

Learning Through Noticing: Theory and Experimental Evidence in Farming*

Rema Hanna

Sendhil Mullainathan

Harvard University, NBER and BREAD

Harvard University and BREAD

Joshua Schwartzstein

Dartmouth College

March 9, 2014

Abstract

Existing learning models suggest that the availability and informativeness of data determine the pace of learning. However, in learning to use a technology, there are often a staggering number of potentially important input dimensions. People with limited attention must choose which dimensions to attend to and subsequently learn about from available data. We use this model of “learning through noticing” to shed light on stylized facts about technology adoption and use. We show how agents with a great deal of experience may persistently be off the production frontier, simply because they failed to notice important features of the data that they possess. The model also allows for predictions on when these learning failures are likely to occur. We test some of these predictions in a field experiment with seaweed farmers. The survey data reveal that these farmers do not attend to pod size, a particular input dimension. Experimental trials suggest that farmers are particularly far from optimizing this dimension. Furthermore, consistent with the model, we find that simply having access to the experimental data does not induce learning. Instead, behavioral changes occur only after the farmers are presented with summaries that highlight previously unattended-to relationships in the data.

*We thank Daniel Benjamin, Matthew Gentzkow, Marina Halac, Andy Newman, Ted O’Donoghue, Tim Ogden, Matthew Rabin, and Andrei Shleifer for helpful comments, as well as seminar participants at the Behavioural Decision Theory Conference, Berkeley, BREAD, BU, Chicago Booth, Cornell, Duke, Harvard/MIT, the NBER Organizational Economics Meeting, and SITE.

1 Introduction

Many production functions are not known *ex ante*. Instead, they are learned, both from personal experiences (Gittins 1979, Arrow 1962, Jovanovic and Nyarko 1996, Foster and Rosenzweig 1995) and from those of others (Besley and Case 1993, 1994, Banerjee 1992, Bikhchandani et al. 1992, Munshi 2004, Conley and Udry 2010). While diverse, existing learning models share a common assumption: the key input for learning is informative data.

Many examples, however, defy this assumption. For many years, doctors had the data that they needed to prevent operating room infections, but the importance of a sterile operating room was not recognized until the germ theory of disease was accepted (Nuland 2004, Gawande 2004). Indian textile manufacturers failed to adopt key management practices, such as having an uncluttered factory floor, despite exposure to natural variation that points to their importance (Bloom et al. 2013). Even experienced teachers do not adopt the best teaching practices (Allen et al. 2011). These examples all imply that learning is not just about the data that you possess, but what you *notice* in those data. In fact, we may not learn from the data that we ourselves generate.

In this paper, we use Schwartzstein’s (2014) model of selective attention to build a model of learning through noticing and test its predictions in a field experiment with seaweed farmers.¹ The model highlights an important constraint on learning. A farmer planting a crop faces a slew of potential features that might affect production – crop spacing, the time of planting, the amount and timing of the water employed, the pressure of the soil on the seedlings, and so on. He cannot possibly attend to everything (Kahneman 1973): his attention is limited (or effortful), while the number of potentially important variables is large. Since he can only learn about the dimensions that he notices (or attends to), this choice becomes a key input into the learning process.

In the model, the farmer allocates attention in a “Savage rational” way: he optimally chooses

¹While the model presented in this paper is specific to issues related to technology adoption and use, Schwartzstein (2014) presents a general model of belief formation when agents are selectively attentive. The approach of modeling economic agents as responding only to a subset of available information dates back to at least Simon (1955). Nelson and Winter (1982) consider how bounded rationality (or “evolutionary processes”) impact technological change. For more recent approaches to modeling limited attention in economic settings, see Sims (2003), Bordalo et al. (2012, 2013), Koszegi and Szeidl (2013), and Gabaix (2013).

what to attend to as a Bayesian, given his prior beliefs and the costs of paying attention. Consistent with the above examples, even with substantial experience – and even with lots of data at their fingertips – farmers need not be at the productivity frontier. Despite the (subjective) optimality assumption, an interesting feedback loop arises: a farmer who initially believes that a truly important dimension is unlikely to matter will not attend to it, and consequently he will not learn whether it does matter. This failure to learn stems from not focusing on the part of those data that could contradict a false belief. When the technology is incongruent with the farmer’s priors, the losses from such sub-optimization can be arbitrarily large.

While other models of incomplete learning also accommodate the idea that people can persistently be far from the productivity frontier, our model predicts where these failures should occur. Learning failures should be concentrated on dimensions where the agents report ignorance, i.e., where they cannot answer key questions about what they precisely do (or have done) along those dimensions. Most other theories of mis-optimization (arising from over-confidence, false beliefs, unawareness, etc.) have little to say about this “failure to notice” variable.

To test the model’s predictions, we conducted a field experiment with seaweed farmers in Indonesia. Seaweed is farmed by attaching strands (“pods”) to lines submerged in the ocean. As in the model, a large number of dimensions affect yield. To look for failures to notice, we directly asked farmers about their own production techniques. Farmers are quite knowledgeable and have clear opinions about many dimensions: almost all farmers had an opinion about the length of their typical line, the typical distance between their lines, the optimal distance between their knots (i.e., pods), the optimal distance between lines, and the optimal cycle length. On the other hand, most do not recognize a role for one key dimension: about 85 percent do not know the size of their pods and will not venture a guess about what the optimal size might be.²

To test whether this apparent failure to notice translates into a learning failure, we conducted experiments on the farmers’ own plots, varying both pod size and pod (or “knot”) spacing. On pod

²Note that there are many other dimensions that might be important: For example, the strength of the tide, the time of day, the temperature, the tightness with which pods are attached, the strain of pods used, etc. could matter. In our analysis, we largely focus on two or three dimensions for parsimony, but actual demands on attention are much greater.

spacing, which almost all farmers had an opinion about, our findings suggest that they were close to the optimum. In contrast, on pod size – which few farmers had an opinion about – our findings suggest they were far from it.

Further support for the model comes from examining farmers’ response to the trial. The model suggests that simply participating in the trials may not change the farmers’ behavior with respect to pod size.³ Intuitively, the farmers’ own behavior generates an experiment of sorts every season – random variation in pod sizing – and the trial presents farmers with similar data that they already had access to, but did not learn from. Consistent with this idea, we find little change in pod size following participation in the trial. However, the model suggests an alternative way to induce learning: to provide a summary of the data that explicitly highlights neglected relationships. Consistent with this prediction, we find that farmers changed their production methods after we presented them with the trial data on yield broken down by pod size from their own plot.

Beyond the analysis of seaweed farming, we extend the model to illustrate how greater experience with related technologies can predictably *increase* the likelihood of learning failures and make precise a notion of “complexity” of the current technology that can also induce failures. The potential costs of experience and the role of complexity are consistent with folk wisdom and evidence on technology adoption and use (e.g., Rogers 2010), but to our knowledge have largely remained theoretically unexplored.

The model also has broader implications for understanding which types of interventions can help individuals learn. In most learning models, providing more data induces learning. Our model and experimental results, in contrast, suggest that simply providing more data can have little impact on behavior if data availability is not a first order problem. Our findings shed light on why some agricultural extension activities are ineffective, or only moderately effective in the long run (e.g., Kilby 1962, Leibenstein 1966, Duflo et al. 2008b). At the opposite extreme, our

³Strictly speaking, this is a prediction of the model under the assumption that merely being asked to participate does not, by itself, significantly alter farmers’ beliefs about the importance of pod size, which we had prior reason to believe would be true (as we discuss below). It also relies on an assumption that it is not significantly easier for the farmers to learn relationships from the raw trial data than from the data they are typically exposed to, which also seems plausible given the experimental design.

model and experimental results show how one can induce learning without providing new data by simply providing summaries highlighting previously unattended-to relationships *in the agents' own data*. This result aligns with growing evidence that interventions that encourage agents to attend more closely to available data can profitably impact behavior in diverse contexts ranging from car manufacturing (Liker 2004), to teaching (Allen et al. 2011), to shopkeeping (Beaman et al. 2014).

The paper proceeds as follows. Section 2 presents the baseline model of learning through noticing and develops the empirical predictions. Section 3 describes the experiment, while Section 4 provides its results. Section 5 explores other potential applications and extends the model to develop comparative static predictions on the prevalence of failures to notice and resulting failures to learn. Section 6 concludes.

2 Model

2.1 Setup

We present a stylized model of learning through noticing. We present it for the case of a farmer to make it concrete, but we will reinterpret it for other contexts, such as management or education, in Section 5. A farmer plants the same crop for two periods, $t \in \{1, 2\}$. He works on a continuous sets of plots indexed by $l \in [0, 1]$. To capture the idea that the constraint on learning comes from having many things to pay attention to, his production technology has many dimensions that might matter. Specifically, for each of his plots l , the farmer chooses an N -dimensional vector of inputs, $\mathbf{x} = (x_1, x_2, \dots, x_N)$, where each x_j is drawn from a finite set X_j , and where we denote the set of possible \mathbf{x} by X .

Given the input bundle $\mathbf{x} \in X$, the dollar value of net yield (total yield net of the cost of

inputs) for the farmer from a plot l at time t is:

$$y_{lt} = f(\mathbf{x}|\boldsymbol{\theta}) + \varepsilon_{lt} = \sum_{j=1}^N f_j(x_j|\boldsymbol{\theta}_j) + \varepsilon_{lt}, \quad (1)$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_N)$ is some fixed parameter of the technology (described below) and $\varepsilon_{lt} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ is a mean zero shock that is independent across plots of land and time.⁴

2.1.1 Beliefs

Since the model focuses on how inattention impacts learning, we assume that the farmer initially does not know the parameters of the production function: he must learn $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_N)$ through experience.

The farmer begins with some prior beliefs about $\boldsymbol{\theta}$, where we denote potential values by $\tilde{\boldsymbol{\theta}}$. We assume that while many dimensions could be relevant for production, only a few of them actually are. To capture this, we assume that the farmer attaches prior weight $\pi_j \in (0, 1)$ to input j being relevant and the remaining weight, $(1 - \pi_j)$, to input j being irrelevant. Specifically, the farmer's prior holds that:

$$f_j(x_j|\tilde{\boldsymbol{\theta}}_j) = \begin{cases} 0 & \text{if input } j \text{ is irrelevant} \\ \tilde{\boldsymbol{\theta}}_j(x_j) \sim \mathcal{N}(0, v^2), \text{ i.i.d across } x_j \in X_j & \text{if input } j \text{ is relevant.} \end{cases} \quad (2)$$

We assume $v^2 > 0$ and that the prior uncertainty is independent across dimensions, meaning that knowledge of how to set input j does not help the farmer set input j' . Under these assumptions, when input j is irrelevant, it does not matter how x_j is set. When input j is relevant, with probability one some input level of x_j produces the greatest output. But the farmer does not initially know which one because the exact $\boldsymbol{\theta}_j$ is unknown and must be learned.

⁴Note that l and t appear symmetrically in Equation (1). For the purpose of forming beliefs about the underlying technology ($\boldsymbol{\theta}$), it does not matter whether the farmer learns through many observations across plots at a fixed time, or through many observations over time on a fixed plot. Equation (1) also reflects a simplifying assumption that the payoff function is separable across input dimensions, allowing us to separately analyze the farmer's decisions across dimensions.

2.1.2 Costly Attention

To this standard learning problem, we layer on limited attention. The farmer makes a zero-one decision of whether or not to attend to a given input. Let $\mathbf{a}_t \in \{0, 1\}^N$ denote a vector that encodes which dimensions the farmer attends to in period t , where $a_{jt} = 1$ if and only if he attends to dimension j in period t . For each input level he attends to, he faces a cost e , reflecting the shadow cost of mental energy and time.

Inattention operates in two ways. First, when input j is not attended to, a “default action” is taken, which we capture simply as a uniform distribution over the possible values of that input: X_j . Thus, the farmer’s actions are random in the absence of attention. When a farmer attends to an input, he chooses its level.

Second, if the farmer does not attend to an input, he also does not notice the set level(s). Thus, in period 2, instead of knowing the values x_{j1} which were actually chosen, he only remembers:

$$\hat{x}_{j1} = \begin{cases} x_{j1} & \text{if } a_{j1} = 1 \\ \emptyset & \text{if } a_{j1} = 0. \end{cases} \quad (3)$$

Notationally, we write the full second-period history as $h = (y_{11}, \mathbf{x}_{11})_{l \in [0,1]}$, which is what the farmer would recall if he was perfectly attentive, and $\hat{h} = (y_{11}, \hat{\mathbf{x}}_{11})_{l \in [0,1]}$, as the recalled history.⁵

In our seaweed application, if the farmer does not attend to pod size, he both creates random variation by virtue of not focusing on this dimension when cutting raw seaweed into pods and does not know the know the specific pod sizes that were set, making it harder for him to learn a relationship between pod size and output.

⁵Under the interpretation that y measures the dollar value of yield net of input costs, we are implicitly assuming that the farmer knows the *total* input costs even if he does not attend to *certain* input choices. By analogy, we may know our total monthly spending without knowing our spending by category, e.g., food or clothing.

2.1.3 The Farmer's Problem

The farmer is risk-neutral and maximizes the expected undiscounted sum of net payoffs – yield minus attentional costs – across the two periods. To simplify the presentation by allowing us to drop the l subscript, we restrict the farmer to strategies that are symmetric across plots in a given period, but allow him to mix over input choices. When it does not cause confusion, we abuse notation and let x_{jt} both denote the farmer's choice along dimension j at time t if he attends to that input, and the default uniform distribution over possible input levels when he does not attend.

This simple model captures a basic learning problem. In the first period, the farmer makes choices that trade off the future benefits of experimentation against maximizing current expected payoffs. In the second period, the farmer simply chooses to maximize current expected payoffs.⁶

2.2 Results

We first ask when the farmer will end up at the productivity frontier. For a technology (θ) and given cost of attention (e), let x_j^* denote a (statically) optimal input choice along dimension j . More precisely, x_j^* maximizes $f_j(x_j|\theta_j)$ whenever *dimension j is worth attending to*, i.e., whenever the payoff from setting the optimal input along dimension j and incurring the attentional cost to do so, $\max_{x_j} f_j(x_j|\theta_j) - e$, exceeds the payoff from not attending to the input and thus randomly setting it, $\frac{1}{|X_j|} \sum_{x'_j} f_j(x'_j|\theta_j)$. Whenever dimension j is not worth attending to, x_j^* equals the default uniform distribution over X_j (with an obvious abuse of notation). To make the problem interesting, we assume that there is at least one dimension that is worth paying attention to, as otherwise the farmer necessarily ends up at the frontier.

To understand whether the farmer ends up at the productivity frontier, we focus on second period choices.

⁶The model's assumption of two periods, but many plots, implies that experimentation will occur across plots in a fixed time period, rather than over time. Thus, we can avoid tangential issues that arise in settings with multi-period experimentation, such as considering the degree to which agents are sophisticated in updating given missing information (Schwartzstein 2014). That there are a *continuum* of plots also simplify matters by allowing us to abstract from noise in the learning process; the farmer can in principle perfectly learn the technology in a single period. This will allow us to focus on how inattention, rather than incomplete information, contributes to incomplete learning.

Proposition 1.

1. *When there are no costs of attention ($e = 0$), the farmer learns to optimize every dimension: in the second (terminal) period he chooses $x_{j2} = x_j^*$ for all j .*
2. *With costs of attention ($e > 0$),*
 - (a) *The farmer may not learn to optimize certain dimensions: For every technology θ , there exists a collection of prior weights $\pi_i > 0$, $i = 1 \dots, N$, such that in the second period he chooses $x_{j2} \neq x_j^*$ for some input j .*
 - (b) *Losses from not optimizing are unboundedly large: For every constant $K \in \mathbb{R}^+$, there exists a technology θ and collection of prior weights $\pi_i > 0$, $i = 1 \dots, N$, such that in the second period the farmer chooses $x_{j2} \neq x_j^*$ for some j and, by doing so, loses at least K .*
3. *The farmer does not learn to optimize a dimension only if he did not attend to it in the first period: in the second period he chooses $x_{j2} \neq x_j^*$ only if $a_{j1} = 0$.*

Proof. See Appendix A for all proofs. ■

The first part of Proposition 1 replicates the standard result of learning by doing models that, with no costs of attention, learning is primarily a function of experience and enough of it guarantees optimal technology use (e.g., Arrow 1962, Nelson and Phelps 1966, Schultz 1975). This intuition is invoked – at least implicitly – in arguments that learning by doing is not an important determinant of adoption for old technologies in which individuals have abundant experience (Foster and Rosenzweig 2010).

