Learning by (virtually) doing: Experimentation and belief updating in smallholder agriculture*

Emilia Tjernström † Travis J. Lybbert‡ Rachel Frattarola Hernández§
Juan Sebastian Correa‡

February 20, 2020

Abstract

In much of sub-Saharan Africa, soil quality heterogeneity hampers farmer learning about the returns to different inputs. This can partly explain why we observe limited adoption of improved inputs in the region. We study how Kenyan farmers respond to an interactive app that enables them to discover agricultural input returns on a virtual plot that is calibrated to resemble their own. Farmers update both their beliefs and behaviors after engaging with the virtual learning app. Additionally, farmers revise their beliefs upwards after using the app. In an incentive-compatible experiment, farmers receive an input budget from the research team, which they can allocate across farm inputs. After they play several virtual seasons on the app, they have the opportunity to update these allocations. Farmers revise their input allocations along several dimensions after the virtual learning experience. As evidence that these adjustments emerge from real learning, we show that farmers with the highest predicted returns to lime—an unfamiliar input in this region—increase their lime orders more than others. Our results suggest that engagement with a personalized virtual platform can induce real learning and enhance farmers' beliefs and technology choices.
JEL codes: D84, O12, O13

*We are grateful to Matt Kimball and Tyler Lybbert for their help animating and programming *Mahindi-Master*. We also wish to thank Laura Schechter and Jen Alix-Garcia for helpful comments, David Cammarano and Christopher Kucharik for their assistance with crop modeling choices, as well as seminar audiences at IFPRI, the University of California - Davis, the University of Wisconsin - Madison, the 2018 AAEA Annual Meeting, and the USAID Global Development Lab for helpful comments. We also wish to thank our respondents for their enthusiasm, generosity and patience during the course of this research. This work was supported by the Michigan State University Global Center for Food Systems Innovations (GCFSI) and the Daniel Louis and Genevieve Rustvold Goldy Fellowship. The pre-analysis plan for this study can be found at the Registry for International Development Impact Evaluations (RIDIE).

†University of Sydney, Australia; emilia.tjernstrom@sydney.edu.au

‡University of California, Davis; tlybbert@ucdavis.edu

§Office of Management and Budget; rachelfrattarola@gmail.com; The views expressed in this paper and any opinions and conclusions expressed herein are those of the author(s) and do not necessarily represent the views of the United States Government, the Administration, or the Office of Management and Budget. This paper was not written as part of Rachel Frattarola Hernández's official position at the Office of Management and Budget.

1 Introduction

Three-quarters of poor households across the developing world live in rural areas and rely on agriculture as a source of income and food. Many of these subsistence farm households—especially in sub-Saharan Africa—cultivate their fields without modern inputs, achieving lower yields than they would under optimal management. While the literature has explored various explanations for these meager adoption rates (see Jack 2011 and Magruder 2018 for thorough reviews), incomplete information often surfaces as an important constraint to adoption (Magruder, 2018).

Many pre-requisites precede a household's decision to adopt a new technology, but perhaps the most basic among them is knowledge about (i) the product's existence and (ii) the product's use and profitability. The literature examines a vast array of constraints that prevent households from adopting even when these pre-conditions are fulfilled, but our study design allows us to ignore many of them.[1] Researchers have examined many types of information interventions, but we still know relatively little about how farmers form and update their subjective beliefs about the benefits of new products. This paper aims to fill part of that knowledge gap by studying how farmers respond to an app that lets them experiment with agricultural inputs on virtual plots calibrated to match their own.

We study the effects of customized information embedded in a virtual farming app on farmers' beliefs and behaviors. On average, farmers revise their beliefs about input returns upwards after interacting with the app. Since the sample farmers' beliefs about yields are lower than historical averages, this suggests that the learning is productive. Furthermore, using an incentive-compatible elicitation method, we find that farmers who were *ex ante* expected to benefit from a new, unfamiliar input respond to information about it by allocating more money towards this new input.[2]

[1]Common constraints in the literature include credit constraints, missing insurance markets, and farmer risk aversion. Our study design allows us to largely ignore the impacts of credit and insurance, and we control for risk aversion in our results.

[2]Enumerators were blind to participants' predicted returns to the new technology.

Our study setting is a particularly challenging context for learning. We study fertilizer adoption by maize farmers in Western Kenya, a region that is characterized by substantial heterogeneity in soil quality (Tittonell et al., 2008). This heterogeneity hampers farmer learning in multiple ways: first, information diffusion by central agencies is difficult since regional fertilizer recommendations will be inaccurate for most farmers. Second, while learning-by-doing and learning from others play important roles in the adoption of new technologies (see for example Foster and Rosenzweig 1995 and Conley and Udry 2010), own-experimentation is both risky and costly. Further, each agricultural season typically yields only one observation. In some contexts, farmers can compensate by learning from their neighbors, but heterogeneous environments—such as our study setting—have been shown to limit farmers' ability to learn from others (Munshi, 2004; Tjernström, 2018).

This project tests an interactive "gamified" information intervention designed specifically to overcome these learning limitations. Drawing on design insights from the gamification literature, we designed an accessible app that animates output from a crop model to give farmers proxy-observations for various input combinations. The app, which we call *Mahindi-Master* (*Mahindi* means maize in *kiSwahili*), is played on a researcher-provided tablet. The game simulates yields for a menu of input options based on plot-level soil samples, historical weather data, and crop model outputs. The app allows farmers to experiment with three different fertilizers. The menu of options included one input that most farmers would be familiar with, a second input that only some farmers had experience with, and a third that would be unknown to most participants.

The most common input in the area is diammonium phosphate (DAP), followed by calcium ammonium nitrate (CAN). The third input, lime, is used to reduce soil acidity—a common problem in the area. This input was both unavailable and unfamiliar in the area at the time of our data collection.[3] Because acidic soils prevent maize from absorbing soil

[3]Recently, a couple of NGOs have begun experimenting with lime provision in our study area. At the time of our study, the survey team had to travel to a limestone quarry where a company processes lime. At that time, the company had no plans to sell agricultural lime via local agrodealers.

nutrients (including applied fertilizer), the returns to lime are highest for farmers with acidic soils. At baseline, our sample farmers know little about their soil acidity or how lime affects acidity and enhances the production response to fertilizer. We are therefore particularly interested in examining how the information provided via the app affects beliefs about and purchases of lime.

Under uncertainty, we expect individuals to form beliefs and assign probabilities to potential outcomes (Delavande, 2014). Given the uncertainty surrounding fertilizer returns, farmers should have subjective expectations over the returns to different inputs. We elicit these beliefs before and after farmers' interactions with the app. We then study whether this new information about fertilizer returns leads farmers to revise their expectations. We also analyze whether the *MahindiMaster* interaction induced behavior change by allowing farmers to allocate a researcher-provided budget across the three inputs (DAP, CAN, and lime) during a pre-intervention survey, and then allowing them to re-allocate their budget after interacting with the app.

Our intervention builds on recent work on gamification in educational settings and using virtual reality in experimental settings. Gamification is often defined as the use of game design elements in non-game contexts, and centers around changing the information flow that users receive (Walz and Deterding, 2015). We use gamification primarily to provide new information to farmers and to display the new information in an interface that eases comprehension. Similarly, virtual reality enables researchers to represent tasks in less abstract and more naturalistic manner that better reflects individuals' decision-making reality . Fiore et al. (2009) find that when using a virtual experiment with virtual reality technology, participants' subjective beliefs more closely align with actual risks. We adhere to several gamification principles identified in the literature as best-practice: feedback should be quick, the goals clear, each player's experience should be customized, and users should have the freedom to fail (see for example Simões et al. 2013; Lee and Hammer 2011; Gordon et al.

2013 for discussions of educational gamification design principles).[4]

Compared to a regular farming season, which lasts for months, *MahindiMaster* users receive feedback in less than a minute. The goals are clearly outlined, and each user's simulations are based on their own field and historical weather data from their location. Being "free to fail" is key to our gamification approach. We explicitly aimed to eliminate the cost and risk of on-farm experimentation, and thereby encourage greater exploration and discovery.[5]

This study also relates to a broader literature examining information interventions. This literature spans many different domains and has produced mixed results. A variety of information-only interventions have yielded null effects: Bryan et al. (2014) study households' migration decisions in Bangladesh and find that information about migration opportunities does not affect household behavior. Bettinger et al. (2012) evaluate the effect of providing aid eligibility information to low-income families and find no significant effect. Ashraf et al. (2013) find that information about water purification products does not significantly increase product demand. Marreiros et al. (2017) experimentally study the effect of providing information about online privacy practices on the privacy decisions of participants. While participants in this study were less likely to share personal information after receiving privacy information, their attitudes towards privacy did not change.

In contrast, other research finds that information interventions can be effective but the type of information, targeting, and characteristics of those who receive it matters. Dupas (2011) finds that a national HIV/AIDS curriculum in Kenya had no effect on pregnancy or sexual partners' age, but giving girls information about the relative risk of contracting HIV lowered pregnancy rates and the age gap between partners. Ajayi et al. (2017) study school choice in Ghana, finding that an information intervention targeting both guardians

[4] While there exists little rigorous empirical work on the effects of gamification on learning, the educational innovation literature is beginning to coalesce around some guiding principles.

[5] Since we do not compare our app against a non-gamified alternative, we cannot rigorously test the impact of gamification. Instead, we borrow gamification principles, including rapid feedback, salience and freedom to fail, in the design of the app.

and students, instead of only students, increases guardians' participation in the process of school selection. This participation may increase informed decision-making, thereby decreasing inefficiencies in the school choice process. Wiswall and Zafar (2015) measure students' subjective expectations before and after providing them with an information treatment related to employment and earnings. The authors find that students exposed to the treatment revise their beliefs and that individuals' characteristics affect the degree to which they update. The notion that novel information is more likely to affect beliefs and behavior implies in our setting that information about the unknown agricultural input (lime, and to some extent CAN) may be particularly effective.

