, argue that the theory has a hard time on its own accounting for the fact that (Seed, CK) and (Broadcast, No CK) have comparable volume. Consider (Seed, CK) first. The active sharing story alone suggests that we should see a similar number of conversations in (Seed, CK) and (Seed, No CK)—in both cases seeds think they may have information that others do not, and thus start diffusions of information. It also predicts that we should see many fewer conversations in either of these two than in (Broadcast, No CK)—a treatment where many more people are inclined to initiate an active sharing conversation. What we in fact see is that there are similar numbers of conversations in (Seed, CK) and (Broadcast, No CK), and many fewer conversations in either of those two than in (Seed, No CK). Thus, the seeds in (Seed, CK) trigger a surprisingly large amount of active sharing, and this explanation can work only if the seeds' desire to reach out to others is stronger in (Seed, CK) than in (Broadcast, No CK) or (Seed, No CK).[44] This is not entirely implausible—perhaps the seeds feel pride or responsibility due to being known as one of a small number of informed people and, as a result, try harder to inform people. However, as discussed in Section 5.1.1, in the data, differences in conversations associated with seeds are only a small part of the observed difference between (Seed, No CK) and (Seed, CK), so this also cannot be the entire explanation.[45]

A final observation is that we would expect active sharing to play a larger role in villages where people start out being more informed at baseline. People are better at sharing information when they are more interested in and know more about the policy. In other words, better information going in is a complement to sharing but, if anything, should reduce seeking. In Table Q.1 we split villages based on the average knowledge at baseline. We find that our main effects are much stronger in relatively *uninformed* villages. This supports an account based on demand for rather than supply of information.

6.1.3. *Seeds being more motivated to provide public goods.* A different kind of explanation focuses on the effort of those informed to understand, filter, and communicate the information in a useful way to others. Clearly, knowledge in our context is a public good. One could hypothesize that when a smaller group of people is publicly selected to provide a public good, they should provide more of it than in (Broadcast, CK), where responsibility is diffuse. This, however, is at odds with standard models of public goods; as Banerjee, Iyer, and Somanathan (2007) discuss, a fairly robust prediction of models of public goods states that while enlarging the set of people who are able to contribute to the public good often reduces per capita contributions, it should not markedly reduce aggregate provision in equilibrium. At a more basic level, theories based on intense effort by "deputized" seeds in (Seed, CK) are at odds with the fact that seeds report few extra conversations in those treatments (recall Section 5.1.1). We flesh out these points in Appendix E.2.

6.2. **Classical social learning models.** Standard social learning models and their elaborations are known to generate counterintuitive outcomes. In this subsection, we argue, nonetheless, that these mechanisms are not likely to explain our results.

A first observation is that many canonical "infection-type" models often used to study social learning, which have exogenous engagement in the learning process, share a basic monotonicity property (Bass, 1969; Bailey, 1975; Jackson, 2008; Jackson and Yariv, 2011; Aral and Walker, 2012): if more individuals are seeded with information, the number of

[44]We spell out this argument in full detail in Appendix E.1.

[45]A variant of this theory is that people share information that they are nearly sure that others do not have. This would explain why there are many conversations in (Seed, CK) which is the only case where people are sure that others do not know what they know. However, it cannot explain why there are almost as many conversations in (Broadcast, No CK).

people ultimately informed increases. In Appendix E.3, we discuss a version of this type of model that is most relevant to our setting, inspired by Möbius et al. (2015). We show that if initial endowments of information improve in the sense that they become Blackwell more informative about the state of interest, then the ultimate information of each individual also improves.

Thus, generating the kinds of reversals where adding information harms learning outcomes requires a different approach. One is looking at models where the focus is the quality of information aggregation rather than simply the extent of its diffusion. Another possibility is that an endogenous engagement margin (different from image concerns) plays a role. We consider several such models next.

6.2.1. *Herding models.* Dropping the assumption that agents transmit the original sources of all the pieces of information they convey (or at least a sufficient statistic) brings us into the world of the literature on herding or information cascades (Bikhchandani et al., 1992; Banerjee, 1992), where efficient information aggregation is no longer guaranteed. Is it possible that in this kind of setting, more information sometimes aggregates to worse outcomes? Unfortunately, characterizing the extent of information aggregation and how it depends on parameters in general herding models tends to be very difficult. However, an approach of Lobel and Sadler (2015), which applies to sequential learning in arbitrary conversation networks, can be used to argue why strong "less is more" forces such as those our main model produces are unlikely to be explained by herding models. We flesh out the details of the argument in Appendix E.4, but the basic logic is something like this: consider, for simplicity, a binary decision—say, whether or not to accept certain denominations of currency. Individuals form opinions about this. Differences in private information and in messages received lead to heterogeneity in the strengths of their beliefs about the right decision. Lobel and Sadler (2015) show that in equilibrium, most agents' decisions are at least as good as those decisions taken by those who are "experts"—very sure of the right answer based on private information (i.e. their own understanding) alone. The intuition can be seen most simply in a model where all predecessors are observed. If decisions or asserted opinions were substantially worse than the expert benchmark for arbitrarily late movers, then the well-informed would act against the prevailing view, revealing their superior information, which would persuade others. Remarkably, the same remains true even when agents observe only some of their predecessors, under certain conditions. The main substantive condition is that the network must be connected enough, with everyone having indirect access to many others. An implication of this is that, in this type of model, improving information endowments can only hurt learning if learning was quite good to start with. In other words, models of herding or information cascades will have difficulty explaining how adding information can lead to outcomes in which most people do badly enough—worse than the individual decisions of "well-informed" individuals.[46]

6.2.2. *Costs of remaining engaged in social learning.* Acemoglu et al. (2014) elaborate the basic viral model of information diffusion by positing that people have the option of dropping out of the social learning process at any point of time, due to an opportunity cost of paying attention to it. When people drop out, it reduces what others can hope to learn, and thus precipitates further exit. In that model, under broadcasting, social learning is *improved* by making it common knowledge that many agents are informed. The reason is that this increases the amount of information that any one of them can expect to receive by a given time, and strictly increases their incentives to stay engaged given a person's own level of informedness. Thus, we would predict an upward shift in equilibrium engagement.

6.3. **Some other behavioral possibilities.** There are a number of more *ad hoc* behavioral assumptions that might account for some of our findings. We briefly review a few of them in light of the evidence.

6.3.1. *Curiosity.* We have so far assumed that the only reason for people to seek information is to be able to make better decisions. A potential alternative theory is that when something out of the ordinary happens and piques their interest, they investigate just to find out what is going on. In such a world, even absent signaling concerns, people's interests may be especially piqued in (Seed, CK) because they were told that there are some others who have information; in (Broadcast, CK), there is no scope for such curiosity.

This argument by itself says nothing about why there is much more seeking in (Broadcast, No CK) than in (Broadcast, CK). To explain that, we could add the assumption that those who are informed in (Broadcast, No CK) are trying to find out if others are informed and in the process get into conversations that end up being informative. This still leaves unexplained the fact that in (Broadcast, CK) the high types seek less than low types, and that this is the only treatment where this is the case.

6.3.2. *Mistaken perceptions and overconfidence.* To explain the lack of conversations in (Broadcast, CK)—in the data it is comparable to (Seed, No CK)—one could posit that participants mistakenly believed they understood the facts they were told (although, in fact, they had much room to understand them better). This runs counter to several different pieces of evidence. First, the direct evidence from the knowledge surveys, in which many participants in (Broadcast, CK) admitted ignorance even to us (Panel B of Table 2). This evidence shows

[46]Of course, one could simply posit that common knowledge of people being informed causes people to drop out of the learning process (thinning out the learning network enough that it hurts diffusion) but this simply begs the question of engagement incentives.

that substantial scope for learning remained and people knew that. Second, and more fundamentally, such a theory does not predict less seeking in (Broadcast, CK) than in (Broadcast, No CK), which is what we observe. Indeed, insofar as subjects overcome overconfidence and ask others, the fact they know that others are informed should make it more, rather than less, appealing to ask them for clarifications.

6.3.3. *Inferring the value of the information from the treatment.* Finally, we consider the possibility that agents value the information differently across treatments. One specific story that could match many of our key predictions is that agents thought that the information was of less value when it was distributed to more individuals. In this argument, they might not have even looked at the pamphlets in (Broadcast, CK), throwing them away as "spam." We do not, however, find this very compelling in our context. It goes against the fact that making public announcements—a small 3-wheeler driving around the village with a loud-speaker attached to its top blaring out the message—is the most common way to get information to people in rural India about a possible tornado or other natural disasters.[47] This is (Broadcast, CK) in our language, and people clearly do not assume all such messages are spam. Most people in our baseline also clearly wanted information and felt that neither they nor their neighbors knew enough about the post-demonetization rules (Table 1).

6.4. **Taking stock.** We have presented a number of alternative frameworks in this section. In each case, we have argued that the alternative explanation is either incomplete (and requires additional ad hoc assumptions to fit all the facts) or inconsistent with what we know about the particular context. By contrast, the signaling model provides a fairly simple and unified account of all the rankings. Nevertheless, given the simplicity of our treatment, there may well be alternative behavioral mechanisms or combinations of those we mention above that could rationalize our findings. While we believe that signaling is an important component of our results, our main finding is there is a friction in seeking that varies not only with what people know but also what they think others know. A definitive decomposition of this friction into its ultimate constituents is beyond our scope and remains an important issue for further studies.

7. Conclusion

Social learning happens in part through choices by about whether to ask questions. We show that, consistent with a model of endogenous social learning motivated by perception payoffs, the number of signals and the structure of common knowledge matter considerably for the extent of participation in social learning. In particular, we find evidence for a set of

[47]We also show evidence in Appendix Table H.2 that individuals in (Broadcast, CK) did learn the facts *from their pamphlets*, but learned nothing more.

clear rankings of policies consistent with image concerns deterring seeking, but not with a frictionless model of engagement in social learning. When looking at targeted seeding, going from no common knowledge to common knowledge increases conversations, but the opposite is true for broadcast strategies. Moreover, conversations actually decline when, holding common knowledge fixed, more people are provided with information. This increase or decline in conversation volume is accompanied by a corresponding increase or decline in knowledge about the rules as well as quality of choice. Thus, the success of an information intervention depends crucially on how the design and how it affects endogenous communication.

Our model of signaling concerns provides a mechanism that can explain both why the "more is more" logic holds when it does, as well as the reversals in the data. The forces in the model are consistent with villagers' reports of their experiences in the context of the Indian demonetization.

Of the full set of experimental interventions, two consistently perform well along all the dimensions—conversations, knowledge, and choice—and have comparable benefits to one another: seeding with common knowledge and broadcast without common knowledge. While seeding with common knowledge is straightforward to implement, whether broadcast without common knowledge is a feasible policy depends considerably on context. For example, posters in the village, loud-speakers on three-wheelers, radio, television, newspapers, or the village crier intrinsically contain a common knowledge component. Moreover, it may be difficult for the same entity (e.g., a local government or agricultural extension service) to carry out a non-common knowledge broadcast strategy without it eventually becoming common knowledge. At the same time, one can imagine contexts where flyers or SMS messages are a natural mode of communication, and in those cases it may take a while to become common knowledge.[48] Our results suggest that broadcasting without common knowledge can be quite effective when possible, but implementations may be blunted in their effectiveness if they become closer to broadcasting with common knowledge in terms of how they are perceived. Seeding benefits from common knowledge unambiguously.

Our results are most directly relevant for other settings where individuals need to act on the information somewhat quickly. Often policy information is disseminated in advance of discrete decision points—information about new agricultural technologies disseminated immediately prior to annual planting decisions, information about specific job vacancies that may be filled quickly, or one-off training programs. Our results would also be relevant in settings where information stimulates bursts of conversation in the short run, but then individuals get tired of talking about a topic and move on.

[48]This is perhaps most likely when the message is sent to everyone within a sub-group (for example, health messages—say, about getting tested for diabetes or getting a flu shot—may be targeted to specific age groups or risk categories) and people may not know the boundaries of the group.

In either case, the results have implications for how researchers and policymakers should think about the use of broadcast media versus extension to educate individuals, and how extension should be structured. Indeed, in any setting where contacting all households is feasible, the policymaker can do just as well as generating common knowledge and informing a few seeds as, say, providing messages to all individuals (even without common knowledge). These lessons play out in related work. Banerjee et al. (2021) provides a policy-relevant example where seeding with common knowledge has been successful.[49]

Exploring other tensions between seeding and broadcasting is a promising avenue for future work. Broadcast strategies are inherently more democratic than seeding and may have different implications for information inequality. Moreover, in equilibrium, repeatedly calling upon the same individuals to act as seeds could concentrate power and change the social dynamic. These types of impacts may also affect the ultimate success of seeding strategies.

Finally, policymakers may be able to avoid the types of learning frictions we document by carefully curating the information shared in their campaigns. It is possible that in some applications, simplifying the information to the easy-to-process essential facts could remove the need for network-based aggregation and neutralize the effects of endogenous social engagement frictions. Our results, however, show that merely making messages brief may backfire. The careful curation needed to make such a strategy successful may be costly and time-consuming if communities have heterogeneous needs or if policy implementation varies across locations.

References

Abaluck, J. and J. Gruber (2011): "Choice Inconsistencies Among the Elderly: Evidence from Plan Choice in the Medicare Part D Program," *American Economic Review*, 101, 1180–1210. 3.1

Acemoglu, D., K. Bimpikis, and A. Ozdaglar (2014): "Dynamics of Information Exchange in Endogenous Social Networks," *Theoretical Economics*, 9, 41–97. 6.2.2

Aral, S. and D. Walker (2012): "Identifying Influential and Susceptible Members of Social Networks," *Science*, 337, 337–341. 6.2

Bailey, N. (1975): *The Mathematical Theory of Infectious Diseases and Its Applications.*, Griffin: London. 6.2

[49]In a cross-randomized experiment with 75 policies covering 40% of the state, Banerjee et al. (2021) examine information policies to encourage child immunization. They study whether various incentives, reminders, and seeding policies (none, "information ambassadors," etc.) affect immunizaiton rates. The finding is that the most cost-effective policy is a combination of using seeds picked through the same "gossip survey" as described in this paper, together with light SMS reminders.

Banerjee, A., A. G. Chandrasekhar, S. Dalpath, E. Duflo, J. Floretta, M. O. Jackson, H. Kannan, F. Loza, A. Sankar, A. Schrimpf, et al. (2021): "Selecting the Most Effective Nudge: Evidence from a Large-Scale Experiment on Immunization," *arXiv preprint arXiv:2104.09645*. 7, 49

Banerjee, A., A. G. Chandrasekhar, E. Duflo, and M. O. Jackson (2013): "The Diffusion of Microfinance," *Science*, 341, 1236498. 1

——— (2016): "Gossip: Identifying Central Individuals in a Social Network," *Working Paper*. 1, 14

Banerjee, A., L. Iyer, and R. Somanathan (2007): "Public Action for Public Goods," *Handbook of Development Economics*, 4, 3117–3154. 6.1.3, E.2

Banerjee, A. V. (1992): "A Simple Model of Herd Behavior," *The Quarterly Journal of Economics*, 107, 797–817. 6.2.1

Banerjee, A. V. and E. S. Maskin (1996): "A Walrasian Theory of Money and Barter," *The Quarterly Journal of Economics*, 111, 955–1005. H

Bass, F. M. (1969): "A New Product Growth for Model Consumer Durables," *Management Science*, 15, 215–227. 6.2

Beaman, L., A. BenYishay, J. Magruder, and A. M. Mobarak (2021): "Can Network Theory-Based Targeting Increase Technology Adoption?" *American Economic Review*, 111, 1918–43. 1

Belloni, A., V. Chernozhukov, and C. Hansen (2014): "Inference on Treatment Effects After Selection Among High-Dimensional Controls," *The Review of Economic Studies*, 81, 608–650. 5.1.1

Beshears, J., J. J. Choi, D. Laibson, and B. C. Madrian (2013): "Simplification and Saving," *Journal of Economic Behavior & Organization*, 95, 130–145. 3.1

Bikhchandani, S., D. Hirshleifer, and I. Welch (1992): "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades," *Journal of Political Economy*, 100, 992–1026. 6.2.1

Breza, E., S. Kaur, and N. Krishnaswamy (2020): "Propping Up the Wage Floor: Collective Labor Supply without Unions," *Working Paper*. 17

Breza, E., S. Kaur, and Y. Shamdasani (2017): "The Morale Effects of Pay Inequality," *The Quarterly Journal of Economics*, 133, 611–663. 17

——— (2021): "Labor Rationing," *American Economic Review*, 111, 3184–3224. 17

Bursztyn, L., G. Egorov, and R. Jensen (2019): "Cool To be Smart or Smart to be Cool? Understanding Peer Pressure in Education," *The Review of Economic Studies*, 86, 1487–1526. 2

Bursztyn, L. and R. Jensen (2017): "Social image and economic behavior in the field: Identifying, understanding, and shaping social pressure," *Annual Review of Economics*, 9, 131–153. 2

Butera, L., R. Metcalfe, W. Morrison, and D. Taubinsky (2019): "The Deadweight Loss of Social Recognition," Tech. rep., National Bureau of Economic Research. 2

Cai, J., A. De Janvry, and E. Sadoulet (2015): "Social Networks and the Decision to Insure," *American Economic Journal: Applied Economics*, 7, 81–108. 1

Carvalho, L. and D. Silverman (2017): "Complexity and Sophistication," . 3.1

Chandrasekhar, A. G., B. Golub, and H. Yang (2018): "Signaling, Shame, and Silence in Social Learning," *Working paper, Stanford University*. 2, 4, 2.4, 53, 57, E.2

Chodorow-Reich, G., G. Gopinath, P. Mishra, and A. Narayanan (2020): "Cash and the Economy: Evidence from India's Demonetization," *The Quarterly Journal of Economics*, 135, 57–103. 8

Conley, T. G. and C. R. Udry (2010): "Learning about a New Technology: Pineapple in Ghana," *The American Economic Review*, 35–69. 1

de Hooge, I. E., M. Zeelenberg, and S. M. Breugelmans (2010): "Restore and Protect Motivations Following Shame," *Cognition and Emotion*, 24, 111–127. 4

Galeotti, A. and S. Goyal (2010): "The Law of the Few," *American Economic Review*, 100, 1468–1492. E.2

Gerber, A. S., D. P. Green, and C. W. Larimer (2008): "Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment," *American Political Science Review*, 102, 33–48. 2

Goldfarb, A., R. C. McDevitt, S. Samila, and B. S. Silverman (2015): "The effect of social interaction on economic transactions: Evidence from changes in two retail formats," *Management Science*, 61, 2963–2981. 2

Gruenewald, T., S. Dickerson, and M. Kemeny (2007): "A Social Function for the Self-Conscious Emotions: Social-Self Preservation Theory," in *The Self-Conscious Emotions: Theory and Research*, Guilford Press, 68–87. 4

Jackson, M. O. (2008): *Social and Economic Networks.*, Princeton University Press. 6.2

Jackson, M. O. and L. Yariv (2011): "Diffusion, Strategic Interaction, and Social Structure," in *Handbook of Social Economics*, Elsevier, vol. 1, 645–678. 6.2

Karing, A. (2018): "Social Signaling and Childhood Immunization: A Field Experiment in Sierra Leone," *Working Paper*. 2

Kaur, S., S. Mullainathan, S. Oh, and F. Schilbach (2019): "Does Financial Strain Lower Worker Productivity?" *Working Paper*. 17

Kiyotaki, N. and R. Wright (1989): "On Money as a Medium of Exchange," *Journal of Political Economy*, 97, 927–954. H

Leskovec, J., L. A. Adamic, and B. A. Huberman (2007): “The Dynamics of Viral Marketing,” *ACM Transactions on the Web (TWEB)*, 1, 5. 1

Lobel, I. and E. Sadler (2015): “Information Diffusion in Networks Through Social Searning,” *Theoretical Economics*, 10, 807–851. 6.2.1, E.4