The second part of the proposition in contrast shows that, with costs of attention, failures to notice can lead to failures to learn and these failures can lead to arbitrarily large payoff losses. The basic idea is simple: inattention is self-confirming (Schwartzstein 2014). If the farmer initially falsely believes an input is not worth attending to, he will not notice the information revealing its importance that proves him wrong and will continue not attending to the input even if it is

very important.⁷ As the proof of Proposition 1 makes clear, the farmer fails to learn to optimize dimension j when the dimension matters, but he places sufficiently low prior weight π_j on it mattering. These errors can be arbitrarily large: the constant K need not be in any way proportional to the cost of attention, so even very small attentional costs can produce large losses. The final part of the proposition says that optimization failures come directly from noticing failures in our model. This makes the model testable by placing an empirically verifiable restriction that a farmer fails to optimize on a dimension j only if $\hat{x}_{j1} = \emptyset$ on that dimension.⁸

Proposition 1 thus yields the following testable predictions:

Prediction P1 *Agents may fail to attend to some dimensions.*

Prediction P2 *Agents may persistently choose sub-optimal input levels along some dimensions.*

Prediction P3 *Agents only persistently choose sub-optimal input levels along dimensions they do not attend to, which can be identified by measuring recall.*

To develop further predictions, recall that inattention has two consequences: a failure to precisely set input levels and a failure to notice and remember what they are. These two effects work in different directions. Failing to remember clearly impedes learning. However, failing to precisely set input levels can help learning: when the farmer does not attend to an input, he *inadvertently* experiments and generates valuable data. Since he does not notice this data, he does not capitalize on it. But this distinction has an interesting implication.

Specifically, suppose that in period 2, a farmer could have access to summary statistics about *his own data* from period 1, for example because an extension officer calculates and presents the farmer with such information. What effect would this have? One way to model this is to suppose that, for each input j , the farmer is told the level $\tilde{x}_j^* \in X_j$ that achieved the greatest yield in period 1,

⁷The logic is similar to why individuals can maintain incorrect beliefs about the payoff consequences to actions that have rarely been tried in bandit problems (Gittins 1979) and in self-confirming equilibria (Fudenberg and Levine 1993), and why these incorrect beliefs in turn support sub-optimal decisions.

⁸Similarly, under the assumption that farmers do not measure and consequently randomly select inputs along dimensions they do not pay attention to, identifying dimensions along which farmers persistently use a wide variety of input levels will also predict failures to notice and resulting failures to learn. Likewise, identifying dimensions along which farmers cannot recall the empirical relationship between the input level and the payoff will predict such failures.

as well as the corresponding sample average $\bar{y}_j(\tilde{x}_j^*)$.⁹ Since the cost of tracking the input relative to yield on measure 1 of plots of land is e , it is natural to assume that the cost of attending to a single statistic $(\bar{y}_j(\tilde{x}_j^*), \tilde{x}_j^*)$ is much lower, which for simplicity we will take to equal zero. To emphasize the impact of receiving a summary about data the farmer would generate on his own, suppose that receiving it comes as a complete surprise: when choosing how to farm and what to attend to in the first period he does not anticipate that he will later be provided with this summary. However, suppose that he understands how the summary is constructed, i.e., that he correctly interprets the summary given his prior over θ and the objective likelihood function of the summary.

Proposition 2. *If the farmer has access to summary information $((\bar{y}_j(\tilde{x}_j^*), \tilde{x}_j^*)_{j=1}^N)$ prior to making second period decisions, then he learns to optimize every dimension j : he chooses $x_{j2} = x_j^*$ for all j .*

The intuition behind Proposition 2 is simple: together with \hat{h} , $(\bar{y}_j(\tilde{x}_j^*), \tilde{x}_j^*)$ is a sufficient statistic relative to the farmer's second-period decision along dimension j , i.e., he will make the same choice on that dimension whether he knows all of h or just \hat{h} and $(\bar{y}_j(\tilde{x}_j^*), \tilde{x}_j^*)$.

Proposition 2 demonstrates that, in the model, learning failures do not result from a lack of readily available data (or from a doctrinaire prior that does not allow for learning), but rather from a failure to attend to important details of the data.¹⁰ Summarizing features of the data that the farmer had seen but failed to notice can thus be useful.

The value of summaries *relies* on the farmer facing costs of paying attention and not attending

⁹Specifically, letting $L_j(x_j) = \{l \in [0, 1] : x_{jl1} = x_j\}$ denote the set of plots of land where the farmer uses input level x_j in the first period, the sample average payoff conditional on using that input equals:

$$\begin{aligned} \bar{y}_j(x_j) &= \frac{1}{|L(x_j)|} \int_{l \in L(x_j)} y_{l1} dl = \frac{1}{|L(x_j)|} \int_{l \in L(x_j)} f_j(x_j | \theta_j) + \sum_{k \neq j} f_k(x_{kl1} | \theta_k) + \varepsilon_{l1} dl \\ &= f_j(x_j | \theta_j) + \sum_{k \neq j} f_k(\sigma_{k1} | \theta_k), \end{aligned}$$

where $\sigma_{k1} \in \Delta(X_k)$ denotes the distribution over input levels in X_k implied by the farmer's first period strategy. \tilde{x}_j^* is then defined as a maximizer of $\bar{y}_j(\cdot)$, where any ties are arbitrarily broken.

¹⁰The stark finding of Proposition 2 – that any failure to learn stems solely from failing to extract information from available data, rather than a lack of exposure – relies on the richness of the environment, e.g., that there is a continuum of plots. The key point is more general: Given the data that he generates, the inattentive farmer could benefit from processing those data differently when he did not attend to important input dimensions when forming beliefs.

to some dimensions on his own, as Proposition 1 demonstrates that the farmer will otherwise learn to optimize independent of receiving a summary. In fact, combining the two propositions yields the simple corollary that the farmer will react to the summary only along dimensions he does not attend to in its absence. The model may thus rationalize certain extension or consulting activities that do not help agents collect new data, but rather to record and highlight relationships in those data. For the purpose of the main empirical exercise, we have the following predictions:

Prediction P4 *Agents may fail to optimize on unnoticed dimensions, even though they are generating the data that would allow them to optimize.*

Prediction P5 *Summaries generated from the agents' own data can change their behavior.*

2.3 Comparison with Alternative Approaches

We briefly contrast our model with alternative approaches. First, a model in which agents exogenously fail to optimize along certain dimensions would be simpler, but would problematically be consistent with every kind of learning failure. In contrast, our model allows for testable predictions about when an agent will fail to optimize, for example when he cannot answer questions about what he does. Second, a more extreme form of not noticing –unawareness– could also produce failures to learn, where here unawareness means that an agent simply does not even realize that a dimension exists. Conversely, our model predicts that people may persistently fail to optimize along important dimensions they are aware of when a technology is prior incongruent– failures to learn come from not appreciating the importance of variables instead of from neglecting their existence entirely. For example, in contrast to models of unawareness, doctors were dismissive of practices like hand-washing long after it was hypothesized that there could be empirical relationships between doctors washing their hands and outcomes like maternal deaths (Nuland 2004). The germ theory of disease was important for getting doctors to appreciate such relationships. Third, the predictions of our model are distinct from those of “bandit models” (Gittins 1979), models of self-confirming equilibria (Fudenberg and Levine 1993), or “local” models of learning by doing (Conley and Udry 2010). While such models also allow for persistent failures to optimize, in those

models a lack of data explains learning failures. When the binding constraint is instead a failure to notice, the main bottleneck is not one of data collection, but rather data processing.

Finally, models of “rational inattention” (e.g., Sims 2003 or Gabaix 2013) also model attentional costs, but further assume a form of rational expectations in which agents know what is worth attending to, rather than having to learn what to attend to through experience. This assumption implies that the size of learning failures resulting from inattention is bound by attentional costs. Proposition 1 in contrast shows that our model places no such bound: potential losses from not noticing are unboundedly large. Empirically, this seems an important distinction to us, since the interesting failures to optimize are large failures to optimize.

2.4 Applying the Model

The analysis suggests an exercise with the following steps: (i) find a multi-dimensional setting with experienced agents, (ii) predict or measure what agents attend to, (iii) assess whether agents are optimizing, (iv) assess whether agents could achieve a higher payoff given data available to them, and (v) conduct an “attentional intervention.”

For (i), we prefer a setting in which the first order reason behind any incomplete learning is compellingly a failure to notice rather than a lack of available information. Situations with experienced farmers, using mostly old technologies (e.g., fertilizer) fits this description; situations in which a new technology has just been introduced (e.g., hybrid seeds in the 1920s) may be more fruitfully analyzed through the lens of more traditional models of technology adoption.

For (ii), we would like to collect data that can be used to predict or measure which dimensions of the production process agents do or do not pay attention to, which, when combined with knowledge of the production function, can be used to predict whether agents are optimizing. Such data can include survey responses detailing agents’ beliefs about how they have set inputs along various production dimensions in the past (allowing agents to reveal their knowledge on those inputs), their beliefs about what constitutes best practices, and data on how they actually set inputs.

For (iii) and (iv), we want to collect or generate data that can be used to analyze whether

agents are optimizing given data available to them, and whether they do a poorer job optimizing dimensions they appear not to pay attention to. To perform this exercise, Proposition 2 suggests that it may suffice to analyze data that the agent generates herself.

For (v), we want to perform an intervention that involves presenting information in a way that helps agents learn relationships that they would not learn on their own, and examine whether the intervention affects agents' behavior and resulting payoffs. From Proposition 2, such an intervention can involve providing a summary of how different input choices along unattended-to dimensions (identified by (ii)) have empirically influenced the payoff-relevant output, given data agents in principle have available to them.

The seaweed experiment, detailed below, follows the steps laid out above and tests predictions **P1-P5**. The model generates testable predictions beyond **P1-P5** that can be explored by following additional exercises. Section 5 details some of these predictions and exercises.

3 The Seaweed Experiment

3.1 Setting

Our experiment takes place with seaweed farmers in the Nusa Penida district in Indonesia. One reason we became interested in seaweed farming is because it shares key features with the ideal setting that we discussed in Section 2.4: it involves experienced agents who have had many opportunities to learn – they have grown seaweed since the early 1980s, with many crop cycles in each year – but where the production technology involves many dimensions.

Most farmers in the area that we study follow what is called “the bottom method”: in each plot, the farmers drive wooden stakes in the shallow bottom of the ocean, and then attach lines across the stakes. They then take raw seaweed from the last harvest and cut them into pods.¹¹ The farmers then plant these pods at a given interval on the lines. After planting, they tend their crops

¹¹Most farmers initially grew a variety called spinosum, but some have moved to a different strain called cottonii, due to some combination of buyer advice and government and NGO extension programs.

(remove debris, etc.). About 35 to 40 days later, they harvest the seaweed, dry it, and then sell it to local buyers.

While seemingly straightforward, this process requires decisions on many different dimensions along the way, ranging from whether to use the bottom method or other methods, to how long to wait before harvesting, and even to where and how to dry it. We focus primarily on three dimensions. We explore the farmers' decisions on the distance between lines and distance between pods, which influence how much sunlight and nutrients the pods have access to, as well as the degree to which they are exposed to waves. Additionally, we look at the size of the pods that the farmers plant, which may influence the growth of the pods for numerous reasons; for example, bigger seedlings may result in higher yields in still water, but may be more likely to break (or be lost completely) in ocean locations that face significant waves.

There were additional reasons why we became interested in this process, including that it shares essential features with farming other crop types, where the many different decisions over inputs add up to determine yields, making it plausible that insights from studying this process could generalize. Prior to our study, there were already indications that some farmers may not have been optimizing, as their methods differed from local extension officers' recommended practices.

3.2 Experimental Design

From June 2007 to December 2007, we administered a survey that can be used to predict which dimensions of the seaweed production process the farmers pay attention to (see Appendix Figure 1 for the project timeline). From a census of about 2706 farmers located in seven villages (24 hamlets), commissioned by us in 2006, we drew a random sample of 500 farmers for the baseline survey, stratified by hamlet. Out of these, 489 were located and participated in the baseline survey (see Appendix Figure 2). The baseline survey consisted of two parts: (1) a questionnaire that covered demographics, income, and farming methods, and (2) a "show and tell" where the enumerators visited the farmers' plots to measure and document actual farming methods (see Section 3.3 for the data description).

From the list of farmers that participated in the baseline survey, we randomly selected 117 farmers to participate in an experimental trial to determine the optimal pod size for one of their plots (stratified by hamlet), as well as the optimal distance between pods for a subset of those farmers. This trial allows us to analyze whether farmers are not optimizing, and whether they do a poorer job optimizing dimensions they seem not to pay attention to. The trials occurred between July 2007 – March 2008, shortly after the baseline was conducted for each farmer. Farmers were told that the enumerators would vary the seaweed production methods across ten lines within one of their plots, with the farmers' assistance, and that they would be presented with the results afterwards. All of the farmers that we approached participated in the trials and were compensated for doing so in two ways. First, we provided the necessary inputs for planting the lines and guaranteed a given income from each line so that the farmers would at least break even. Second, we provided a small gift (worth \$1) to each farmer to account for their time.

Participating farmers were randomly assigned into one of two sub-treatments: *sort* (65 farmers) and *weight* (52 farmers). The sort sub-treatment was built around the idea that farmers had substantial variation in pod size even within their own plots (see Appendix Figure 3 for the distribution of sizes within farmers' own plots in this sub-treatment). Given this variation, we wanted to understand whether farmers could achieve a higher yield by systematically using a specific size within the support of those already used. Farmers were asked to cut pods as they usually would and then the pods were sorted into three groups. Working with the farmers, the enumerators attached the pods into the lines by group (3 lines with small pods, 4 lines with medium pods, and 3 lines with large pods). The lines were then planted in the farmer's plot.

Despite the wide range of pod sizes used within a given farmer's plot, it is still possible that he could do better by moving to a size outside that range. The weight sub-treatment was designed to address this issue by testing a broader set of initial pod sizes. To generate variation, the pod weights were initially set at 40g to 140g (in intervals of 20g) for the first few trials. However, in order to better reflect the ranges of weights used by the farmers, the weights were changed to 90g-180g for

spinosum and 90g-210g for cottonii (both in intervals of 30g).¹² The pods of different sizes were randomly distributed across about 10 lines, with the enumerator recording the placement of each pod. The farmers were present for the trials and saw where each pod was planted on the lines.

In the weight subtreatment, we also tested whether farmers optimized distance between pods. We compared two distances, 15cm and 20cm, since the average distance between pods at baseline was around 15cm and past technical assistance programs had suggested larger spacing.

All farmers were told to normally maintain their plots. The enumerators returned to reweigh the seedlings twice while in the sea: once around day 14, and again around day 28. On around day 35, the seaweed was harvested and weighed for a last time. In addition to being present for the planting, farmers were present for and helped with the weighing, harvesting, and recording of results. Thus, they had access to all of the raw information generated by the trials.

From April to May, 2008, we conducted the first follow-up surveys, which were designed to test whether farmers changed any of their methods after participating in a trial. These changes would have happened in the cycle after the trial: farmers would have had time to incorporate anything they learned on their own from the trial into the next cycle. Surveys were conducted with a subsample of 232 farmers, which included all of the farmers who participated in the trials, as well as an additional set of farmers who were randomly selected from the baseline as a control group; 231 farmers completed the survey.

From May to June, 2008, the enumerators gave each farmer a summary table that provided information on their returns from different methods and highlighted which method had the highest yields.¹³ The enumerators talked through the results with the farmers, describing the average pod size that they typically used and the difference between that size and the optimal one (note that the optimal pod weight for sorting trials was the average of the optimal size conditional on the optimum being small, medium or large). For the weight trials, they were also told whether the optimal distance between pods was 15cm or 20cm. Appendix Figure 4 provides examples.

¹²We ensured that the actual weights of the pods that were planted were within 10 grams of the target weight, and so it is best to think of these as bins within these averages. This adds some noise to the weight categories, biasing us against finding different effects by weight.

¹³We worked with a local NGO to design a simple, easy to understand table to summarize the trial results.

About two months after we gave the results to the farmers (July – August 2008), we conducted a second follow-up survey to learn if they changed their methods as a result of having received their trial results, allowing us to examine whether presenting farmers with summaries highlighting empirical relationships uncovered in readily available data affects their behavior. Out of the original 232 farmers, 221 were found.

3.3 Data, Sample Statistics and Randomization Check

3.3.1 Data

Baseline Survey: The baseline consisted of two parts. First, there was a household questionnaire, in which we collected detailed information on demographics (e.g. household size, education), income, farming experience, and current farming practices – labor costs, capital inputs, technologies employed, difference in methods based on seasonality and plot location, crop yields, etc. – as well as beliefs on optimal farming methods. We additionally collected data on both learning and experimentation. The “learning” questions focused on issues such as where farmers gain their knowledge on production methods and where they go to learn new techniques. The “experimentation” questions focused on issues like whether farmers ever experiment with different techniques and, if yes, with which sorts of techniques, and whether they change their production methods all in one shot or via a step-wise process.

Second, we conducted a “show and tell” to document the farmers’ actual production methods. During the show and tell, we collected information on the type of line used, the sizes of a random sample of pods, the distances between a random sample of seedlings, the distances between seaweed lines, and so forth.

Experimental Trial Results: We compiled data from each of the experimental trials, documenting each plot location, the placement of pods within a plot, and pod sizes at each of the three weighings. Thus, we can compute the yield and the return from each pod.

Follow-up Surveys: Both follow-up surveys (one after the trial and one after the summary data was seen) collected information on farmers’ self-reported changes in farming techniques. We also conducted the “show and tell” module to measure actual changes in techniques.

3.3.2 Baseline Sample Statistics and Randomization Check

As Panel A of Table 1 illustrates, most farmers (83 percent) were literate. Many had been farming for about 18 years, with about half reporting that they learned how to farm from their parents (Panel B). About a third of the sample grew the cottoni strand at baseline. The “show and tell” revealed that the mean pod size was about 105 grams, while both the distance between pods and between lines was about 15 cm.