Given the limited success of many information-only interventions, work in this literature often explores response constraints, including potential disconnects between the information provided and the target behavior change. In some cases, this may be because the information provided was not new, and therefore failed to induce belief updating. Huffman et al. (2007) conduct a willingness to pay (WTP) experiment related to genetically modified foods. They find that more informed participants had a lower WTP than did uninformed participants. Conversely, uninformed participants were the most affected by new information. Attanasio and Kaufmann (2017) analyze survey data on young Mexican beneficiaries of *Jóvenes con Oportunidades*. The survey includes respondents' subjective expectations regarding the returns to education. Respondents believe that education will improve their labor market outcomes, and higher expectations about the returns to college are associated with higher probabilities of college enrollment.

Another potential explanation for the lack of response to information-only interventions is that individuals do not deem the information useful, even if it is new to them. We expect that farmer valuation of what they learn from the app to increase with the relevance to their context. In our case, we expect information about lime to be differentially effective depending on baseline soil acidity. Households with acidic soils (whose soils are unproductive without added lime) would be expected to react more strongly to the intervention. However, the empirical literature thus far suggests that this expectation is not always borne out. For example, Hoffman (2016) examines the acquisition of costly information by industry experts. The study finds that individuals' WTP is actually larger for *less* valuable signals (Hoffman, 2016; Ambuehl and Li, 2018). Ambuehl and Li (2018) also find that individuals undervalue more informative information and that the degree of belief updating differs across individuals.

The heterogeneous responses found in the literature may also help explain the attenuated impact of information interventions. Some studies focus on (over)confidence as an explanation. Dessí and Zhao (2018) hypothesize that overconfidence is endogenous, and that differences in the stability of an environment may affect the degree of overconfidence since overconfidence is not as beneficial in a stable environment as in a dynamic environment. The authors find that more stable countries have lower self-confidence measures, and argue that overconfidence should lead people to invest more in projects in dynamic environments but invest less in stable environments. Barham et al. (2018) examine the influence of two other individual characteristics on US farmers' technology adoption behaviors. They find that receptiveness to advice can either slow or speed up adoption, depending on agents' innate cognitive ability. We investigate how a few dimensions of individual characteristics (confidence and overconfidence) affects farmers' responsiveness to the *MahindiMaster* advice. In line with past findings, we find that farmers whom we classify as overconfident update their beliefs less than do appropriately-confident farmers.

While we do not have measures of cognitive ability, we also explore heterogeneity along the dimension of farming knowledge. Farmers who have more correct questions on a farming knowledge quiz tend to update their beliefs more than less knowledgeable farmers. Further, farmers who choose to answer "I don't know" to more quiz questions tend to update their beliefs more, suggesting that it does matter whether an individual knows what they do not know.

2 Data

2.1 Sample

Our sample consists of 200 farmers from 19 villages in Western Kenya. These farmers were drawn from the rosters of a three-wave panel survey of 1,800 farm households. The households had previously taken part in a randomized controlled trial (RCT) of a newly-introduced hybrid seed variety. The RCT was designed to measure how access to a new hybrid variety affected yields and household incomes. This earlier study employed a research design that provided farmers in treatment villages with information about the seeds, as well as sample seed packs for experimentation.[6] A second intervention, randomized at the household level provided farmers with high-quality fertilizer. Based on the findings of that earlier study, we expect that our sample farmers may have above-average baseline fertilizer adoption rates.

To put the experimental design and data components of this study in context, the top panel of Figure 1 shows the main agricultural seasons in Kenya. The rainfall distribution in most of Kenya is bimodal, with a main season beginning in March (with the harvest occurring in August or September) and a short season between October and December-January. For both the RCT and for the *MahindiMaster* pilot, data collection and intervention timing centered around these agricultural seasons. In particular, the current study took place prior to the planting of the main season in 2017, so that participating farmers would be able to apply their chosen inputs in the upcoming main season.

The bottom panel of Figure 1 presents the timing of the app-based pilot intervention as well as the timing of the earlier RCT and the original household panel data collection. Throughout the paper, we refer to data coming from the RCT panel baseline survey as "baseline data." In contrast, we refer to all survey questions asked before farmers interacted

[6] Carter et al. (2019) report the results of this study: receiving samples of a regionally appropriate hybrid maize seed variety increased hybrid adoption and maize yields, but with important heterogeneity by geographical area.

with the app as pre-game data, and those elicited after the interaction as post-game data.

For budgetary reasons, we chose a convenience sample of villages in Western Kenya, south of Lake Victoria. The participants were drawn from the same rosters as the panel survey. The original sample was determined in 2013, by randomly choosing households with these villages from a complete listing of households, proportional to the size of the village within a circle drawn around a seed company demonstration plot.[7] Some of the *MahindiMaster* villages were in the RCT treatment villages, i.e. received sample packs of a maize hybrid; others were originally in control. Similarly, some households were randomly selected for the original fertilizer treatment arm, while others did not receive fertilizer in the original RCT. Given that our sample is a subset of a larger RCT, we can use information on previous fertilizer and hybrid seed use as well as past yields in our analysis.

2.2 Experimental design

In February and March 2017, enumerators visited the 19 villages in the sample and invited the households to a central location in the village. Farmers completed the pre-game questionnaire, which included a module on confidence and risk preferences. Farmers also allocated their input budget across DAP, CAN, and lime (henceforth called the pre-game allocation) and subsequently had the opportunity to play *MahindiMaster*. Enumerators helped facilitate an initial one-on-one session, in which they walked farmers through some pre-defined screens in the app that illustrated how to navigate the game. Once farmers felt comfortable using the tablets and the app, they continued playing in semi-private.

Post-game, participants answered a number of questions measuring respondents' perceptions of the game and elicited farmers' post-game budget allocation, as well as post-game yield beliefs. Farmers then received their chosen post-game input amounts. Participants did not discuss the game with other farmers until they had made their final selection. Given the short amount of time that lapsed between participants' pre-game input selections and the

[7] For more details on the sampling and the RCT, please see Carter et al. (2019).

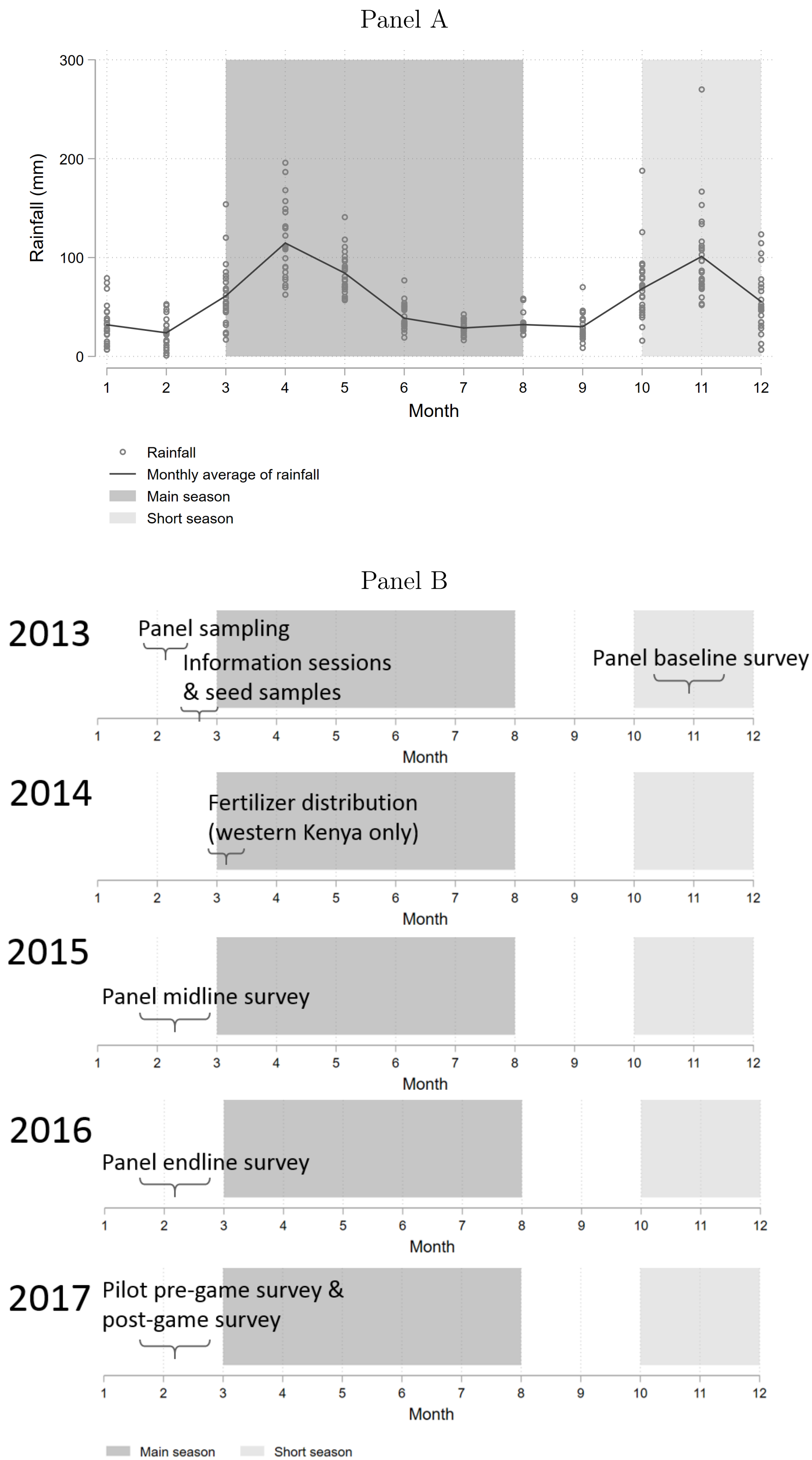

Figure 1: Rainfall distribution in sample area and data collection timeline

post-game, we believe that the likelihood is minimal that any factors other than the game would influence the post-game selection, allowing us to attribute order updates to farmers' interactions with the game.

MahindiMaster is based on crop model simulations using the software DSSAT. Briefly, DSSAT uses soil characteristics, rainfall, temperature, and solar radiation as inputs to simulate crop yields (in our case maize yields) under different fertilizer types and application rates. We calibrated yields under a large combination of DAP, CAN, and lime rates, as well as under three different weather scenarios (low, medium, and good). In total, each household has a potential 972 simulations that they can observe within the game. Appendix A provides more details on our use of the DSSAT model and related assumptions.