Miller, G. and A. M. Mobarak (2014): “Learning About New Technologies Through Social Networks: Experimental Evidence on Nontraditional Stoves in Bangladesh,” *Marketing Science*, 34, 480–499. 1

Möbius, M., T. Phan, and A. Szeidl (2015): “Treasure Hunt: Social Learning in the Field,” NBER Working Paper No. 21014. 6.2

Niehaus, P. (2011): “Filtered Social Learning,” *Journal of Political Economy*, 119, 686–720. 7

Pennington, C. R., D. Heim, A. R. Levy, and D. T. Larkin (2016): “Twenty Years of Stereotype Threat Research: A Review of Psychological Mediators,” *PloS one*, 11, e0146487. 4

Perez-Truglia, R. and U. Troiano (2018): “Shaming Tax Delinquents,” *Journal of Public Economics*, 167, 120–137. 2

Ryan, B. and N. C. Gross (1943): “The Diffusion of Hybrid Seed Corn in Two Iowa Communities.” *Rural Sociology*, 8, 15. 1

Samuelson, P. A. (1958): “An Exact Consumption-Loan Model of Interest With or Without the Social Contrivance of Money,” *Journal of Political Economy*, 66, 467–482. H

Simonsohn, U. (2021): “[99] Hyping Fisher: The Most Cited 2019 QJE Paper Relied on an Outdated STATA Default to Conclude Regression p-values Are Inadequate,” . 20

Spencer, S. J., C. Logel, and P. G. Davies (2016): “Stereotype Threat,” *Annual Review of Psychology*, 67, 415–437, pMID: 26361054. 4

Steele, C. and J. Aronson (1995): “Stereotype Threat and The Intellectual Test-Performance of African-Americans,” *Journal of Personality and Social Psychology*, 69, 797–811. 4

Wallace, N. (1980): “The Overlapping Generations Model of Fiat Money, in Models of Monetary Economies (J. Kareken and N. Wallace, eds.), 49(82),” *Federal Reserve Bank of Minneapolis*. H

Figures

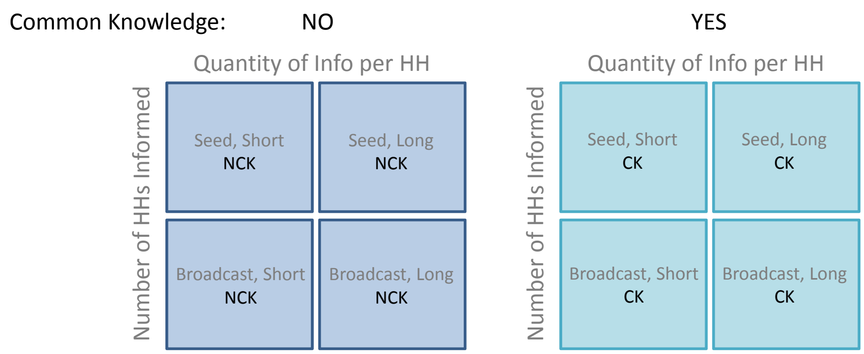


Figure 1. Experimental Design

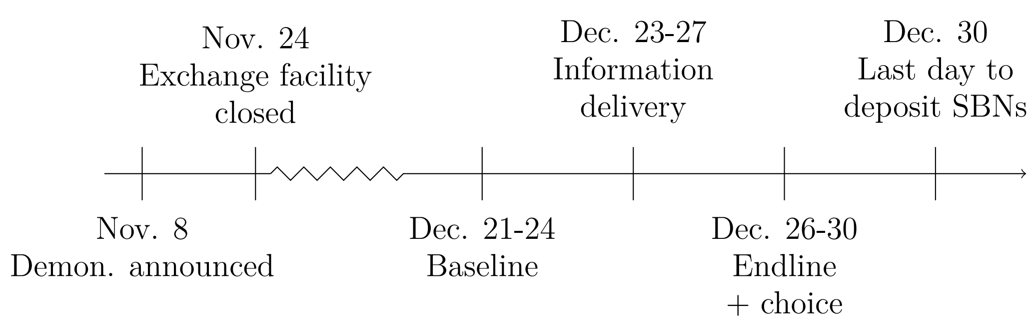


Figure 2. Intervention Timeline

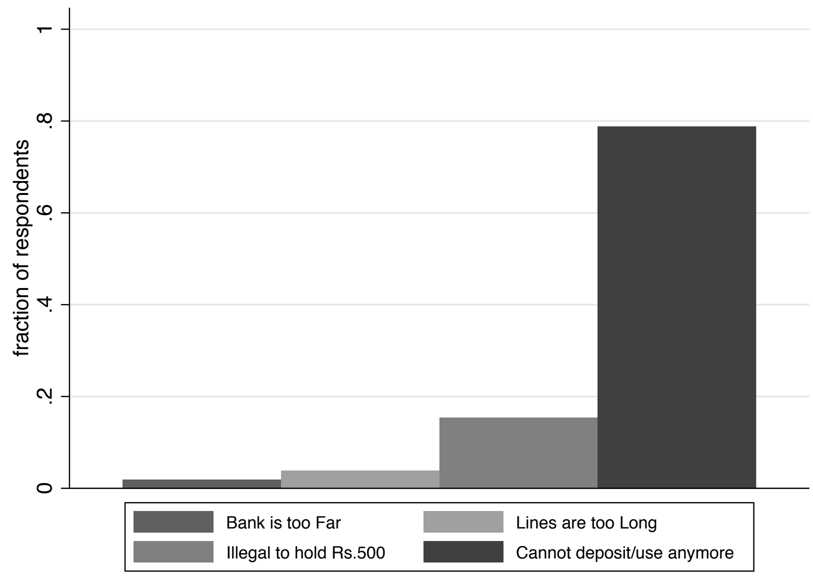
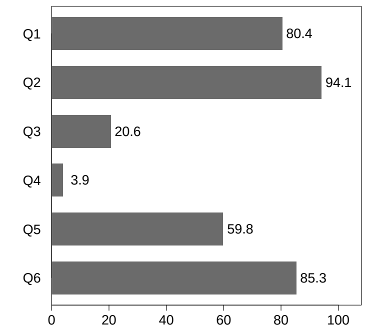


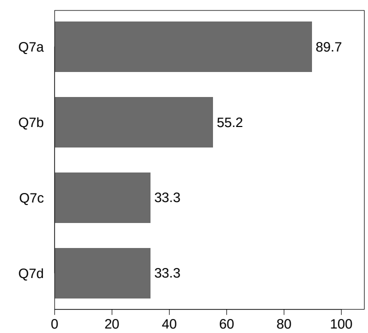
Figure 3. Why did you not choose 500?

1. Were you confused about any of of these changes?
2. Do you think anyone in village your were confused about these changes?
3. Do you think you completely understood how much to exchange, where to exchange, how to exchange, till when you could exchange, application process etc.?
4. Do you think everyone in your village completely understood how much to exchange, where to exchange, how to exchange, till when you could exchange, application process etc.?
5. After the policy was introduced, did you ever hesitate to ask someone from your village about the note-ban policy because you were concerned about what they might think about you?
6. After the policy was introduced, did you ever hesitate to ask **an acquaintance** from your village about the note-ban policy because you were concerned about what they might think about you?



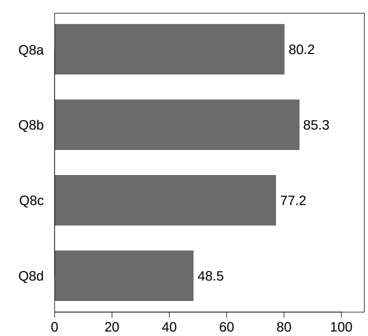
Subfigure A

7. If yes, did you hesitate because you were concerned they would think you are:
   (a) dumb?
   (b) irresponsible?
   (c) lazy?
   (d) dealing in black money?



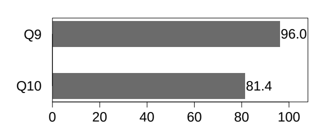
Subfigure B

8. If someone from your vil1age asks about the note-ban policy in December after it was heavily broadcasted on TV, do you think people would think he is:
   (a) dumb for not understanding even after being broadcast?
   (b) irresponsible for not checking earlier?
   (c) lazy?
   (d) dealing in black money?



Subfigure C

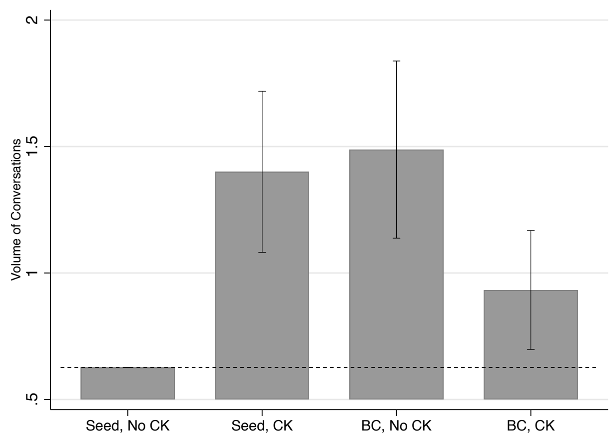
9. In December, since the news about the note-ban policy was being heavily broadcasted on TV, do you think it was the responsibility of people in your village to know everything/completely about the note-ban policy?
10. In December after being heavily broadcasted on TV, do you think some people in your village reduced asking about the note-ban policy even though they were confused because they were scared/worried that they would be judged as dumb/lazy/irresponsible/or dealing in black money?



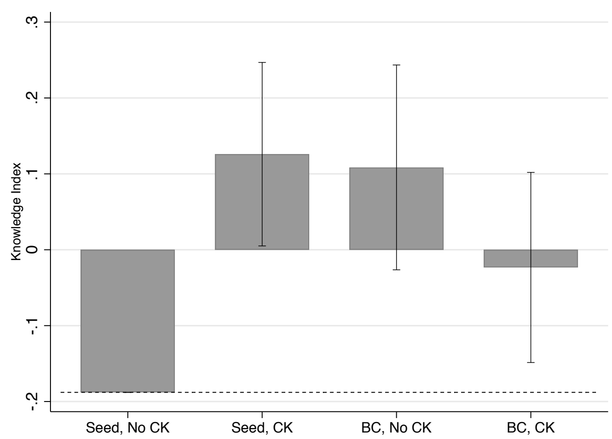
Subfigure D

Note : The sample consists of 102 randomly sampled respondents across 4 villages in Karnataka.

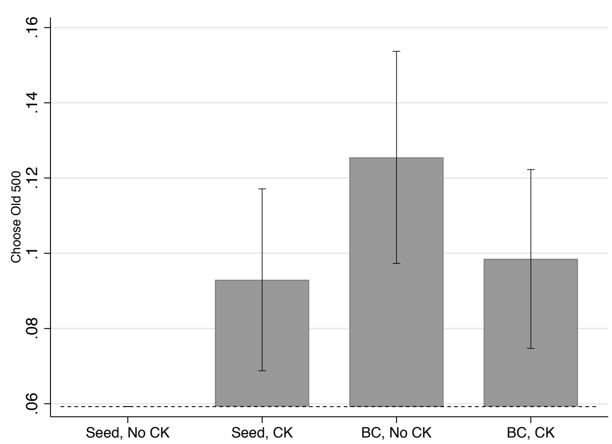
FIGURE 4. Survey Results: Percent of respondents answering "Yes" to each question/subquestion



(A) Volume of conversations



(B) Knowledge error



(C) Chose old 500

FIGURE 5. Raw Data: Core Experimental Outcomes

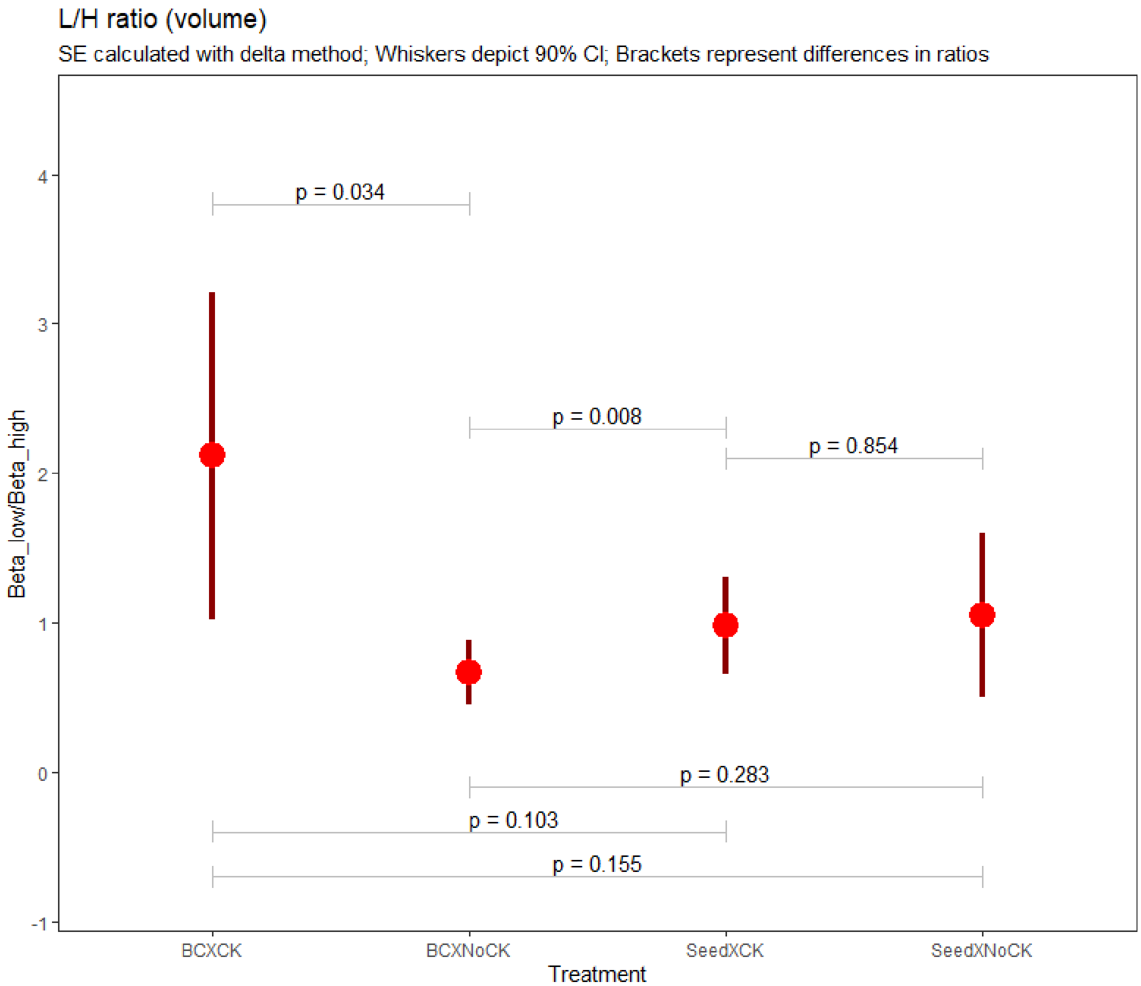


FIGURE 6. Treatment Effect Ratios of Low Ability to High Ability

This figure illustrates relative seeking rates of Low versus High ability agents for each treatment group, with ratios constructed using coefficients from regressing an indicator for having any conversation on an indicator for high ability, treatment indicators, and their interactions, controlling only for subdistrict fixed effects (see F.2 for regression results).

Tables

Table 1. Summary Statistics

	mean	sd	obs
Female	0.32	(0.47)	1082
SC/ST	0.50	(0.50)	1082
Age	39.18	(11.88)	1079
Casual laborer	0.21	(0.41)	1082
Farmer: landed	0.16	(0.37)	1082
Domestic work	0.16	(0.37)	1082
Farmer: sharecropper	0.09	(0.29)	1082
Unemployed	0.02	(0.14)	1082
Bank account holder	0.89	(0.31)	1078
Literate	0.80	(0.40)	1047

Notes: This table gives summary statistics on the endline sample used for analysis.

Table 2. Baseline Error Statistics

Panel A: Error rates

	mean	sd	obs
10 rupees coin	0.15	(0.36)	965
General currency	0.17	(0.38)	965
Over-the-counter exchange	0.25	(0.44)	965
Exchange locations other than banks	0.50	(0.50)	966
Weekly withdrawal limits from bank accounts	0.78	(0.41)	965
Withdrawal limits on Jan Dhan accounts	0.87	(0.33)	965
Daily withdrawal limits on ATMs	0.90	(0.30)	965

Panel B: Incidence of "don't know" responses

	mean	sd	obs
General currency	0.01	(0.11)	966
Exchange locations other than banks	0.30	(0.46)	966
Weekly withdrawal limits from bank accounts	0.33	(0.47)	966
Withdrawal limits on Jan Dhan accounts	0.78	(0.41)	966
Daily withdrawal limits on ATMs	0.32	(0.47)	966

Notes: Panel A gives error rates on knowledge about demonetization in the baseline sample. Panel B gives the incidence of "don't know" responses for the relevant questions. All respondents giving a "don't know" response were asked to make their best guess of the response.

Table 3. Baseline covariate balance

	Means				Pairwise Differences p-values					
	(1) Seed, No CK	(2) Seed, CK	(3) Broadcast, No CK	(4) Broadcast, CK	(5) SNCK - SCK	(6) SNCK - BCNK	(7) SNCK - BCK	(8) SCK - BNCK	(9) SCK - BCK	(10) BNCK - BCK
Beyond 40kms of urban center	.14	.21	.1	.22	.39	.53	.35	.13	.93	.11
Within 5kms of urban center	.31	.4	.35	.31	.41	.73	1	.63	.39	.72
Standardized entry time	-.12	.1	.02	-.21	.23	.49	.65	.71	.13	.3
Survey date	3.55	3.64	3.7	3.76	.54	.26	.12	.64	.36	.63
New strata	.09	.07	.05	0	.83	.53	.05	.67	.05	.09
Female	.32	.25	.33	.39	.25	.91	.29	.17	.02	.29
Literate	.8	.8	.82	.78	.89	.75	.6	.66	.74	.41
Bank account holder	.91	.86	.85	.93	.27	.1	.56	.9	.16	.04
Age	40.01	40.06	38.27	38.24	.97	.12	.15	.14	.16	.98

Notes: Table compares (Seed, No CK), (Seed, CK), (Broadcast, No CK), and (Broadcast, CK) across whether the village is very rural, peri-urban, time of entry for endline survey, date of entry, whether the village was reassigned, gender of subject, literacy of subject, whether the subject has a bank account, and age of subject. Columns 1-4 present means by covariate in the four treatment cells aforementioned, in that order. Columns 5-10 present p-values of pairwise comparisons of differences in means across cells.

Table 4. Bank Summary Statistics

	median	mean	sd	obs
Actual wait time at banks (mins)	10.00	11.86	(7.87)	51
Perceived wait time at banks (mins)	15.00	17.06	(22.13)	32
Nearest Bank (mins)	20.00	19.84	(9.88)	63

Notes: This table gives actual wait time at banks near our sample villages. On the last day on which SBNs were accepted, we surveyed as many banks as possible near the study villages. Our enumerators made it to 51 banks, where employees were surveyed. It also gives perceived wait time and perceived time taken to reach the nearest bank by a sub-sample of the endline respondents.

Table 5. Engagement in social learning, knowledge and decision-making

VARIABLES	(1) Volume of conversations	(2) # secondary conversations	(3) # primary conversations	(4) Knowledge	(5) Chose 500
CK	0.644	0.437	0.207	0.0313	0.0482
	(0.310)	(0.255)	(0.103)	(0.0126)	(0.0223)
	[0.0378]	[0.0871]	[0.0441]	[0.0129]	[0.0304]
Broadcast	0.709	0.521	0.188	0.0280	0.0676
	(0.349)	(0.316)	(0.124)	(0.0140)	(0.0266)
	[0.0422]	[0.0986]	[0.130]	[0.0461]	[0.0109]
Broadcast × CK	-1.493	-1.113	-0.380	-0.0505	-0.109
	(0.520)	(0.435)	(0.186)	(0.0189)	(0.0386)
	[0.00411]	[0.0105]	[0.0412]	[0.00764]	[0.00478]
Observations	1,078	1,078	1,078	1,082	1,067
Number of groups	0	0	0	0	0
Seed, No CK Mean	0.627	0.490	0.137	0.566	0.0592
CK + BC × CK = 0 p-val	0.0172	0.0262	0.239	0.154	0.0385
BC + BC × CK = 0 p-val	0.0254	0.0352	0.110	0.0572	0.104
CK = BC p-val	0.861	0.790	0.878	0.789	0.489

Notes: All columns control for randomization strata (subdistrict) fixed effects. Other controls for each column selected with PDS Lasso from date and time of entry into the village, caste category of the treatment hamlet, distance from the village to an urban center, and respondent-level controls such as age, gender, literacy and potential seed status. Standard errors (clustered at the village-level) are reported in parentheses and p-values are reported in brackets.