In Appendix Table 1, we provide a randomization check across the control and both sub-treatment groups. We test for differences across the groups on ten baseline demographic and farming variables. As illustrated in Columns 4 through 6, out of 30 comparisons we consider, 3 are significant at the 10 percent level, consistent with chance.

4 Experimental Results

4.1 Results from the Baseline Survey and the Experimental Trial

A basic implication of the theoretical framework is that some farmers will not keep track of input dimensions that influence returns. Table 2 presents baseline survey responses for 489 farmers. Panels A and B document self-reported current and optimal methods. We present the percentage of farmers who were unable to provide an answer in Column 1, and provide means and standard deviations of self reports in Column 2 (conditional on answering).

4.1.1 Result 1: Farmers only attend to certain dimensions of the production function

Consistent with prediction **P1**, a vast majority of farmers were inattentive to some input dimensions, particularly pod size. Following Proposition 1, we measure farmers’ attention by eliciting

self reports on current practices. Eighty six percent could not provide an answer for their current pod size at baseline (Column 1 of Panel A), while 87 percent of farmers did not even want to hazard a guess about what the optimal pod size should be (Panel B).¹⁴ Since many farmers fail to notice key facts about their pod sizes, it is perhaps unsurprising that a broad range of sizes is observed both across (see Figure 1) and within (see Appendix Figure 3) farmers.

On the other hand, farmers were more attentive to other input dimensions. They appeared attentive to both the distance between knots that secure the pods to a line and the distance between lines. Unlike pod size, most farmers (98 to 99 percent) provided an answer for their distance between lines (Panel A) and similarly many had an opinion about the optimal distance between both the knots and lines (Panel B). Given that the farmers appear relatively more attentive to the distance between knots and lines than pod size, we might expect that the actual distances employed would exhibit less variance than the pod sizes. Indeed, the coefficient of variation for the distance between lines (0.13) and pods (0.10) is much smaller than that for pod size (0.27), indicating that the practices for these inputs are relatively less variable across farmers.¹⁵

4.1.2 Results 2 and 3: Farmers are not optimizing, and seem to do a relatively poor job optimizing unattended-to dimensions

Predictions **P2** and **P3** hold that farmers may be off the productivity frontier and that learning failures will be concentrated along dimensions they fail to attend to; given the above results, this suggests that farmers will be especially far from optimizing pod size. For the 117 farmers that participated in the experimental trials, we have data on optimal pod size. Table 3 summarizes the

¹⁴The enumerators kept reporting to us that many farmers could not answer these questions, even when probed.

¹⁵About 20 percent of farmers reported having ever made a change in any aspect of farming other than a change in strain at baseline (details available from authors). This small percentage indicates that farmers may not experiment very much on their own and that learning from other sources (e.g., parents) may contribute to farmers' beliefs about optimal practices.

This small percentage may appear at odds with the broad range in pod sizes observed within farmers (Appendix Figure 3). However, one interpretation is that farmers do not view this range as reflecting varying practices, but rather the stable practice of not carefully measuring size. Put differently, it is consistent with the notion that the within-farmer variation does not reflect active experimentation.

predicted percentage income gain from moving away from initial pod sizes.¹⁶ Panel A reports this estimate for farmers in the sort treatment, providing information on both the predicted gain a farmer could achieve by changing the size of each pod from his baseline average to the best performing size, as well as the predicted gain by changing the size of each pod from the worst performing size to the best among sizes used at baseline. Panel A also provides information on p-values from F-tests of the null that yield does not vary in pod sizes used at baseline for farmers in the sort treatment. Panel B then presents the predicted income gain a farmer could achieve by changing the size of each pod from his baseline average to the best performing size in the weight treatment, as well as information on p-values from F-tests of the null that yield does not vary in pod sizes used in this treatment. We provide the estimated median across farmers in Column 1, and provide the confidence interval of the estimate in Column 2.¹⁷

On net, the results indicate that farmers are potentially forgoing large income gains by not noticing and optimizing pod size.¹⁸ In the sort treatment, the median estimated percentage income gain by moving from the average to the best performing size is 7.06, while the median estimated gain by moving from the worst to the best is 23.3 (Panel A). In the weight treatment, the estimated gain by moving from the baseline average to the best size is 37.87 (Panel B). The potential gains are comparable to estimates of the gains to switching from the lower-yielding spinosum to the higher-yielding cottonni strain, where many farmers were induced to switch strains due to a combination of buyer advice and extension services.¹⁹ The gains are also large when compared to Indonesia's transfer programs: Alatas et al. (2012) find that the unconditional cash transfer program (PKH) is equivalent to 3.5-13 percent of the poor's yearly consumption, while the rice subsidy program is equivalent to 7.4 percent.

¹⁶We do not have follow-up data on income or yields; we compute predicted changes to income based on results from the trials. To do so, we make several strong assumptions. First, we assume that past seaweed prices are consistent with the future ones; this may be unrealistic as the prices may fall if all farmers increase their yields. Second, we assume that farmers do not change other methods (have fewer cycles, harvest earlier, etc.) if their yields change. Thus, this evidence should be viewed more as *suggestive*, rather than causal.

¹⁷Given the wide heterogeneity in the results, the median is likely a better measure of what is typical than the mean.

¹⁸In the sort treatment, about half the farmers were told that their most productive bin was their largest bin, while about 30 percent were told that it was the smallest. In the weight treatment, about 55 percent were told that they should increase their initial sizes.

¹⁹See, for example, <http://www.fao.org/docrep/x5819e/x5819e06.htm>, Table 6.

Given the wide heterogeneity in returns, illustrated in Figure 2, many individual farmers could even potentially increase their incomes by much more than what is typical across the farmers. Most strikingly, the gains from the sort treatment suggest that farmers would have done much better by systematically using a specific size *within the support of those already used*. This fact indicates that it is unlikely that farmers' failure to optimize purely reflects a failure of experimentation, and is consistent with prediction **P4** that farmers fail to optimize given their own data when they do not attend to important dimensions.²⁰

Turning to the precision of these estimates, each trial had around 300 pods per farmer so we have a reasonable number of observations to calculate these returns. In the sort treatment, we estimated a regression of yield on size dummies (small, medium, large) for each farmer separately, where the median p-value from F-tests of the null that pod size does not matter among sizes used at baseline – i.e., that the coefficients on the dummies are all equal – is .01 across farmers (Table 3, Panel A). Figure 3 presents the distribution of these p-values across farmers. While there is some variability across farmers in the precision with which we can reject the null that pod size does not matter among sizes used at baseline, p-values bunch in the range [0, .01]. The story is even clearer in the weight treatment, where, for every farmer, we can reject the null that pod size does not matter at a .01 significance level (Table 3, Panel B). In fact, for every farmer, the p-value from the F-test is estimated at 0 up to four decimal places.

Farmers appear to perform better in setting their distance between pods – a dimension they seemed to notice at baseline. Results from the weight treatment indicate that, for 80 percent of farmers, the optimal distance between pods was 15cm. Given that most farmers were at 15cm to begin with, these data suggest that very few farmers would do better by changing to 20cm.²¹

Overall, the findings suggest that many farmers failed to notice pod size and were not optimizing size, while many farmers noticed distance between pods and may have been optimizing

²⁰However, comparing the gains-from-switching estimates across sort and weight treatments indicates that farmers may have done even better by moving to a size outside of the support of those already used, which can be interpreted as suggesting that a lack of experimentation also contributes to a failure to optimize.

²¹Also, given the apparent heterogeneity in the optimal size across farmers (Appendix Figure 5), this suggests that there is more heterogeneity across farmers in the optimal size than the optimal distance between pods.

distance (at least within the support of distances that we tested). These results – consistent with predictions **P2** and **P3** – suggest that inattention contributes to a failure to optimize and hinders learning by doing. We next analyze the farmers’ response to the trial to further examine prediction **P4** and to test prediction **P5**.

4.2 Results Following the Experimental Trial

The model suggests farmers should respond more to the trial plus the summary than to trial participation by itself. Specifically, consistent with prediction **P4**, we may not expect trial participation by itself to have large effects on future behavior since farmers’ own behavior generated an experiment of sorts every season – their failure to notice size created random variation in pod sizing – and, in effect, the trial presents farmers with data that is similar to what they already had access to but incompletely learned from. However, consistent with prediction **P5**, farmers should be more likely to respond when they are presented with a summary of the trial findings, as the summary is easier to process.²²

We begin by exploring the effects of participating in a trial on production outcomes. For each farmer i in hamlet v , we estimate the following model:

$$Y_{ivt} = \beta_0 + \beta_1 F1_t + \beta_2 F2_t + \beta_3 Trial_{iv} + \beta_4 Trial_{iv} \cdot F1_t + \beta_5 Trial_{iv} \cdot F2_t + \alpha_v + \eta_{ivt}, \quad (4)$$

where Y_{ivt} is the production choice at time t , $F1_t$ is an indicator variable that denotes the first follow-up after the experimental trial, $F2_t$ is an indicator variable that denotes the second follow-up after the summary findings were presented to farmers, and $Trial_{iv}$ is an indicator variable that

²²The prediction that farmers will respond more to the trial plus the summary than to the trial by itself implicitly relies in part on an assumption that simply being asked to participate in the trial does not significantly draw farmers’ attention to pod size – i.e., being asked to participate does not lead farmers to significantly update the probability they place on pod size mattering, π_{size} . While similar assumptions may not hold in other contexts (e.g., Zwane et al. 2011), it appears reasonable in this context. Few farmers in other parts of Indonesia previously took up NGO advice on pod size. Indeed, this was one of the reasons that we became interested in running the current experiment in the first place. In fact, very few farmers at baseline (roughly 10 percent) indicated that they would change their farming methods in response to an NGO or government recommendation, or in response to advice from a friend (Appendix Table 2), while far more farmers (roughly 40 percent) indicated that they would change their practice in response to results from other plots. These results suggest a hesitation among these farmers to take advice at face value.

denotes trial participation. We also include a hamlet fixed effect, α_v , as the randomization was stratified along this dimension.²³ There are two key parameters of interest: β_4 provides the effect of participating in the trial prior to obtaining the summary of the findings, while β_5 provides the effect after the summary is provided.

Table 4 presents these results. In Columns 1 and 2, the outcome of interest is the self-reported measure of whether the farmer has made any changes in his production techniques. In Column 1, we report the coefficient estimates from Equation (4); in Column 2, we report the estimates from a model that additionally includes farmer fixed effects. Columns 3 and 4 replicate the analysis in the first two columns, but with the enumerator measured pod size as the outcome. We estimate all models using OLS and all standard errors are clustered by farmer.

4.2.1 Results 4 and 5: Participating in the trial by itself had little effect on farmers' decisions, while presenting farmers with a summary of the trial's findings was more effective

Consistent with predictions **P4** and **P5**, just participating in the trial had little impact on subsequent production decisions, while being presented with a summary of the results was effective. We do not observe a significant effect on self-reported changes to farming techniques from participating in the trial, prior to when farmers received the summarized results (Table 4, Column 1). However, about 16 percent more farmers reported changing a technique after receiving the results, which is about one and a half times the mean of the dependent variable (Column 1). Adding farmer fixed effects does not significantly alter the coefficient (Column 2). Note, however, that a test comparing β_4 and β_5 fails to reject the null hypothesis of equality at conventional levels (p-value = 0.1483 in Column 1 and p-value= 0.2209 in Column 2). It is possible that some of the results from this self-reported measure may be driven by farmers wanting to please the enumerators after participating in the trial, though this is unlikely as the control group also received regular visits from enumerators to both survey and measure their farming practices. We next turn to the enumerator measured

²³The inclusion of a hamlet fixed effect does not significantly influence the results.

results, which are less likely to suffer from this type of bias.

We do not observe a significant change in pod size from just participating in the trial, i.e. prior to receiving the summaries (Columns 3 and 4).²⁴ After receiving the summary, however, the treatment group increased pod size by about 7 grams (on average) relative to the control. This is significant at the 10 percent level in the basic specification (Column 3) and the positive sign is consistent with the average trial recommendation. However, while the coefficient estimate remains roughly the same (7.3 grams) when including farmer fixed effects, the significance level falls below conventional levels (p -value = 0.14) due to an increase in the standard error (Column 4). Nevertheless, we reject the null that the coefficients are equal ($\beta_4 = \beta_5$) with p -values of 0.0033 (Column 3) and 0.0154 (Column 4). While farmers did not appear to attend to pod size prior to the trials, providing summary information on their optimal pod size seems to have changed their behavior.²⁵

We next separately explore the impact of participating in the sort and weight treatments. Specifically, we modify the basic model to include separate dummy variables for participating in the sort and weight treatments and interact these variables with the indicators for follow-up status. Table 5 presents the results. The outcome of interest is the self-reported measure of whether the farmer has made any changes in his production techniques in Columns 1 and 2, enumerator-measured pod size in Columns 3 and 4, and enumerator-measured distance between pods in Columns 5 and 6 (recall that we only experimented with this distance in the weight sub-treatment). The findings are again consistent with the predictions of the model. Columns 1-4 present results that are similar to what we found in the model that did not distinguish between the treatments: trial participation by itself did not appear to influence farmers' decisions, while partic-

²⁴There is a statistically significant negative coefficient on "After Trial" and "After Summary Data" in Columns 3 and 4, suggesting that, on average, control farmers used larger pods at baseline. Common shocks to productivity could be responsible for such a trend. For example, since pods are cut from raw seaweed from the previous harvest it is possible that common shocks led to high yields in the harvest before the baseline, which in turn led to bigger pods even if farmers did not attend to – nor precisely measure – them.

²⁵In Appendix Table 3, we disaggregate the results by whether farmers were told to increase or decrease pod size in the follow-ups. To do so, we interact the interaction of the treatment and follow-up variables with an indicator variable that denotes whether the farmer was told to increase their pod size. We observe that farmers who were told to increase pod size did so both after the first and second follow-up. However, it is possible that this is simply capturing the fact that if farmers are randomizing with respect to pod size, those who "should go bigger" are those who had abnormally low draws of pod size at baseline. Thus, in expectation, those farmers would in fact go bigger the next period even if they continue to randomize, making these results difficult to interpret.

ipation plus receiving the summary appeared to impact both self-reported production techniques (Columns 1 and 2) as well as enumerator-measured pod sizes (Columns 3 and 4), though the last effect is statistically significant only in the sort treatment.²⁶

Finally, we find no effect of trial participation on enumerator-measured distance between pods, even after being presented with the summary (Columns 5 and 6). This is consistent with the model: Unlike pod size, farmers appeared to previously notice distance, had beliefs on the optimal distance, and tended to be at the optimum (at least within the support of distances tested in the trial). As a result, we would not expect large changes in distance as a result of either participating in the trial or receiving its results. Note, however, that while this result is consistent with the model, the insignificant result on distance could also be driven by the smaller sample size.

4.3 Alternative Explanations

While the experimental findings track the theory, other explanations are possible. First, perhaps we missed important costs and the farmers' pod size strategy is optimal. For example, perhaps carefully cutting pods to a particular size is costly in terms of labor effort. However, if farmers believed that they were at the optimum, there would be no reason for them to react to the treatment. Their reactions suggest that they felt that they were not optimizing based on the trial data.

A second explanation is that maybe the farmers are simply innumerate or face computational constraints. They could not learn from participating in the trial – or even from the natural data variation – because they could not perform the necessary calculations.²⁷ However, 83 percent of farmers are literate. The fact that they perform well on distance also suggests that numeracy is not the main constraint or it would be problematic along this dimension as well.

Third, perhaps the variation in size is too small to detect (at least without a measuring technology, e.g., a scale). This too seems unlikely: pod size variation is large with the average range

²⁶We also tested whether treatment effects differed by years of seaweed farming (experience) and education; we do not observe differences. However, one possibility is that we do not have sufficient variation in these variables or sufficient power size to adequately test for such heterogeneous treatment effects.

²⁷Recent experimental evidence suggests that computational constraints can impede learning in laboratory games, providing another rationale for why summaries can be effective (Fudenberg and Peysakhovich 2013).

of sizes used by a given farmer in the sort subtreatment being 39 grams, or roughly a third of the average size. This is roughly equivalent to the size of 8 nickels or 19 grapes. Such variation is likely detectable by sight or feel, especially considering the acuity required for many agricultural tasks, like applying fertilizers or pesticides.²⁸ Variation in yield across the range of sizes typically used is similarly large, as indicated by the large implied percentage income changes presented in Table 3. Converting these numbers into percentage gains in grams, the median percentage gain to moving from the worst to the recommended size is over 30 percent in the sort subtreatment, for example. Finally, we saw that farmers do not react to the data generated by the experimental trials, even though they were present for (and helped with) the weighing and recording of results.²⁹

Perhaps the biggest stumbling block for these (and other) explanations is the link between failures to learn and self-reported knowledge. The theory makes a clear prediction: failures to learn will be centered on dimensions that individuals lack knowledge of, as in the case of pod size. While other explanations may explain learning failures— even from experiments that people participate in – it is less clear why those failures would be linked to what they notice. Limited attention appears to be the most plausible explanation of this key fact.