The simulation results then get translated into *MahindiMaster*, via a Unity-based Android application. The app first selects a specific farmer and then animates the simulation results. This enables farmers to receive tailored yield information in an accessible way. Farmers make choices within the game about the type of fertilizer, the amount of fertilizer, and rainfall scenario. The application animates key steps in the maize-growing process: planting, two separate fertilizer applications (DAP and CAN should be applied at different stages of the growing cycle), rainfall, and crop growth for each fictional season and then displays the expected yields. Together with the yields, the app also displays the cost per unit of harvest to enable profitability calculations. At the end of each fictional season, farmers can make a new fertilizer choice and simulate a new season.

The game play had a specific structure, in which farmers could choose levels of DAP (the most familiar fertilizer) in all rounds of the game. After three rounds, CAN became available, and lime was introduced after five rounds. The weather realizations were random for the first seven rounds, but after the seventh round, farmers also had a choice of weather scenario. Farmers were able to choose DAP and CAN in 25kg-intervals ranging from 0 to 125kg. Recommended lime application rates are much larger, so farmers selected lime in 250kg intervals ranging from 0 to 2000kg. Farmers had to play a minimum of nine fictional

(a) Input selection

(b) Fertilizer application

(c) Seed planting

(d) Harvest summary screen

Figure 2: Screenshots from *MahindiMaster* gameplay

seasons. After the required rounds, they could continue playing for as long as they wanted.

Once farmers decided that they wanted to stop playing, they would play the final round. The final round was supposed to reflect farmers' updated fertilizer orders (these orders were also recorded separately by the enumerators when they handed out the chosen inputs). As farmers interacted with *MahindiMaster*, the app recorded the number of rounds played, the random weather scenarios seen by farmers, their in-game input choices (type and amount), and their weather choices.

Figure 2 shows screenshots from *MahindiMaster*. Figure 2a shows the selection screen when only DAP is available, 2b shows the animated hand applying fertilizer (which gets deposited into a hole, made with a stick on the previous screen), 2c shows the same hand depositing the seeds into the hole, and 2d shows the summary screen at the end of the season. The summary screen displays the yield that the selection resulted in, the rainfall scenario, and the realized cost per 90kg bag of maize (a common unit for measuring harvests).

2.3 Data sources

2.3.1 Panel baseline data

We use some information from the original panel baseline survey as control variables in our analysis. During the panel baseline in 2013, enumerators asked farmers about their hybrid seed and fertilizer use over the five years preceding the survey (for both the main and the short season). We use this information to create a measure of farmers' past experience with fertilizer. Specifically, we create two variables that measure the number of main and short seasons that farmers used fertilizer during the five years preceding the panel baseline survey.

2.3.2 Soil information

We collected soil samples from each household in October 2016. Our enumerators received training on soil sample collection from an ISO-certified soil testing lab in Nairobi, Kenya, who also carried out the sample analysis. We analyzed both macro and micro-nutrients in our samples, providing data on the soil's pH, cation exchange capacity (CEC), electric conductivity, organic matter, element levels (e.g. nitrogen, phosphorus), as well as micronutrients such as boron.

We additionally have soil sample data from 2014 on the full panel sample, and during the earlier RCT all households received a printout of their soil's measurements. The printout further contained the soil testing lab's fertilizer recommendations for a set target yield for each field. While the research team simplified that information as much as possible, farmers do not appear to have understood most of the information. On the one hand, the anecdotes about the information being largely unused inspired this project. On the other, since all our sample farmers had already received their soil information, it helps boost our interpretation of the *MahindiMaster* intervention as a gamification and learning intervention, rather than a pure information intervention.[8]

[8] The fact that the research team provided the input package was also likely important in observing behavior change. Harou et al. (2018) report results from an RCT of soil information provision and find that information alone does not alter farmer investments in fertilizer. Only the treatment group that received

2.3.3 Pre- and post-game data

Subjective expectations

Before and after farmers interacted with the app, enumerators elicited farmers' subjective expectations about yields under different input combinations for the maize field from which we collected soil samples. We followed standard methods for eliciting subjective expectations, which have by now been used in a variety of developing-country settings.[9] The first distribution was based on farmers' expected maize harvest if they applied no fertilizer at all. To obtain the lower bound of the support, we asked what yield they would expect to get in the worst year that they could imagine; the upper bound was similarly based on the best year that they could imagine. The data collection took place on tablets, which produced five bins based on these reported maximum and minimum harvests.

Enumerators then gave farmers 20 beans (or maize kernels) to allocate across the bins. The enumerators explained to participants that the number of beans allocated to each bin represented the number of years out of the next twenty years that they thought their maize harvest would fall within that interval.

We repeated this elicitation for farmers' "normal" fertilizer application, i.e., what they apply in a representative year. We finally repeated the procedure for a yield distribution that combined "normal" fertilizer application rates with the addition of lime (henceforth the "fertilizer-plus-lime" belief distribution). During a pre-pilot, we found that farmers found it difficult to provide their beliefs about lime alone, as most of them were unfamiliar with the input. This joint belief elicitation allowed participants to express a flat or diffuse prior by saying that their beliefs did not differ from the "normal" fertilizer application rates.

To compute the mean and variance of farmers' subjective yield distributions, we fit a distribution using 5 points from the elicitation procedure.[10] To fit the distribution, we

a combination of liquidity and soil information update their behavior in a manner consistent with existing plot-specific soil nutrient limitations.

[9]See Delavande et al. (2011a), Delavande et al. (2011b), and Delavande (2014) for an overview of eliciting subjective expectations in developing countries.

[10]For respondents who were unable to state their subjective expectations about yields under fertilizer-

input the right endpoint for each bin and the probability assigned to each bin by the farmer (the number of beans divided by 20) and input this information into a Matlab function. The function creates a CDF based on the probabilities. The endpoints for each bin become points on the $x-$ and $y-$axis in the cumulative probability.

Subjective probabilities imply bounds on the subjective yield distribution but do not identify the actual distribution without additional assumptions.[11] Our goal is to fit a respondent-specific parametric distribution to estimate the first two moments of the distribution. We choose the log-normal family of distributions since log-normal tends to fit the underlying variable (yields) reasonably well and is a common modeling assumption in the literature. Note that our results are not sensitive to this assumption. We therefore fit a respondent-specific lognormal CDF to five data points (the right endpoints for the bins), using nonlinear least squares. We use the parameters from the fitted distribution to calculate the mean and variance of the subjective yields distribution for each farmer.

Farmer knowledge, confidence, and overconfidence

We elicit confidence and farming knowledge as part of the pre-game survey using a 10-question quiz related to general maize farming knowledge.[12] After answering the questions, farmers also had to guess how many questions they believed that they answered correctly. We compare the number of questions that farmers answered correctly to the number of questions farmers believed that they answered correctly. If farmers got the same number of quiz questions correct as they reported believing they got correct (plus or minus two questions), we consider the farmer "appropriately confident." If the farmer believed that she answered more questions correctly than she did and the difference is greater than two, we consider the farmer "overconfident." If the farmer believes that she answered fewer questions

plus-lime, we fill in the missing values with the fertilizer-only distribution.

[11] As Dominitz and Manski (1997) point out, given the small number of "bins" typically used to elicit subjective probabilities, we cannot confidently determine whether this distribution fits the shape of the elicited probabilities the best. More bins would be too cognitively taxing, so this is an inherent trade-off in the method.

[12] Our measure of confidence most closely relates to Moore and Healy (2008)'s "overestimation" definition of confidence.

correctly than she did and the difference is greater than one, we would consider the farmer "underconfident." Only two farmers are classified as underconfident in the sample, so our analysis uses a dummy variable that equals one if the farmer is overconfident and zero otherwise.

We elicit risk through a series of non-incentivized gambles and survey questions measuring subjective risk attitudes after farmers have interacted with the app. We choose these simpler methods of risk elicitation as recent evidence from Senegal suggests that rural populations often do not understand sophisticated risk elicitation methods very well (Charness and Viceisza, 2016). In the non-incentivized gambles, farmers chose between two bags that contain balls of different values. The first bag contains one ball worth 3000 Kenyan Shillings (KES). The second bag contains two balls: one ball worth 5000KES, and a second ball that varies in value across rounds. In the first round, the "random" ball is worth 2500KES. Its value decreases by 500KES until the farmer chooses the first bag, or until the "random" ball is worth 500KES. Based on this elicitation procedure, we are able to rank farmers on risk preferences. We also ask farmers a series of questions about their willingness to take risk in general and on their farm. These "stated" risk questions are based on those used in the German Socio-Economic Panel but simplified for the developing country context.[13]

2.3.4 Incentive-compatible input orders

As part of the experimental session, we gave farmers an experimental budget of 5000 KES to allocate across three different fertilizers: DAP, CAN, and lime. We refer to this allocation as farmers' "fertilizer order" since the research team gave participants the exact amounts that they had ordered upon finishing the post-game questionnaire. As is common in experiments, the actual value of this budget was scaled down by a fixed factor that was known to farmers. After being scaled to real Kenyan Shillings (KES), the inputs provided were roughly sufficient

[13] Dohmen et al. (2011) find that the general risk survey questions are strongly correlated with responses from incentivized lotteries.

to plant an experimental plot of 10x10 meters.[14] Farmers placed their pre-game orders before interacting with the app, and then had the option to update their orders after they interacted with the game (post-game order). Since the research team provided farmers with fertilizer according to the scaled order, farmers had an incentive to truthfully report their preferences. We are especially interested in whether farmers update their orders after playing the game, which would suggest an updating of beliefs about the returns to different fertilizers.

3 Descriptive statistics

3.1 Farmer characteristics

Panel A in Table 1 presents descriptive statistics from our sample. Some measures are from the panel baseline survey (fertilizer and seed experience), others are from the pre-game questionnaire (sampled plot size, quiz performance, and confidence). The soil sample data describes the 2016 soil sample results. On average, before any interactions with the survey team, farmers used fertilizer and hybrid seeds during two of the five long rain seasons. This suggests that some farmers are familiar with fertilizer use, but also that many do not use the input on a regular basis. Roughly half of farmers report using DAP in a modal year, and 40% report using CAN. None of the sample farmers report having used lime, confirming our understanding that lime is relatively unknown and/or inaccessible in this region.