Table 6. Heterogeneity by length of information

VARIABLES	(1) Volume of conversations	(2) # Secondary conversations	(3) # Primary conversations	(4) Knowledge	(5) Chose 500
CK	0.811	0.680	0.131	0.0208	0.0551
	(0.481)	(0.397)	(0.160)	(0.0158)	(0.0393)
	[0.0918]	[0.0865]	[0.414]	[0.186]	[0.161]
Broadcast	0.973	0.667	0.306	0.0266	0.0789
	(0.534)	(0.474)	(0.213)	(0.0167)	(0.0350)
	[0.0687]	[0.160]	[0.152]	[0.110]	[0.0243]
Long	-0.0795	0.00630	-0.0858	-0.0127	-0.00757
	(0.361)	(0.320)	(0.127)	(0.0171)	(0.0292)
	[0.826]	[0.984]	[0.498]	[0.458]	[0.795]
Broadcast × CK	-2.207	-1.608	-0.599	-0.0534	-0.144
	(0.719)	(0.613)	(0.258)	(0.0242)	(0.0545)
	[0.00215]	[0.00874]	[0.0202]	[0.0275]	[0.00833]
CK × Long	-0.357	-0.479	0.122	0.0171	-0.0163
	(0.550)	(0.467)	(0.191)	(0.0249)	(0.0499)
	[0.515]	[0.304]	[0.522]	[0.490]	[0.744]
Broadcast × Long	-0.575	-0.317	-0.258	-0.000731	-0.0261
	(0.665)	(0.603)	(0.227)	(0.0256)	(0.0534)
	[0.387]	[0.599]	[0.256]	[0.977]	[0.625]
Broadcast × CK × Long	1.428	0.991	0.437	0.00785	0.0713
	(0.790)	(0.716)	(0.276)	(0.0373)	(0.0769)
	[0.0706]	[0.167]	[0.114]	[0.833]	[0.354]
Observations	1,078	1,078	1,078	1,082	1,067
Seed, No CK, Short Mean	0.523	0.385	0.138	0.564	0.0374
CK + BC × CK = 0 p-val	0.00428	0.0233	0.0305	0.0803	0.0121
BC + BC × CK = 0 p-val	0.0138	0.0205	0.0589	0.118	0.117
CK = BC p-val	0.787	0.978	0.463	0.735	0.543
CK × Long + BC × CK × Long=0 p-val	0.0783	0.339	0.0190	0.347	0.311
CK × Long = BC × Long p-val	0.776	0.800	0.170	0.468	0.858
CK + BC × CK + CK × Long + BC × CK × Long=0 p-val	0.448	0.283	0.537	0.695	0.451
BC + BC × CK + BC × Long + BC × CK × Long=0 p-val	0.330	0.408	0.475	0.254	0.503

Notes: All columns control for randomization strata (subdistrict) fixed effects. Other controls for each column selected with PDS Lasso from date and time of entry into the village, caste category of the treatment hamlet, distance from the village to an urban center, and respondent-level controls such as age, gender, literacy and potential seed status. Standard errors (clustered at the village-level) are reported in parentheses and p-values are reported in brackets.

Appendix A. Timeline of Rule Changes

Nov-08	Rs. 500 and Rs. 1000 notes shall have their legal tender withdrawn wef midnight Nov 8
	Closure of ATMs from Nov 9th to Nov 11th
	All ATM free of cost of dispensation
	ATM machine withdrawal limit: Rs. 2000 per day per card (till Nov. 18th); Rs. 4000 thereafter
Nov-09	Re-Calibration of ATMs to dispense Rs. 50 and Rs. 100 notes
	Withdrawal of Rs. 2000 limit per day per card
	Cash withdrawals could be made from Banking Correspondents and Aadhar Enabled Payment Systems
Nov-10	Rs. 4000 or below could be exchanged for any denomination at banks
	Max deposit for an account without KYC: Rs. 40000
	Cash withdrawal per day: Rs. 10,000; with a limit of Rs. 20,000 in one week
Nov-13	Limit for over the counter withdrawal: Rs. 4500
	Daily withdrawal on debit cards: Rs. 2500
	Weekly withdrawal limit: Rs. 24,000
	Daily limit of Rs. 10,000: withdrawn
	Separate queues for senior citizens and disabled
Nov-14	Waivers of ATM customer charge
	Current account holders: Withdrawal limits Rs. 50,000 with notes of mostly Rs. 2000
Nov-17	Over the counter exchange of notes limited to Rs. 2000
	PAN card is mandatory for deposits over Rs. 50,000, or opening a bank account
Nov-20	Withdrawal of ATM: limit unchanged at Rs. 2500
Nov-21	Cash withdrawal for wedding: Rs. 2,50,000 for each party for wedding before Dec. 30th, for customers with full KYC
	60 day extra for small borrowers to repay loan dues
	Limit of Rs. 50,000 withdrawal also extended to overdraft, cash credit account (in addition of current account - Nov-14)
	Farmers can purchase seeds with the old Rs. 500 notes
Nov-22	Prepaid payment instruments: limit extended from Rs. 10,000 to Rs. 20,000 in order to push electronic payment systems
	For wedding payments: a list must be provided with details of payments for anyone to whom a payment of more that 10,000 is to be made for wedding purposes
Nov-23	SBNs not allowed to deposit money in Small Saving Schemes
Nov-24	No over the counter exchange of SBNs wef midnight Nov-24
	Only the old Rs. 500 notes will be accepted till Dec. 15th in the following places: government school or college fees, pre-paid mobiles, consumer co-op stores, tolls for highways
Nov-25	Weekly withdrawal limit: Rs. 24,000 (unchanged)
	Foreign citizens allowed to exchange Rs. 5000 per week till Dec 15th

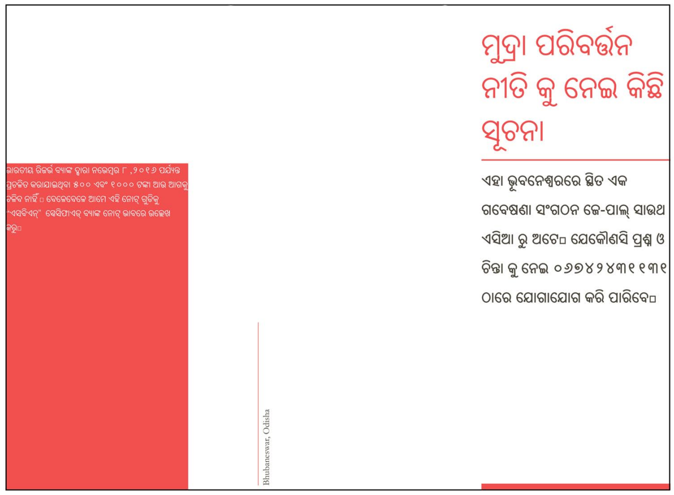
Nov-28	Relaxation in norms of withdrawal from deposit accounts of deposits made in legal tender note wef Nov-29
Nov-29	For account holders of Pradhan Mantri Jan Dhan Yojana: limit of Rs. 10,000 withdrawal per month for full KYC customers; Rs. 5000 with customers with partial KYC
Dec-02	Aadhaar-based Authentication for Card Present Transactions
Dec-06	Relaxation in Additional Factor of Authentication for payments upto Rs. 2000 for card network provided authentication solutions
Dec-07	Old Rs. 500 notes can only be used for purchase of railway tickets till Dec. 10th
Dec-08	OTP based e-KYC allowed
Dec-16	Pradhan Mantri Garib Kalyan Deposit Scheme Issued wef Dec 17
	Foreign citizens allowed to exchange Rs. 5000 per week till Dec 31st
	Merchant discount rate for debit card transactions revised
	No customer charges to be levied for IIMPS, UPI, USSD
Dec-19	SBNs of more than Rs. 5000 to be accepted only once till Dec 30th to full KYC customers
Dec-21	The limit of Rs. 5000 deposit not applicable to full KYC customers
Dec-26	60 day extra for short term crop loans
Dec-29	Additional working capital for MSEs
Dec-30	Closure of the scheme of exchange of Specified Bank Notes
	PPI guideline (issued Nov 22) extended
	ATM machine withdrawal limit: Rs. 4500 per day per card
Dec-31	Grace period for non-present Indians for SBN exchange at RBI
Jan-03	Allocation changes to cash in rural areas
	Foreign citizens allowed to exchange Rs. 5000 per week till Jan 31
Jan-16	ATM limit extended to Rs. 10,000 per day per card
	Current account withdrawal limits extended to 1,00,000

Appendix B. List of Facts

Chapter 1: DEPOSITING OR TENDERING SPECIFIED BANK NOTES	1. The old Rs. 500 and Rs.1000 notes will be accepted at bank branches until 30/12/2016. If you deposit more than Rs. 5,000 then you will have to provide a rationale for why you didnt deposit the notes earlier. 2. You will get value for the entire volume of notes tendered at the bank branches / RBI offices. 3. If you are not able to personally visit the branch, you may send a representative with a written authority letter and his/her identity proof with tendering the notes. 4. Banks will not be accepting the old Rs.500 and Rs. 1000 notes for deposits in Small Saving Schemes. The deposits canbe made in Post Office Savings accounts. 5. Quoting of PAN is mandatory in the following transactions: Deposit with a bank in cash exceeding Rs. 50,000 in a single day; Purchase of bank drafts or pay orders or bankers cheques from a bank in cash for an amount exceeding Rs. 50,000 in a single day; A time deposit with a Bank or a Post Office; Total cash deposit of more than Rs. 2,50,000 during November 09 to December 30th, 2016
Chapter 2: EXCHANGING SPECIFIED BANK NOTES	1. The over the counter exchange facility has been discontinued from the midnight of 24th November, 2016 at all banks. This means that the bank wont exchange the notes for you anymore. You must first deposit them into an account. 2. All of the old Rs.500 and Rs. 1,000 notes can be exchanged at RBI Offices only, up to Rs.2000 per person. 3. Until December 15th, 2016, foreign citizens will be allowed to exchange up to Rs. 5000 per week. It is mandatory for them to have this transaction entered in their passports. 4. Separate queues will be arrangedfor Senior Citizens and Divyang persons, customers with accounts in the Bankand for customers for exchange of notes (when applicable).
Chapter 3: CASH WITHDRAWAL AT BANK BRANCHES	1. The weekly limit of Rs. 20,000 for withdrawal from Bank accounts has been increased to Rs. 24,000. The limit of Rs. 10,000 per day has been removed. 2. RBI has issued a notification to allow withdrawals of deposits made in the valid notes (including the new notes) on or after November 29, 2016 beyond the current limits. The notification states that available higher denominations bank notes of Rs. 2000 and Rs. 500 are to be issued for such withdrawals as far as possible. 3. Business entities having Current Accounts which are operational for last three months or more will be allowed to draw Rs. 50,000 per week. This can be done in a single transaction or multiple transactions. 4. To protect innocent farmers and rural account holders of PMJDY from money launders, temporarily banks will: (1) allow account holders with full KYC to withdraw Rs. 10,000 in a month;(2) allow account holders with limited KYC to withdraw Rs.5,000 per month, withthe maximum of Rs.10,000 from the amount deposited through SBN after Nov 09,2016 5. District Central Cooperative Banks (DCCBs) will also facilitate withdrawals with the same limits as normal banks.
Chapter 4: ATM WITHDRAWALS	1. Withdrawal limit increased to Rs. 2,500 per day for ATMs that have been recalibrated to fit the new bills. This will enable dispensing of lower denomination currency notes for about Rs.500 per withdrawal. The new Rs. 500 notes can be withdrawn 2. Micro ATMs will be deployed to dispense cash against Debit/Credit cards up to the cash limits applicable for ATMs. 3. ATMs which are yet to berecalibrated, will continue to dispense Rs. 2000 till they are recalibrated.
Chapter 5: SPECIAL PROVISIONS FOR FARMERS	1. Farmers would be permitted to withdraw up to Rs. 25,000 per week in cash from their KYC compliant accounts for loans. These cash withdrawals would be subject to the normal loan limits and conditions. This facility will also apply to the Kisan Credit Cards (KCC). 2. Farmers receiving payments into their bank accounts through cheque or other electronic means for selling their produce, will be permitted to withdraw up to Rs.25,000 per week in cash. But these accounts will have to be KYC compliant. 3. Farmers can purchase seeds with the old bank notes of 500 from the State or Central Govenment Outlets, Public Sector Undertakings, National or State Seeds Corporations, Central or State Agricultural Universities and the Indian Council of Agricultural Research (ICAR), with ID proof.

	4. Traders registered with APMC markets/mandis will be permitted to withdraw up to Rs. 50,000 per week in cashfrom their KYC compliant accounts as in the case of business entities. 5. The last date for payment of crop insurance premium has been extended by 15 days to 31st December,2016.
Chapter 6: SPECIAL PROVISIONS FOR WEDDINGS	1. In the case of a wedding, one individual from the family (parent or the person themselves) will be able to withdraw Rs. 2,50,000 from a KYC compliant bank account. PAN details and self-declaration will have to be submitted stating only one person is withdrawing the amount. The girls and the boys family can withdraw this amount separately. 2. The application for withdrawal for a wedding has to be accompanied by the following documents: An application form; Evidence of the wedding, including the invitation card, copies of receipts for advance payments already made, such as Marriage hall booking, advance payments to caterers, etc.; A declaration from the person who has to be paid more than Rs. 10,000 stating that they do not have a bank account, anda complete list of people who have to be paid in cash and the purpose for the payment.
Chapter 7: OTHER DETAILS	1. In Odisha, Panchayat offices can be used for banking services in areas where banks are too far or banking facilities are not available. 2. You can use NEFT/RTGS/IMPS/InternetBanking/Mobile Banking or any other electronic/ non-cash mode of payment. 3. Valid Identity proof is any of the following: Aadhaar Card, Driving License, Voter ID Card, Pass Port, NREGA Card, PAN Card, Identity Card Issued by Government Department, Public Sector Unit to its Staff. 4. You may approach the control roomof RBI on Telephone Nos 022-22602201 22602944 5. The date for submission of annual life certificate has been extended to January 15, 2017 from November for all government pensioners 6. As of December 15, 2016, specified bank notes of only Rs. 500 can no longer be used for the following: Government hospitals and pharmacies, railway and government bus tickets, consumer cooperative stores, government and court fees, government School fees, mobile top-ups, milk booths, crematoria and burial grounds, LPG gas cylinders, Archaelogical Survey of India monuments, utilities, toll payments

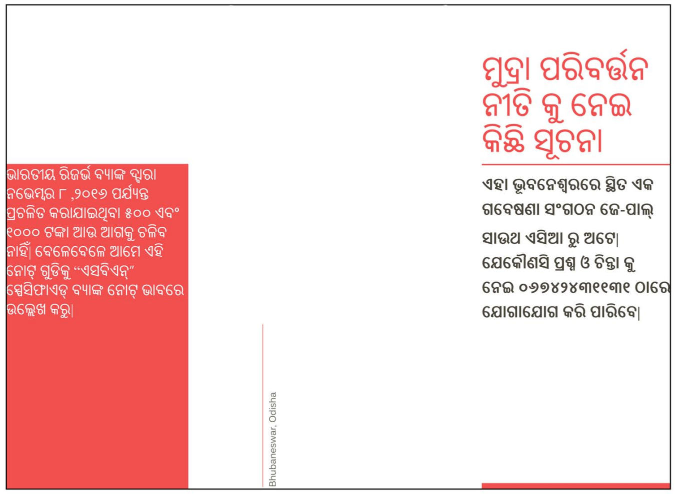
Appendix C. Example Pamphlet Excerpts



(A) Front

(B) Back

Figure C.1. Short pamphlet (2 facts)



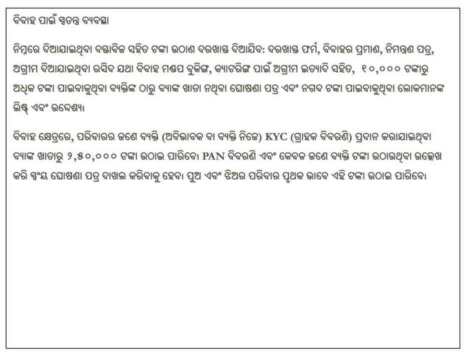
ଭାରତୀୟ ରିଜର୍ଭ ବ୍ୟାଙ୍କ ଦ୍ଵାରା ନଭେମ୍ବର ୮ ,୨୦୧୬ ପର୍ଯ୍ୟନ୍ତ ପ୍ରଚଳିତ କରାଯାଇଥିବା ୫୦୦ ଏବଂ ୧୦୦୦ ଟଙ୍କା ଆଉ ଆଗକୁ ଚଳିବ ନାହିଁ| ବେଳେବେଳେ ଆମେ ଏହି ନୋଟ୍ ଗୁଡିକୁ "ଏସବିଏନ୍" ସ୍ପେସିଫାଏଡ୍ ବ୍ୟାଙ୍କ ନୋଟ୍ ଭାବରେ ଉଲ୍ଲେଖ କରୁ|

Bhubaneswar, Odisha

ମୁଦ୍ରା ପରିବର୍ତ୍ତନ ନୀତି କୁ ନେଇ କିଛି ସୂଚନା

ଏହା ଭୁବନେଶ୍ଵରରେ ସ୍ଥିତ ଏକ ଗବେଷଣା ସଂଗଠନ ଜେ-ପାଲ୍ ସାଉଥ ଏସିଆ ରୁ ଅଟେ| ଯେକୌଣସି ପ୍ରଶ୍ନ ଓ ଚିନ୍ତା କୁ ନେଇ ୦୬୭୪୨୪୩୧୧୩୧ ଠାରେ ଯୋଗାଯୋଗ କରି ପାରିବେ|

(A) Front

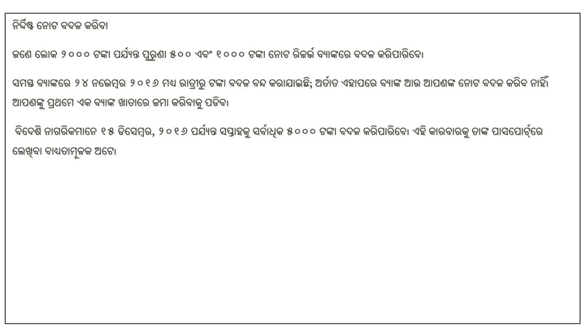


ବିବାହ ପାଇଁ ସ୍ଵତନ୍ତ୍ର ବ୍ୟବସ୍ଥା

ନିମ୍ନରେ ଦିଆଯାଇଥିବା ଦସ୍ତାବିଜ ସହିତ ଟଙ୍କା ଉଠାଣ ଦରଖାସ୍ତ ଦିଆଯିବ: ଦରଖାସ୍ତ ଫର୍ମ, ବିବାହର ପ୍ରମାଣ, ନିମନ୍ତ୍ରଣ ପତ୍ର, ଅଗ୍ରୀମ ଦିଆଯାଇଥିବା ରସିଦ ଯଥା ବିବାହ ମଣ୍ଡପ ବୁକିଙ୍ଗ, କ୍ୟାଟରିଙ୍ଗ ପାଇଁ ଅଗ୍ରୀମ ଇତ୍ୟାଦି ସହିତ, ୧୦,୦୦୦ ଟଙ୍କାରୁ ଅଧିକ ଟଙ୍କା ପାଇବାକୁଥିବା ବ୍ୟକ୍ତିଙ୍କ ଠାରୁ ବ୍ୟାଙ୍କ ଖାତା ନଥିବା ଘୋଷଣା ପତ୍ର ଏବଂ ନଗଦ ଟଙ୍କା ପାଇବାକୁଥିବା ଲୋକମାନଙ୍କ ଲିଷ୍ଟ୍ ଏବଂ ଉଦ୍ଦେଶ୍ୟ।