5 Extensions

5.1 Other Applications

Though the model was cast as one of farming, it more generally applies to tasks where agents learn which input choices x maximize payoff-relevant output y . In this section, we sketch a few other applications. The goal is not to argue definitively that the model explains a particular set of

²⁸Indeed, research on perception indicates that people can detect much smaller differences. For example, laboratory evidence on weight perception suggests that people can detect changes that amount to at least 2% of an initial weight. As an illustration, people can accurately detect the difference between 100g and 102g and the difference between 200g and 204g (Teghtsoonian 1971).

²⁹While farmers knew they would be presented with the summary information – which would have attenuated any incentive to attend to the relationship between pod size in yield in the raw trial data – they also knew there would be a lag in receiving the summary, meaning that they still had *some* incentive to attend if they thought pod size mattered, as this could allow them to improve their practices in the interim.

stylized facts. Instead, it is to highlight facts consistent with the model, and to suggest how future experiments or data collection exercises could test our interpretation. Note that a prerequisite to applying a model of noticing is to be more precise about the possible, granular decisions that go into production. The use of “fertilizer” cannot be summarized by “dollars spent on fertilizer.” Instead, the individual choices that need to be made, such as when it is applied, how to space applications, and so on need to be specified. Put simply, a model of noticing only matters when there are details to notice.

Management

In an attempt to maximize product quality (y), the managerial process involves many nuanced choices (x) – exactly how to monitor worker effort, what to look for to determine whether a machine needs preventive maintenance, whether to worry about the cleanliness of the shop floor, and so on. Bloom et al. (2013) examine managerial choices for a set of Indian textile plants. They give free consulting advice to a randomly selected subset. The advice is interesting because it does not suggest any new physical technology, but simply helps managers think about their production processes differently. In the simplest economic frameworks, where firms optimize given available technologies (at least given sufficient experience), we would expect this type of advice to have little to no impact. Instead, Bloom et al. (2013) find a large, 17 percent, increase in productivity in the treated plants.

Why does consulting help so much? This puzzle is magnified considering that much of the advice simply uses data that were already available to the manager. One piece of advice, for example, was to record defect rates by design. Another was to clean up trash from the shop floor. However, floor cleanliness naturally varies over time and the relationship between cleanliness and output could have been seen in the plant’s own data. Our interpretation is that managers could have acted on this relationship, if only they had *known to look*. The managers were stuck in noticing traps.³⁰

³⁰The famous Toyota problem-solving system also appears to be based in part on the idea that managers often did not notice important relationships in existing data (Liker, 2004). The “father” of the Toyota Problem Solving System,

A simple twist on this experiment, suggested by Proposition 1, would allow us to test this interpretation. In seaweed, the tell-tale sign was that farmers did not even know their pod size. One could similarly ask the manager questions about the features suggested in the consultation, prior to providing the advice. For example, does the manager know how often the factory floor is cleaned? Our model suggests that optimization failures are concentrated on dimensions where he would answer “I don’t know.” This approach would also allow an analyst to predict heterogeneity of treatment effects of the consulting advice, since plant managers may not have the same priors and thereby fail to notice different features.

Of course, consulting advice can come in many forms. Our theory suggests that, in principle, consultants could simply provide advice on what managers should pay attention to. However, Bloom et al.’s (2013) findings suggest potential limitations to such an approach: they find that diagnostic phases – in which each firm’s current practices are recorded and they are given recommendations for change – may not be as effective as also having implementation phases – in which some of the recommended practices are also demonstrated and fine-tuned with the help of consultants. Why? While this question is somewhat beyond the model’s scope, we conjecture that issues of trust likely come into play. Trust may mean a fear of deception or simply a tacit recognition that production functions may differ over place and time. The theoretical framework suggests a role for simple communication (to influence π_j), but only when the receivers trust the advice enough so that it substantially moves their beliefs on what to attend. In practice, that level of trust may not always be present across different settings.

Education

A teacher who aims to maximize student achievement (y) needs to consider many variables (\mathbf{x}): the intensity with which he covers various subjects, the ways in which he tests students, how he

Taiichi Ohno, would require trainees to stand in a circle drawn on the floor of a plant and observe highly routinized jobs (e.g., install a headlamp) over and over again until they could devise improvements. This is arguably all about attention: the “Ohno circles” do not encourage workers to gather new data, but rather they ask workers to look more carefully – to pay greater attentional costs – to extract more information from already available data. (Thanks to Tim Ogden for suggesting this example.)

interacts with students, etc. Here also, recent studies show that even experienced teachers are not on the educational production frontier. An evocative study is Allen et al.'s (2011), in which teachers review video recordings of their own classes, with an expert pointing out various details. This simple intervention showed large effects on measured student achievement in the year following completion of the intervention. The effect size is equivalent to moving a student from the 50th to the 59th percentile in test scores.

Our interpretation is that this intervention pointed out relationships in the teachers' own data that they failed to notice. For example, a teacher might have neglected how the precise way in which he handles students' questions appears to impact their engagement. Of course, in the process of reviewing video tapes, experts may have also communicated new information (such as "other teachers do this"), thereby doing more than simply pointing out regularities in the teachers' own data.³¹ A way to extend the intervention to test whether inattention underlies the original failure to optimize would be to first survey teachers about their own class behaviors. One could then videotape teachers both pre- and post-intervention, code up their practices on various dimensions, and see whether the intervention improves their teaching on the dimensions that they appear not to attend to as measured by pre-intervention surveys.

Extension

Agricultural extension services demonstrate profitable technologies on farmers' own plots to encourage adoption. Duflo, Kremer and Robinson (2008b) provide an example of how ineffective extension services can be. Farmers who observed a fertilizer trial on their own plot or a neighbor's plot showed modest increases in fertilizer use, but this did not last.

In our framework, a problem of extension is that farmers may not know what to notice while watching demonstrations. We illustrate this with a simple extension of the model in Appendix B. The result is that farmers can (correctly) believe from the demonstration that the technology

³¹However, Allen et al. themselves describe a major part of the intervention as having teachers "observe [video clips of] his or her behavior and student reactions and to respond to consultant prompts by noting the connection between the two" (Allen et al. 2011, page 1035).

is profitable but falsely believe that they learned how to use it. Indeed, they may not notice the dimensions necessary for proper use.³² The result can be a pattern of adoption and decay, like that found in the Duflo et al. (2008b) study. Farmers give up on using fertilizer because it does not produce the yields they thought it would.³³ The model further suggests a way of testing this mechanism: by eliciting farmers' beliefs about the demonstrator's actions along various input dimensions. The model predicts a greater decay effect when farmers do not accurately recall what the demonstrator did along essential dimensions.

Surgery

A final example comes from considering a surgeon who aims to maximize the post-operation health of a patient (y) through a multitude of choices (x), including her effort (e.g., how hard she concentrates, how much time she spends on the surgery), her alertness (e.g., the time of day the operation is scheduled, how many surgeries she does in a row), and how she interacts with the rest of the operating team (e.g., does she make sure everybody knows each others' name, do they discuss the case prior to operation). It is natural that she may not attend to some important factors.

Here, the evidence suggests that introducing checklists can reduce the incidence of complications. For example, Haynes et al. (2009) find that introducing a 19-item surgical safety checklist in eight hospitals reduced the rate of death in the 30 days following noncardiac surgery from 1.5 percent before the checklist was introduced to 0.8 afterwards, and more general inpatient complications from 11 to 7 percent. Checklists surely deal with simple forgetfulness: people forget a step

³²In a different context, Beaman et al. (2013) provide evidence that is potentially consistent with this hypothesis. They find that women rice farmers in Mali who received fertilizer increased output, but did not significantly increase profits. This is in contrast to Duflo et al. (2008a) who find that use of similar amounts of fertilizer significantly increases profits. A difference in the designs of their studies is that extension agents from a partner NGO helped farmers apply the fertilizer in Duflo et al.'s (2008a) study, while farmers were provided with fertilizer and brief training, but not assistance in applying the fertilizer, in Beaman et al.'s (2013). Despite training, Beaman et al.'s (2013) farmers may have suboptimally applied the fertilizer.

³³Similarly, Hanna, Duflo, and Greenstone (2012) explore take-up of improved cooking stoves. While having significant effects on smoke exposure for women in the first year, households persistently do not take up the stoves and use declines over time. Despite training and demonstrations that the stoves use less fuel, almost 40 percent of households that ever took up a stove claim that it uses the same amount or more fuel; this lack of accurate recall of information households should have learned is suggestive that inattention may be one of the reasons why take-up remained low.

that they know they should take. However, they may also help counteract selective inattention. Take a common checklist item: asking surgeons to ensure that all team members are introduced by name and role prior to skin incision. This can facilitate team communication, but when faced with more direct “medical” details, it is easy to imagine that surgeons under-value and fail to notice this one. The checklist can force attention on this detail. To explore this hypothesis, it would be useful to measure surgeons’ beliefs about the importance of the differing checklist items. The model suggests that the checklists’ benefits in part stem from including items that surgeons may believe are less important than they turn out to be.

5.2 Predicting Learning Failures

Our approach has been to bring empirical rigor to the model by exploiting the prediction that failures to optimize should go hand-in-hand with *measured* failures to notice. This largely leaves open the question of how we might *ex ante* predict failures to notice and resulting failures to learn based on features of technologies or environments. In the context of seaweed farming, for example, could we have *ex ante* predicted that farmers would notice the distance between pods, but not pod size? While not meant to be conclusive, here we sketch out some possible ideas.

First, agents’ past experiences with technologies can create blinders. Appendix B considers an extension of the model where the farmer sequentially uses different technologies and his beliefs about whether an input dimension is likely to be important depends on his experiences with earlier technologies. The key result is that experience has a potential cost: previous experience with a similar technology may “teach” the farmer to attend to the “wrong” things.³⁴ In the context of seaweed farming, for example, other agricultural experiences may have played a role in why

³⁴Rogers (2010) presents an example on the introduction of tractors in northern India, taken from Carter (1994). Tractors replaced bullocks as a way to power farms and provide transportation, but these tractors typically quickly broke down. Farmers did not perform routine maintenance, such as cleaning air filters and replacing oil filters. Moreover, they placed blankets over the tractors’ hoods in cold weather, which may lead the engines to overheat. They did this despite a foreign consultant who advised the farmers on maintenance routines. One possibility for the farmers’ persistent (mis)-behavior is that they transferred knowledge from their prior experience: bullocks did not require air filters to be cleaned or oil filters to be changed, but they did need to be protected from harsh winter weather. The model makes the further prediction that, all else equal, farmers would be more likely to learn to properly use tractors if they did not previously use a different technology, like bullocks, as a source of power.

farmers attend to distance between pods, but not pod size: it could be that the plant seed's size typically does not affect yield, while the distance between seeds does.

Second, the complexity of the technology – which we equate with the the number of dimensions N – creates greater demands on noticing. Suppose technologies are drawn from some distribution where a given input j matters with independent probability p . Further, suppose the agent's priors are drawn from some distribution such that the farmer believes input j is likely to be important ($\pi_j = \pi^H \in (0, 1)$) with independent probability q and likely to be unimportant ($\pi_j = \pi^L < \pi^H$) with the remaining probability. Then, the likelihood that the farmer initially believes that some important input is unlikely to matter is $[1 - (pq + (1 - p))^N]$, which is increasing in N and tends towards 1 as $N \rightarrow \infty$. Intuitively, the agent will miss at least one important input as the number of inputs increases.

Third, greater noise in the relationship between the input and outcome can make the person less likely to attend to the input since any given observation is less informative about the systematic part of the relationship. The empirical work above suggests more heterogeneity in optimal pod size than optimal distance between pods across farmers. While slightly outside the formal model that abstracts from social learning, such “noise” implies a lower value of attending to data from neighbors' plots and a greater attentional cost of learning the importance of size.³⁵

Finally, inputs that are naturally recorded are plausibly more likely to be attended-to and related to the outcome. In other words, some dimensions need neither attention nor memory to recollect. For example, even at the time of harvest, the distance between pods is still easily observable. The original pod size, however, can no longer be recovered simply by looking at the lines at harvest time. The farmer would have to remember the pod's size from the start of the season to relate it to the outcome.

³⁵While greater noise can also matter in other models since it impacts the speed with which relationships can be learned (e.g., Munshi 2004), it may matter much more when people are inattentive: greater noise can completely shut down learning by making it (subjectively) not worthwhile to attend.

6 Conclusion

In this paper, we propose an alternative hypothesis for learning failures: they stem not only from insufficient data, but also from individuals insufficiently attending to key features of data that they do possess. This perspective has important implications for how we think about the role of experience in learning, particularly challenging the standard intuition that experience guarantees effective technology use: while experience with a technology leads to improved performance along noticed dimensions, it can have little impact along other dimensions that are important, but neglected. Experience with a related technology can even have detrimental effects on learning when the input dimensions that are important across the technologies importantly fail to overlap. The model similarly provides insights into educational interventions, suggesting they are useful not only for new technologies, but also for existing technologies when there are indications that people are insufficiently attentive to key aspects of production. It also suggests ways of improving these interventions: there can be large benefits from moving away from just providing more data to also helping individuals understand the relationships in the data they that already have.

We test the model in the context of seaweed farming, showing that the farmers fail to optimize along input dimensions that they do not notice, but that helping them “see” relationships along those dimensions impacts their input choices. Looking forward, the model provides a framework for future empirical work in this area. It highlights the benefits of studies aimed at better understanding what predicts inattention to some dimensions while suggesting factors to look for, including previous experience with incongruent technologies. Similarly, research could further explore what gets people to start noticing important dimensions, for example the conditions under which communication between asymmetrically informed parties leads to better outcomes.

At the broadest level, the paper suggests a more nuanced view of human capital. Human capital is not summarized by exposure to data or experience: Embodied in individuals is information about what to notice and what to neglect.

References

- Alatas, V., A. Banerjee, R. Hanna, B. Olken, and J. Tobias**, “Targeting the Poor: Evidence from a Field Experiment in Indonesia,” *American Economic Review*, 2012, *102* (4), 1206–40.
- Allen, J., R. Pianta, A. Gregory, A. Mikami, and J. Lun**, “An Interaction-based Approach to Enhancing Secondary School Instruction and Student Achievement,” *Science*, 2011, *333* (6045), 1034–1037.
- Arrow, K.**, “The Economic Implications of Learning by Doing,” *The Review of Economic Studies*, 1962, *29* (3), 155–173.
- Banerjee, A.**, “A Simple Model of Herd Behavior,” *Quarterly Journal of Economics*, 1992, *107* (3), 797–817.
- Beaman, L., D. Karlan, B. Thuysbaert, and C. Udry**, “Profitability of Fertilizer: Experimental Evidence from Female Rice Farmers in Mali,” *American Economic Review*, 2013, *103* (3), 381–86.
- _____, **J. Magruder, and J. Robinson**, “Minding Small Change: Limited Attention Among Small Firms in Kenya,” *Journal of Development Economics*, *Forthcoming*, 2014.
- Besley, T. and A. Case**, “Modeling Technology Adoption in Developing Countries,” *American Economic Review*, 1993, *83* (2), 396–402.
- _____ **and** _____, “Diffusion as a Learning Process: Evidence from HYV Cotton,” *Mimeo*, 1994.
- Bikhchandani, S., D. Hirshleifer, and I. Welch**, “A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades,” *Journal of Political Economy*, 1992, pp. 992–1026.
- Bloom, N., B. Eifert, A. Mahajan, D. McKenzie, and J. Roberts**, “Does Management Matter: Evidence from India,” *Quarterly Journal of Economics*, 2013.

Bordalo, P., N. Gennaioli, and A. Shleifer, “Salience Theory of Choice Under Risk,” *Quarterly Journal of Economics*, 2012.

____, _____, and _____, “Salience and Consumer Choice,” *Journal of Political Economy*, 2013, 121 (5).

Carter, T., “The Process of Change: Tools for the Change Agent,” *Report, National Dairy Development Board*, 1994.

Conley, T.G. and C.R. Udry, “Learning About a New Technology: Pineapple in Ghana,” *American Economic Review*, 2010, 100 (1), 35–69.

Duflo, E., M. Kremer, and J. Robinson, “How High are Rates of Return to Fertilizer? Evidence from Field Experiments in Kenya,” *American Economic Review*, 2008, 98 (2), 482–488.

____, _____, and _____, “Why Don’t Farmers Use Fertilizer? Experimental Evidence from Kenya,” *Mimeo*, 2008.

Foster, A.D. and M.R. Rosenzweig, “Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture,” *Journal of Political Economy*, 1995, pp. 1176–1209.

____ and _____, “Microeconomics of Technology Adoption,” *Annual Reviews of Economics*, 2010, 2 (1), 395–424.

Fudenberg, D. and A. Peysakhovich, “Recency, Records and Recaps: The Effect of Feedback on Behavior in a Simple Decision Problem,” *Mimeo*, 2013.

____ and **D. Levine**, “Self-Confirming Equilibrium,” *Econometrica*, 1993, pp. 523–545.

Gabaix, X., “A Sparsity-Based Model of Bounded Rationality,” *Mimeo*, 2013.

Gawande, A., “On Washing Hands,” *New England Journal of Medicine*, 2004, 350 (13), 1283–1286.