The average pH on farmers' fields is 6.45, which is within the recommended range for maize farming (5.8 to 7). The lowest-pH plots, however, are substantially below the optimal range and well into the range where yields are likely starkly reduced from the acidity. Soil Cation Exchange Capacity (CEC) ranges from a low of around 6 to a high of almost 70, indicating that we have a wide range of soil types in the sample. For example, sandy soils tend to have low CEC values, and the measure increases roughly with the amount of clay,

[14]Farmers could also choose cash instead of fertilizer, but no one chose to receive cash instead, which may reflect the fact that many farmers face relatively high transaction costs when purchasing inputs for use on-farm.

silt and organic matter present in the soil. Low-CEC soils hold on to less nutrients and water and fertilizer risks leaching out rapidly.

Most farmers performed poorly on the farming quiz. On average, participants answered less than three out of ten questions correctly and no respondents answered every quiz question correctly. Furthermore, farmers stated that they did not know the answer to roughly half the questions. Based on our metric of overconfidence, almost 60 percent of the sample is classified as overconfident.

Figure 3 shows that many participants played more rounds than was required. Further, Panel B summarizes farmers' behavior during their interactions with the app. Across the rounds in the game, farmers experimented with an average of 31.8 kg/acre of DAP, a common fertilizer in the area. Farmers chose to apply some amount of DAP in almost 80% of seasons played. Farmers are on average less familiar with CAN, and they applied an average of 23.8 kg/acre across game rounds. Note that since farmers were only able to begin experimenting with CAN in round 3 of the game, this is an underestimate of how much they applied conditional on CAN being available as those rounds get recorded as zero here. Once CAN became available, farmers applied non-zero amounts of CAN in 81% of the rounds.

Further, we can see that farmers generally experimented a fair bit with lime, a new and unfamiliar input to most farmers. On average, they applied almost 120 kg of lime across all rounds, and experimented with some amount of lime in 60% of the rounds in which lime was available. Farmers played an average of 10.6 rounds, of which 9 were mandatory. Some farmers played the game without fertilizer—perhaps as a form of ground-truthing exercise. We also observe that farmers, when allowed to choose the type of weather to simulate, choose the good rainfall scenario a majority of the time. The rainfall scenario variable takes on a value of 3 for the good weather scenario, 2 for median, and 1 for the poor weather scenario. All in all, these descriptive game-play results indicate that farmers were willing to experiment with unfamiliar fertilizer inputs, and that many farmers played beyond the required rounds, suggesting interest in the game and in the information presented.

Table 1: Baseline characteristics and game interactions

Panel A: Farmer and plot characteristics	Mean	Std Deviation	Min	Max
No. of seasons used fertilizer (long rains)	2.06	2.29	0	5
No. of seasons used hybrids (long rains)	2.10	2.15	0	5
Uses any DAP in a 'normal' year	0.49	0.50	0	1
Uses any CAN in a 'normal' year	0.38	0.49	0	1
Uses any lime in a 'normal' year	0	0	0	0
pH of sampled plot	6.45	0.70	4.93	8.65
CEC of sampled plot	25.3	16.3	5.91	68.6
Sampled plot size (acres)	1.15	0.92	0.13	5
No. farming quiz questions correct	2.93	1.12	0	6
No. farming quiz questions=don't know	4.87	2.04	0	9
Overconfident (0/1)	0.59	0.49	0	1

Panel B: Game play	Mean	Std Deviation	Min	Max
Amount of DAP (kg) across rounds	31.8	17.1	7.14	95
Amount of CAN (kg) across rounds	23.8	14.2	0	85
Amount of Lime (kg) across rounds	118.2	95.1	0	611.1
Yields (kg/acre) obtained in game	1343.8	498.4	419.3	2765.3
Share of rounds with DAP	0.78	0.23	0.20	1
Share of rounds with CAN (conditional on avail.)	0.80	0.26	0	1
Share of rounds with lime (conditional on avail.)	0.60	0.31	0	1
Share of rounds with no fertilizer	0.046	0.062	0	0.25
Random rainfall scenario (1=poor, 2=normal, 3=good)	2.01	0.30	1.25	2.80
Chosen rainfall scenario (1=poor, 2=normal, 3=good)	2.70	0.44	1	3
Rounds played	10.6	2.93	1	19
Observations	175			

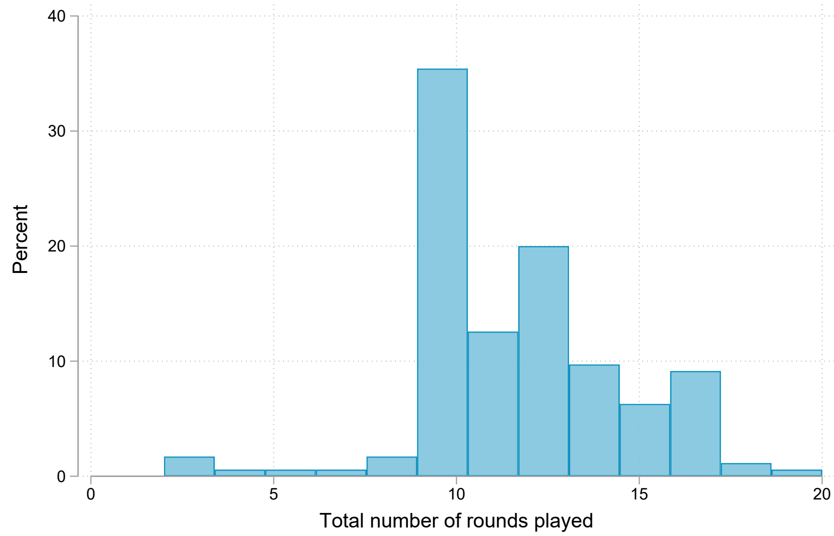

Figure 3: Number of rounds that experimental participants played

3.2 Beliefs

This section provides descriptive statistics for the subjective beliefs data. Some readers may wonder whether smallholder farmers really understand abstract concepts like probability distributions—even when elicited using visual aids. We hope that this section will help convince such readers that these measures are both reasonable and useful.

Panel A of Table 2 compares the upper and lower bounds of the subjective belief distribution (the smallest and largest of the five bins) to the observed maize yields from the panel survey. On average, the lower and upper bounds are reasonable. The upper bounds with fertilizer are still well within reasonable yield ranges. We can also note that the lower and upper bounds increase as farmers begin to consider fertilizer and lime. Similarly, after interacting with the game, farmers tend to revise both the upper and lower bounds, but the minimum lower bound remains zero or close to zero, which is reassuring since no amount of fertilizer can protect completely against severe weather shocks, for example.

In panel B we can see that the vast majority of respondents placed at least some beans in all bins. In other words, they allocated positive probability mass across the support. Additionally, Table 3 shows the distribution of probability across the support. As we might expect if our assumption that yield beliefs follow a log-normal distribution, generally the middle bins tend to have more mass, although in some cases the upper bins have more mass on average.

Table 2: Descriptive statistics on yield beliefs

Panel A: Bounds on belief distribution	Observations	Mean	s.d.	Min	Max
Pre-game					
No fertilizer					
Lower bound (kg/acre)	175	125.5	129.5	0	720
Upper bound (kg/acre)	175	408.0	283.7	9	1440
Fertilizer					
Lower bound (kg/acre)	121	269.4	230.3	0	1200
Upper bound (kg/acre)	121	788.1	446.3	162	2185.7
Fertilizer and lime					
Lower bound (kg/acre)	4	626.3	451.1	135	1200
Upper bound (kg/acre)	4	1605	942.5	990	3000
Post-game					
Fertilizer					
Lower bound (kg/acre)	175	309.7	285.3	0	1440
Upper bound (kg/acre)	175	1078.6	681.1	144	3420
Fertilizer and lime					
Lower bound (kg/acre)	60	435.5	348.1	0	1500
Upper bound (kg/acre)	60	1545.9	1087.3	162	5400
Avg. maize yields					
(kg/acre)	158	484.0	244.3	76.8	1590.5

Panel B: Share of bins with probability mass > 0	Observations	Mean	s.d.	Min	Max
Pre-game					
No fertilizer	175	4.81	0.66	1	5
Fertilizer	121	4.84	0.61	1	5
Fertilizer and lime	4	5	0	5	5
Post-game					
Fertilizer	175	4.83	0.61	1	5
Fertilizer and lime	60	4.78	0.78	1	5

Table 4 shows how yield beliefs correlate with soil characteristics and past yield reports. Columns (1) and (2) show correlations between no-fertilizer yields, (3) and (4) examine beliefs about yields with fertilizer application, and the last four columns follow the same pattern

Table 3: Probability distribution across bins

	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
Pre-game					
No fertilizer	0.14	0.15	0.20	0.23	0.29
Fertilizer	0.14	0.15	0.19	0.25	0.27
Fertilizer and lime	0.10	0.16	0.35	0.21	0.18
Post-game					
Fertilizer	0.13	0.14	0.19	0.25	0.28
Fertilizer and lime	0.13	0.16	0.18	0.22	0.32

Table 4: Pre-game expectations, correlations with baseline characteristics

	Mean (no fert.)	Mean (no fert.)	Mean (fert.)	Mean (fert.)	CV (no fert.)	CV (no fert.)	CV (fert.)	CV (fert.)
pH	6.03	5.71	-38.41	-40.19	-0.00	-0.00	-0.00	-0.01
	(28.95)	(28.61)	(42.80)	(42.96)	(0.01)	(0.01)	(0.00)	(0.00)
CEC	3.62*	3.70	1.36	1.84	0.00	0.00	0.00	0.00
	(2.11)	(2.54)	(2.13)	(2.39)	(0.00)	(0.00)	(0.00)	(0.00)
Mean yield	0.27***	0.27***	0.51***	0.50***	-0.00	-0.00	-0.00	-0.00
	(0.10)	(0.10)	(0.15)	(0.15)	(0.00)	(0.00)	(0.00)	(0.00)
Proportion of seasons with fertilizer		14.73		88.60		0.01		0.01
		(159.36)		(137.33)		(0.02)		(0.01)
Intercept	46.21	34.52	515.01*	444.51	0.69***	0.68***	0.66***	0.66***
	(211.53)	(272.11)	(274.15)	(298.98)	(0.04)	(0.04)	(0.03)	(0.04)
R^2	0.10	0.10	0.16	0.16	0.03	0.03	0.02	0.02
N	112	112	113	113	112	112	113	113