ବିବାହ କ୍ଷେତ୍ରରେ, ପରିବାରର ଜଣେ ବ୍ୟକ୍ତି (ଅବିଭାବକ ବା ବ୍ୟକ୍ତି ନିଜେ) KYC (ଗ୍ରାହକ ବିବରଣି) ପ୍ରଦାନ କରାଯାଇଥିବା ବ୍ୟାଙ୍କ ଖାତାରୁ ୨,୫୦,୦୦୦ ଟଙ୍କା ଉଠାଇ ପାରିବେ। PAN ବିବରଣି ଏବଂ କେବଳ ଜଣେ ବ୍ୟକ୍ତି ଟଙ୍କା ଉଠାଉଥିବା ଉଲ୍ଲେଖ କରି ସ୍ଵଂୟ ଘୋଷଣା ପତ୍ର ଦାଖଲ କରିବାକୁ ହେବ। ପୁଅ ଏବଂ ଝିଅର ପରିବାର ପୃଥକ ଭାବେ ଏହି ଟଙ୍କା ଉଠାଇ ପାରିବେ।

(B) Page 1/8



ନିର୍ଦ୍ଦିଷ୍ଟ ନୋଟ ବଦଳ କରିବା

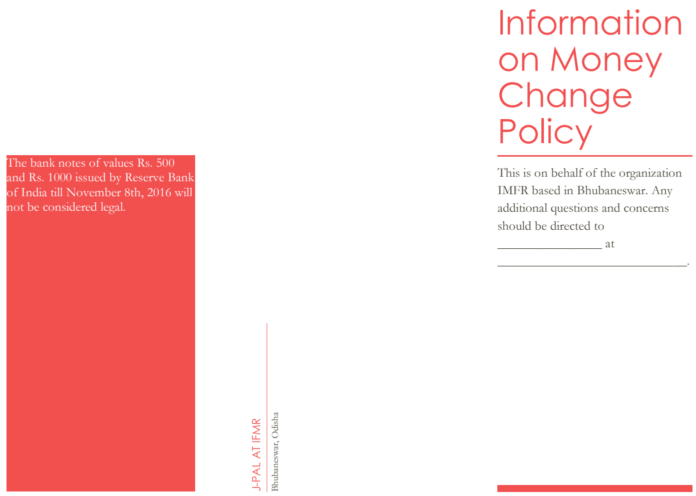
ଜଣେ ଲୋକ ୨୦୦୦ ଟଙ୍କା ପର୍ଯ୍ୟନ୍ତ ପୁରୁଣା ୫୦୦ ଏବଂ ୧୦୦୦ ଟଙ୍କା ନୋଟ ରିଜର୍ଭ ବ୍ୟାଙ୍କରେ ବଦଳ କରିପାରିବେ।

ସମସ୍ତ ବ୍ୟାଙ୍କରେ ୨୪ ନଭେମ୍ବର ୨୦୧୬ ମଧ୍ୟ ରାତ୍ରୀରୁ ଟଙ୍କା ବଦଳ ବନ୍ଦ କରାଯାଇଛି; ଅର୍ଥାତ ଏହାପରେ ବ୍ୟାଙ୍କ ଆଉ ଆପଣଙ୍କ ନୋଟ ବଦଳ କରିବ ନାହିଁ। ଆପଣଙ୍କୁ ପ୍ରଥମେ ଏକ ବ୍ୟାଙ୍କ ଖାତାରେ ଜମା କରିବାକୁ ପଡିବ।

ବିଦେଶି ନାଗରିକମାନେ ୧୫ ଡିସେମ୍ବର, ୨୦୧୬ ପର୍ଯ୍ୟନ୍ତ ସପ୍ତାହକୁ ସର୍ବାଧିକ ୫୦୦୦ ଟଙ୍କା ବଦଳ କରିପାରିବେ। ଏହି କାରବାରକୁ ତାଙ୍କ ପାସପୋର୍ଟ୍‌ରେ ଲେଖିବା ବାଧ୍ୟତାମୂଳକ ଅଟେ।

(C) Page 2/8

FIGURE C.2. Long pamphlet (24 facts)



The bank notes of values Rs. 500 and Rs. 1000 issued by Reserve Bank of India till November 8th, 2016 will not be considered legal.

J-PAL AT IFMR
Bhubaneswar, Odisha

Information on Money Change Policy

This is on behalf of the organization IMFR based in Bhubaneswar. Any additional questions and concerns should be directed to _______________ at ___________________________.

Figure C.3. Front Page of Pamphlets, English Translation

Appendix D. Detailed signaling model

D.1. **The model.** Consider a set N of agents (the village). The model focuses on the choice of a single decision-maker, $\mathrm{D} \in N$ of whether to seek or not.

D.1.1. *Timing.* The timing of the interaction is as follows:

(1) (a) The policymaker privately chooses a breadth of dissemination

$$\mathfrak{b} \in \{\text{Broadcast}, \text{Seed}, \text{None}\}.$$

The prior probability of breadth $\mathfrak{b}$ is $\beta_{\mathfrak{b}} \in (0,1)$. Conditional on $\mathfrak{b} = \text{Broadcast}$, all members of the village N receive facts. Conditional on $\mathfrak{b} = \text{Seed}$, a nonempty, proper subset S of individuals is randomly drawn to be informed.

(b) The policymaker sends a public signal (which reaches all members of N)

$$\mathfrak{p} \in \{\text{CK:Broadcast}, \text{CK:Seed}, \text{No CK}\}.$$

When a "CK:$\mathfrak{b}$" announcement is made, it is always the case that the breadth is in fact $\mathfrak{b}$. If no "CK:$\mathfrak{b}$" signal is sent, that is necessarily common knowledge; we call that outcome the No CK signal, which, practically, is an absence of such a public announcement. Under breadth $\mathfrak{b}$, the probability of a CK:$\mathfrak{b}$ announcement is $\chi_{\mathfrak{b}} \in (0,1)$.

(2) If $\mathfrak{b} \in \{\text{Broadcast}, \text{Seed}\}$, then with certainty the facts mechanically reach the Town Square.

(3) The decision-maker, $\mathrm{D} \in N$, privately learns his incremental value of getting additional information beyond the facts he received. He then decides whether to go to the Town Square to seek information about the facts delivered. D's decision is denoted by

$$d \in \{0 \text{ (Not Seeking)}, 1 \text{ (Seeking)}\}.$$

(4) An Observer in the Town Square sees whether D has come to seek information, and updates his belief about D's type.[50]

A treatment in our experiment may be summarized by a pair $\mathfrak{t} = (\mathfrak{b}, \mathfrak{p})$, the breadth of dissemination and the public signal.

The interpretation of the Town Square is that there are locations in the village (a store, tea shop, etc.) where exchange of information takes place and where the local news of the day can be accessed. There, individuals interested in learning about an issue can participate in conversations about it.

[50]We will discuss beliefs about D's type more below.

This model abstracts from important forces, such as social learning outside the Town Square and the dependence of learning and signaling on others' seeking decisions. To some extent such forces can be captured in parameters of this simple model; for instance, the extent of social learning may affect the probability that information is in the Town Square. In Section 6, we consider some models with richer social learning.

Types and payoffs. The payoff that D experiences from seeking depends on (i) what information there is to gain by going to the Town Square, compared to the information D already has; (ii) non-learning costs and benefits of going to the Town Square, such as the cost of time or the possibility of running into a friend; (iii) reputational payoffs depending on what people may infer about D based on his decision to go to the Town Square. This subsection introduces the primitives we use to model these considerations.

We posit that D has a privately known ability type $a \in \{\mathrm{H}, \mathrm{L}\}$, with prior probabilities $\pi_{\mathrm{H}}, \pi_{\mathrm{L}} \in (0, 1)$, respectively.[51] We will assume these are generic.[52] Let $I_{\mathrm{D}} \in \{0, 1\}$ denote whether D has received facts from the policymaker. This occurs if $\mathfrak{b} =$ Broadcast or if $\mathfrak{b} =$ Seed and $\mathrm{D} \in S$. Let $I_{\mathrm{S}} \in \{0, 1\}$ denote whether there is information in the Town Square. The information is present ($I_{\mathrm{S}} = 1$) when $\mathfrak{b} =$ Broadcast or Seed, and absent otherwise.

With this notation in hand, we introduce quantities capturing (i) and (ii) above: the direct (i.e., non-reputational) payoffs of Seeking and Not Seeking. The random variable $V^{(I_{\mathrm{D}}, I_{\mathrm{S}})}(1)$ is the direct payoff of Seeking when the informational states are $(I_{\mathrm{D}}, I_{\mathrm{S}})$, while $V^{(I_{\mathrm{D}})}(0)$ is the direct payoff (which can be positive or negative) of not seeking when the seeker's information is I_{D}. The realizations of these V quantities for all their arguments – $\{V^{(I_{\mathrm{D}}, I_{\mathrm{S}})}(1)\}_{I_{\mathrm{D}}, I_{\mathrm{S}}}$ and $\{V^{(I_{\mathrm{D}})}(0)\}_{I_{\mathrm{D}}}$ – are known to D at stage (4), the time he makes his decision.

The following random variable, whose prior distribution we call $F_a^{(I_{\mathrm{D}}, I_{\mathrm{S}})}$, represents the incremental direct payoff gain to seeking:

$$\Delta^{(I_{\mathrm{D}}, I_{\mathrm{S}})} := V^{(I_{\mathrm{D}}, I_{\mathrm{S}})}(1) - V^{(I_{\mathrm{D}})}(0) \sim F_a^{(I_{\mathrm{D}}, I_{\mathrm{S}})}. \tag{D.1}$$

Crucially, the V random variables, and hence the random variable $\Delta^{(I_{\mathrm{D}}, I_{\mathrm{S}})}$, have distributions that depend on D's ability type. Because of this, if seeking decisions provide information about $\Delta^{(I_{\mathrm{D}}, I_{\mathrm{S}})}$, they can signal D's ability.

In addition to the direct payoff, D receives a reputational, or *perception*, payoff. If D chooses to seek and goes to the Town Square, this choice will be observed by some other villagers, who may make inferences about D's ability.

For a simple model of how D values others' assessment of him, we posit that, in the Town Square, there is an agent called the Observer (O), drawn uniformly at random from the

[51]The ability random variable is independent of all others in the model except those defined below that explicitly condition on it.

[52]That is, drawn from a measure absolutely continuous with respect to the Lebesgue measure.

village. This Observer sees D's decision of whether to seek or not. Because this person is also in the village, she has her own information, a realization I_{O}. (Thus, for example, when a broadcast has disseminated information to everyone, the Observer has received the information, too.) We assume D does not know in advance who may observe his decision to seek, and therefore does not condition the seeking decision on the realized identity of the Observer. The perception payoff enters D's utility function additively, as a term

$$\lambda \mathbb{P}(a = \mathrm{H} \mid d, \mathfrak{p}, I_{\mathrm{O}}),$$

where λ is a positive number. Note that the Observer is conditioning on everything she knows: the decision he observes D taking, the public signal, and the Observer's own information about the state. The idea behind the perception payoff is that D is better off when other villagers assess D's ability to be high – for example, because in that case those villagers are more likely to collaborate with D later.[53]

D's total payoff given seeking decision d is, therefore,

$$u^{(I_{\mathrm{D}}, I_{\mathrm{S}})}(d) = V^{(I_{\mathrm{D}}, I_{\mathrm{S}})}(d) + \lambda \mathbb{P}(a = \mathrm{H} \mid d, \mathfrak{p}, I_{\mathrm{O}}). \tag{D.2}$$

It will be useful to write the difference

$$u^{(I_{\mathrm{D}}, I_{\mathrm{S}})}(1) - u^{(I_{\mathrm{D}}, I_{\mathrm{S}})}(0) = \Delta^{(I_{\mathrm{D}}, I_{\mathrm{S}})} - \lambda \Pi \tag{D.3}$$

where $\Delta^{(I_{\mathrm{D}}, I_{\mathrm{S}})}$ is defined in (D.1) and

$$\Pi = \mathbb{P}(a = \mathrm{H} \mid d = 0, \mathfrak{p}, I_{\mathrm{O}}) - \mathbb{P}(a = \mathrm{H} \mid d = 1, \mathfrak{p}, I_{\mathrm{O}}). \tag{D.4}$$

D will take expectations over the perception payoffs in making his decision. In turn, the posterior belief that other villagers have about ability is endogenous: it depends on the seeking behaviors for both types, which depend on their payoffs. This leads us to an examination of the equilibria of the game.

D.1.2. *Equilibrium: Definition and basic observations.* We study a Bayesian equilibrium of this game. A strategy of D determines beliefs of the Observer – i.e. $\mathbb{P}(a = \mathrm{H} \mid d, \mathfrak{p}, I_{\mathrm{O}})$ – for both values $d = 1, 0$.[54] That, in turn, determines D's incentives, since he cares about perceptions.

A strategy for D is a map that gives a decision d as a function of the tuple of all realizations D knows at the time of his decision – ability a, public signal $\mathfrak{p}$, own information state I_{D}, and the values $V^{(I_{\mathrm{D}}, I_{\mathrm{S}})}(d)$ across decisions d and pairs $(I_{\mathrm{D}}, I_{\mathrm{S}})$. However, the decision can

[53] Foundations for this assumption are discussed in Chandrasekhar et al. (2018).

[54] As usual, the equilibrium can be given a population interpretation: there is a population of D's, who have different draws of private information, and the Observer is inferring the attributes of a particular D in view of the population's behavior.

actually be simplified: in any rational strategy, D will seek if and only if his expectation of his direct gain $\Delta^{(I_\mathrm{D},I_\mathrm{S})}$ exceeds his expectation of the perception benefit of not seeking, Π, which in equilibrium is a known number.[55]

An equilibrium strategy is characterized by these conditions: (i) D seeks if and only if his expectation of $\Delta^{(I_\mathrm{D},I_\mathrm{S})}$ is at least his expectation of $\lambda\Pi$; (ii) the beliefs about ability a in (D.4) are consistent with (i) and Bayes' rule.

If each $F_a^{(I_\mathrm{D},I_\mathrm{S})}$ has no atoms – an assumption we will maintain – then an equilibrium can be described essentially completely by specifying a cutoff for D to seek: how high D's expected value of $\Delta^{(I_\mathrm{D},I_\mathrm{S})}$ has to be in order to choose $d = 1$. The cutoff, which we call $\underline{v}(\mathfrak{p}, I_\mathrm{D})$ only depends on the public signal $\mathfrak{p}$ and on I_D and, as a function of these, it is commonly known in equilibrium.[56]

D.1.3. *Assumptions.*

Payoffs. We now discuss assumptions on the distribution of $\Delta(\cdot,\cdot)$. First, for technical convenience, we will maintain the assumption that the support of $F_a(I_\mathrm{D}, I_\mathrm{S})$ includes the positive reals, for all values of a and $(I_\mathrm{D}, I_\mathrm{S})$.

Next, we make assumptions on how different abilities value information.

P1 (a) For any $(I_\mathrm{D}, I_\mathrm{S})$, the distribution $F_\mathrm{L}^{(I_\mathrm{D},I_\mathrm{S})}$ first-order stochastically dominates $F_\mathrm{H}^{(I_\mathrm{D},I_\mathrm{S})}$.

A low-ability D always has at least as much to gain from seeking as a high-ability one, all else equal.

(b) For all values of I_S, the ratio $\frac{1-F_\mathrm{L}^{(I_\mathrm{D},I_\mathrm{S})}(v)}{1-F_\mathrm{H}^{(I_\mathrm{D},I_\mathrm{S})}(v)}$ is strictly increasing in I_D for any v.

For any cutoff, having a value of information above that cutoff signals low ability more when D is informed ($I_\mathrm{D} = 1$) than when D is not informed ($I_\mathrm{D} = 0$).

Assumption P1(a) reflects that a low-ability D needs more help to figure out the content of information. It ensures that seeking is (weakly) a signal of low ability, because for any cutoff D uses, the low-ability type is (weakly) more likely to exceed it. Assumption P1(b) imposes some structure on that signal, as described above.

Our next assumption imposes structure on how the informational states of D and of the Town Square affect the payoffs of seeking.

P2 (a) $F_a^{(I_\mathrm{D},1)}(v) < F_a^{(I_\mathrm{D},0)}(v)$ for all $v \geq 0$ and all values of a and I_D.

Regardless of ability and own signal, seeking is (in the stochastic sense) strictly more beneficial when there is information in the Town Square.

[55]D's decision does not depend on his private ability type a. The reason is as follows: Given $\Delta^{(I_\mathrm{D},I_\mathrm{S})}$, D's ex post direct gain to seeking, (D.3), does not depend on his private ability type. Because his ability type is unobservable, the reputational payoff cannot depend on it, either.

[56]We make the innocuous tie-breaking assumption that the seeker seeks if and only if $\Delta^{(I_\mathrm{D},I_\mathrm{S})} \geq \underline{v}(\mathfrak{p}, I_\mathrm{D})$.

(b) $F_a^{(0,1)}$ first-order stochastically dominates $F_a^{(1,1)}$ for both values of a.
The direct benefit of seeking is weakly greater when one is uninformed, assuming there is information in the Town Square.

Our final assumption allows us to obtain non-parametric comparisons of seeking rates across treatments; though stronger than necessary, it enables the use of monotone comparative statics arguments.

P3 For any $(I_{\mathrm{D}}, I_{\mathrm{S}})$, the ratio $\frac{1-F_{\mathrm{H}}^{(I_{\mathrm{D}},I_{\mathrm{S}})}(v)}{1-F_{\mathrm{L}}^{(I_{\mathrm{D}},I_{\mathrm{S}})}(v)}$ is strictly decreasing in v for all $v \geq 0$.

This is a regularity condition on the distribution of values of seeking which is satisfied if, for example, F_{L} and F_{H} are stochastically ordered normal distributions centered to the left of zero. Economically, this means that the higher is the cutoff for seeking, the worse is the inference about D's ability if D chooses to seek. This condition is useful because it enables us to use the techniques of monotone comparative statics to study how $\underline{v}(\mathfrak{p})$, the cutoff for seeking, varies across treatments.

Beliefs. In our description of the timing of the game, we did not make any assumptions about how S, the set of seeded individuals, is drawn. We now make two assumptions on individuals' beliefs that restrict this distribution, which we will need in some, but not all, of our results.

B1 For any $i \in N$, the probability $\mathbb{P}(i \in S)$ is between $1/n$ and $\overline{k}/n$ for some constant $\overline{k}$.

B2 For any two individuals i and j, there is a constant C so that the conditional probability $\mathbb{P}(i \in S \mid j \in S, \mathfrak{b} = \text{Seed})$ is at most $C\mathbb{P}(i \in S \mid \mathfrak{b} = \text{Seed})$.

These assumptions say that there are not too few or too many seeds, and from the perspective of any j, individual i's membership in the seed set S is not too correlated with j's own.

D.2. **Analysis and results.**

D.2.1. *Dependence of seeking rates on treatment.* In general the model may have multiple equilibria.[57] However, under our assumptions (the key one being P3) the game has some nice structure. In particular, as the cutoffs[58] $\underline{v}(\mathfrak{p}, I_{\mathrm{D}})$ increase, incentives to seek decrease monotonically for all realizations of private information. (This occurs because, loosely speaking, seeking becomes a worse signal.) Because the resulting game of incomplete information then has a supermodular structure, we can identify an equilibrium that has maximum seeking in a strong sense: for every realization of D's private information, there is more seeking in that

[57]For more on this multiplicity, see Chandrasekhar et al. (2018).
[58]Introduced in Section D.1.2 above.

equilibrium than in any other. This equilibrium will always be stable under best-response dynamics, and call this the *maximum equilibrium*.[59]

Let $s(\mathfrak{t})$ be the probability, in the maximum equilibrium, that D chooses $d = 1$ (Seeking) in treatment $\mathfrak{t} = (\mathfrak{b}, \mathfrak{p})$ – for example $\mathfrak{t} =$ (Seed, CK:Seed). This is an ex ante probability: we integrate over all ability types, information realizations, etc. We focus on this statistic because it is one that is observed in our experiments. Now we can state the two main propositions yielding our predictions.