- Gittins, J.C.**, “Bandit Processes and Dynamic Allocation Indices,” *Journal of the Royal Statistical Society. Series B (Methodological)*, 1979, pp. 148–177.
- Hanna, R., E. Duflo, and M. Greenstone**, “Up in Smoke: the Influence of Household Behavior on the Long-run Impact of Improved Cooking Stoves,” *Mimeo*, 2012.
- Haynes, A, T Weiser, W Berry, S Lipsitz, A Breizat, E Dellinger, T Herbosa, S Joseph, P Kibatala, M Lapitan et al.**, “A Surgical Safety Checklist to Reduce Morbidity and Mortality in a Global Population,” *New England Journal of Medicine*, 2009, 360 (5), 491–499.
- Jovanovic, B. and Y. Nyarko**, “Learning by Doing and the Choice of Technology,” *Econometrica*, 1996, 64 (6), 1299–1310.
- Kahneman, D.**, *Attention and Effort*, Prentice-Hall Inc., 1973.
- Kilby, P.**, “Organization and Productivity in Backward Economies,” *Quarterly Journal of Economics*, 1962, 76 (2), 303–310.
- Koszegi, B. and A. Szeidl**, “A Model of Focusing in Economic Choice,” *Quarterly Journal of Economics*, 2013.
- Leibenstein, H.**, “Allocative Efficiency vs. X-Efficiency,” *American Economic Review*, 1966, 56 (3), 392–415.
- Liker, J.**, *The Toyota Way*, Esensi, 2004.
- Munshi, K.**, “Social Learning in a Heterogeneous Population: Technology Diffusion in the Indian Green Revolution,” *Journal of Development Economics*, 2004, 73 (1), 185–213.
- Nelson, R. and S. Winter**, *An Evolutionary Theory of Economic Change*, Harvard University Press, 1982.
- Nelson, R.R. and E.S. Phelps**, “Investment in Humans, Technological Diffusion, and Economic Growth,” *American Economic Review*, 1966, 56 (1/2), 69–75.

Niehaus, P., “Filtered Social Learning,” *Journal of Political Economy*, 2011, 119 (4), 686–720.

Nuland, S.B., *The Doctors’ Plague: Germs, Childbed Fever, and the Strange Story of Ignac Semmelweis*, WW Norton & Company, 2004.

Rogers, E., *Diffusion of Innovations*, Simon and Schuster, 2010.

Schultz, T.W., “The Value of the Ability to Deal with Disequilibria,” *Journal of Economic Literature*, 1975, 13 (3), 827–846.

Schwartzstein, J., “Selective Attention and Learning,” *Journal of the European Economic Association*, 2014, Forthcoming.

Simon, H., “A Behavioral Model of Rational Choice,” *The Quarterly Journal of Economics*, 1955, 69 (1), 99–118.

Sims, C.A., “Implications of Rational Inattention,” *Journal of Monetary Economics*, 2003, 50 (3), 665–690.

Teghtsoonian, R., “On the Exponents in Stevens’ Law and the Constant in Ekman’s Law,” *Psychological Review*, 1971, 78 (1), 71–80.

Zwane, A.P., J. Zinman, E. Van Dusen, W. Pariente, C. Null, E. Miguel, M. Kremer, D.S. Karlan, R. Hornbeck, X. Giné et al., “Being Surveyed can Change Later Behavior and Related Parameter Estimates,” *Proceedings of the National Academy of Sciences*, 2011, 108 (5), 1821–1826.

Table 1: Baseline Demographic Characteristics and Farming Practices

	Mean (1)	Standard Deviation (2)	N (3)
<i>Panel A: Demographic Characteristics</i>			
Monthly Per Capita Expenditures (Indonesian Rph)	369543	348368	487
Age of HH Head (Years)	43.08	11.87	474
Number of Assets	8.09	3.23	487
HH Head is Literate	0.83	0.38	480
<i>Panel B: Seaweed Farming Practices</i>			
Years Farming Seaweed	18.36	7.15	475
Learned to Farm Seaweed from Parents	0.50	0.50	487
Has a Loan from Person to Whom Sells Seaweed	0.28	0.45	353
Number of Days in Previous Cycle	36.74	7.75	487
Mean Distance Between Lines at Baseline (Enumerator Measured; cm)	15.47	1.96	486
Mean Distance Between Pods at Baseline (Enumerator Measured; cm)	15.20	1.47	486
Mean Pod Size at Baseline (Enumerator Measured; grams)	105.74	28.72	487
Cottoni Strand	0.34	0.47	487

Notes: This table provides sample statistics on the farmers' demographic characteristics and seaweed farming practices from the baseline survey.

Table 2: Baseline Survey Responses on Process and Practices

	Percent Unable to Provide Answer (1)	Perceived Mean (2)
<i>Panel A: Self-Reported Current Production Methods</i>		
Typical Pod Size (grams)	86%	118.11 [57.01]
Typical Length of Line (cm)	2%	5.05 [1.04]
Typical Distance Between Lines (cm)	1%	16.49 [3.14]
<i>Panel B: Self-Reported Optimal Production Methods</i>		
Optimal Pod Size (grams)	87%	148.26 [248.45]
Optimal Distance Between Knots (cm)	2%	15.97 [2.84]
Optimal Distance Between Lines (cm)	2%	16.39 [3.01]
Optimal Cycle Length (days)	1%	37.43 [7.14]

Notes: This table provides sample statistics on 489 farmers' responses from the baseline survey. Standard deviations are in brackets.

Table 3: Estimated Percent Income Gain from Switching to Trial Recommendations

	Median	95 Percent Confidence Interval
	(1)	(2)
<i>Panel A: Sort Treatment Group</i>		
Gain to Moving from Average to Recommendation	7.06	[2.92 ,14.19]
Gain to Moving from Worst to Recommendation	23.3	[19.00, 28.18]
P-Value From F-Test of Equality of Coefficients	0.01	
<i>Panel B: Weight Treatment Group</i>		
Gain to Moving from Average to Recommendation	37.87	[23.60, 58.86]
P-Value From F-Test of Equality of Coefficients	0	

Notes: In the sort treatment, the F-tests come from separate farmer-level regressions of yield on the three size dummies (small, medium, large), where the null is that the coefficients on the dummies are equal. The F-tests in the weight treatment are constructed analogously.

Table 4: Effect of Participating in the Trial on Self-Reported Techniques and Measured Pod Size

	Changed Farming Techniques		Pod Size (Grams)	
	(1)	(2)	(3)	(4)
Trial Participation	-0.084 (0.051)		-2.184 (3.610)	
After Trial	-0.146 (0.048)***	-0.148 (0.057)**	-11.333 (3.003)***	-11.661 (3.578)***
After Summary Data	-0.145 (0.050)***	-0.150 (0.061)**	-13.587 (2.896)***	-13.859 (3.496)***
Trial Participation * After Trial	0.072 (0.060)	0.079 (0.071)	-2.051 (4.411)	-1.550 (5.306)
Trial Participation * After Summary Data	0.162 (0.069)**	0.171 (0.084)**	6.951 (4.095)*	7.316 (4.982)
Hamlet Fixed Effects	X		X	
Farmer Fixed Effects		X		X
Observations	684	684	684	684
<i>Testing the Equality of Coefficients on Trial Participation * After Trial and Trial Participation * After Summary Data</i>				
F-Statistic	2.10	1.51	8.79	5.95
P-Value	0.148	0.221	0.003	0.015
<i>Mean of Dependent Variable for the Control Group:</i>				
After Trial	0.10	0.10	97.68	97.68
After Summary Data	0.11	0.11	95.39	95.39

Notes: This table provides the coefficient estimates of the effect of treatment on farming methods after the trial (follow-up 1) and after observing the summary data (follow-up 2), conditional on baseline farming methods. The trial participation dummy indicates that the farmer belongs in either the sort or weight treatment group. Changed Farming Techniques includes self-reported changes in pod size and distances, while pod size is enumerator measured. All regressions are estimated using OLS and standard errors are clustered at the farmer level. Statistical significance is denoted by: *** p<0.01, ** p<0.05, * p<0.10.

Table 5: Effect of Trial Participation in Sort Versus Weight Treatments

	Changed Farming Techniques		Pod Size (Grams)		Distance Between Pods (Cm)	
	(1)	(2)	(3)	(4)	(5)	(6)
Sort * After Trial	0.089 (0.065)	0.100 (0.077)	3.944 (4.461)	4.657 (5.310)		
Weight * After Trial	0.052 (0.075)	0.053 (0.089)	-9.257 (6.610)	-8.929 (7.882)	0.289 (0.328)	0.304 (0.387)
Sort * After Summary Data	0.141 (0.075)*	0.153 (0.091)*	10.908 (4.418)**	11.768 (5.286)**		
Weight * After Summary Data	0.187 (0.095)*	0.192 (0.114)*	2.185 (5.819)	2.093 (7.002)	0.226 (0.303)	0.172 (0.362)
Hamlet Fixed Effects	X		X		X	
Farmer Fixed Effects		X		X		X
Observations	684	684	684	684	499	499
<i>Mean of Dependent Variable for the Control Group:</i>						
After Trial	0.10	0.10	97.68	97.68	15.39	15.39
After Summary Data	0.11	0.11	95.39	95.39	15.27	15.27

Notes: This table provides the coefficient estimates of the effect of the different treatments on farming methods after the trial (follow-up 1) and after observing the summary data (follow-up 2), conditional on baseline farming methods. Changed Farming Techniques includes self-reported changes in pod size and distances, while pod size and distance between pods are enumerator measured. All regressions are estimated using OLS and standard errors are clustered at the farmer level. Statistical significance is denoted by: *** p<0.01, ** p<0.05, * p<0.10.

Figure 1A: Distribution of Baseline Pod Sizes (in Grams)

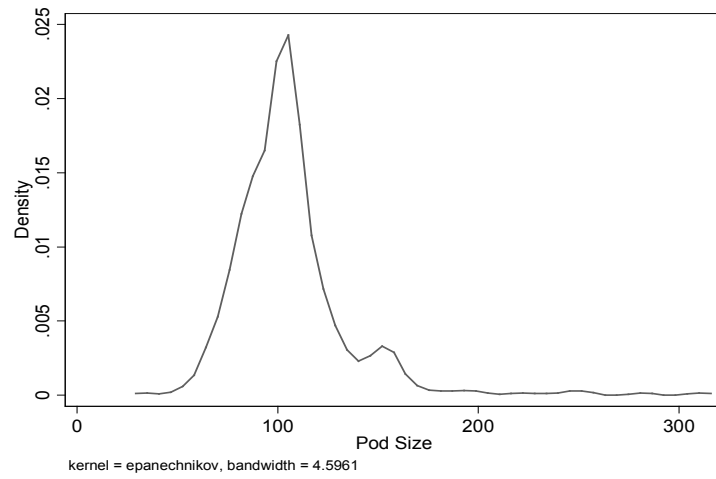


Figure 1B: Distribution of Baseline Pod Sizes for Cottoni Growers

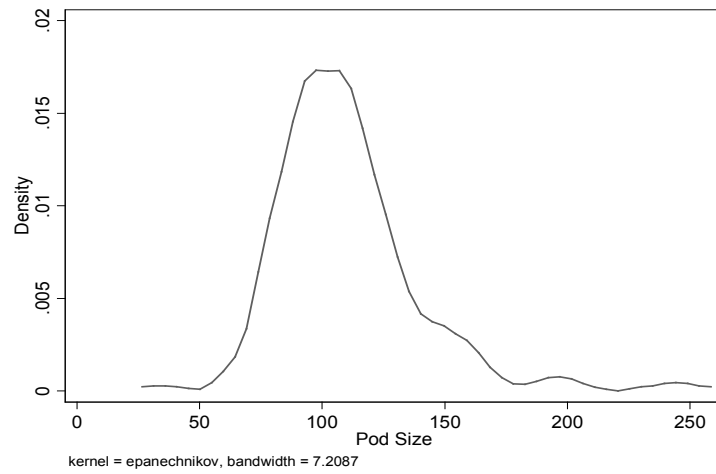


Figure 1C: Distribution of Baseline Pod Sizes for Spinosim Growers

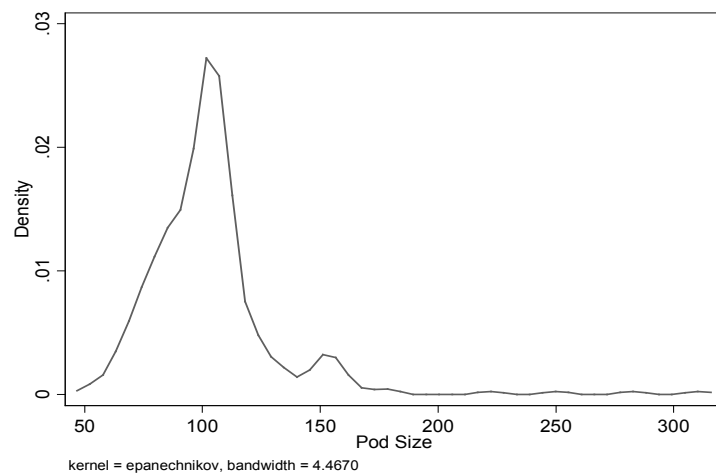
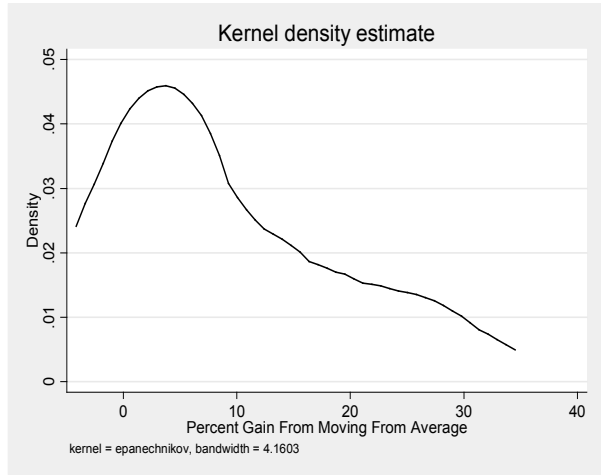
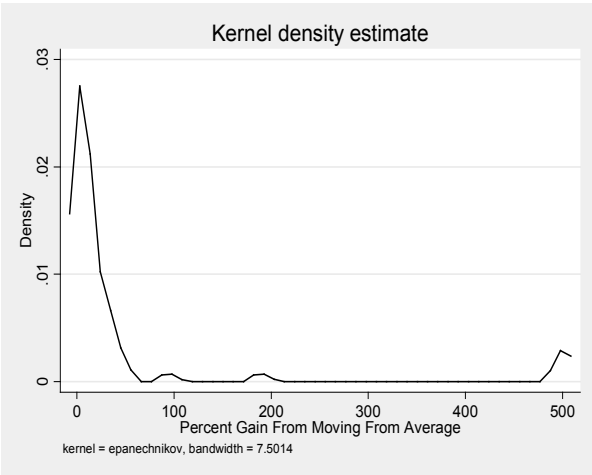
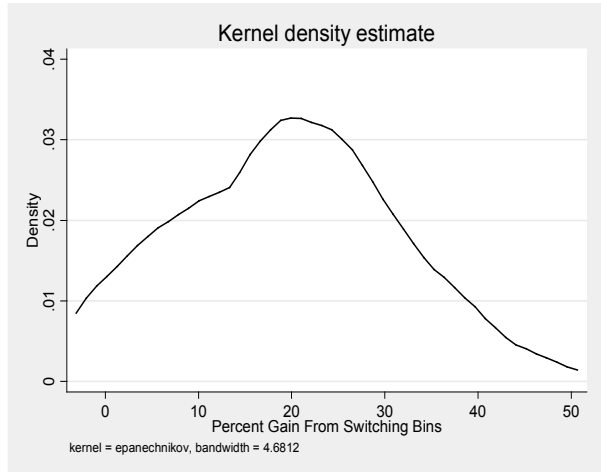
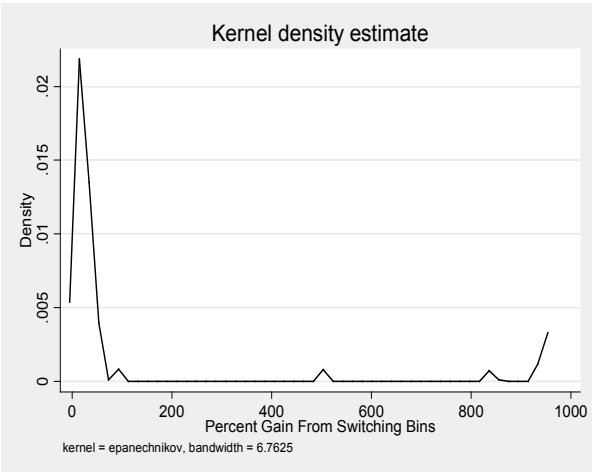


Figure 2: Distribution of the Estimated Percent Income Gain from Switching to Trial Recommendations