Notes: Dependent variable in columns (1) and (2) is mean yield expectation without fertilizer; in (3) and (4) it is the mean of yield expectations with fertilizer; (4)-(7) show the coefficient of variation for yield beliefs without fertilizer and with fertilizer, respectively.
Standard errors in parentheses
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

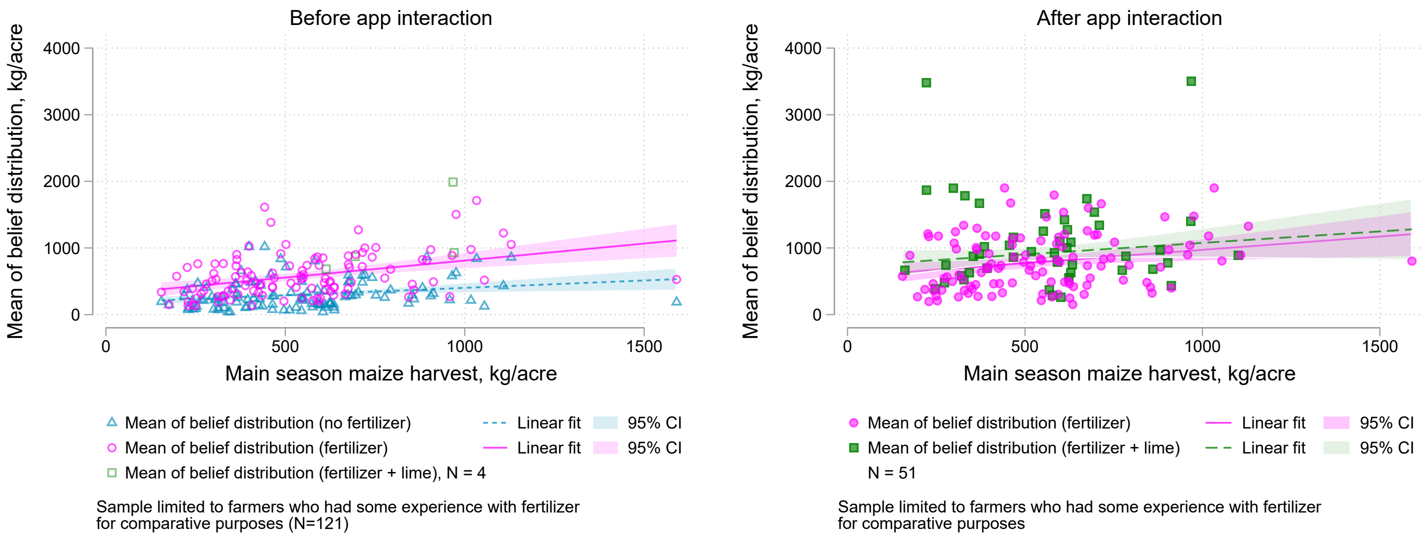

Figure 4: Subjective yield beliefs vs. main season yields

for the coefficient of variation. While the coefficient of variation does not seem to correlate with any of the observables, we can see that farmers who on average reported greater maize yields in the earlier panel survey tend to believe that their yields will be greater. While neither causal nor surprising, this correlation is broadly reassuring that the beliefs reflect the underlying reality.

Figure 4 displays this correlation in a different format. The left hand side shows that yield beliefs are positively correlated with main season yields—both for beliefs under fertilizer and no fertilizer, although the slope is flatter for the no-fertilizer correlation. After interacting with the game, farmers' beliefs have shifted up. The correlation has also become a bit noisier. Figure 5 shows a similar correlation but with pH instead of past yields. Here, the correlations are a bit less clear, although it appears that farmers with more acidic soils believe that yields under their normal fertilizer regime are greater than those with more basic soils. On the right hand side, after farmers have interacted with the app, we can see that the fertilizer + lime beliefs on average lie above those for fertilizer alone for farmers whose soils are acidic, while the two are not distunguishable at higher pH levels.

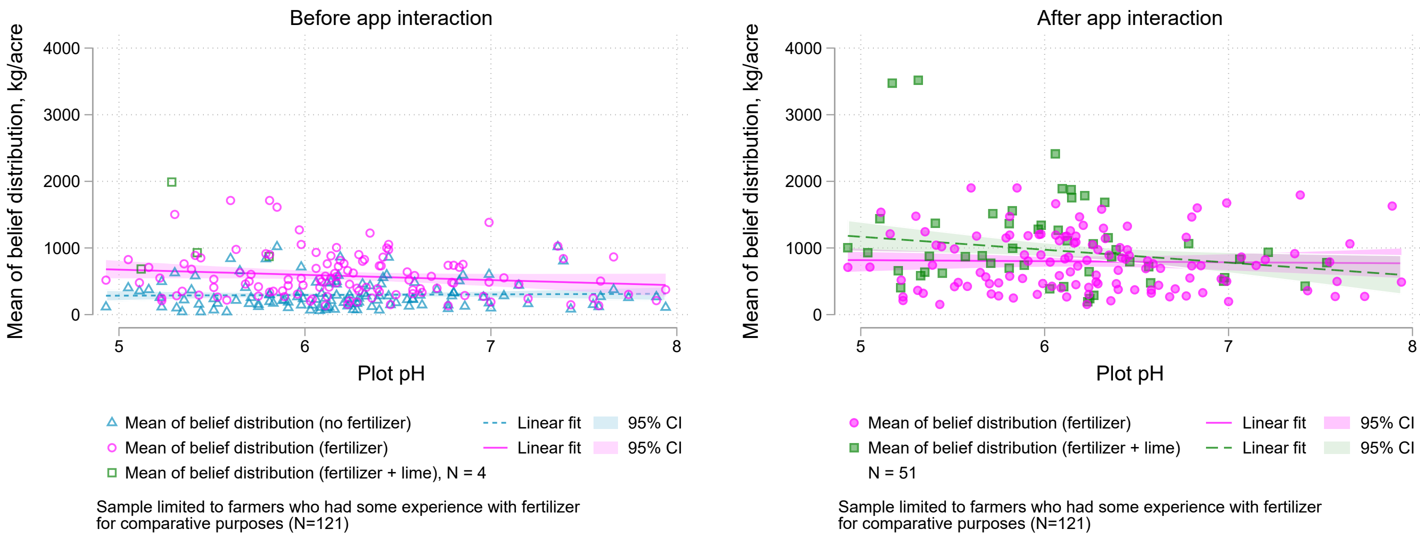

Figure 5: Subjective yield beliefs vs. maize plot soil pH

4 Empirical strategy

For this study, we registered a pre-analysis plan outlining hypotheses that focus on how farmers would update their beliefs about fertilizer and optimal inputs after playing *Mahindi-Master* and how individual farmers' characteristics could affect how they interact with the game and their subsequent belief updating. We also pre-registered a number of descriptive hypotheses (low-weight and low-prior hypotheses in the vocabulary of Anderson and Magruder (2017)).[15] We focus here three main groups of outcomes: formation and evolution of beliefs, changes to fertilizer orders, and experimentation within the app.

4.1 Outcomes

Formation and evolution of beliefs: Data on the formation and evolution of beliefs was elicited through subjective expectations. Before farmers interacted with the app, we elicited their expectations about the returns to fertilizer and fertilizer-plus-lime, which enables us to construct their prior distribution from which we calculate the mean and coefficient of

[15]For transparency, the full set of hypotheses and the results of tests of those hypotheses can be found in the supplementary materials (Appendix D contains the online appendix).

variation. Similarly, we also elicit farmers' expectations after interacting with the game from which we construct the posterior distribution and calculate the mean and coefficient of variation. Another outcome of interest is how beliefs evolve which we measure using the percentage change in the mean of the subjective expectation distributions before and after interacting with the game.

Changes in behavior: While we are interested in the updating of beliefs, we also want to know if farmers change their real-world behavior in response to any belief updating. We use the incentive compatible fertilizer orders as a measure of farmer behavior. Since farmers place orders before and after interacting with the app, we are able to measure the effect that the intervention had on farmer behavior. We measure changes to orders as the percentage change in the value of each fertilizer ordered before and after interacting with the game. Since many farmers did not order lime before interacting with the app, we also calculate the difference in the value of lime ordered (post-pre) instead of the percentage change. Given that lime is the main technology of interest, we also estimate the effect on the quantity of lime ordered. Since DAP is the most well-known fertilizer, we examine changes in the share of DAP in the order as well as the share of lime.

Experimentation within the app: We use the following outcome variables to measure experimentation within the game: share of rounds played with a positive amount of each input and whether the farmer played multiple final rounds. The share of rounds in the game played with familiar and unfamiliar inputs enables us to gauge the degree to which farmers experiment with new inputs. In the final round, farmers make their final fertilizer choice, which will be their fertilizer order. They are able to simulate yields with this choice and modify their choice as many times as desired. We measure this decision as a dummy variable that equals one if farmers play more than one final round and zero otherwise.

4.2 Estimation

We do not have a control group for the experiment and instead look at variation across farmers in our sub-sample and within-person beliefs in the periods before and after interacting with *MahindiMaster*. Beyond budgetary constraints, there are several reasons behind our reliance on a within-participant design: first, participants may react to the presence of other villagers. While there are numerous definitions of the Hawthorne effect, the definition in the Oxford English Dictionary, as cited by Levitt and List (2009), states that it encapsulates "an improvement in the performance of workers resulting from a change in their working conditions, and caused either by their response to innovation or *by the feeling that they are being accorded some attention*" (emphasis ours). Having your neighbors observe you while you interact with the app could distort behavior.

In addition, we worried that potential control farmers might react to being placed in the control group, which could cause John Henry effects (whereby control group individuals change their behavior in response to knowing their control group status). In addition to potential effects on the control group's behavior, we worried that participants may also change what types of inputs they experiment with (or other important behaviors) in anticipation of receiving questions by the control group after the experiment.