PROPOSITION **1.** Under Assumptions P1–P3:

(a) s(Broadcast, No CK) $>$ s(Broadcast, CK);

(b) s(Seed, CK) $>$ s(Broadcast, CK).

The proof of this and all other propositions appears in Section D.3 of the Appendix. We give the key ideas of the argument in the next subsection. An important corollary that we use in testing the mechansim is:

COROLLARY **1.** Under Assumptions P1–P3, the type-dependent seeking probabilities satisfy the following inequalities

(a) s_a(Broadcast, No CK) $>$ s_a(Broadcast, CK) for each ability type a

(b) $\frac{s_L(\text{Broadcast, CK})}{s_H(\text{Broadcast, CK})} > \frac{s_L(\text{Broadcast, No CK})}{s_H(\text{Broadcast, No CK})}$

The second proposition relies on assumptions about beliefs, ranking the amount of communication in the Seed treatments.

PROPOSITION **2.** Under Assumptions P1–P3 and B1–B2, and assuming $\overline{k}/n$ is small enough, it holds that s(Seed, CK) $>$ s(Seed, No CK).

Finally, the prediction that requires the most structure is:

PROPOSITION **3.** Take Assumptions P1–P3 and B1–B2, and, fixing all other parameters, suppose the following three quantities are small enough: (i) $\overline{k}/n$; (ii) β_{Seed}; and (iii) $\frac{1-\chi_{\text{Broadcast}}}{(\overline{k}/n)^2}$. Then s(Broadcast, No CK) $>$ s(Seed, No CK).

Intuition behind the Propositions. We now explain the key forces behind each of the main predictions entailed in the propositions above.

Proposition 1

(a) (Broadcast, No CK) has more seeking than (Broadcast, CK). In both cases, D's assessment of direct payoffs is the same: since $I_{\text{D}} = 1$, D knows that $I_{\text{S}} = 1$. In the

[59]Making another selection, such as the *minimum* equilibrium, which also exists, would not change the analysis or the results. Of course, this selection point is moot if equilibrium is unique; conditions for uniqueness are available upon request.

(Broadcast, CK) treatment, O is certain that D is informed, and D knows this. It is in that case that signaling concerns are the strongest they could be, by Assumption P1(b). In (Broadcast, No CK) the signaling effect is weaker, because some probability is placed on D not being informed. Thus, there is more seeking under (Broadcast, No CK).

(b) (Seed, CK) has more seeking than (Broadcast, CK):

Considering the signaling contribution to payoffs: for any given cutoffs, we can write the beliefs of the Observer conditional on $d = 1$ (given either value of $\mathfrak{p}$) as a convex combination over values of I_{D}. The term corresponding to $I_{\mathrm{D}} = 1$ is the same across the two treatments. This is the only term with a positive weight in the (Broadcast, CK) treatment. The term corresponding to $I_{\mathrm{D}} = 0$ involves a weakly greater posterior that $a = \mathrm{H}$ by Assumption P1. Thus, signaling concerns are smaller in (Seed, CK).

Turning now to the direct payoffs, $I_{\mathrm{S}} = 1$ is known in both cases. By Assumption P2(b), the value of seeking is greater for the uninformed, who are at least as prevalent in the Seed treatment. Thus, direct payoffs are greater there.

Proposition 2

First, under (Seed, CK), D is certain that information is in the Town Square, which by P2 shifts up the expected direct value of seeking relative to (Seed, No CK) by at least some positive amount. Now we turn to signaling concerns. Condition on $I_{\mathrm{D}} = 0$ (which is the case with high probability under Seed, since $\overline{k}/n$ is small by assumption). In this case, D is nearly certain that O is uninformed. Conditioning on $I_{\mathrm{O}} = 0$, by the same token, O is nearly certain that D is uninformed. Thus, signaling concerns are very similar to the case in which it is common knowledge that D is uninformed.

Proposition 3

For the argument behind Proposition 3, we need a lemma, which we state somewhat informally. It follows immediately from Bayes' rule.[60]

Lemma **1.** Under the assumptions of Section D.1.3, suppose that $(1 - \chi_{\text{Broadcast}})$ is small enough relative to $(\overline{k}/n)^2$. Then conditional on $\mathfrak{p} = \text{No CK}$ and any realizations of I_{D} and I_{O}, the probability that $\mathfrak{b} = \text{Broadcast}$ is negligibly small.

Now we can establish the proposition. Concerning the direct benefit: in (Seed, No CK), when D receives no information ($I_{\mathrm{D}} = 0$), the fact that β_{Seed} is small means that his expectations approximate those when $I_{\mathrm{S}} = 0$. In contrast, in (Broadcast, No CK), given that

[60]Consider an observer who knows that $I_{\mathrm{D}} = I_{\mathrm{O}} = 1$ and that $\mathfrak{p} = \text{No CK}$. His posterior likelihood ratio that $\mathfrak{b} = \text{Broadcast}$ has occurred versus $\mathfrak{b} = \text{Seed}$ is of order $(1 - \chi_{\text{Broadcast}})/(\overline{k}/n)^2$. Thus if this is small, then even this observer will consider Broadcast unlikely.

$I_{\mathrm{D}} = 1$, the breadth $\mathfrak{b}$ is in $\{\text{Broadcast}, \text{Seed}\}$ (i.e., not equal to "None") and information is certain to be in the Town Square ($I_{\mathrm{S}} = 1$). By Assumption P2, seeking is more valuable in this case.

Turning now to signaling concerns, the key step is to rule out the possibility that the observer under (Broadcast, No CK) assumes that since he has a signal, so does everyone else (i.e. the state is Broadcast). This is where we make use of the face that because there is no public announcement, by Lemma 1, O will be nearly certain that $\mathfrak{b} \neq$ Broadcast. Because $\overline{k}/n$ is small, he will also be nearly certain that D is not a seed. To sum up, O will believe I_{D} holds with high probability. Thus, signaling concerns are therefore almost the same in the two cases.

The proof formalizes these ideas using monotone comparative statics.

Comments on modeling choices. We close this subsection with some brief comments on our modeling choices. One choice we make is to assume that the Observer is not the source of the information that is available in the Town square. An alternative would have been to have the person asked for information to also be the Observer, thus merging the roles of the source T and O. However, this raises a variety of challenging modeling decisions: do we explicitly model the aggregation of information by this person? What if she herself is unable to process the signal she received? How are signaling concerns affected by the fact that she may be able to infer, based on the number of people coming to her, what the $(\mathfrak{b}, \mathfrak{p})$ realization is? Another direction would be to more realistically model a Town Square where there are many different people, and now the information D gets is obtained by talking to a member of this population, drawn according to some distribution. Aggregation of information in the Town Square would now have to be modeled explicitly, which presents considerable complications; there will also be potential for bilateral signaling, both by Seekers and Advisers. Our modeling abstracts from these complications to get at what we believe are the essential phenomena, though models addressing these richer concerns may be interesting in their own right.

D.2.2. *Knowledge and choice quality in equilibrium.* Propositions 1 and 2 focus on the rates of seeking – which, in the experiment, we measure by the amount of conversation. But our experiments also consider other outcomes: knowledge about demonetization and choice quality. To study these using our theory, we analyze the expected direct payoff

$$p(\mathfrak{t}) = \mathbb{E}[V^{(I_{\mathrm{D}}, I_{\mathrm{S}})}(d) \mid \mathfrak{t}]$$

in a given treatment $\mathfrak{t}$. This is the value of information gross of signaling concerns. Again, it is pooled over ability types and information realizations. Consider the comparisons of Propositions 1 and 2. When I_{D} is held fixed, the rankings are just as in that proposition:

Corollary **2.** Under the conditions of Proposition 2,

(a) $p(\text{Broadcast, No CK}) > p(\text{Broadcast, CK})$
(b) $p(\text{Seed, CK}) > p(\text{Seed, No CK})$

Note that in both (a) and (b), D's information endowment is the same. In (a), the proof of Proposition 1 shows that the direct value is the same on both sides of the inequality, while the signaling concerns are smaller on the left-hand side, furnishing the conclusion. In (b) the proof of Proposition 2 shows that the signaling concerns are no greater while the incremental value of information is appreciably higher.

When the comparison of two given treatments also involves changes in I_{D}, the comparisons are not as immediate. However, we will now discuss, somewhat informally, what is needed for the remaining rankings of knowledge and decision quality to parallel those that were derived for s above:

- $p(\text{Seed, CK}) > p(\text{Broadcast, CK})$
- $p(\text{Broadcast, No CK}) > p(\text{Seed, No CK})$ under the assumptions of Proposition 3.

For the first item, let us consider how the inequality could possibly be reversed relative to the corresponding item in Proposition 1. For a reversal, it would have to be that the base level of knowledge possessed by agents in (Broadcast, CK) is enough to make them better off even if signaling concerns deter seeking. The reversal would therefore *not* happen if we assume: (a) low-ability types who don't seek make decisions approximately as if they were uninformed, and (b) there are enough low-ability types. In that case, seeking rates become pivotal to the welfare of enough of the population; knowledge and choice quality then move in tandem with seeking rates.

The condition needed for the second ranking is similar. If we assume that β_{Seed} is small, then, as we argued in Proposition 3, the expected incremental direct benefit of seeking ($\Delta^{(I_{\mathrm{D}},I_{\mathrm{S}})}$) is very close to its expectation under $I_{\mathrm{D}} = I_{\mathrm{S}} = 0$. Under (Broadcast, No CK), it is much higher, while signaling concerns are very similar across the two cases. Thus equilibrium welfare must also be higher for those types who need to seek in order to do better than their uninformed welfare.

D.3. **Proofs.**

D.3.1. *Preliminaries for Proof of Main Proposition.* Introduce an *index* $\omega \in (0,1)$ for the type of the decision-maker D. This index is drawn uniformly from [0,1]. By the assumption of no atoms, we can view $\Delta^{(I_{\mathrm{D}},I_{\mathrm{S}})}$ as a continuous increasing function $(0,1) \to \mathbb{R}$. Moreover, by P2, we may assume that, pointwise, $\Delta^{(I_{\mathrm{D}},1)}(\omega) > \Delta^{(I_{\mathrm{D}},0)}(\omega)$ and $\Delta^{(0,1)}(\omega) \geq \Delta^{(1,1)}(\omega)$. This uses the standard coupling for random variables ordered by stochastic dominance.

Recall the payoff difference formula (D.3)

$$u^{(I_{\mathrm{D}},I_{\mathrm{S}})}(1) - u^{(I_{\mathrm{D}},I_{\mathrm{S}})}(0) = \Delta^{(I_{\mathrm{D}},I_{\mathrm{S}})} - \lambda\Pi,$$

where Π is the signaling penalty. For any $\mathfrak{p}$, a strategy profile in which D is best-responding can be summarized by a vector of interior cutoffs $\boldsymbol{c} = (c(\mathfrak{p}, I_{\mathrm{D}}))_{I_{\mathrm{D}}}$ such that D seeks given I_{D} if his index ω is above $c(\mathfrak{p}, I_{\mathrm{D}})$, and does not seek if his index is below $c(\mathfrak{p}, I_{\mathrm{D}})$. (Interiority is guaranteed by the assumption that the distributions of Δ in each case have full support.)

We may now write the right-hand side of (D.3) as

$$W^{(I_{\mathrm{D}}, I_{\mathrm{S}})}(\omega; \boldsymbol{c}) = \Delta^{(I_{\mathrm{D}}, I_{\mathrm{S}})}(\omega) - \lambda \Pi(\boldsymbol{c}).$$

Here $\Delta^{(I_{\mathrm{D}}, I_{\mathrm{S}})}(\omega)$ is increasing in ω and $\Pi(\boldsymbol{c})$ is increasing in $\boldsymbol{c}$ by P3.

Define $W^{(I_{\mathrm{D}}, \mathfrak{p})}(\omega)$ to be the expectation of $W(\omega)$ given public signal $\mathfrak{p}$ and a realization of I_{D}. Define the analogous notation for Δ.

Because λ is a finite constant, cutoffs given both values of I_{D} are guaranteed to be in some compact subset $\mathcal{C} \subseteq (0, 1)$ irrespective of strategies; so we will restrict attention to this subset from now on in studying equilibria.[61]

For each $\mathfrak{p}$ and each ω, the payoff advantage $W^{(I_{\mathrm{D}}, \mathfrak{p})}(\omega)$ of seeking is monotone decreasing in the cutoff vector $\boldsymbol{c}$, so this is a supermodular game. In particular, a minimum equilibrium cutoff profile (which corresponds to maximum seeking) exists. We now state two results which follow from the supermodular structure of the game:

Fact **1.** The following hold:

- SM1 If $W^{(I_{\mathrm{D}}, \mathfrak{p})}(\omega; \boldsymbol{c})$ strictly increases for each $\omega, \boldsymbol{c} \in \mathcal{C}$ and I_{D} then the minimum cutoff $\boldsymbol{c}$ strictly decreases in each component.
- SM2 Let $\iota_{\mathfrak{p}}$ be the ex ante probability of $I_{\mathrm{D}} = 1$ given $\mathfrak{p}$. Then, for each $\mathfrak{p}$, the maximum equilibrium cutoff $c(\mathfrak{p}, 0)$ is continuous in $\iota_{\mathfrak{p}}$ at $\iota_{\mathfrak{p}} = 0$ for generic priors $(\pi_{\mathrm{H}}, \pi_{\mathrm{L}})$.

The first part, SM1, is a standard monotone comparative statics fact. The second, SM2, is argued as follows. Define a reaction function $r_{\iota_{\mathfrak{p}}} : \mathcal{C}^2 \to \mathcal{C}^2$ mapping any cutoffs $\boldsymbol{c}$ to the best-response cutoffs when the Observer updates assuming the cutoffs $\boldsymbol{c}$. Because the distribution of $\Delta^{(I_{\mathrm{D}}, I_{\mathrm{S}})}$ has full support, inferences of the Observer depend arbitrarily little on the behavior of $I_{\mathrm{D}} = 1$ types as $\iota_{\mathfrak{p}} \downarrow 0$. Thus, the reaction functions $r_{\iota_{\mathfrak{p}}}$ may be bounded within an arbitrarily narrow band of the reaction functions r_0. Thus, for generic parameters (guaranteeing that r is transversal to the hyperplane $(x, y) \mapsto (x, y)$ at the equilibrium), the equilibrium will be continuous in $\iota_{\mathfrak{p}}$.

D.3.2. *Proof of Proposition 1.*

(a) (Broadcast, No CK) has more seeking than (Broadcast, CK). In both cases, $W^{(I_{\mathrm{D}}, \mathfrak{p})}(\omega)$: since $I_{\mathrm{D}} = 1$, D knows that $I_{\mathrm{T}} = 1$.

[61] To show the cutoff does not get arbitrarily close to 0 in ω space, we can simply note that each function $\Delta^{(I_{\mathrm{D}}, \mathfrak{p})}(\omega)$ is negative below some $\omega > 0$. Because $\Pi \geq 0$, cutoffs cannot occur in the region where W is negative.

Now we turn to signaling concerns. Denote by $\mathcal{I}_{\mathrm{D}}$ all the information D has when making his decision. Write

(D.5) $$\mathbf{E}^{\mathrm{D}}\left[\Pi(\boldsymbol{c}) \mid \mathcal{I}_{\mathrm{D}}\right] = \xi \mathbb{P}_{\boldsymbol{c}}(a = \mathrm{H} \mid d = 1, \mathfrak{p}, I_{\mathrm{D}} = 1) + (1-\xi)\mathbb{P}_{\boldsymbol{c}}(a = \mathrm{H} \mid d = 1, \mathfrak{p}, I_{\mathrm{D}} = 0).$$

This says that D's interim expectation of perception payoffs can be written as a convex combination (involving a weight ξ that depends on $\mathcal{I}_{\mathrm{D}}$) of conditional probabilities of $a = \mathrm{H}$ *given* the value of I_{D}. The probabilities assessed by O depend on the cutoffs used, hence the subscripts $\boldsymbol{c}$. Note that under (Broadcast, CK), $\xi = 1$, while under (Broadcast, No CK), ξ is not 1 because the probability of Seeding is positive and the seed set S is a proper (strict) subset of N. Now, by P1(b), the first probability (the one being multiplied by ξ) is smaller than the second probability (the one being multiplied by $1-\xi$), by P1(b). This formalizes the claim that signaling concerns could not be greater than they are in the (Broadcast, CK) case. Applying SM1 finishes the proof.

(b) (Seed, CK) has more seeking than (Broadcast, CK).

Considering the signaling contribution to payoffs: for any given cutoffs, just as in (a), we can write the update of the Observer (given either value of $\mathfrak{p}$) as a convex combination conditioning on values of I_{D}. The term corresponding to $I_{\mathrm{D}} = 1$ is the same across the two treatments, and the term corresponding to $I_{\mathrm{D}} = 0$ involves a strictly lower posterior that $a = \mathrm{H}$. Only the first term is nonzero in the (Broadcast, CK) treatment, while both contribute in the (Seed, CK) treatment. Turning now to the direct payoffs, $I_{\mathrm{S}} = 1$ is known in both cases. By Assumption P2(b), $\Delta^{(0,1)}(\omega) \leq \Delta^{(1,1)}(\omega)$ for every ω.

Applying SM1 to the two W functions gives the result.

D.3.3. *Proof of Proposition 2.* First, under (Seed, CK), D is certain that information is in the Town Square, while under (Seed, No CK) this probability is strictly less than 1 assuming $I_{\mathrm{D}} = 0$. Thus $\Delta^{(0,\mathrm{CK:Seed})}(\omega)$ is pointwise strictly greater than $\Delta^{(0,\mathrm{No\ CK})}(\omega)$. By compactness of $\mathcal{C}$, it is strictly greater for all $\omega \in \mathcal{C}$, by at least a positive quantity $\nu > 0$.

Now we turn to signaling concerns. Condition first on $I_{\mathrm{D}} = 0$. By the argument given in the main text, once $\overline{k}/n$ is small enough, in the decomposition of (D.5) the weight on the $I_{\mathrm{D}} = 1$ term under either value of $\mathfrak{p}$ is arbitrarily small. Thus, the difference between signaling payoffs under $\mathfrak{p} = $ No CK and under $\mathfrak{p} = $ CK:Seed is less than ν. Thus we see $W^{(0,\mathfrak{p})}$ strictly increases pointwise for each $\omega, \boldsymbol{c} \in \mathcal{C}$ when we move from $\mathfrak{p} = $ No CK to $\mathfrak{p} = $ CK:Seed.

Because the realizations with $I_{\mathrm{D}} = 1$ become very unlikely (by smallness of $\overline{k}/n$), we can apply SM2 to finish the proof.

D.3.4. *Proof of Proposition 3.* We now state a formal version of Lemma 1, whose proof follows by Bayes' rule.

LEMMA 1. Fix any $\epsilon > 0$. Then there is a δ (depending on this ϵ) so that if $(1-\chi_{\text{Broadcast}}) < \delta(\overline{k}/n)^2$, then conditional on $\mathfrak{p} =$ No CK and any realizations of I_{D} and I_{O}, the probability that $\mathfrak{b} =$ Broadcast is at most ϵ.

Now, to prove the proposition in several steps. First, we will show that (Seed, No CK) has a level of seeking arbitrarily close to the one when it is common knowledge that $I_{\text{S}} = 0$ and $I_{\text{D}} = 0$.