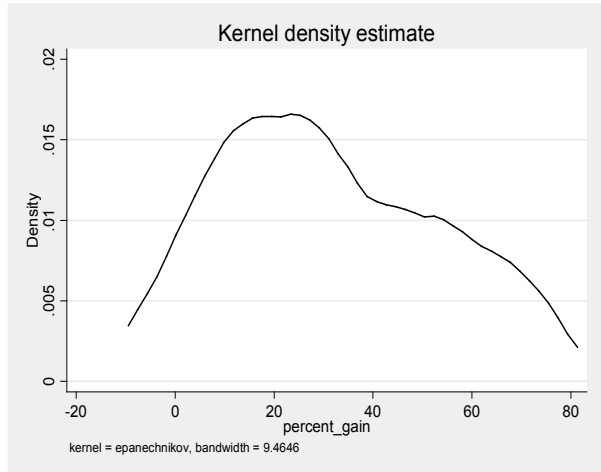
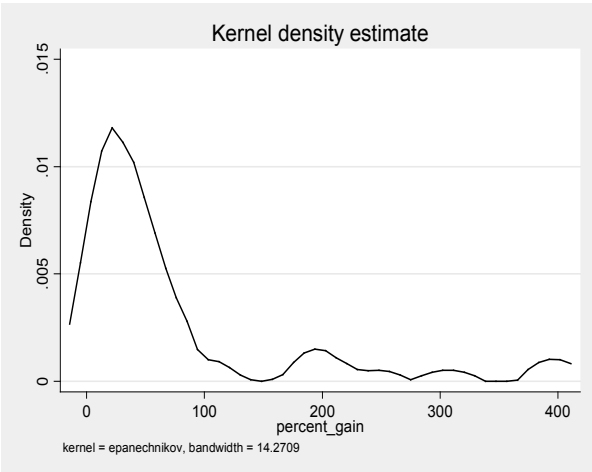
Panel A: Percent Gain to Moving from the Baseline Average to the Recommended Bin in the Sort Treatment



Panel B: Percent Gain to Moving from the Lowest Performing to the Recommended Bin in the Sort Treatment

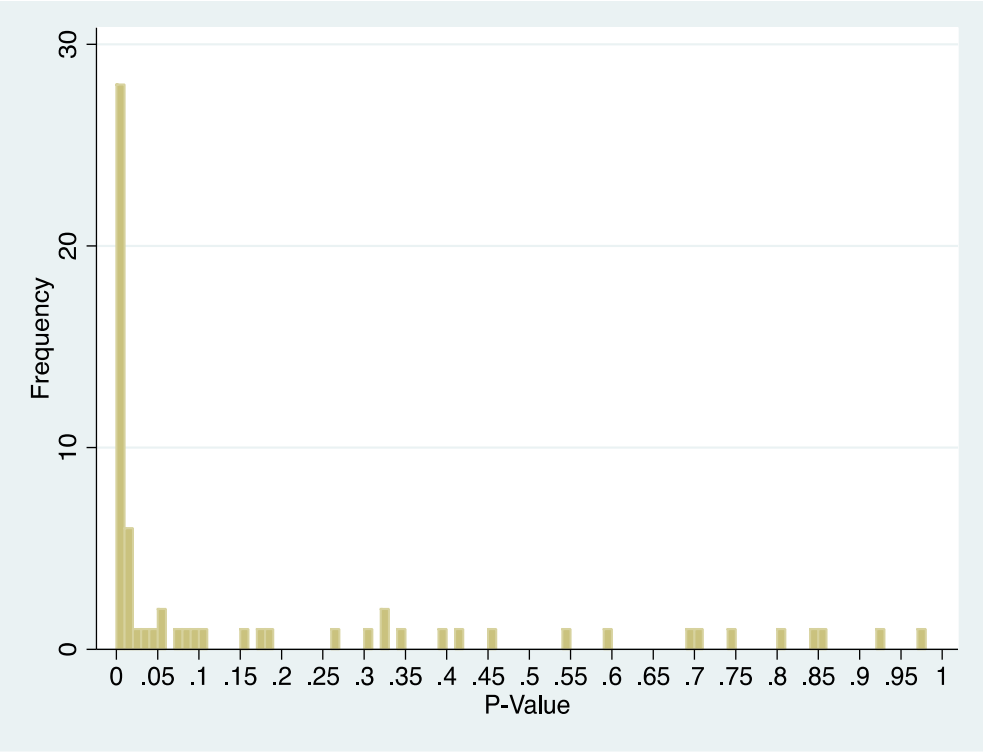


Panel C: Percent Gain to Moving from the Baseline Average to the Recommended Bin in the Weight Treatment



Notes: The first column of figures provides the full set of returns, while the second column focuses on the bottom 80 percent of the sample.

Figure 3: Distribution of P-Values from Farmer-Level Tests of Whether Pod Size Matters in the Sort Treatment



Online Appendix

A Proofs

A.1 Notation and Preliminary Lemmas

Before turning to the proofs of propositions, we introduce other notation and useful lemmas.

To highlight the role that mixed strategies play in some of the analysis, let $\sigma_{jt} \in \Delta(X_j)$ denote the farmer's input choice along dimension j at time t . Now define

$$v_{jt} = v_j(\sigma_{jt}, a_{jt}; \theta) = \sum_{x'_j} \sigma_{jt}(x'_j) f_j(x'_j | \theta_j) - e \cdot a_{jt}$$

to equal net payoffs along dimension j in period t as a function of the farmer's decisions that period and the (unknown) technology θ , where σ_{jt} necessarily takes on the default uniform distribution when $a_{jt} = 0$. We denote the default distribution by σ_j^d . The farmer's flow payoff in period t then equals $v_t = \sum_j v_{jt}$, so what he is maximizing is the expected value (under his prior) of $v_1 + v_2$. We denote the maximum attainable payoff given technology θ as $v^*(\theta) = \max_{\sigma, \mathbf{a}} v(\sigma, \mathbf{a}; \theta) = \sum_j v_j^*(\theta_j)$, where each $v_j^*(\theta_j)$ is defined in the obvious manner, and arguments that attain this maximum as $(\sigma^*(\theta), \mathbf{a}^*(\theta))$. For a final bit of notation, denote the total loss in profits from period- t farming practices relative to the (static) optimum by $L_t = v^*(\theta) - v_t$ and the loss along dimension j as $L_{jt} = (v_j^*(\theta_j) - v_{jt})$.

An optimal strategy of the farmer entails that in the second (terminal) period, for every input dimension j he selects (σ_j, a_j) to maximize $E[v_j(\sigma_j, a_j; \tilde{\theta}_j) | \hat{h}]$, given any recalled history \hat{h} . Let $(\sigma_j(\hat{h}), a_j(\hat{h}))$ denote such an optimal choice of (σ_j, a_j) . It will be useful to compare $(\sigma_j(\hat{h}), a_j(\hat{h}))$ to the farmer's optimal terminal period decisions if he (hypothetically) perfectly knew the underlying technology, θ .³⁶

Lemma A.1.

1. *If the farmer attended to input j in the first period and followed a first period strategy in which he placed positive probability on each input in X_j on every plot of land, then he learns to optimally set input j given the underlying technology θ : If $a_{j1} = 1$ and $\sigma_{j1}(x'_j) > 0$ for all $x'_j \in X_j$, then $(\sigma_j(\hat{h}), a_j(\hat{h})) = (\sigma_j^*(\theta), a_j^*(\theta))$.*
2. *If the farmer did not attend to input j in the first period, then he will not attend to input j in the second period, and consequently may not optimally set input j given the underlying technology θ : If $a_{j1} = 0$ and $e > 0$, then $(\sigma_j(\hat{h}), a_j(\hat{h})) = (\sigma_j^d, 0)$.*

Proof. 1. Suppose the farmer attended to input j in the first period and followed a first period strategy under which $\sigma_{j1}(x'_j) > 0$ for all $x'_j \in X_j$. Then for all $x_j \in X_j$, the farmer can calculate the empirical average gain from using x_j relative to not attending, $\bar{g}_j(x_j)$, which reveals the population

³⁶The notation ignores knife-edge cases where the farmer is indifferent between various options in the second period. To handle these cases, just consider an arbitrary tie-breaking rule.

expected static gain from using x_j relative to not attending:

$$\bar{g}_j(x_j) = (\bar{y}_j(x_j) - e) - \frac{1}{|X_j|} \sum_{x'_j} \bar{y}_j(x'_j) = (f_j(x_j|\theta_j) - e) - \frac{1}{|X_j|} \sum_{x'_j} f_j(x'_j|\theta_j).$$

The farmer's posterior will thus place probability 1 on $\tilde{\theta}$ satisfying $(f_j(x_j|\tilde{\theta}_j) - e) - \frac{1}{|X_j|} \sum_{x'_j} f_j(x'_j|\tilde{\theta}_j) = (f_j(x_j|\theta_j) - e) - \frac{1}{|X_j|} \sum_{x'_j} f_j(x'_j|\theta_j)$ for all $x_j \in X_j$.

The farmer then optimally sets

$$\begin{aligned} a_j(\hat{h}) &= \begin{cases} 1 & \text{if } \max_{x'_j \in X_j} \bar{g}_j(x'_j) > 0 \\ 0 & \text{otherwise} \end{cases} \\ &= a_j^*(\theta). \end{aligned}$$

Further, for recalled histories satisfying $a_j(\hat{h}) = 1$, he sets $\sigma_j(\hat{h})[x_j^*(\hat{h})] = 1$, where $x_j^*(\hat{h}) \in \arg \max_{x'_j \in X_j} \bar{g}_j(x'_j)$, which implies that $\sigma_j(\hat{h}) = \sigma_j^*(\theta)$.

2. For any \hat{h} consistent with not attending to input j in the first period, there exists a constant $K \in \mathbb{R}$ such that

$$E[f_j(x'_j|\tilde{\theta}_j)|\hat{h}] = E[\tilde{\theta}_j(x'_j)|\hat{h}] = K, \quad \forall x'_j \in X_j. \quad (5)$$

Equation (5) follows from the fact that the farmer's marginal prior distributions over $\tilde{\theta}_j(x'_j)$ are the same across $x'_j \in X_j$, as are the likelihood functions $\Pr(\hat{h}|\tilde{\theta}_j(x'_j))$ given any \hat{h} consistent with not attending to input j .

From (5), the farmer's terminal-period expected payoff function along dimension j satisfies

$$E[v_j(\sigma_j^d, 0)|\hat{h}] = K = E[v_j(x_j, 1)|\hat{h}] + e$$

$\forall x_j \in X_j$ given any \hat{h} consistent with the farmer not attending to input j , which implies that $(\sigma_j(\hat{h}), a_j(\hat{h})) = (\sigma_j^d, 0)$ for each \hat{h} when $e > 0$. ■

The first part of Lemma A.1 considers what happens when the farmer attends to input j in the first period and follows a first period strategy in which he places positive probability on each input in X_j on every plot of land, and shows that in this case the farmer learns to optimally set input j . This result leaves open the question of what happens if the farmer attends to input j , but follows a strategy in which he does not experiment with certain inputs in X_j . The next result shows that we need not concern ourselves with such strategies: among strategies where the farmer attends to input j in the first period, it is without loss of generality to confine attention to strategies where he experiments with each input in X_j on a positive measure of plots of land.

Lemma A.2. *If it is optimal for the farmer to attend to input j in the first period, then he optimally follows a strategy under which $\sigma_{j1}(x'_j) > 0$ for all $x'_j \in X_j$.*

Proof. The farmer's expected payoff if he attends to input j equals

$$E[v_{-j1} + v_{-j2}] + E[v_j(\sigma_{j1}, 1; \tilde{\theta}_j) + v_j(\sigma_{j2}, a_{j2}; \tilde{\theta}_j)] \leq E[v_{-j1} + v_{-j2}] + E[-e + v_j^*(\tilde{\theta}_j)],$$

where $v_{-jt} \equiv \sum_{i \neq j} v_{it}$, and the upper bound follows from the fact that $E[\tilde{\theta}_j(x'_j)] = 0 \forall x'_j$ and, for all $\tilde{\theta}_j$, $v_j(\sigma_{j2}, a_{j2}; \tilde{\theta}_j) \leq v_j^*(\tilde{\theta}_j)$ by the definition of $v_j^*(\tilde{\theta}_j)$. But, by Lemma A.1.1, this upper bound is attained through a strategy under which $\sigma_{j1}(x'_j) > 0$ for all $x'_j \in X_j$ (and clearly cannot be attained through any strategy under which $\sigma_{j1}(x''_j) = 0$ for some $x''_j \in X_j$, since, in this case, $\sigma_j(\hat{h}) \neq \sigma_j^*(\theta_j)$ with positive probability). ■

In the remaining proofs and extensions, we will make use of the following measure of the degree to which a dimension is relevant.

Definition A.1. Input dimension j is K -relevant to using the technology θ if

$$\max_{\tilde{x}_j} f_j(\tilde{x}_j | \theta_j) - \frac{1}{|X_j|} \sum_{x'_j} f_j(x'_j | \theta_j) > K.$$

Otherwise, dimension j is K -irrelevant to using the technology.

Input dimension j is K -relevant to using the technology if optimizing the dimension yields an expected payoff of at least K relative to selecting the input along that dimension at random. We say that a dimension is *relevant* when it is 0-relevant.

A.2 Proofs of Propositions

Proof of Proposition 1. Because the payoff function is separable across input dimensions and the prior uncertainty is independent across dimensions, we can separately analyze the farmer's decisions across dimensions. For a given input j , the farmer's first period decision boils down to choosing between:

1. Not attending to input j in the first period.
2. Attending to input j in the first period, and following a strategy under which $\sigma_{j1}(x'_j) > 0$ for all $x'_j \in X_j$.

(By Lemma A.2, we don't need to consider strategies in which the farmer attends to input j in the first period and sets $\sigma_{j1}(x'_j) = 0$ for some x'_j .)

If the farmer does not attend to input j in the first period, then his expected payoff along dimension j equals:

$$\begin{aligned} E[v_j(\sigma_j^d, 0; \tilde{\theta}_j) + v_j(\sigma_{j2}, a_{j2}; \tilde{\theta}_j)] &= 2E[v_j(\sigma_j^d, 0; \tilde{\theta}_j)] \\ &= 2E \left\{ \frac{1}{|X_j|} \sum_{x'_j} \tilde{\theta}_j(x'_j) \right\} \\ &= 0, \end{aligned} \tag{6}$$

where the first equality follows from Lemma A.1.2 and the last from the fact that $E[\tilde{\theta}_j(x'_j)] = 0$ for all $x'_j \in X_j$.

Conversely, if the farmer attends to input j in the first period and follows a strategy under which $\sigma_{j1}(x'_j) > 0$ for all $x'_j \in X_j$, then his expected payoff along dimension j equals:

$$\begin{aligned}
E[v_j(\sigma_{j1}, 1; \tilde{\theta}_j) + v_j(\sigma_{j2}, a_{j2}; \tilde{\theta}_j)] &= E[v_j(\sigma_{j1}, 1; \tilde{\theta}_j) + v_j^*(\tilde{\theta}_j)] \\
&= -e + E[v_j^*(\tilde{\theta}_j)] \\
&= -e + \pi_j \cdot E[v_j^*(\tilde{\theta}_j) | j \text{ relevant}] \\
&= -e + \pi_j \cdot \{ \Pr(j \text{ } e\text{-relevant} | j \text{ relevant}) \cdot E[v_j^*(\tilde{\theta}_j) | j \text{ } e\text{-relevant}] \\
&\quad + \Pr(j \text{ } e\text{-irrelevant} | j \text{ relevant}) \cdot E[v_j^*(\tilde{\theta}_j) | j \text{ } e\text{-irrelevant}] \} \\
&= -e + \pi_j \cdot \{ \Pr(j \text{ } e\text{-relevant} | j \text{ relevant}) \cdot E[v_j^*(\tilde{\theta}_j) - \bar{v}_j(\tilde{\theta}_j) | j \text{ } e\text{-relevant}] \}, \tag{7}
\end{aligned}$$

where the first equality follows from Lemma A.1.1, the second from the fact that $E[\tilde{\theta}_j(x'_j)] = 0$ for all $x'_j \in X_j$, the third from the observation that $v_j^*(\theta_j) = 0$ if j is irrelevant, the fourth from expanding the expression using Bayes' rule, and the final from subtracting off $0 = E[v_j(\sigma_j^d, 0; \tilde{\theta}_j)] \equiv E[\bar{v}_j(\tilde{\theta}_j)]$ and observing that $E[v_j^*(\tilde{\theta}_j) - \bar{v}_j(\tilde{\theta}_j) | j \text{ } e\text{-irrelevant}] = 0$. Note that $E[v_j^*(\tilde{\theta}_j) - \bar{v}_j(\tilde{\theta}_j) | j \text{ } e\text{-relevant}] > 0$ (by the definition of e -relevance), so the right hand side of Equation (7) is increasing in π_j , and tends towards $-e$ as $\pi_j \rightarrow 0$.

By comparing Equations (6) and (7), we see that the farmer chooses to attend to input j if and only if the right hand side of Equation (7) is positive. This is clearly true when $e = 0$ and $\pi_j > 0$ (recall that $E[v_j^*(\tilde{\theta}_j) - \bar{v}_j(\tilde{\theta}_j) | j \text{ } e\text{-relevant}] > 0$), so in this case the farmer attends to all input dimensions in the first period. Combined with Lemma A.1.1, this establishes part 1 of the proposition.