While some of the above concerns—notably the John Henry effects—could have been alleviated with a placebo treatment, the placebo would not address all of them. For example, if farmers spoke afterwards and learned that some had received farming information and others had not, our concerns about the spread of poor-quality information could still be warranted.

In sum, we decided that the costs outweighed the benefits. We believe that the evidence regarding the types of farmers (low pH) who exhibited the largest changes in input orders supports the notion that it is not purely driven by experimenter effects. Given that enumerators were blind to the pH status of farmers' soils, we feel that it supports our interpretation of the results.

To evaluate the effect of the intervention, we run a regression of the post-treatment outcome on the pre-treatment outcome and farmer traits. Specifically, we estimate the following equation using Ordinary Least Squares (OLS):

$$Post_i = \alpha + \beta Pre_i + \gamma Trait_i + \epsilon_i \tag{1}$$

In addition, given that the returns to lime are different across pH levels, we consider a quasi-experiment where we treat farmers with low soil pH as the treatment group then run a difference-in-differences framework to analyze the effect of the treatment on these farmers. Maize is best grown in slightly acidic soil, and since lime increases soil pH, the application of lime would only be beneficial for farmers who have more acidic soil. Therefore, the group of farmers with lower pH levels would have the highest returns from using lime, the main new technology that we introduce in the app. We estimate two sets of regressions to analyze the effect of soil pH, allowing first for a quadratic relationship with pH, and secondly a more flexible specification with indicator variables for set ranges of pH.

$$Post_i = \alpha + \beta Pre_i + \gamma_1 pH_i + \gamma_2 pH_i^2 + u_i \tag{2}$$

$$Post_i = \alpha + \beta Pre_i + \phi_k \sum_{k=1}^{5} pH_i^k + u_i \tag{3}$$

where the pH_i^k are dummy variables that indicate whether farmer i's pH is in one of five ranges of pH.[16]

[16] These ranges are pH < 5.5, pH$\in (5.5, 6)$, pH$\in (6, 6.5)$, pH$\in (6.5, 7)$, and pH> 7.

5 Results

5.1 Average effect of intervention on beliefs and behavior

Our main results focus on the effect of the intervention on beliefs and behavior. Table ?? presents the raw data: farmers' fertilizer orders in kilograms pre- and post-game. Farmers update their fertilizer orders. The top panel shows mean values in kilograms, while the bottom shows the mean value of the order in Kenyan Shillings (KES). On average, we observe that DAP orders decrease after the intervention and that lime orders increase, with both changes being statistically significant. Before playing the game, 49 farmers placed a non-zero lime order, while 67 farmers placed a non-zero lime order after playing the game. Forty-three of the 67 farmers who ordered lime after the game ordered no lime at all before playing the game. This suggests that some farmers may have chosen to order lime before interacting with the game purely to experiment with it, and then changed their minds after interacting with the app. This switch would also be consistent with the app encouraging learning.

Table 5: Fertilizer Orders

	Mean(pre)	Mean(post)	t-statistic
DAP (kg)	36.9	26.5	-6.45
CAN (kg)	40.1	40.5	0.13
Lime (kg)	54.9	96.8	2.26
DAP (KHS)	2448.1	1713.1	-5.92
CAN (KHS)	1965.2	2050.6	0.53
Lime (KHS)	345	612.9	2.18

Note: Differences based on clustering at the village level

We can similarly test for differences in the pre- and post-game means of farmers' subjective expectation distributions for both fertilizer and fertilizer-plus-lime (see Table ??). After farmers interact with *MahindiMaster*, they update their beliefs about the returns to fertilizer. On average, farmers' belief revisions are upwards, as seen by the higher mean yields for fertilizer and fertilizer-plus-lime. The differences in means before and after game-play for fertilizer as well as fertilizer-plus-lime are statistically significant. We also test if the post-mean for fertilizer (8.64) is statistically different from the post-mean for fertilizer-plus-lime (9.74) (t-statistic: 4.26). Pre-game, these means are not statistically distinguishable (t-statistic: 0.88). Both the upward revisions to the means of the belief distributions and the divergence of fertilizer versus fertilizer-plus-lime are consistent with what we would expect, given our priors that farmers generally have sub-optimally low beliefs about fertilizer and little knowledge of the benefits of lime application.

Table 6: Subjective expectations pre- and post-game

	Mean(pre)	Mean(post)	t-statistic
Mean (no fertilizer)	3.31		
Mean (fertilizer)	5.47	8.64	7.95
Mean (fertilizer + lime)	5.51	9.74	9.06
CV (fertilizer)	0.64	0.64	1.55
CV (fertilizer + lime)	0.64	0.64	1.72

Note: Differences based on clustering at the village level

As discussed in Section 4, we are particularly interested in understanding whether the observed updating behavior reflects learning. We estimate this using a difference-in-differences method based on the hypothesis that households with high *ex ante* expected returns to lime would respond more strongly to interacting with the game. In results available from the authors, we can see that a "production function" estimation shows that farmers with low pH received strong signals in the game about the returns to lime. Specifically, we regressed within-game yields in a given round on the chosen fertilizer levels, weather dummies, and interacted the chosen input levels with baseline pH. The interaction term between lime and pH is negative and significant across most specifications, suggesting that the game indeed revealed different marginal returns to lime for different types of farmers.

Table 7 shows the regression estimates of the effect of soil pH on lime orders and lime order updating. Columns (1) and (2) show results when the dependent variable is the total amount of lime ordered after the game. In columns (3) and (4), the outcome variable is the share of the post-game order allocated to lime, and columns (5) and (6) show results for

Table 7: Post-game lime orders by soil pH

Dependent variable:	Amt. (kg)	Amt. (kg)	Share	Share	(post-pre)	(post-pre)
pH	-431.68***		-61.60***		-1877.06	
	(152.85)		(23.27)		(1433.93)	
	[0.014]		[0.021]		[0.332]	
pH^2	28.06**		4.06**		119.28	
	(11.35)		(1.74)		(104.39)	
	[0.031]		[0.041]		[0.396]	
5.5 < pH <= 6		-33.43		-5.83		288.33
		(43.31)		(6.24)		(416.17)
		[0.508]		[0.458]		[0.508]
6 < pH <= 6.5		-128.53***		-17.86***		-278.88
		(39.65)		(5.89)		(396.65)
		[0.006]		[0.008]		[0.508]
6.5 < pH <= 7		-133.37***		-18.20***		-414.52
		(40.63)		(6.03)		(419.64)
		[0.006]		[0.008]		[0.455]
pH > 7		-153.79***		-20.71***		-450.95
		(40.03)		(5.99)		(408.56)
		[0.002]		[0.005]		[0.400]
Kg lime (pre)	0.07	0.06				
	(0.08)	(0.08)				
Lime share (pre)			0.07	0.05		
			(0.10)	(0.11)		
Intercept	1696.28***	201.23***	239.34***	28.15***	7350.35	486.67
	(511.75)	(37.59)	(77.55)	(5.61)	(4880.31)	(379.58)
R^2	0.16	0.18	0.13	0.15	0.05	0.07
N	167	167	167	167	167	167
Mean of dep. var:	98.41	98.41	13.78	13.78	270.24	270.24

Notes: Dependent variable for (1) and (2) is amount of lime in kilograms ordered post-game, for (3) and (4) it is the share of the value of the post-game order allocated to lime, and for (5) and (6) it is the difference in the value of the order allocated to lime (post-pre).
The numbers in square brackets are q-values. They denote the expected proportion of false positives that we would incur if we reject the null hypothesis of the coefficient above it equals zero (based on considering all the pH-related hypothesis tests in this table.
The omitted pH category is pH < 5.5.
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

the difference in the value of the order post-game minus pre-game. The results consistently suggest that farmers with lower-pH plots allocate larger budget shares to lime and order more lime in terms of value and quantity. The results are not significant in columns (5) and (6), but the patterns are similar to the other regressions.

Figure 6: Average share of post-game order allocated to inputs, by soil pH

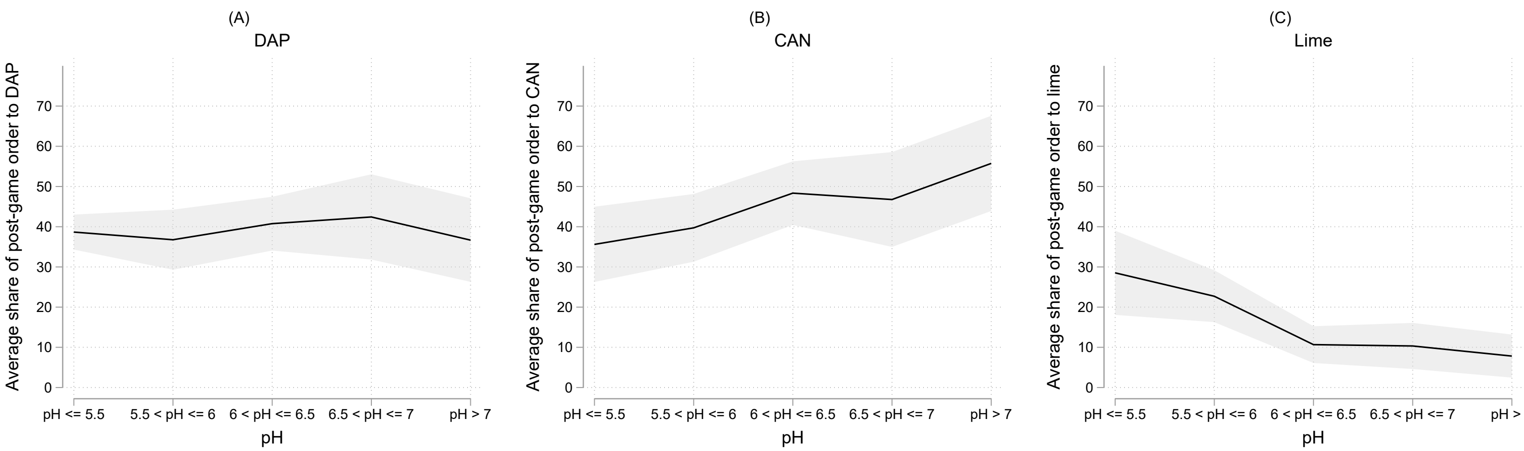

Figure 6 plots the predictive margins (marginal means) of the share of post-orders allocated to the different inputs for different bins of soil pH.[17] The interpretation is that the solid line shows the average share of post-orders that would be allocated to a particular input if all the farmers in the sample had a soil pH in that bin. If this average is the same across pH bins, then farmers do not seem to respond to pH in making their decisions about that paritcular input. We observe a negative gradient for lime in panel C, suggesting that farmers whose soil pH is acidic order more lime as a share of their total order, and that farmers who were already within the suitable range for maize order less lime. This suggests that most of our effects on lime order changes are due to the "right" farmers ordering more lime. We would not necessarily expect the same kind of gradient for DAP or CAN, and panels A and B of Figure 6 show the share of post-game orders allocated to DAP and CAN, respectively. DAP post-game orders do not seem to vary with pH, and the budget share allocated to CAN

[17]The plotted results correspond to column (4) in Table 7 for lime, and the analogous results for DAP and CAN, which can be found in Appendix Table C1

after the intervention slightly increases with pH (panel B). This latter gradient may suggest that low-pH farmers substitute between CAN and lime.