Consider (Seed, No CK). Condition on $I_{\text{D}} = 0$. When D receives no information ($I_{\text{D}} = 0$), the fact that β_{Seed} is small means that his expectations approximate those when $I_{\text{S}} = 0$. Thus, his direct benefits as a function of ω are arbitrarily close to $\Delta^{(0,0)}$ on the compact set $\mathcal{C}$. Moreover, in (Seed, No CK), conditioning on $I_{\text{D}} = 0$, D is certain that $\mathfrak{b} \neq$ Broadcast, and thus (because the probability of seeding is small) he believes that $I_{\text{O}} = 0$ with high probability, and thus signaling concerns are uniformly bounded by an arbitrarily small number on $\mathcal{C}$. By the full support assumption on $\Delta^{(0,0)}$, it follows that for any cutoffs, there is an arbitrarily small measure of ω for which the decision differs from the case where Π is exactly zero. Finally, applying SM2 shows that the conclusion extends even when we take into account the $I_{\text{D}} = 1$ realizations.

Now consider (Broadcast, No CK), every realized D is certain that $I_{\text{S}} = 1$ and thus assesses the direct benefits to be greater than his $I_{\text{D}} = 0$ counterpart, by an amount bounded away from 0, as in Proposition 2. Fourth, under (Broadcast, No CK), signaling concerns are negligible, as follows. By the lemma, conditional I_{D}, D is nearly certain that $\mathfrak{b} \neq$ Broadcast. The probability of $\mathfrak{b} =$ Seed is small. Putting these facts together, D is also nearly certain that $I_{\text{O}} = 0$. Thus, in the decomposition of (D.5) the weight on the $I_{\text{D}} = 1$ term under either value of $\mathfrak{p}$ is arbitrarily small. Continuing from that point just as in the proof of Proposition 2, we conclude that signaling concerns are negligible. Thus, seeking rates are as if it is common knowledge that $I_{\text{S}} = 1$ and $I_{\text{D}} = 0$.

By P2, there is more seeking when it is common knowledge that $I_{\text{S}} = 1$ and $I_{\text{D}} = 0$ than when it is common knowledge that $I_{\text{S}} = 0$ and $I_{\text{D}} = 0$ (this follows by a simple comparison of direct payoffs without any signaling concerns).

Appendix E. Alternative Models

E.1. **Details on active sharing with image concerns.** we will now argue that the theory has a harder time accounting for the fact that (Broadcast, No CK) and (Seed, CK) have comparable seeking rates. Consider (Seed, CK) and let $\bar{c}$ denote the average number of conversations that are caused by a given seed's being informed. In order to explain the success of (Seed, CK) through sharing alone, $\bar{c}$ must be fairly large: each seed's sharing must be directly or indirectly responsible for considerable conversations (for example, through spontaneous "did you hear" sharing and resharing).[62] On the other hand, evidence from (Seed, No CK) suggests that $\bar{c}$ is actually small. The reason is that each informed agent in (Seed, No CK) is just like any villager in (Broadcast, No CK), and we know from the high volume in (Broadcast, No CK) that in this treatment individuals are willing to bring up the information at a substantial rate. Thus, a substantial fraction of seeds should be willing to initiate the process of diffusion in (Seed, No CK), and each of these should then lead to about $\bar{c}$ conversations. But, contrary to this prediction, we see volume in (Seed No, CK), comparable to villages where we did not intervene at all.

E.2. **Supply Effects: Information as a Public Good.** The core model of Chandrasekhar, Golub, and Yang (2018) and its application to our setting focuses on seeking effort or endogenous participation in learning. A different kind of explanation focuses on the effort of those informed to understand, filter, and communicate the information in a useful way to others. The simplest framework to capture this is a model of public goods provision and free-riding. This class of model has been studied extensively in a development context, and we rely on arguments from Banerjee, Iyer, and Somanathan (2007) to explain why supply-side effects are unlikely to explain our results.

A robust point within such public goods models is that enlarging the set of people who are able to provide a public good should not, in equilibrium, reduce its aggregate provision. Indeed, if anything provision should slightly increase, which is contrary to our empirical results.

For a simple model, consider a situation where those initially given information have the opportunity to provide the public good of processing and disseminating it to others. There are n agents, and each of those informed believes that k in total are able to contribute. Every i who has information invests an effort $z_i \geq 0$ in transmitting. Their payoffs are given by

$$U_i(z_1, \ldots, z_n) = V\left(\sum_i z_i\right) - cz_i.$$

[62] Recall that seeds make up 10% of the population, while (Broadcast, No CK) and (Seed, CK) have similar volume.

Here V is an increasing, smooth function with $V'(z)$ tending to 0 at large arguments z, and $c > 0$ is a cost parameter. Those who are unable to contribute are constrained to $z_i = 0$ and are passive. The key fact, which is formalized for instance by Banerjee, Iyer, and Somanathan (2007), is that at any equilibrium with some people contribution, for those contributing we have

$$V'\left(\sum_i z_i\right) = c, \tag{E.1}$$

so the aggregate level of contribution cannot depend on n or k. The intuition is simple: the free-riding problem is self-limiting, at least in the sense of aggregate (though not per-person) provision. If more agents try to free-ride, then others have more reason to provide the good. A similar force is present in the network model of Galeotti and Goyal (2010): there, endogenously, networks form so that only a few people provide the public good but everyone can access it, and a larger number of potential providers does not make for less provision.

If agents have a private benefit term in their utility function, $v_i(z_i)$, where v is increasing and $v'(z_i) > c$ for $z_i \in [0, \delta)$, then as long as there are sufficiently many agents who can provide the public good, the amount provided will be at least $k\delta$—a lower bound which is increasing in k. A similar argument applies if only some agents have such a v term.

Thus, natural public goods theories do not predict a decrease in the amount of overall provision, and thus in overall learning, as k (the number of potential providers) increases. One can, of course, elaborate these models with stochastic k and idiosyncratic c_i, but the basic intuition described above is quite robust.

One further supply-side effect to consider is one of social obligation. If the seeds are publicly "deputized," as they are in the CK treatment, each may face stronger incentives to provide information relative to a situation in which provision opportunities are diffuse. Though this is outside a basic public goods model, our evidence on seed effort does not support this hypothesis.

E.2.1. *Application to Experiment.* The number of people, k, who can contribute is either $k = 5$ or $k = n$. Under common knowledge, this matches up with the beliefs agents hold, so in this sense the simple model is faithful to the experiment. Thus, the basic public goods theory predicts (contrary to the demand-side theory) that holding CK fixed and moving from Seed to Broadcast should not hurt aggregate provision.

When common knowledge is not present, agents will have beliefs about k. But as long as their beliefs about k are reasonably consistent (e.g., agents have common priors about it), the essence of the above argument goes through: a stochastic version of (E.1) still holds, and changes in beliefs about k alone should not lead to large swings in provision.

This model is inconsistent with our empirical findings for several reasons. First, aggregate provision of effort cannot decline, as established above. If the number of people a typical subject in our random sample conversed with measures conversational effort, this means that the number of conversations for the average person must not decline. Column 1 of Table 5 shows that, conditional on common knowledge, going from $k = 5$ to $k = n$ corresponds to a 61% decline in the number of conversations ($p = 0.029$), which means that aggregate contribution to conversations must be decreasing.

Second, the model suggests that the amount of value being generated cannot decline, since after all otherwise a given individual would have an incentive to put in some more effort to gain more marginal benefit. Here, we can measure this either through knowledge or choice quality. Turning to Table 5, recall that columns 1 (for knowledge) and 2 (for choice) show robust declines in aggregate social learning and quality of choice when we go from $k = 5$ to $k = n$ under common knowledge ($p = 0.0621$ and $p = 0.104$).

E.3. **Tagged Information Aggregation.** There is an undirected graph $G = (N, E)$ of potential communication opportunities, corresponding to the social network with nodes N and edges E. At time 0, agents are endowed with certain information, the realization of a random variable S_i. (In our application, this represents one's degree of understanding of the information delivered in the intervention.) At each discrete time $t = 1, 2, \ldots$ a subset $E_t \subseteq E$ of agents who can communicate is realized randomly.[63] We make no assumptions on this process: it may involve arbitrary correlations, etc. If agents i and j are able to communicate at time t, they send each other messages, with the $i \to j$ message $m_{ij,t}$ reaching its destination with probability $p_{ij,t}$. Again, we make no assumptions on these numbers. Critically, information is "tagged." This means that at time t, agent i's information, $I_{i,t}$, consists of a set of signals labeled by their origin (formally, a set of pairs (k, S_k)). When agent i sends a message to j, the message reveals his whole information set I_t, which then is incorporated into j's information. Consider any improvement in initial information—making the profile of initial signal random variables $(S_i)_{i\in N}$ more informative in the Blackwell sense to obtain a new profile $(\widetilde{S}_i)_{i\in N}$. Then, holding fixed the parameters of the model, at any time t and for any agent i, the information $\widetilde{I}_{i,t}$ dominates $\widetilde{I}_{i,t}$.[64]

E.4. **Herding model.** We briefly review the notation of the Lobel and Sadler (2015) model, paraphrasing their Section 2. Agents, indexed by natural numbers n which correspond to the time they move, sequentially make choices $x_n \in \{0, 1\}$, which can be thought of making the correct choice or statement about the new currency. Agents receive a positive payoff from

[63]We omit formal notation for the probability space in the background.

[64]Formally, if we order information sets by containment, then under this order $\widetilde{I}_{i,t}$ first-order stochastically dominates $I_{i,t}$.

matching the state $\theta \in \{0,1\}$, and zero otherwise. In contrast to the tagging model, this is a maximally coarsened mode of communication. Each individual, when acting, observes two things: a private signal $s_n \in \mathcal{S}$, and the actions of a set of predecessors $B(n)$, which may be drawn with randomness. This allows us to encode network structure into the model. Private signals are conditionally independent given the true state θ.

Lobel and Sadler (2015) show that in equilibrium, the decisions of all sufficiently late-moving agents (those with high n) are at least as good as those decisions that would be made based on s_n alone, for the most informative possible realizations of s_n. To state this more formally, they define the private belief p_n as the belief about θ induced by n's signal, and define the strongest possible private beliefs to be the extreme points of the support of p_n, which they denote by $\underline{\beta}$ and $\overline{\beta}$. So, more formally, Lobel and Sadler (2015) show that the decisions of all sufficiently late-moving agents achieve essentially the utility that would be achieved by getting one of the strongest possible private signals. This requires some conditions on the network structure. The simplest of these (in their Theorem 1) is that individuals' neighborhoods are independent, and each late-moving agent has paths of observation leading back to arbitrarily many prior movers' choices.

Though in the sequential social learning model, equilibrium outcomes may be nonmontonic in signal endowments, the Lobel-Sadler lower bound described above is monotonic in signal endowments: when we make everyone's initial information better, the $\underline{\beta}$ and $\overline{\beta}$ become more extreme (corresponding to stronger signals and better decisions) and the lower bound is strengthened.

ONLINE APPENDIX: NOT FOR PUBLICATION

Appendix F. Heterogeneity by Ability

Table F.1. Ability-Treatment Interactions

VARIABLES	(1) Volume of conversations	(2) # Secondary conversations	(3) # Primary conversations	(4) Knowledge	(5) Chose 500
H	-0.131	-0.00957	-0.121	0.0104	0.00347
	(0.253)	(0.237)	(0.0784)	(0.0134)	(0.0289)
	[0.605]	[0.968]	[0.122]	[0.441]	[0.904]
CK	0.606	0.468	0.138	0.0452	0.0213
	(0.371)	(0.315)	(0.134)	(0.0157)	(0.0304)
	[0.102]	[0.137]	[0.301]	[0.00409]	[0.483]
Broadcast	0.620	0.545	0.0749	0.0454	0.0559
	(0.388)	(0.352)	(0.127)	(0.0168)	(0.0401)
	[0.110]	[0.122]	[0.555]	[0.00697]	[0.163]
Broadcast×CK	-0.730	-0.678	-0.0520	-0.0741	-0.0800
	(0.579)	(0.505)	(0.222)	(0.0234)	(0.0527)
	[0.207]	[0.179]	[0.815]	[0.00150]	[0.130]
H×CK	0.0940	-0.0981	0.192	-0.0220	0.0186
	(0.442)	(0.393)	(0.159)	(0.0168)	(0.0444)
	[0.832]	[0.803]	[0.228]	[0.192]	[0.675]
H×Broadcast	0.685	0.363	0.322	-0.0229	0.00782
	(0.544)	(0.528)	(0.124)	(0.0176)	(0.0476)
	[0.208]	[0.492]	[0.00913]	[0.192]	[0.869]
H×Broadcast×CK	-1.614	-1.051	-0.562	0.0394	-0.0250
	(0.742)	(0.675)	(0.264)	(0.0232)	(0.0653)
	[0.0297]	[0.119]	[0.0330]	[0.0892]	[0.702]
Observations	944	944	944	948	935
Seed, No CK, Not H Mean	0.492	0.373	0.119	0.551	0.0480
CK + BC × CK = 0 p-val	0.761	0.560	0.621	0.0747	0.178
BC + BC × CK = 0 p-val	0.794	0.703	0.902	0.0598	0.486
CK = BC p-val	0.976	0.847	0.633	0.990	0.429
H × CK + H × BC × CK = 0 p-val	0.0127	0.0402	0.0845	0.266	0.897
H × CK = H × BC	0.328	0.421	0.424	0.952	0.836
CK + BC × CK + H × CK + H × BC × CK = 0 p-val	0.00121	0.00323	0.0562	0.476	0.0880
BC + BC × CK + H × BC + H × BC × CK = 0 p-val	0.00418	0.00518	0.141	0.364	0.235

Notes: All columns control for randomization strata (subdistrict) fixed effects. Other controls for each column selected with PDS Lasso from date and time of entry into the village, caste category of the treatment hamlet, distance from the village to an urban center, and respondent-level controls such as age, gender, literacy and potential seed status. Standard errors (clustered at the village-level) are reported in parentheses and p-values are reported in brackets.

Table F.2. Ability-Treatment Group Interactions

VARIABLES	(1) Any Conversation
H	-0.00511
	(0.0320)
	[0.873]
Seed × CK	0.139
	(0.0517)
	[0.00762]
Broadcast × No CK	0.109
	(0.0595)
	[0.0679]
Broadcast × CK	0.129
	(0.0497)
	[0.0102]
H ×Seed× CK	0.00973
	(0.0583)
	[0.868]
H × Broadcast × No CK	0.111
	(0.0597)
	[0.0642]
H × Broadcast × CK	-0.118
	(0.0569)
	[0.0400]
Observations	944

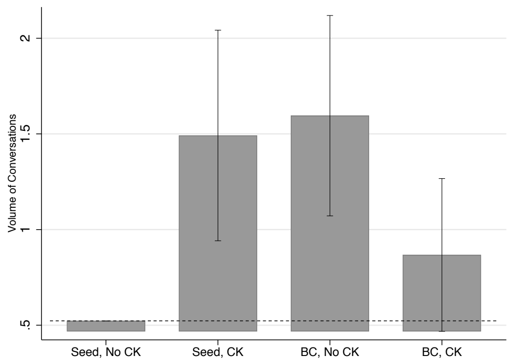
Notes: This table regresses a binary indicator of whether or not the individual had any conversation on their ability interacted with their treatment group. These results are used to generate Figure 6; for interpretability, this regression only controls for randomization strata (subdistrict) fixed effects. Standard errors (clustered at the village-level) are reported in parentheses and p-values are reported in brackets.

Appendix G. Long vs. Short Treatments

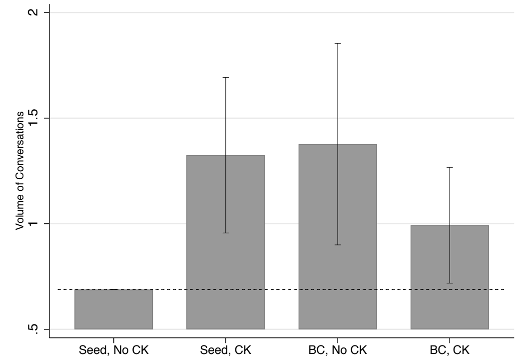
Table G.1. Short vs. Long

VARIABLES	(1) Volume	(2) Knowledge	(3) Chose 500
Long	-0.286	-0.00652	-0.0193
	(0.244)	(0.00927)	(0.0176)
	[0.241]	[0.482]	[0.274]
Observations	1,078	1,082	1,067
Short Mean	1.136	0.583	0.0954

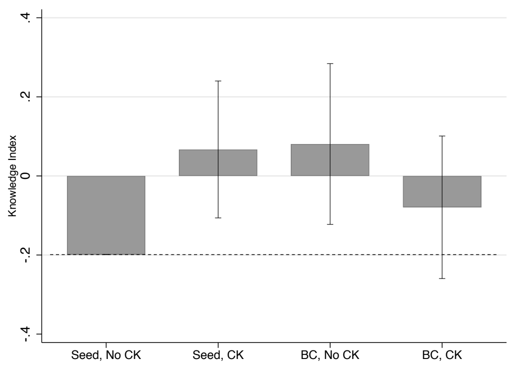
Notes: All columns control for randomization strata (subdistrict) fixed effects. Other controls for each column selected with PDS Lasso from date and time of entry into the village, caste category of the treatment hamlet, distance from the village to an urban center, and respondent-level controls such as age, gender, literacy and potential seed status. Standard errors (clustered at the village-level) are reported in parentheses and p-values are reported in brackets.



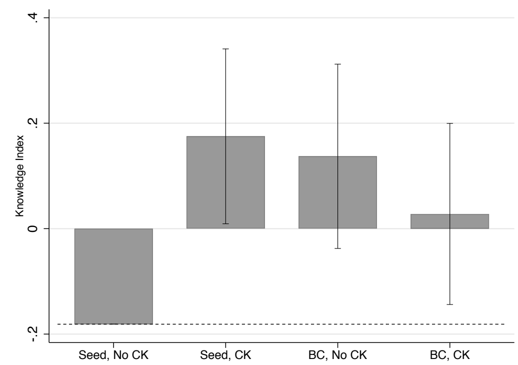
(A) Volume of conversations: Short



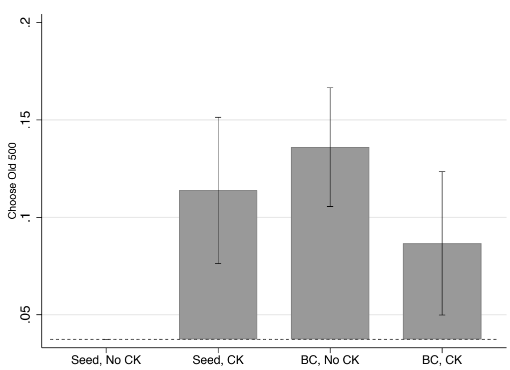
(B) Volume of conversations: Long



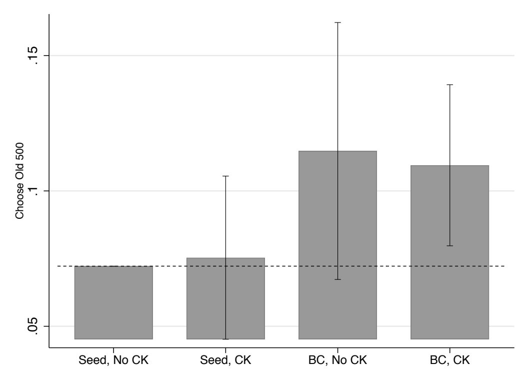
(C) Knowledge error: Short



(D) Knowledge error: Long



(E) Chose old 500: Short



(F) Chose old 500: Long

Figure G.4. Raw Data: Core Experimental Outcomes by Pamphlet Length

Appendix H. Other choice and knowledge metrics

Recall that because we randomized content, we have variation in whether the questions we ask about in the endline were actually provided to the villagers and also how relevant the information was. Table H.1 looks at whether facts are more likely to be known if (a) they were actually the ones provided in the information pamphlet to the village and (b) whether they were ex-ante deemed to be more useful to villagers. This would tell us whether there were complementarities and filtering occuring in the social learning process. The analysis is conducted on a person-fact level. Thus, it is a panel of the respondent's answers to each of the 34 facts asked in the endline survey.