For part 2a, recall the assumption that at least one dimension, say j , is worth attending to, i.e., is e -relevant. For π_j sufficiently low, the right-hand-side of Equation (7) is negative and the farmer does not attend to input j in the first period. By Lemma A.1, this implies that the farmer does not optimize along dimension j , where the loss from not optimizing equals

$$L_{j2} = \left(\max_{\tilde{x}_j} f_j(\tilde{x}_j | \theta_j) - e \right) - \frac{1}{|X_j|} \sum_{x'_j} f_j(x'_j | \theta_j) > 0$$

and the inequality follows from the assumption that j is e -relevant.

For part 2b, fix some $K \in \mathbb{R}^+$. Under the assumption that at least one dimension j is worth attending to, there exists a dimension j where $|X_j| \geq 2$. We proceed by establishing that there exists a θ' under which j is e -relevant and $L_{j2} > K$, which implies that the farmer loses at least K from not optimizing when π_j is sufficiently small. Consider some θ' , where

$$\theta'_j = (K + e, \underbrace{-K - e - \varepsilon, -K - e - \varepsilon, \dots, -K - e - \varepsilon}_{|X_j| - 1 \text{ times}}),$$

given an $\varepsilon > 0$. Plugging this θ'_j into the above formula for L_{j2} yields

$$\begin{aligned} L_{j2} &= K + e - e - \frac{1}{|X_j|} (K + e - (|X_j| - 1) \cdot (K + e + \varepsilon)) \\ &= K + \frac{1}{|X_j|} [\varepsilon + (|X_j| - 2)_+ \cdot (K + e + \varepsilon)] \\ &> K, \end{aligned}$$

where $(\cdot)_+$ denotes an operator where $(Y)_+ = Y$ if $Y \geq 0$ and $(Y)_+ = 0$ otherwise.

For part 3, $a_{j1} = 1$ implies that $\sigma_{j1}(x'_j) > 0$ for all $x'_j \in X_j$ (by Lemma A.2), which implies that the farmer learns to optimize dimension j (by Lemma A.1.1). ■

Proof of Proposition 2. First consider dimensions that the farmer attended to in the first period. The farmer learns to optimize such dimensions by Lemmas A.1 and A.2.

Next consider dimensions that the farmer did not attend to in the first period. Because the farmer did not attend to these dimensions, he uses each input in X_j infinitely often in the first period: $a_{j1} = 0$ implies that $\sigma_{j1} = \sigma_j^d$, where $\sigma_j^d(x'_j) > 0$ for all $x'_j \in X_j$. As a result, the empirical average gain from using x_j relative to not attending, $\bar{g}_j(x_j)$, reveals the population expected static gain from using x_j relative to not attending:

$$\bar{g}_j(x_j) = (\bar{y}_j(x_j) - e) - \frac{1}{|X_j|} \sum_{x'_j} \bar{y}_j(x'_j) = (f_j(x_j|\theta_j) - e) - \frac{1}{|X_j|} \sum_{x'_j} f_j(x'_j|\theta_j).$$

While the farmer cannot compute $\bar{g}_j(x_j)$ for every x_j , he need only compute $\bar{g}_j(\tilde{x}_j^*)$ to make optimal decisions in the second period, which he can derive from the summary information together with \hat{h} : $\bar{g}_j(\tilde{x}_j^*) = (\bar{y}_j(\tilde{x}_j^*) - e) - \bar{y}$, where $\bar{y} = \int y_{l1} dl = f_j(\sigma_j^d|\theta_j) + \sum_{k \neq j} f_k(\sigma_{k1}|\theta_k)$ equals the empirical average yield across all plots of land. The farmer's posterior will thus place probability one on $\tilde{\theta}$ satisfying $(\max_{x_j} f_j(x_j|\tilde{\theta}_j) - e) - \frac{1}{|\tilde{X}_j|} \sum_{x'_j} f_j(x'_j|\tilde{\theta}_j) = \bar{g}_j(\tilde{x}_j^*) = (\max_{x_j} f_j(x_j|\theta_j) - e) - \frac{1}{|X_j|} \sum_{x'_j} f_j(x'_j|\theta_j)$ and $\tilde{x}_j^* \in \arg \max_{x_j} f_j(x_j|\tilde{\theta}_j)$, where $\tilde{x}_j^* \in \arg \max_{x_j} f_j(x_j|\theta_j)$.

The farmer then optimally sets

$$\begin{aligned} a_j &= \begin{cases} 1 & \text{if } \bar{g}_j(\tilde{x}_j^*) > 0 \\ 0 & \text{otherwise} \end{cases} \\ &= a_j^*(\theta) \end{aligned}$$

and, when $a_j = 1$, sets $x_j = \tilde{x}_j^* = x_j^*(\theta)$. ■

B Extensions

This appendix elaborates on the extensions and results we refer to in Section 5.

B.1 Further Setup and Concepts

Before presenting these extensions, we place a bit more structure on the model and introduce some useful concepts.

To fix ideas, we restrict $\pi_j \in \{\pi^L, \pi^H\}$ for each j , where $0 < \pi^L < \pi^H < 1$. We interpret $\pi_j = \pi^H$ as a situation where the farmer believes that input j is likely to be relevant, or likely that his input choice along this dimension will impact his payoff in an *a priori* unknown manner, and $\pi_j = \pi^L$ as a situation where he believes it is likely that input j is *not* relevant, or likely that his input choice along this dimension will not impact his payoff.

We will say that a technology is *prior congruent* if the features that matter for using the technology line up well with the features that the farmer thinks should matter, and is *prior incongruent* otherwise. Formally:

Definition B.1. The technology is *prior congruent* if $\pi_j = \pi^H$ for all dimensions j that are *e*-relevant to using the technology. On the other hand, when $\pi_j = \pi^L$ for some dimension j that is *e*-relevant to using the technology then the technology is *prior incongruent*.

The role of this definition will be clearer after we state the following simplifying assumption. To limit the number of cases considered, it will be useful to focus on parameter values such that the farmer chooses to attend to an input in the first period when he places sufficiently large prior weight on the importance of that input. This is true, for example, whenever the cost of attending (e) is not too large.

Assumption B.1. Parameter values are such that the solution to the farmer's problem satisfies $a_{j1} = 1$ if $\pi_j = \pi^H$.

Combined with the definition of prior congruence, this assumption implies that a technology is prior congruent if and only if the farmer initially attends to all input dimensions that are worth attending to when optimally set.³⁷ Whether the technology is prior congruent predicts whether the farmer will learn to optimize the technology.

B.2 Extension 1: Sequential Technologies

To explore how experience with related technologies may influence whether the farmer learns to optimize some target technology, consider an extension of the model where a farmer sequentially uses different technologies $\tau = 1, 2, \dots$ for two periods each. His behavior over the two periods for a fixed technology is as described by our baseline model, and his prior over the importance of variable j given technology τ , $\pi_j^\tau \in \{\pi^L, \pi^H\}$, depends on his experience with previous technologies.

³⁷Thus, a technology is prior congruent even if the farmer initially attends to some variables that are not worth attending to, so long as he attends to all variables that are worth attending to: Our notion of prior congruence ignores biases of commission. This is natural when the focus is primarily on understanding long-run behavior (as is ours), but other notions are relevant in understanding short-run behavior since it is costly for the farmer to attend to relationships that ultimately do not matter for profits.

Specifically, suppose that, when using a technology for the first time, the farmer has a prior belief that input dimension j is important if and only if he attended to that dimension the period before when he used the directly preceding technology:

$$\text{For } \tau > 1, \pi_j^\tau = \pi^H \text{ if and only if } a_{j2}^{\tau-1} = 1. \quad (8)$$

To complete the description of the sequential technology extension of the model, we need to specify initial conditions for the farmer's prior; that is, his prior for technology $\tau = 1$. It is intuitive that the farmer's prior is initially symmetric across input dimensions, reflecting the idea that the farmer must learn which dimensions are important. To keep the model interesting, we further assume that he starts off believing that every dimension is likely to be important:³⁸

$$\pi_j^1 = \pi^H \text{ for all } j. \quad (9)$$

To gain intuition for how previous experience can influence whether a given technology is prior congruent, we first highlight a feature of the baseline model:

Lemma B.1. *For all τ , π_j^τ , and $e > 0$, $a_{j2}^\tau = 1$ only if dimension j is e -relevant for technology τ .*

Proof. Suppose dimension j is *not* e -relevant for technology τ . There are two cases to consider. First, if $a_{j1}^\tau = 1$, then $a_{j2}^\tau = a_j^*(\theta^\tau) = 0$ by Lemmas A.1.1 and A.2. Conversely, if $a_{j1}^\tau = 0$, then $a_{j2}^\tau = 0$ by Lemma A.1.2. ■

Lemma B.1 says that, for a fixed technology, the farmer will eventually not attend to any dimension that is not worth attending to. The intuition is straightforward: If the farmer initially attends to a dimension that is not worth attending to, he will learn to stop; if he initially does not, he will continue not to attend. From condition (8), an immediate corollary is that a technology $\tau > 1$ is prior congruent only if all input dimensions j that are e -relevant for technology τ were also e -relevant for technology $\tau - 1$, since the farmer will initially place low weight on the importance of any input dimension that he stopped attending to when using the last technology. We get a stronger result when we assume that e is sufficiently large that the prior influences what a farmer attends to:

Proposition 3. *Suppose that the farmer sequentially uses technologies $\tau = 1, 2, \dots$ and his prior satisfies conditions (8) and (9). For sufficiently large e , technology $\tau > 1$ is prior congruent if and only if all input dimensions j that are e -relevant for that technology are also e -relevant for every technology $1, 2, \dots, \tau - 1$.*

Proof. Consider a technology $\tau > 1$. For the “if” direction, suppose that all input dimensions that are e -relevant for τ are also e -relevant for technologies $1, 2, \dots, \tau - 1$. The goal is to show that $\pi_j^\tau = \pi^H$ for any dimension j that is e -relevant for technology τ . Consider any one such dimension, k . Combining the assumption that $\pi_k^1 = \pi^H$ (Equation (9)) with Assumption B.1 implies that $a_{k1}^1 = 1$. Since k is e -relevant it further follows that $a_{k2}^1 = a_k^*(\theta_k^1) = 1$ by Lemmas A.1 and A.2, which implies that $\pi_k^2 = \pi^H$ by Equation (8). Iterating this argument yields the desired conclusion.

³⁸Otherwise, experience with other technologies will not matter: Depending on the level of e , the farmer will either learn to optimize every technology (when e is sufficiently low that he initially attends to everything) or no technologies (when e is sufficiently large that he initially attends to nothing).

For the “only if” direction, suppose that for some technology $\tilde{\tau} < \tau$ there is some dimension j that is not e -relevant for $\tilde{\tau}$, but is e -relevant for τ . Since j is not e -relevant for $\tilde{\tau}$, $a_{j2}^{\tilde{\tau}} = 0$ by Lemma B.1, which implies that $\pi_j^{\tilde{\tau}+1} = \pi^L$ by Equation (8). For e sufficiently large, this then implies that $a_{j1}^{\tilde{\tau}+1} = 0$ from inspecting Equation (7) in the proof of Proposition 1, which in turn implies that $a_{j2}^{\tilde{\tau}+1} = 0$ by Lemma A.1.2, and finally that $\pi_j^{\tilde{\tau}+2} = \pi^L$ by Equation (8). Iterating this argument implies that $\pi_j^{\tilde{\tau}} = \pi^L$, which implies that technology τ is not prior congruent. ■

Proposition 3 says that a technology is prior congruent if and only if input dimensions that are e -relevant to using the current technology were also e -relevant to using *every* preceding technology. The farmer starts off attending to everything, but stops attending to a dimension the first time he encounters a technology under which it is not worth attending to.

Proposition 3 implies path dependence, where the same technology can be prior congruent or incongruent, depending on the technologies the farmer previously encountered. A more unique implication is that experience with related technologies may not be beneficial—and may even be harmful— if the relevant inputs differ from those of the current technology. Proposition 3 implies the particularly stark result that, all else equal, the same technology is *less* likely to be prior congruent if it occurs later in the sequence $\tau = 1, 2, \dots$ since this gives a farmer more opportunities to stop attending to a given input dimension.³⁹

B.3 Extension 2: Providing Demonstrations

A common intervention for improving productivity is to provide demonstrations. What is the impact of demonstrations when farmers are inattentive?

To incorporate demonstrations in our model, suppose now that there is an additional period 0, where the farmer does not farm, but rather observes a demonstration by an individual with knowledge of the underlying production function. We will call this individual a *best-practice demonstrator*. The best practice demonstrator optimally farms the farmer’s plot, i.e., he follows $(\sigma^*(\theta), \mathbf{a}^*(\theta))$, and it is common knowledge that the demonstrator knows how to optimize the technology; the farmer’s prior over the demonstrator’s strategy is thus derived from his prior over θ . We are sidestepping (potentially interesting) issues that could arise from the farmer being uncertain about whether the strategy followed by the demonstrator would be optimal for him either because of uncertainty over whether the demonstrator optimizes or over whether his plot is different.

Proposition 4. *Consider a farmer who learns from a best-practice demonstrator in period 0 and then decides how to farm in periods 1 and 2.*

1. *If there are no costs of attention ($e = 0$), then the farmer learns to optimize the technology from watching the demonstrator: $L_1 = L_2 = 0$.*

³⁹While this discussion highlights a potential cost of experience, the model points to other benefits. Experience can teach the farmer to ignore variables that truly are unimportant across many technologies, which reaps benefits the first period the farmer uses a new technology. Further, as emphasized above, experience is always helpful for a *fixed* technology: the farmer will attain a higher payoff the second period he uses a given technology as compared to the first.

2. If there are costs of attention ($e > 0$), then the farmer may not learn to optimize the technology from watching the demonstrator: For sufficiently large e , $L_1 = L_2 = 0$ only if the technology is prior congruent.

Proof. To establish the first part of the proposition, consider the case where there are no costs of attention. In this case, the farmer learns $(\sigma^*(\theta), \mathbf{a}^*(\theta))$ by watching the demonstrator. Consequently, he farms according to $(\sigma^*(\theta), \mathbf{a}^*(\theta))$ in periods 1 and 2, which implies that $L_1 = L_2 = 0$.

For the second part, assume that e is sufficiently large that the farmer will choose not to attend to the demonstrator's choice along dimension j , $(\sigma_j^*(\theta_j), a_j^*(\theta_j))$, whenever $\pi_j = \pi^L$.⁴⁰ Suppose further that the technology is prior incongruent, so there is at least one dimension, j , that the farmer does not attend to when watching the demonstrator, but is e -relevant. Because the farmer does not attend to what the demonstrator does along dimension j , he does not learn $(\sigma_j^*(\theta_j), a_j^*(\theta_j))$ by watching the demonstration; instead, because his prior is symmetric across $\tilde{\theta}_j(x'_j)$, $x'_j \in X_j$, he will remain indifferent between input levels after watching the demonstrator: There exists a $K \in \mathbb{R}$ such that $E[\tilde{\theta}_j(x'_j)|\hat{h}_1] = K$ for all $x'_j \in X_j$, where \hat{h}_1 denotes the farmer's recalled history at the start of period 1. Consequently, in the first period the farmer either will not attend to input j , in which case $L_1 > 0$, or will attend and set $\sigma_{j1}(x'_j) > 0$ for some $x'_j \neq x_j^*(\theta_j)$, in which case $L_1 > 0$ as well.⁴¹ ■

Proposition 4 indicates that demonstrations do not ensure that farmers learn to use the production technology when they face costs of paying attention and the technology is prior incongruent. In such cases, farmers will be unable to faithfully replicate practices they saw in the demonstration.⁴²

The model may also help understand why demonstration trials can have little long-run impact on technology adoption, as has been observed in some empirical contexts (Kilby 1962, Leibenstein 1966, Duflo et al. 2008b). To briefly incorporate the technology adoption decision into our model, suppose that in periods $t = 1, 2$, the farmer faces the additional decision of whether to farm at all, where his outside option is to earn \bar{v} if he chooses not to farm. Assume that farming is profitable when inputs are optimally set, $v^*(\theta) > \bar{v}$, so a farmer who knows best practices will choose to farm. Then, in the standard case where $e = 0$, farmers will necessarily choose to farm in periods 1 and 2 after observing the demonstration: In this case, the demonstration teaches the farmers that

⁴⁰Note that we can find such a range of e consistent with the farmer choosing to attend when $\pi_j = \pi^H$ since the net benefit to attending is lower when $\pi_j = \pi^L$ (as opposed to $\pi_j = \pi^H$), is strictly decreasing in e , and drops below 0 when e is sufficiently large (fixing π_j).

⁴¹To see why the farmer follows a strategy in which $\sigma_{j1}(x'_j) > 0$ for some $x'_j \neq x_j^*(\theta_j)$ if he attends, suppose the farmer instead follows a strategy in which he sets $a_{j1} = 1$ and $\sigma_{j1}(x_j^*(\theta_j)) = 1$. Given his first period beliefs, his expected payoff from following this strategy equals $E[v_{-j1} + v_{-j2} + K - e + v_{j2}|\hat{h}_1] < E[v_{-j1} + v_{-j2} + K - e + v_j^*(\tilde{\theta}_j)|\hat{h}_1]$ (since $E[v_{j2}|\hat{h}_1] < E[v_j^*(\tilde{\theta}_j)|\hat{h}_1]$ if the farmer follows such a strategy). But he could instead get the higher expected payoff by following a strategy in which $a_{j1} = 1$ and $\sigma_{j1}(x'_j) > 0$ for all $x'_j \in X_j$ since this alternative strategy guarantees that he would learn to optimize dimension j since he would then experiment with each input in X_j infinitely often.