5.2 Effect heterogeneity

We are also interested in understanding whether some farmers are more likely to respond to the intervention. To this end, we examine whether farmers with differing farming ability are differentially likely to update their orders. It is not clear how farming ability should interact with new information. If highly-skilled farmers already had precise information about the returns to different inputs, we might expect them to respond less to the new information. If instead, the better farmers (as measured by the quiz) are better-informed more generally, but lack plot-specific information on returns, then we might expect them to be better equipped to update in response to the information.

Table 8 presents regression results analyzing these hypotheses. The dependent variable is the change in fertilizer orders and the variable of interest is farming quiz knowledge. The first three columns control only for the number of quiz questions that a farmer answered correctly, and it is negatively associated with DAP updating and positively associated with changing lime orders. Farmers who answer more questions correctly do not order different amounts of DAP or lime before interacting with the game, so the differences here are coming from the updating, not from the pre-game levels, which we also control for. However, this suggests that there is some correlation between belief updating and farming knowledge. Once we control for the number of questions that a farmer answered "Don't know" to in columns (4)-(6), the coefficient on changes in lime orders becomes insignificant. One possible interpretation of these results, overall, is that better-informed farmers (as measured by the quiz) are more responsive to novel information.

We can similarly examine whether farming ability correlates with the amount of belief updating that takes place after interacting with the app. Table 9 shows the results from this analysis, with the dependent variable being either the percentage change in the mean

Table 8: Orders and farming ability

	DAP (post-pre)	CAN (post-pre)	Lime (post-pre)	DAP (post-pre)	CAN (post-pre)	Lime (post-pre)
No. questions correct	-146.28*	141.73	147.41***	-179.00**	147.71	87.20
	(75.59)	(90.86)	(53.17)	(89.97)	(105.05)	(71.11)
No. questions='Don't know'				-32.13	5.89	-58.53
				(55.50)	(62.14)	(43.64)
DAP value (pre)	-0.78***			-0.79***		
	(0.15)			(0.15)		
CAN value (pre)		-0.97***			-0.97***	
		(0.20)			(0.20)	
Lime value (pre)			-0.87***			-0.89***
			(0.10)			(0.10)
Intercept	1600.28***	1585.49***	178.10	1868.42***	1543.86**	641.24
	(472.20)	(569.52)	(168.53)	(632.92)	(672.79)	(398.78)
R^2	0.13	0.15	0.33	0.13	0.15	0.34
N	158	158	158	158	158	158

Notes: Dependent variable for (1) and (4) is the difference in value of the order allocated to DAP, for (2) and (5) is the difference in value of the order allocated to CAN, and for (3) and (6) is the difference in the value of the order allocated to lime (post-pre).
Standard errors in parentheses
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

subjective yields for fertilizer (columns 1 and 3) or for fertilizer-plus-lime. Consistent with the results from Table **??**, we can see that farmers with more correct quiz answers seem to be more responsive to the intervention. This result is robust to controlling for the number of quiz questions that farmers knew that they did not know (columns 3 and 4).

Appendix B reports results for similar regressions in which we allow the extent of belief updating to vary by past experience with improved inputs, and by confidence. The evidence on confidence is somewhat sensitive to how we define confidence. A dummy variable for being overconfident is strongly associated with less belief updating for both distributions (fertilizer and fertilizer-plus-lime). The simple difference between the number of questions that farmers thought they got right and the number they did get right is also negatively correlated with belief updating of both subjective belief distributions. For past fertilizer use, we find that farmers who had used fertilizer more in the years prior to the baseline survey update their beliefs substantially less. We do not see much of a relationship between past experience and changes in the dispersion of the belief distributions (as measured by the coefficient of variation).

In a non-preregistered analysis found in Table 10, we additionally run an instrumental variable regression. We exploit the fact that rainfall simulations were random for the first seven app "seasons." When examining correlations, we noticed that the random variation in rainfall scenarios—specifically the fraction of good vs. bad rainfall scenarios—correlated quite strongly with their choices within the game. The analysis in Table 10 relies on an exclusion restriction that warrants some skepticism: that the (random) rainfall scenarios do not directly affect participants' belief revisions. If we are willing to believe this restriction, then we can use the random rainfall variation to instrument for in-game experimentation.

In particular, the 2SLS regression that we present here suggests that the proportion of rounds that a farmer plays with lime (instrumented for using random weather draws) appears to influence the extent to which she updates her beliefs. We find this intriguing, although we caution the reader to interpret these findings with the appropriate amount of skepticism.

5.3 What Influences Experimentation?

Finally, we are interested in exploring the factors that influence how farmers interact with the game. Do farmers experiment more with unfamiliar inputs, now that it is relatively costless to do so? Do they continue experimenting more with it if they find that it has positive returns? We see in Figure 7 that the pH of a farmer's field is not strongly associated with the share of rounds played with either DAP or CAN (panels A and B). In contrast, the pH of a farmer's field has a strong negative association with the share of rounds in the game played with a positive amount of lime, as shown by the steep negative gradient on panel C. While most farmers experimented with lime in at least one round, this suggests that the farmers whose fields most needed lime also experimented more with the input in the game—perhaps to get a clearer sense of the shape of the marginal returns to the input.

Table 9: Belief revisions

	%Δ Mean fert.	%Δ Mean fert.	%Δ Mean fert.	%Δ Mean fert+lime	%Δ Mean fert+lime	%Δ Mean fert+lime
No. correct	16.44*	42.92***	27.71**	33.65***	65.53***	51.84***
	(9.03)	(13.12)	(11.94)	(12.24)	(18.41)	(17.41)
No. q's = Don't know		24.64***	19.48***		29.66***	25.02***
		(6.19)	(5.45)		(8.09)	(7.53)
Overconfident (0/1)			-55.90**			-50.33
			(24.12)			(31.35)
Intercept	44.57*	-153.16***	-50.18	20.84	-217.19***	-124.48
	(23.72)	(57.94)	(56.18)	(28.57)	(78.61)	(80.19)
R^2	0.01	0.09	0.11	0.04	0.10	0.11
N	175	175	175	175	175	175
Mean of dep. var:	92.77	92.77	92.77	119.49	119.49	119.49
Exploratory analysis:						
No. correct	27.71**	29.37**	30.20**	51.84***	47.45***	47.45***
	(11.94)	(12.33)	(14.01)	(17.41)	(17.52)	(17.52)
No. q's = Don't know	19.48***	23.36***	10.41	25.02***	29.02***	29.02***
	(5.45)	(6.01)	(7.12)	(7.53)	(8.49)	(8.49)
Overconfident (0/1)	-55.90**	-54.47**	-39.73*	-50.33	-48.83	-48.83
	(24.12)	(23.20)	(22.29)	(31.35)	(30.51)	(30.51)
No. rounds		-5.93**	-3.79		-6.91*	-6.91*
		(2.87)	(2.58)		(4.12)	(4.12)
Share w/ CAN		102.13**	91.57*		134.61**	134.61**
		(46.79)	(47.36)		(52.97)	(52.97)
Share w/ lime		-59.00	-31.35		-16.73	-16.73
		(36.62)	(32.09)		(50.88)	(50.88)
pH			36.58**			
			(17.52)			
Share of seasons w/ fert.			-2.17			
			(4.89)			
Experience w/ DAP			-7.95			
			(38.80)			
Experience w/ CAN			-57.34**			
			(24.29)			
Land size (acres)			-15.69			
			(9.88)			
Intercept	-50.18	-52.80	-221.00	-124.48	-149.84*	-149.84*
	(56.18)	(60.05)	(137.91)	(80.19)	(88.30)	(88.30)
R^2	0.11	0.14	0.22	0.11	0.14	0.14
N	175	175	175	175	175	175
mean						

Notes: Dependent variable is the percentage change in mean of the subjective yields distribution for fertilizer and fertilizer + lime, where noted.
Standard errors in parentheses
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 10: 2SLS regression: effect of experimenting with lime on belief updating

Share of rounds w/ lime	73.68	143.04**
	(52.40)	(60.00)
pH	62.15***	47.68**
	(18.69)	(20.24)
Intercept	-351.72***	-273.08*
	(132.33)	(143.22)
R^2	0.07	0.03
N	175	175
Kleibergen-Paap F-stat	35.60	35.60

Notes: Dependent variable: percentage change in mean of the subjective yields distribution for fertilizer and fertilizer + lime. Share of rounds played with lime is instrumented using the share of rounds that the respondent received good and bad rainfall draws (during rounds when they had no choice).)
Standard errors in parentheses
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Appendix C shows the regressions associated with Figure 7.

We also analyze how farmers' previous experience with inputs affects how they interact with the game. What share of rounds do people play with unfamiliar inputs, and does confidence affect this? Tables 11-13 show the how confidence correlates with the share of rounds played with familiar and unfamiliar inputs. Across the input types and specifications, we observe that farmers who answered more quiz questions correctly experiment more with all the inputs (at the extensive margin). This may be a signal of more sophisticated experimentation aimed at understanding the interaction of inputs (as opposed to the returns to a single input at a time) or of a greater general propensity to experiment. Confidence does not appear to be strongly associated with the share of rounds played with DAP or CAN, regardless of how we specify confidence. For lime, the percentage difference between the number of quiz questions that a farmer believes she answered correctly and the number that she actually answered correctly is negative and significant. We do not read too much into this since the magnitude of the point estimate is small and the other specifications are insignificant.