In columns 1 and 2, for facts that were not provided and not useful respectively, we see that neither (Seed, CK) nor (Broadcast, No CK) is distinguishable from (Seed, No CK). However, when we look at the effect on knowledge of facts that were provided during information delivery, adding Common Knowledge to the Seed treatment increases knowledge by 15.5% (column 1, $p = 0.014$). Under no Common Knowledge, Broadcast increases knowledge by 13.6% (column (1), $p = 0.0345$) relative to Seed. Similarly, in column 2 we see that holding useful facts fixed, (Seed, CK) increases knowledge by 6.8% ($p = 0.008$) and (Broadcast, No CK) increases knowledge by 6.1% ($p = 0.0345$), compared to (Seed, No CK). We can conclude that the core effects on aggregation are being driven by facts that were provided during information delivery and facts that were deemend useful.

Next we turn to the fact that even if the subject rejected the Rs. 500 in favor of a 3-5 day IOU for either Rs. 200 in non-demonetized notes or Rs. 200 worth of dal, we know which they picked. Table H.3 explores this. Column 1 looks at a regression where the outcome variable is a dummy for picking the dal option. We can see that relative to (Seed, No CK), adding common knowledge considerably reduces the probability of selecting dal which corresponds to a 15.6% decline ($p = 0.135$). We also see a 14% decrease in the probability of selecting dal when going from (Seed, No CK) to (Broadcast, No CK) ($p = 0.138$). The interaction of broadcast with common knowledge has a large point estimate but is extremely noisy, however.

Note that the above says nothing about where the mass that moves away from dal ends up going. In columns 2 and 3, we present the results of a multinomial logit, where the omitted category is dal and the first column is Rs. 200 relative to dal and the second is Rs. 500 relative to dal. We see that going to (Seed, CK) from (Seed, No CK) leads to a 3.4pp increase in the probability of selecting the IOU for Rs. 200 in cash instead of dal, relative to a mean rate of selection of Rs. 200 of 40.8% ($p = 0.285$). However we cannot detect any broadcast or broadcast interacted with common knowledge effects. When we compare the choice of Rs. 500 relative to dal, the resulting marginal changes in the probability of picking

Table H.1. Heterogeneity in knowledge

VARIABLES	(1) Knowledge (Told)	(2) Knowledge (Useful)
CK	-0.0239	-0.0352
	(0.0282)	(0.0669)
	[0.396]	[0.599]
Broadcast	-0.0189	-0.0325
	(0.0270)	(0.0658)
	[0.486]	[0.622]
Told/Useful	-0.0840	0.0750
	(0.0410)	(0.0488)
	[0.0419]	[0.126]
Broadcast × CK	0.0160	0.117
	(0.0390)	(0.0941)
	[0.682]	[0.216]
CK × Told/Useful Facts	0.112	0.0661
	(0.0596)	(0.0686)
	[0.0614]	[0.336]
BC × Told/Useful Facts	0.0962	0.0606
	(0.0575)	(0.0676)
	[0.0962]	[0.371]
BC × CK × Told/Useful Facts	-0.125	-0.163
	(0.0852)	(0.0975)
	[0.145]	[0.0957]
Observations	36,788	36,788
Seed, No CK, Untold/Not useful Mean	0.569	0.457
CK + CK × Told/Useful = 0 p-val	0.0140	0.00829
BC + BC × Told/Useful = 0 p-val	0.0345	0.0345

Notes: All columns control for randomization strata (subdistrict) fixed effects. They also control for date and time of entry into the village, caste category of the treatment hamlet and distance from the village to an urban center. Respondent-level controls include age, gender, literacy and potential seed status. Column (1) displays effects on knowledge if the fact being asked about was told during information delivery. Column (2) displays effects on knowledge if the fact being asked about is a useful fact or not. Standard errors (clustered at the village-level) are reported in parentheses and p-values are reported in brackets.

Rs. 500 look much like our main results: a 4.7pp increase when we move to (Seed, CK), a 6.9pp increase when we move to (Broadcast, No CK), and a relative decline of 4.1pp when going from (Broadcast, No CK) to (Broadcast, CK), all on a base rate of picking Rs. 500 at 5.9%.

Recall that we had two successful information dissemination strategies: (Seed, CK) and (Broadcast, No CK). We find that in the former, but not the latter, we also see movement away from dal in favor of Rs. 200 in cash. This suggests that at least some part of the

Table H.2. Did the Broadcast, Common Knowledge Group Learn Anything?

VARIABLES	(1) Volume	(2) Knowledge Index	(3) Knowledge Panel (Told)	(4) Chose 500
Broadcast x Common Knowledge	-0.119	0.0129	0.0654	-0.00805
	(0.242)	(0.0134)	(0.0350)	(0.0227)
	[0.623]	[0.337]	[0.0633]	[0.723]
Observations	1,078	1,082	36,788	1,067
Mean: Seed x No CK, Non-seed HH	0.868	0.580	0.489	0.0677

Notes: Regressions compare outcomes for the Broadcast, Common Knowledge treatment relative to the Seed, No Common Knowledge treatment. The regression coefficient only includes households that were not potential seeds. All columns control for randomization strata (subdistrict) fixed effects. They also control for date and time of entry into the village, caste category of the treatment hamlet and distance from the village to an urban center. Respondent-level controls include age, gender, literacy and potential seed status. Columns (1), (2), and (4) use the same specifications as Table ??. Column (3) considers a respondent x question panel and focuses only on knowledge of the facts that were told in the respondent's village. Standard errors (clustered at the village-level) are reported in parentheses and p-values are reported in brackets.

misinformation involved decreased confidence in Rs. 100 notes as well, because otherwise Rs. 200 in cash should dominate dal.

Finally, because the dal, equivalent cash, and Rs. 500 are welfare-ordered, in that order, we have in column 4 an ordinal logit which shows again that (Seed, CK) and (Broadcast, No CK), relative to (Seed, No CK) improve outcomes in choice quality.

Table H.3. Other choice outcomes

VARIABLES	(1) OLS Chose dal	(2) Multinomial Logit Chose 200	(3) Multinomial Logit Chose 500	(4) Ordinal Logit Choice
CK	-0.0832	0.257	0.700	0.377
	(0.0554)	(0.241)	(0.357)	(0.208)
	[0.135]	[0.285]	[0.0496]	[0.0699]
Broadcast	-0.0756	0.124	0.932	0.398
	(0.0507)	(0.223)	(0.340)	(0.193)
	[0.138]	[0.578]	[0.00611]	[0.0396]
Broadcast $\times$ CK	0.0887	-0.117	-1.170	-0.523
	(0.0782)	(0.332)	(0.464)	(0.297)
	[0.258]	[0.724]	[0.0116]	[0.0780]
Observations	1,067	1,067	1,067	1,067
Seed, No CK Mean	0.533	0.408	0.059	
CK + BC $\times$ CK = 0 p-val	0.914	0.539	0.126	0.451
BC + BC $\times$ CK = 0 p-val	0.826	0.978	0.467	0.567

Notes: All columns control for randomization strata (subdistrict) fixed effects. They also control for date and time of entry into the village, caste category of the treatment hamlet and distance from the village to an urban center. Respondent-level controls include age, gender, literacy and potential seed status. Standard errors (clustered at the village-level) are reported in parentheses and p-values are reported in brackets.

Our study was certainly not designed to quantify the costs and benefits of demonitization in India. However, by studying misinformation and its remedies during the SBN deposit window, a few, more modest lessons emerge. First, we show that in the context of rural Orissa, while basic policy knowledge was near-universal, individuals still had a poor grasp on some of the most basic policy rules at baseline. This suggests that there was substantial room for improvement in the quality of outreach between the policy makers and villagers. Second, in our experiment, we show that decisions are impacted by the provision of information. Individuals in treatments that lead to better community wide knowledge of the policy do change their incentivized choices and are more likely to recognize that an old Rs. 500 note is more valuable than Rs. 200 in the days before the deadline. Moreover in the some treatment conditions associated with improved knowledge, namely (Seed, CK), individuals are more likely to choose currency over commodities of equivalent face value. This result suggests that a portion of the individuals preferring lentils over cash in our benchmark, non-intervention villages were likely doing so out of a loss of confidence in paper money. This observation relates back to the foundational macroeconomic literature on fiat money (Samuelson, 1958; Kiyotaki and Wright, 1989; Banerjee and Maskin, 1996; Wallace, 1980) and suggests that sowing confusion about the government's intervention in the currency undermines trust.

Appendix I. Heterogeneous Communication by Potential Seeds

Table I.1. How much more do potential seed households speak?

VARIABLES	(1) Volume of conversations	(2) # secondary conversations	(3) # primary conversations
Seed HH	0.608	0.0893	0.518
	(0.842)	(0.403)	(0.469)
	[0.470]	[0.824]	[0.269]
CK	0.514	0.317	0.198
	(0.296)	(0.248)	(0.100)
	[0.0823]	[0.202]	[0.0491]
Broadcast	0.723	0.543	0.180
	(0.357)	(0.328)	(0.104)
	[0.0425]	[0.0972]	[0.0818]
Broadcast × CK	-1.365	-1.060	-0.305
	(0.497)	(0.423)	(0.171)
	[0.00605]	[0.0121]	[0.0748]
Seed HH × CK	1.317	1.231	0.0853
	(1.477)	(1.128)	(0.622)
	[0.373]	[0.275]	[0.891]
Seed HH × BC	-0.497	-0.698	0.201
	(1.135)	(0.607)	(0.802)
	[0.661]	[0.250]	[0.802]
Seed HH × BC × CK	-0.926	0.0952	-1.021
	(1.839)	(1.480)	(0.894)
	[0.615]	[0.949]	[0.253]
Observations	1,078	1,078	1,078
Seed, No CK, Non-seed HH Mean	0.627	0.490	0.137
CK + BC × CK = 0 p-val	0.0134	0.0135	0.392
BC + BC × CK = 0 p-val	0.0381	0.0368	0.304
BC = CK p-val	0.0114	0.0275	0.0454

Notes: All columns control for randomization strata (subdistrict) fixed effects. Other controls for each column selected with PDS Lasso from date and time of entry into the village, caste category of the treatment hamlet, distance from the village to an urban center, and respondent-level controls such as age, gender, literacy and potential seed status. Standard errors (clustered at the village-level) are reported in parentheses and *p*-values are reported in brackets.

Table I.1 looks at how the volume of conversations changed by treatment, and in particular whether there was differential conversation participation by "seed households" relative to the others. Specifically, this allows us to ask if part of the positive effect on communication in (Seed, CK) relative to (Seed, No CK) is coming from the seed household itself putting in more effort and having more conversations. We remind the reader that every village (even broadcast treatments) has a set of "seed households." This is because the seeds were chosen using responses to the gossip survey that was conducted at baseline in each village.

In Table I.1, we see that our main results hold for the households that are not seeds: (1) adding common knowledge to seeding increases conversations, (2) broadcasting information to all households without common knowledge raises conversations relative to seeding, (3) broadcasting information to all households reduces conversations if there is common knowledge, and (4) adding common knowledge to broadcasting reduces conversations.

Turning to the seed households, there is a noisily estimated 1.3 increase in the conversation count for a Seed in CK relative to No CK ($p = 0.39$). If anything, this entirely comes from secondary conversations, and one cannot statistically reject an effect size of 0. Note that there is a 0.5 increase in conversations per random non-seeded households. This means that in a village of 50 households, there will be 23 extra conversations. If every seeded household gained 1.3 conversations, then this explains 6.5 or 29% of the increase in conversations. (Even if we assume that there are double the coefficient's number, so 13 conversations, this at best would only explain 56% of the increase in conversations.) Finally, note that by column 3, because the effect is not coming from primary conversations (i.e., purposeful seeking or advising behavior), any increase in seed conversations does not appear to be driven by the seed actively going out to explain the information to others, nor others actively seeking out the seeds. Taken together, this suggests that a primary driver of information aggregation here comes from decentralized conversations among non-seeds.

Appendix J. Heterogeneity by respondent gender

We present our main results in Table J.1, allowing for heterogeneous treatment effects by respondent gender. While women have fewer conversations about demonetization and know less overall, we find no evidence of treatment heterogeneity.

It is important to exercise caution when interpreting these results. Recall, our sampling strategy did not target a representative sample by gender; the survey enumerators simply asked to speak with any adult household member who was available at that time. For example, the types of women who responded to the surveys are likely to be more representative of female headed households and may be less subject to restrictive gender norms.

Table J.1. Engagement in social learning, knowledge and decision-making

VARIABLES	(1) Volume of conversations	(2) # secondary conversations	(3) # primary conversations	(4) Knowledge	(5) Chose 500
CK	0.535	0.303	0.232	0.0321	0.0659
	(0.358)	(0.296)	(0.137)	(0.0140)	(0.0295)
	[0.135]	[0.305]	[0.0898]	[0.0220]	[0.0253]
Broadcast	0.709	0.497	0.212	0.0209	0.0844
	(0.459)	(0.417)	(0.154)	(0.0151)	(0.0309)
	[0.122]	[0.233]	[0.167]	[0.167]	[0.00625]
Broadcast × CK	-1.551	-1.094	-0.457	-0.0538	-0.141
	(0.632)	(0.543)	(0.221)	(0.0198)	(0.0458)
	[0.0141]	[0.0439]	[0.0389]	[0.00649]	[0.00207]
CK × Female	0.425	0.484	-0.0591	-0.00573	-0.0523
	(0.526)	(0.436)	(0.163)	(0.0190)	(0.0525)
	[0.419]	[0.267]	[0.717]	[0.763]	[0.320]
Broadcast × Female	-0.00638	0.0722	-0.0785	0.0198	-0.0503
	(0.458)	(0.402)	(0.141)	(0.0197)	(0.0481)
	[0.989]	[0.858]	[0.577]	[0.315]	[0.296]
Broadcast × CK × Female	0.0592	-0.153	0.212	0.0129	0.0934
	(0.720)	(0.627)	(0.232)	(0.0276)	(0.0724)
	[0.934]	[0.808]	[0.361]	[0.640]	[0.197]
Female	-0.999	-0.842	-0.157	-0.0597	0.0180
	(0.304)	(0.247)	(0.108)	(0.0137)	(0.0320)
	[0.00100]	[0.000665]	[0.148]	[1.28e-05]	[0.574]
Observations	1,078	1,078	1,078	1,082	1,067
Seed, No CK Mean	0.893	0.711	0.183	0.580	0.0513
CK + BC × CK = 0 p-val	0.0275	0.0498	0.168	0.0890	0.0168
BC + BC × CK = 0 p-val	0.0309	0.0594	0.0783	0.00593	0.0542

Notes: All columns control for randomization strata (subdistrict) fixed effects. Other controls for each column selected with PDS Lasso from date and time of entry into the village, caste category of the treatment hamlet, distance from the village to an urban center, and respondent-level controls such as age, gender, literacy and potential seed status. Standard errors (clustered at the village-level) are reported in parentheses and p-values are reported in brackets.

Appendix K. Instrumenting for treatment assignment

Typically a village has one SCST hamlet and one GOBC hamlet. In conducting our intervention in a small sample of 16 villages, our field staff visited the wrong hamlet. However, we did an endline in these "missed" hamlets, which were intended to receive the treatment, as well though with a slightly smaller random sample. Here we present our main results where we only look at the set of hamlets originally that should have received treatments. We instrument for actual treatment assignment with intended treatment assignment.

Table K.1 and K.2 present versions of our main results with this IV strategy. We see that all our main results essentially go through.

Table K.1. Engagement in social learning

VARIABLES	(1) IV Volume of conversations	(2) IV # secondary conversations	(3) IV # primary conversations
CK	0.679	0.459	0.220
	(0.327)	(0.268)	(0.108)
	[0.0379]	[0.0872]	[0.0423]
Broadcast	0.888	0.619	0.269
	(0.379)	(0.340)	(0.140)
	[0.0190]	[0.0688]	[0.0551]
BC × CK	-1.720	-1.236	-0.484
	(0.548)	(0.459)	(0.199)
	[0.00171]	[0.00702]	[0.0152]
Observations	1,068	1,068	1,068
Seed, No CK Mean	0.651	0.514	0.137
CK + BC × CK = 0 p-val	0.00496	0.0149	0.0870
BC + BC × CK = 0 p-val	0.0199	0.0313	0.0771

Notes:All columns control for randomization strata (subdistrict) fixed effects. Other controls for each column selected with PDS Lasso from date and time of entry into the village, caste category of the treatment hamlet, distance from the village to an urban center, and respondent-level controls such as age, gender, literacy and potential seed status. CK, Broadcast and BC×CK are instrumented with CK in intended hamlet, Broadcast in intended hamlet and BC×CK in intended hamlet. Only outcomes from intended treatment hamlets are used. CK, Broadcast and BC×CK are instrumented with CK in intended hamlet, Broadcast in intended hamlet and BC×CK in intended hamlet. Standard errors (clustered at the village-level) are reported in parentheses and p-values are reported in brackets.

Table K.2. Knowledge and decision-making

VARIABLES	(1) IV Knowledge	(2) IV Chose 500
CK	0.0419	0.0464
	(0.0128)	(0.0225)
	[0.00103]	[0.0390]
Broadcast	0.0328	0.0656
	(0.0147)	(0.0278)
	[0.0263]	[0.0182]
BC × CK	-0.0639	-0.110
	(0.0196)	(0.0400)
	[0.00110]	[0.00611]
Observations	1,073	1,057
Seed, No CK Mean	0.564	0.0557
CK + BC × CK = 0 p-val	0.115	0.0393
BC + BC × CK = 0 p-val	0.00926	0.0933

Notes: All columns control for randomization strata (subdistrict) fixed effects. Other controls for each column selected with PDS Lasso from date and time of entry into the village, caste category of the treatment hamlet, distance from the village to an urban center, and respondent-level controls such as age, gender, literacy and potential seed status. CK, Broadcast and BC×CK are instrumented with CK in intended hamlet, Broadcast in intended hamlet and BC×CK in intended hamlet. Only outcomes from intended treatment hamlets are used. CK, Broadcast and BC×CK are instrumented with CK in intended hamlet, Broadcast in intended hamlet and BC×CK in intended hamlet. Standard errors (clustered at the village-level) are reported in parentheses and *p*-values are reported in brackets.

Appendix L. Dropping villages from new subdistrict

From our original sample we added 16 new villages from a new subdistrict. Unfortunately, the reassignment was not randomly done, which we discuss at length in Online Appendix M. To deal with this, here we repeat our main results dropping the set of 16 villages that were assigned a new subidstrict. Tables L.1 and L.2 show that all of our main results go through.

Table L.1. Engagement in social learning

VARIABLES	(1) Volume of conversations	(2) # secondary conversations	(3) # primary conversations
CK	0.596	0.389	0.206
	(0.325)	(0.268)	(0.109)
	[0.0670]	[0.146]	[0.0582]
Broadcast	0.690	0.496	0.194
	(0.357)	(0.322)	(0.129)
	[0.0535]	[0.124]	[0.131]
Broadcast × CK	-1.447	-1.064	-0.383
	(0.530)	(0.443)	(0.188)
	[0.00634]	[0.0165]	[0.0419]
Observations	1,020	1,020	1,020
Seed, No CK Mean	0.685	0.536	0.150
CK + BC × CK = 0 p-val	0.0181	0.0275	0.239
BC + BC × CK = 0 p-val	0.0342	0.0477	0.118

Notes: All columns control for randomization strata (subdistrict) fixed effects. Other controls for each column selected with PDS Lasso from date and time of entry into the village, caste category of the treatment hamlet, distance from the village to an urban center, and respondent-level controls such as age, gender, literacy and potential seed status. Standard errors (clustered at the village-level) are reported in parentheses and p-values are reported in brackets. Villages from newly added strata are not included in this sample.