⁴²In principle, the best-practice demonstrator can explain his strategy to the farmer either before or after the demonstration. However, the demonstrator makes many input choices and communicating these choices to the farmer is costly (Niehaus 2011). He may only want to communicate choices along dimensions he believes the farmer is predisposed not to attend to and are relevant for the task at hand. The effectiveness of communication is then increasing in the degree to which the demonstrator has knowledge of the farmer's mental model; i.e., of what he does and does not attend to.

farming is profitable relative to the outside option when inputs are optimally set, and further how to optimally set inputs. Matters are more complicated when farmers face costs of paying attention.

To illustrate, consider a simple example where there are two input dimensions and the production technology is such that it is profitable when both x_1 and x_2 are optimally set, but not when the farmer does not attend to one of the inputs: $v^*(\theta) > \bar{v}$, but $v(\sigma, \mathbf{a}; \theta) < \bar{v}$ whenever $a_j = 0$ for some $j = 1, 2$. In period 0, the best practice demonstrator will then set $x_1 = x_1^*(\theta)$ and $x_2 = x_2^*(\theta)$ on each plot of land, which will on average yield $f(\mathbf{x}^*(\theta)|\theta) = v^*(\theta) + 2 \cdot e$. If the technology is prior congruent, which in this context means $(\pi_1, \pi_2) = (\pi^H, \pi^H)$ since each input dimension is e -relevant, then the farmer will learn everything necessary to make optimal decisions in periods 1 and 2 from the demonstration (so long as parameter values are such that the farmer attends to the demonstrator's input choices): he will learn $\mathbf{x}^*(\theta)$ and $v^*(\theta)$ with certainty. Consequently, he will choose to farm in periods 1 and 2 and will optimally set inputs 1 and 2.

On the other hand, suppose the technology is prior incongruent since the farmer initially believes it is unlikely that the second dimension is relevant, $(\pi_1, \pi_2) = (\pi^H, \pi^L)$, and the farmer then only attends to the demonstrator's input choice along the first dimension. In this case, the demonstration will not teach the farmer how to optimally farm: while the farmer learns how to optimally set the first input and the expected yield given optimal input choices, $f(\mathbf{x}^*(\theta)|\theta)$, he will remain uncertain about how to set the second input since he did not attend to the demonstrator's choice along that dimension. However, while uncertain, if the farmer initially attached low probability to the second dimension being relevant, $\pi^L \approx 0$, then this uncertainty will be minimal: he is (mistakenly) very confident that it is optimal not to attend to that dimension. He consequently chooses to farm in the second period, expecting to earn a payoff much higher than his outside option with probability close to one, but is disappointed with certainty: he believes that his payoff will equal $f(\mathbf{x}^*(\theta)|\theta) - e = v^*(\theta) + e > \bar{v}$ with high probability if he sets $x_1 = x_1^*(\theta)$ and does not attend to the second input, but he in fact earns a payoff less than \bar{v} since $v(\sigma, \mathbf{a}|\theta) < \bar{v}$ whenever $a_2 = 0$. Since his expected payoff from farming is now at most \bar{v} following this disappointment, the farmer will choose to stop farming in the next, terminal, period.

This example indicates why demonstration trials can have a small impact on adoption decisions: even if people recognize from the demonstration that technologies are profitable when used optimally, they may not come away with knowledge on how to use the technology optimally if it is prior incongruent. Because of this, farmers may not consistently use the technology following the demonstration even though they "saw" how to use it optimally. In fact, farmers may initially try the technology but then give it up when they realize they have not learned to use it profitably.⁴³ Summarizing:

Proposition 5. *Consider a farmer who learns from a best-practice demonstrator in period 0 and then decides whether and how to farm in periods 1 and 2, where the technology is profitable when used optimally, but not profitable if the farmer does not attend to a dimension that is e -relevant:*

⁴³If the technology is prior incongruent and the farmer does not attend to input dimensions he thinks are unlikely to be relevant both when watching the demonstration in period 0 and when trying the technology for the first time on his own in period 1, then he will not learn to optimize the technology ($L_2 > 0$). However, by comparing his average yield with the average yield of the demonstrator, he will learn that he is not optimizing the technology: He will know he does not know, but will not know what he does not know. In a richer model with more than two periods, this can induce him to experiment, though this desire may be limited if there are many potential inputs that he could attend to. It will be further limited if the demonstration takes place on another plot and the farmer places some weight on that plot differing from his own.

$v^*(\theta) > \bar{v}$, but $v(\sigma, \mathbf{a}|\theta) < \bar{v}$ whenever $a_j = 0$ for some dimension j that is e -relevant. Suppose the farmer chooses to farm in period 1, but does not attend to input dimensions satisfying $\pi_j = \pi^L$ both when watching the demonstration and when he farms in period 1. Then, if the technology is prior incongruent:

1. The farmer will be disappointed with what he earns from farming: he expects to earn over $2 \cdot \bar{v}$ combined across periods 1 and 2, but instead earns less than $2 \cdot \bar{v}$.
2. The farmer will give up farming: the farmer chooses not to farm in period 2.

Proof. If the farmer decides to farm in period 1, this means that he expects to earn at least $2\bar{v}$ in total across periods 1 and 2, since can earn $2\bar{v}$ by not farming in either period. While such a farmer expects to earn at least $2\bar{v}$, he earns less than \bar{v} in the first period whenever the technology is prior incongruent and he does not attend to dimensions satisfying $\pi_j = \pi^L$ in periods 0 and 1, given the assumption that $v(\sigma, \mathbf{a}; \theta) < \bar{v}$ whenever $a_j = 0$ for some e -relevant dimension j . This means that this farmer in fact earns less than $2\bar{v}$ across periods 1 and 2 so long as he gives up farming in the second period.

To complete the proof we must then verify part (2) of the proposition; that the farmer indeed finds it optimal not to farm in period 2. Since, by assumption, the farmer does not attend to input dimensions satisfying $\pi_j = \pi^L$ both when watching the demonstration and when farming himself in period 1, he does not learn to optimize those dimensions and remains indifferent between different input choices along those dimensions. As a result, he cannot expect to earn more than $v_1 = v(\sigma_1, \mathbf{a}_1; \theta) < \bar{v}$ by farming in the second (terminal) period, and stops. ■

An implication of the analysis is that finding that a demonstration has little impact on long-run beliefs or behavior does *not* necessarily imply that farmers have little left to learn or even that the demonstration provided them with little informative data. Rather, when the technology is prior incongruent, it could reflect that the demonstration did not sufficiently alter farmers' initially mistaken beliefs about which input dimensions matter.

Appendix Table 1: Randomization Check

	Mean, by Treatment Group			Differences		
	Control (1)	Sort (2)	Weight (3)	Col 1 - Col 2 (4)	Col 1 - Col 3 (5)	Col 2 - Col 3 (6)
ln(Monthly Per Capita Income)	12.43 (1.37)	12.42 (0.72)	12.67 (0.76)	0.233 (0.164)	-0.016 (0.156)	-0.249 (0.138)*
HH Head is Literate	0.79 (0.41)	0.84 (0.37)	0.88 (0.32)	0.093 (0.059)	0.050 (0.060)	-0.043 (0.064)
Number of Assets	8.36 (3.15)	7.95 (3.19)	8.28 (3.06)	-0.079 (0.511)	-0.409 (0.494)	-0.330 (0.580)
Age of HH Head	43.23 (12.22)	43.95 (12.14)	43.54 (11.33)	0.304 (1.937)	0.718 (1.905)	0.414 (2.192)
Years Farming	18.00 (6.96)	18.94 (7.06)	18.27 (7.06)	0.275 (1.183)	0.937 (1.103)	0.662 (1.330)
Parents Farmed Seaweed	0.47 (0.50)	0.55 (0.50)	0.53 (0.50)	0.063 (0.083)	0.081 (0.078)	0.019 (0.093)
Loans from Someone Sells to	0.31 (0.46)	0.33 (0.48)	0.22 (0.42)	-0.085 (0.088)	0.026 (0.085)	0.111 (0.097)
Farms Cottoni	0.34 (0.47)	0.33 (0.47)	0.47 (0.50)	0.135 (0.082)*	-0.008 (0.074)	-0.144 (0.091)
Pod Size at Baseline	109.01 (28.79)	102.83 (22.57)	112.78 (38.00)	3.777 (5.850)	-6.174 (3.884)	-9.951 (5.930)*
Number of Days of Previous Cycle	36.61 (8.24)	36.00 (7.31)	37.08 (6.54)	0.463 (1.178)	-0.612 (1.190)	-1.075 (1.281)

Notes: This table provides a check on the randomization. Columns 1 - 3 provide the mean and standard deviation of each baseline characteristic for the control group, sort group, and weight group, respectively. Columns 4 - 6 give the difference in means (and standard errors) between the noted experimental groups. Statistical significance is denoted by: *** p<0.01, ** p<0.05, * p<0.10.

Appendix Table 2: Reasons Why Farmer May Want to Try a New Method

	Mean	Standard Deviation	N
	(1)	(2)	(3)
Would Not Want to Make any Changes	0.04	0.18	482
Own Initiative	0.02	0.13	482
Pest or Price	0.02	0.13	482
Advice from Friend	0.10	0.30	482
NGO or Government Recommendation	0.11	0.31	482
Seeing Results on the Plots of Other Farmers	0.39	0.49	482

Notes: This table provides the baseline survey responses on the reasons why a farmer may try a new method.

Appendix Table 3: Effect of Treatment, by Recommendation

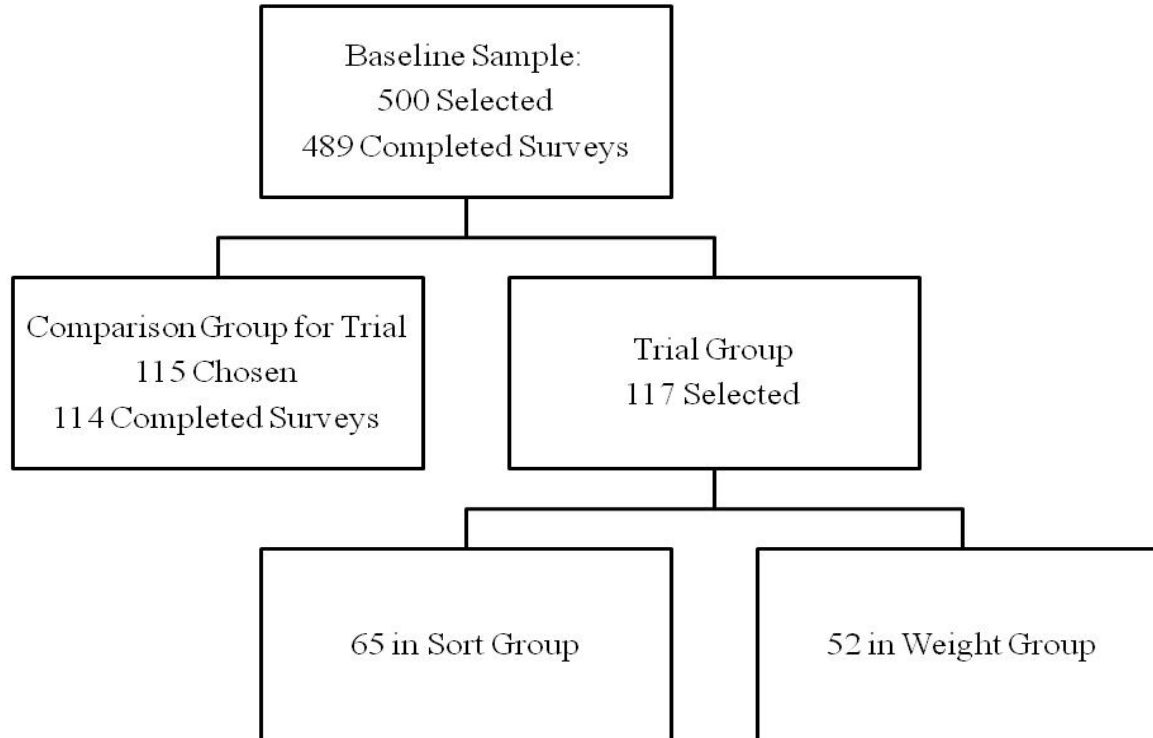
	Pod Size (Grams)	
	(1)	(2)
Increase	-21.091 (4.807)***	-50.976 (1.851)***
Increase * Trial Participation * After Trial	27.264 (6.223)***	27.060 (7.627)***
Increase * Trial Participation * After Summary Data	19.502 (5.808)***	19.242 (7.149)***
Hamlet Fixed Effects	X	
Farmer Fixed Effects		X
Observations	675	675

Notes: This table provides the coefficient estimates of the effect of treatment on farming methods after the trial (follow-up 1) and after observing the summary data (follow-up 2), conditional on baseline farming methods and disaggregated by whether the farmers were recommended to increase or decrease their pod size when observing the summary data. The trial participation dummy indicates that the farmer belongs in either the sort or weight treatment group. Increase is an indicator variable for being told to increase size. All regressions are estimated using OLS and include the main effects as well as the double interactions, and standard errors are clustered at the farmer level. Statistical significance is denoted by: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

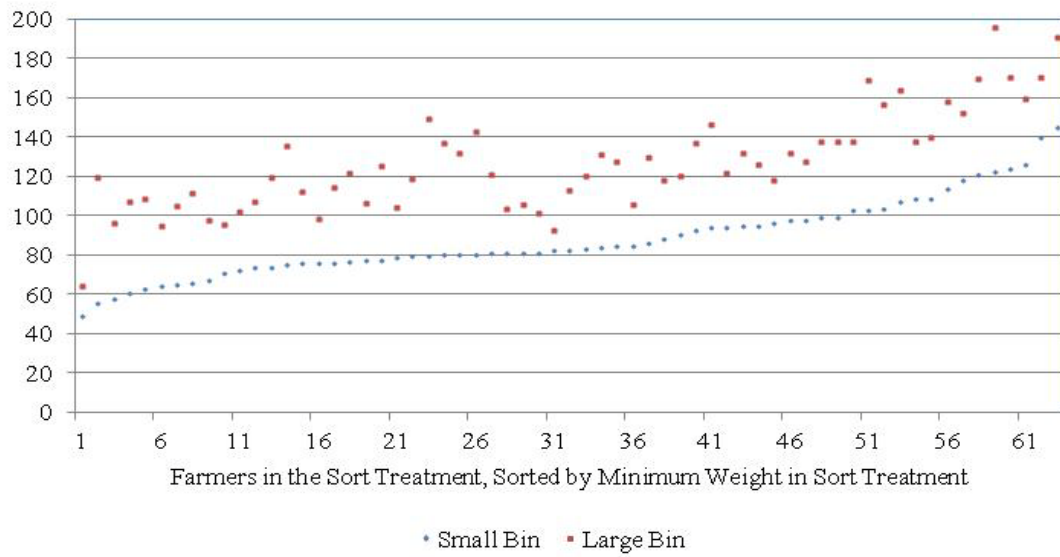
Appendix Figure 1: Experimental Design

	Jun 2007	Jul 2007	Aug 2007	Sept 2007	Oct 2007	Nov 2007	Dec 2007	Jan 2008	Feb 2008	Mar 2008	Apr 2008	May 2008	Jun 2008	Jul 2008	Aug 2008
Baseline	■	■	■	■	■	■	■								
Trials		■	■	■	■	■	■	■	■	■					
Follow-Up 1											■	■			
Follow-Up 2														■	■

Appendix Figure 2: Sample Design



Appendix Figure 3



Notes: This figure documents the within farmer variation in pod sizes at baseline. For each farmer, it shows the average size within the small and large bin in the sort treatment, where farmers are sorted according to the average size in their small bin.

Appendix Figure 4: Examples of Summarized Trial Results

A. Weight Example

Pod Size	Distance	#Pods per line	Initial investment	Return per line
40	15	33	1650	4510
40	20	26	1300	3553
60	15	33	2310	1517
60	20	26	1820	1195
80	15	33	2970	1871
80	20	26	2340	1474
100	15	33	3630	1904
100	20	26	2860	1500
120	15	33	4290	597
120	20	26	3380	470
140	15	33	4950	1574
140	20	26	3900	1240

Currently

Pod Weight: 152.5

Distance: 15

Recommendation:

Pod Weight: 40

Distance: 15

B. Sort Example

Line Type	Distance	Average Pod Weight (g)	Return per pod (g)	Average pods per line	Return per line (g)
Large		129.92	167.65		5716.76
Medium	14	98.34	155.51	34.1	5302.92
Small		86.18	158.82		5415.88

Recommendation Switch to **large** pod size, with average weight **129.92g**.

Appendix Figure 5: Distribution of Recommended Pod Sizes in the Weight Treatment