Table L.2. Knowledge and decision-making

VARIABLES	(1) Knowledge	(2) Chose 500
CK	0.0367	0.0531
	(0.0128)	(0.0229)
	[0.00404]	[0.0207]
Broadcast	0.0274	0.0732
	(0.0142)	(0.0269)
	[0.0537]	[0.00657]
Broadcast × CK	-0.0539	-0.116
	(0.0190)	(0.0388)
	[0.00451]	[0.00283]
Observations	1,024	1,009
Seed, No CK Mean	0.562	0.0534
CK + BC × CK = 0 p-val	0.204	0.0340
BC + BC × CK = 0 p-val	0.0249	0.0934

Notes: All columns control for randomization strata (subdistrict) fixed effects. Other controls for each column selected with PDS Lasso from date and time of entry into the village, caste category of the treatment hamlet, distance from the village to an urban center, and respondent-level controls such as age, gender, literacy and potential seed status. Standard errors (clustered at the village-level) are reported in parentheses and p-values are reported in brackets. Villages from newly added strata are not included in this sample.

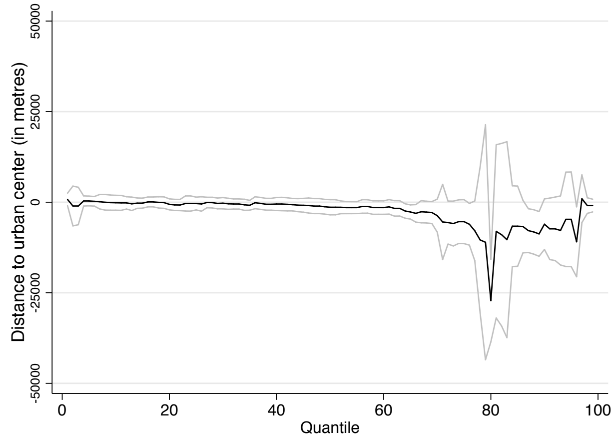
Appendix M. Status Quo Appendix

We also attempted to get 30 villages of data where we did not intervene whatsoever and instead only collected endline data. We call these the “status quo” villages. Unfortunately, these villages are not entirely comparable to our core set. “Status quo” villages are considerably more likely to be peri-urban/neighboring a city, larger in size, more educated, and due to survey logistics were surveyed much closer to the deadline. This was due to the following implementation failures: (1) mechanically, survey teams were less familiar with the “status quo” villages because no treatment was delivered, and unfortunately, they went to these villages after intervention villages. This both pushed the visits closer to the deadline and later in any given day; (2) a share of initially selected “status quo” villages were dropped and the replacements were not randomly drawn from a list of a villages in a subdistrict, placing them city-adjacent; (3) there was geographic imbalance in the initial randomization between “status quo” and intervention villages. Therefore, we do not include these along with the analysis.

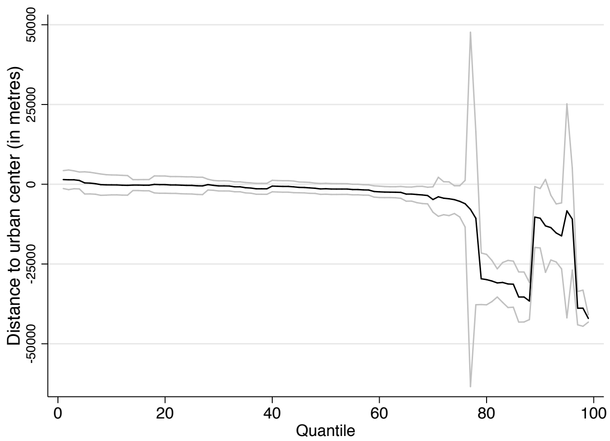
We can include “status quo” in a regression analysis to compare it to our other treatments, but we need to keep in mind that this is observational, and relies on controlling for the distribution of distance from cities, survey timing, etc. That means when we compare to “status quo” we should interpret it with caution. When we do this, we find suggestive evidence that the number of conversations between “status quo” villages and (Seed, No CK) is similar, while (Seed, CK) exceeds “status quo”. Our information and choice analysis have commensurate estimates, but results are noisier.

Recall that the goal of the paper is to understand how changes to the seeding structure affect endogenous participation and subsequent knowledge and choice. The “status quo” treatment cell is unnecessary for accomplishing this.

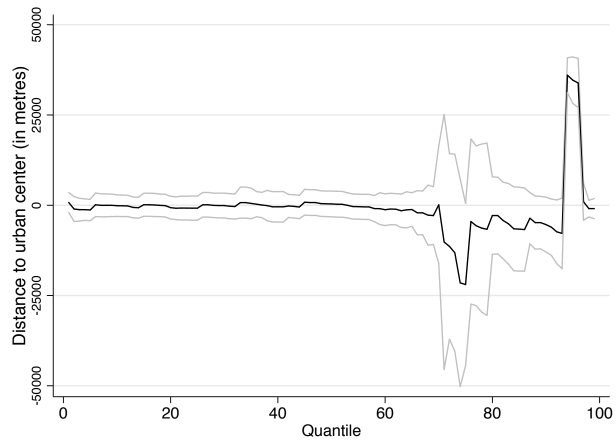
We begin by looking at the distance distributions for the “status quo” and intervention villages. Figure M.1, Panels A, B, and C present coefficients from a quantile regression of distance from urban center against “status quo”, conditional on caste of the hamlet. Panel A conditions on caste, and Panels B and C consider only data from GOBC and SC/ST, respectively. We see that “status quo” hamlets are much more likely to be considerably closer to an urban center particularly in the tail of the distribution.



(A) Controlling for hamlet caste



(B) Only General caste hamlets



(C) Only SC/ST hamlets

FIGURE M.1. Distance to urban center: status quo vs. treated

Table M.1. Imbalance: status quo vs. treated

VARIABLES	(1) Beyond 40kms of urban center	(2) Within 5kms or urban center	(3) Standardized entry time	(4) Survey day	(5) New strata	(6) Female	(7) Literate	(8) Has bank account	(9) Age	(10) Surveyed seed	(11) Surveyed seed
Control	-0.106	0.137	0.312	0.214	0.0488	-0.0223	-0.0349	-0.0101	0.937	0.0326	0.0232
	(0.0508)	(0.105)	(0.175)	(0.109)	(0.0601)	(0.0574)	(0.0427)	(0.0409)	(0.972)	(0.0230)	(0.0104)
	[0.0380]	[0.193]	[0.0764]	[0.0511]	[0.417]	[0.699]	[0.414]	[0.805]	[0.336]	[0.158]	[0.0266]
Observations	1,242	1,242	1,248	1,241	1,248	1,248	1,209	1,244	1,239	1,248	1,248
Treated Mean	0.166	0.345	-0.0539	3.660	0.0536	0.323	0.800	0.890	39.18	0.0518	0

Notes: Columns (1) and (2) are covariates describing distance from the village to an urban center. Column (10) is a dummy for if respondent was a potential seed. Column (11) is a dummy for if respondent was a potential controlling for if the household being surveyed was a potential seed household. Standard errors (clustered at the village-level) are reported in parentheses and p-values are reported in brackets.

Table M.1 presents information analogous to our prior balance table, to show that “status quo” is often imbalanced. Column 1 shows that these villages are much less likely to be very rural, defined as beyond 40km from the nearest city: 6% instead of 16% ($p = 0.038$). Column 2 shows that these villages are 13.7pp likely to be peri-urban, within 5km of a city ($p = 0.193$). These distance imbalances come from several issues. In the original randomization, we were unlucky and had some imbalance. This was compounded by the “status quo” villages not being drawn randomly from a list of villages in the replacement subdistrict (10% of the sample fall into this category and were all within the 61th percentile of distance to an urban center in the treatment distance distribution).

Column 3 and 4 look at time of entry. We see that they were much more likely to be visited later in the day (0.312 standard deviations later, $p = 0.076$) and later during the study period (0.2 days later, $p = 0.05$). The time of day matters because it can affect the composition of which members of which households are home (for instance whether they are working in the field or in town or are home). Furthermore, status quo villages are much more likely to be done about half a day later than the treatment villages.

Columns 5 - 9 show no detectable difference in terms of likelihood of being replaced, a female subject being surveyed, a literate subject being surveyed, the subject having a bank account, nor age. Columns 10 and 11 do show that the respondent is more likely to be a seed, and conditional on interviewing a seed household, the seed himself is more likely to be interviewed.

Table M.2. Experiment Outcomes: status quo vs. treated

VARIABLES	(1) Volume of conversations	(2) # secondary conversations	(3) # primary conversations	(4) Knowledge	(5) Chose 500
Seed	0.00619	0.0483	-0.0421	-0.0202	-0.0115
	(0.455)	(0.409)	(0.134)	(0.0183)	(0.0335)
	[0.989]	[0.906]	[0.753]	[0.272]	[0.732]
Seed × CK	0.688	0.342	0.346	0.0303	0.0399
	(0.345)	(0.276)	(0.125)	(0.0146)	(0.0296)
	[0.0471]	[0.216]	[0.00600]	[0.0392]	[0.180]
Broadcast	0.519	0.352	0.167	0.00244	0.0584
	(0.523)	(0.479)	(0.157)	(0.0160)	(0.0306)
	[0.323]	[0.464]	[0.289]	[0.879]	[0.0577]
Broadcast × CK	-0.854	-0.621	-0.233	-0.0144	-0.0421
	(0.442)	(0.408)	(0.159)	(0.0155)	(0.0290)
	[0.0547]	[0.130]	[0.144]	[0.354]	[0.149]
Observations	1,190	1,190	1,190	1,194	1,179
Status Quo Mean	1.116	0.939	0.177	0.588	0.0793
Seed + Seed × CK = 0 pval	0.128	0.325	0.0231	0.478	0.370
BC + BC × CK = Seed + Seed × CK	0.00294	0.0167	0.00576	0.119	0.725

Notes: All columns control for randomization strata (subdistrict) fixed effects. They also control for date and time of entry into the village, caste category of the treatment hamlet and distance from the village to an urban center. Respondent-level controls include age, gender, literacy and potential seed status. Standard errors (clustered at the village-level) are reported in parentheses and p-values are reported in brackets.

Against this backdrop, Table M.2 presents the main regressions of our paper, bringing in the status quo villages as well, as the omitted category. We are controlling for entry time, survey date, flexibly for distance, caste of hamlet, whether it was replaced, and subdistrict fixed effects. We find similar results to our main results. In column 1 we look at total volume of conversations. As one would have thought, (Seed, No CK) is not appreciably different from status quo, since we only handed out 5 pamphlets and there was no common knowledge of this. Meanwhile, (Seed, CK) is statistically distinguishable from (Seed, No CK), and corresponds to a 0.688 increase in the number of people spoken to relative to status quo ($p = 0.128$). We see that going from status quo to (Broadcast, No CK) leads to a large increase in the number of people spoken to, though this is not statistically distinguishable from zero ($p = 0.323$). However, we can precisely say that adding common knowledge to broadcast reduces the conversation rate relative to (Broadcast, No CK) ($p = 0.055$). And we also see that conditional on common knowledge, going from seeding to broadcast reduces conversations ($p = 0.003$). These same patterns largely hold in columns 2 and 3 across secondary and primary conversations, as well as in columns 4 and 5 across knowledge and choice.

Taken together, the evidence suggests that when controlling for sources of imbalance and failures in execution, status quo mostly behaves like (Seed, No CK), whereas (Seed, CK) and (Broadcast, No CK) perform better on conversation and choice metrics.

Appendix N. Attrition

Table N.1 presents p-values from a regression at the village level, among the 237 villages in our baseline, of whether a village dropped out of the study before endline on treatment assignment. We conduct all pairwise comparisons among (Seed, No CK), (Seed, CK), (Broadcast, No CK), (Broadcast, CK), and Status Quo. We find there is no differential attrition of village by treatment assignment. The attrition rates respectively are 7.4%, 5.66%, 5.77%, 2.1%, and 6.25%.

Table N.1. Attrition

SNCK - SCK	SNCK - BNCK	SNCK - BCK	SCK - BNCK	SCK - BCK	BNCK - BCK	SNCK - SQ	SCK - SQ	BNCK - SQ	BCK - SQ
.72	.74	.2	.98	.35	.34	.91	.84	.93	.39

Notes: p-values listed from pairwise comparisons of attrition rates.

Appendix O. Effect on Joint Distribution of Conversations and Information Quality

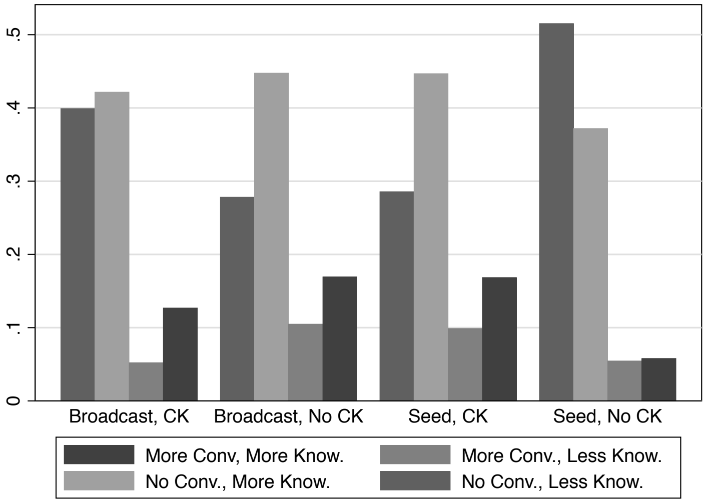
Here we look at how the joint distribution of conversations and information quality move. Table O.1 presents multinomial logistic regressions. In column 1, the outcome variable takes on values of "Conversations and High Knowledge", "Conversations and Low Knowledge," "No Conversations and High Knowledge," and "No Conversations and Low Knowledge". Therefore we look at whether as we move across treatments, for instance from (Seed, No CK) to (Seed, CK), whether the mass moves towards the joint outcome of both conversations going up and quality of information going up. This provides suggestive evidence consistent with social learning. Column 2 repeats the exercise but where information quality in this case is measured by whether the respondent chose the Rs. 500 note. Figure O.1 presents the same results with raw data.

We find that going from (Seed, No CK) to (Seed, CK) leads to a large increase in the mass of respondents who both have more conversations and have higher information quality (measured by knowledge and choice). The same is the case when comparing (Seed, No CK) to (Broadcast, No CK). However, we see that (Broadcast, No CK) is differentially less likely to both increase knowledge and conversations together, and more likely to push mass into the no conversations cells. This is consistent with a story wherein (Seed, CK) and (Broadcast, No CK) both encourage engagement in social learning whereas (Broadcast, No CK) discourages social learning.

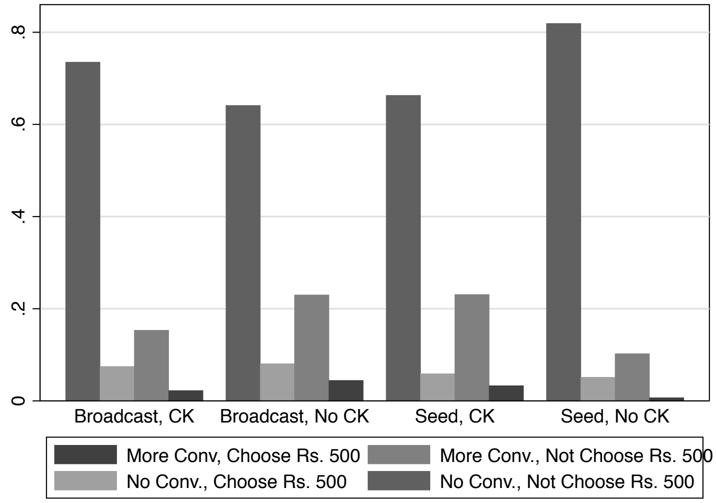
Table O.1. Joint distribution of conversations and information quality

	(1) Knowledge	(2) Rs. 500
Convo_Knowledge		
CK	1.603	1.682
	(0.330)	(0.799)
	[1.18e-06]	[0.0352]
Broadcast	1.648	1.963
	(0.416)	(0.867)
	[7.57e-05]	[0.0236]
Broadcast × CK	-2.351	-2.858
	(0.552)	(1.043)
	[2.02e-05]	[0.00614]
Convo_NoKnowledge		
CK	1.190	1.052
	(0.422)	(0.261)
	[0.00480]	[5.71e-05]
Broadcast	1.114	1.011
	(0.474)	(0.296)
	[0.0188]	[0.000640]
Broadcast × CK	-2.281	-1.661
	(0.667)	(0.405)
	[0.000622]	[4.06e-05]
NoConvo_Knowledge		
CK	0.775	0.350
	(0.279)	(0.362)
	[0.00542]	[0.333]
Broadcast	0.889	0.693
	(0.326)	(0.358)
	[0.00634]	[0.0530]
Broadcast × CK	-1.292	-0.791
	(0.439)	(0.532)
	[0.00324]	[0.137]
Observations	1,082	1,067
Convo, Knowledge: CK + BC × CK = 0 p-val	0.115	0.0342
Convo, Knowledge: BC + BC × CK = 0 p-val	0.0564	0.125
Convo, No Knowledge: CK + BC × CK = 0 p-val	0.0253	0.0503
Convo, No Knowledge: BC + BC × CK = 0 p-val	0.0113	0.0148
No Convo, Knowledge: CK + BC × CK = 0 p-val	0.130	0.288
No Convo, Knowledge: BC + BC × CK = 0 p-val	0.138	0.796

Notes: The table presents marginal effects from a multinomial regression on treatment. In each column the outcome variable consists of whether or not the participant had conversations about demonetization with a measure of information quality. In column 1 this measure is whether the participant has above average knowledge on our test. In column 2 this is whether the participant selected the Rs. 50 note. Standard errors (clustered at the village-level) are reported in parentheses and *p*-values are reported in brackets.



(A) Knowledge



(B) Choice of Rs. 500

FIGURE O.1. Joint distribution of conversations and information quality

Appendix P. Impacts on Seeking (Binary)

Table P.1. Engagement in Social Learning

VARIABLES	(1) Any Conversation	(2) Any secondary Conversation	(3) Any primary Conversation
CK	0.138	0.129	0.0561
	(0.0362)	(0.0336)	(0.0260)
	[0.000144]	[0.000125]	[0.0307]
Broadcast	0.135	0.117	0.0608
	(0.0523)	(0.0487)	(0.0331)
	[0.00976]	[0.0164]	[0.0665]
Broadcast × CK	-0.249	-0.218	-0.128
	(0.0688)	(0.0636)	(0.0483)
	[0.000285]	[0.000609]	[0.00802]
Observations	1,082	1,082	1,082
Seed, No CK Mean	0.113	0.0853	0.0512
CK + BC × CK = 0 p-val	0.0398	0.0757	0.0471
BC + BC × CK = 0 p-val	0.00848	0.0122	0.0228

Notes: All columns control for randomization strata (subdistrict) fixed effects. Other controls for each column selected with PDS Lasso from date and time of entry into the village, caste category of the treatment hamlet, distance from the village to an urban center, and respondent-level controls such as age, gender, literacy and potential seed status. Standard errors (clustered at the village-level) are reported in parentheses and p-values are reported in brackets.

Appendix Q. Baseline Village Ability Levels

Table Q.1. Village Ability-Treatment Interactions

VARIABLES	(1) vol
vH	0.151
	(0.428)
	[0.725]
CK	0.555
	(0.353)
	[0.116]
Broadcast	1.826
	(0.692)
	[0.008]
Broadcast × CK	-2.072
	(0.733)
	[0.005]
vH × CK	0.175
	(0.536)
	[0.745]
vH × Broadcast	-1.387
	(0.795)
	[0.081]
vH × Broadcast × CK	0.716
	(0.888)
	[0.420]
Observations	944
Number of groups	0
CK + Broadcast × CK = 0	0.0205
vH × CK + vH × Broadcast × CK = 0	0.215

Notes: Variable vH takes the value of 1 for villages above baseline average ability level. All columns control for randomization strata (subdistrict) fixed effects. Other controls for each column selected with PDS Lasso from date and time of entry into the village, caste category of the treatment hamlet, distance from the village to an urban center, and respondent-level controls such as age, gender, literacy and potential seed status. Standard errors (clustered at the village-level) are reported in parentheses and *p*-values are reported in brackets.