Figure 7: Average share of rounds played with non-zero input amounts, by soil pH

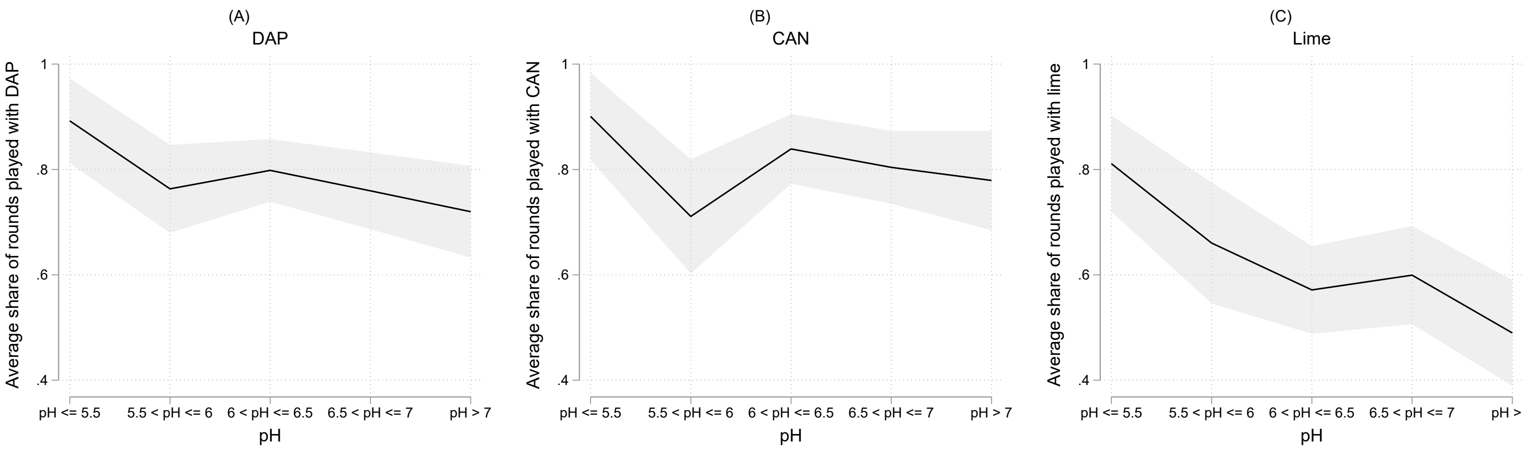

Table 11: Share of Rounds with DAP

	(1)	(2)	(3)
(No. believed correct - No. correct)/ No. believed correct	0.001		
	(0.00)		
No. questions correct	0.080***	0.081***	0.073***
	(0.02)	(0.02)	(0.02)
(No. believed correct - No. correct)		0.014*	
		(0.01)	
Overconfident (0/1)			0.013
			(0.04)
Uses DAP	0.110***	0.100***	0.110***
	(0.03)	(0.03)	(0.03)
Intercept	0.459***	0.446***	0.500***
	(0.07)	(0.07)	(0.07)
R^2	0.22	0.22	0.21
N	173	173	173
Mean of dep. var:	0.78	0.78	0.78

Notes: Dependent variable is the share of rounds played with a positive amount of DAP.
Standard errors in parentheses
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 12: Share of Rounds with CAN

	(1)	(2)	(3)
(No. believed correct - No. correct)/ No. believed correct	0.000		
	(0.00)		
No. questions correct	0.074***	0.072***	0.075***
	(0.02)	(0.02)	(0.02)
(No. believed correct - No. correct)		0.003	
		(0.01)	
Overconfident (0/1)			0.020
			(0.05)
Uses CAN	0.023	0.024	0.022
	(0.04)	(0.04)	(0.04)
Intercept	0.568***	0.576***	0.566***
	(0.07)	(0.08)	(0.08)
R^2	0.10	0.10	0.11
N	173	173	173
Mean of dep. var:	0.81	0.81	0.81

Notes: Dependent variable is the share of rounds played with a positive amount of CAN.
Standard errors in parentheses
$*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

Table 13: Share of Rounds with Lime

	(1)	(2)	(3)
(No. believed correct - No. correct)/ No. believed correct	-0.001***		
	(0.00)		
No. questions correct	0.069***	0.081***	0.086***
	(0.02)	(0.02)	(0.02)
(No. believed correct - No. correct)		-0.010	
		(0.01)	
Overconfident (0/1)			-0.013
			(0.05)
Intercept	0.452***	0.389***	0.350***
	(0.07)	(0.08)	(0.08)
R^2	0.12	0.11	0.10
N	173	173	173
Mean of dep. var:	0.60	0.60	0.60

Notes: Dependent variable is the share of rounds played with a positive amount of CAN.
Standard errors in parentheses
$*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

6 Conclusion

Of the many constraints and impediments to agricultural technology adoption among poor producers, one of the most fundamental is the challenge they face when trying to learn about the properties and performance of a new technology. This learning challenge stems from pronounced heterogeneity in the returns to agricultural technologies. A number of factors drive heterogeneity, including complex interactions with other inputs and local production conditions, stochasticity of the production settings in which these farmers operate, and farmers' own limited ability to proactively experiment in ways that generate useful knowledge about the returns to the technology in their specific case. In such contexts, simply providing information about a technology may be insufficient to overcome information gaps.

In particular, producers' mental models shape their expectations of the production relationships they seek to manage and optimize. If the goal of information interventions is to change expectations, it may therefore be important to induce the kind of learning that changes these mental models. Learning that changes mental models and subsequent behavior demands active engagement and discovery. In this paper, we explore the possibility and efficacy of such engaged learning and discovery through a virtual platform that enables farmers to proactively experiment with agricultural inputs on virtual plots calibrated to match their own.

Using incentive-compatible input orders and subjective expectations elicitation techniques, we find that farmers revise their beliefs about input returns upwards after interacting with a platform designed as a virtual maize farming app. More specifically, farmers with acidic soils who are best positioned to benefit *ex ante* from a new, unfamiliar input discover this opportunity purely as a result of playing repeated and riskless virtual seasons on the app. This encouraging learning response emerges as these likely beneficiaries engage in more intensive, more intentional experimentation with the new input on the virtual platform.

Although this work is but an initial step towards a better understanding of the potential role of virtual learning platforms in complex settings such as smallholder agriculture, it nonetheless raises several intriguing possibilities in broader development contexts. Could the introduction of virtual learning platforms and the application of gamification principles enhance learning and productivity in other contexts? The spread of simple cell phones into rural areas of poor countries opened new options for building adult literacy (Aker et al., 2012). The steady expansion of enhanced ICT services delivered by smart phones and other internet-enabled devices into such places will create new commercial opportunities for improved market access, reduced transaction costs and enhanced information flows and entertainment. As such, these new ICT tools may have relevance well beyond formal classroom applications. By leveraging these opportunities to offer new ways of learning and discovering, gamified virtual platforms with accessible user interface designs could become a potent tool for a wide array of adult education and vocational training objectives.

Such discovery and learning platforms could also directly shape supply chains. Poorly developed supply chains and poorly integrated input and output markets are often prominent constraints to agricultural technology adoption. If a platform like the one we prototype and test in this paper could be integrated into local supply chains for improved agricultural inputs, it might help to stimulate demand for inputs and generate useful information about demand heterogeneity to potential suppliers. Although potentially complex, such coordination and integration with suppliers is a natural extension of the kind of ICT services that are beginning to proliferate among smallholder farmers.

We want to be careful to point out that these results ought to be interpreted as intriguing and a potential reason to explore similar interventions in future work— not as conclusive or with substantial external validity. The study was designed more as a "proof of concept" and many open questions remain, including whether farmers' belief revisions and behavioral changes persists over time. Other interesting questions include whether different versions of the app could be more effective. Some examples of versions that we considered at some point of the design process include more personalized, more competitive, and more pedagogical versions. Perhaps different versions would appeal to different types of farmers. If so, what observable characteristics determine this interaction? What role do risk and time preferences play?

Our results are encouraging. They apply to a specific agricultural input in a particular context, but we leave it for future research to explore whether the underlying principles of engaged, self-directed discovery apply to a broader set of learning contexts. The diffusion of bandwidth and connected devices into previously unconnected places sets the stage for tapping these broader opportunities to shape mental models and the expectations they generate. Since this level of learning is fundamental to improved decision making, these opportunities raise a host of important practical possibilities as well as economic questions with clear policy relevance.

References

Kehinde F Ajayi, Willa H Friedman, and Adrienne M Lucas. The importance of information targeting for school choice. *American Economic Review*, 107(5):638–43, 2017.

Jenny C Aker, Christopher Ksoll, and Travis J Lybbert. Can mobile phones improve learning? Evidence from a field experiment in niger. *American Economic Journal: Applied Economics*, 4(4):94–120, 2012.

Sandro Ambuehl and Shengwu Li. Belief Updating and the Demand for Information. *Games and Economic Behavior*, 109:21–39, May 2018.

Michael L. Anderson and Jeremy Magruder. Split-Sample Strategies for Avoiding False Discoveries. Technical report, National Bureau of Economic Research, 2017. URL http://www.nber.org/papers/w23544.

Nava Ashraf, B. Kelsey Jack, and Emir Kamenica. Information and Subsidies: Complements or Substitutes? *Journal of Economic Behavior & Organization*, 88:133–139, April 2013.

Orazio P. Attanasio and Katja M. Kaufmann. Education Choices and Returns on the Labor and Marriage Markets: Evidence from Data on Subjective Expectations. *Journal of Economic Behavior & Organization*, 140:35–55, August 2017.

Bradford L. Barham, Jean-Paul Chavas, Dylan Fitz, and Laura Schechter. Receptiveness to advice, cognitive ability, and technology adoption. *Journal of Economic Behavior & Organization*, 149:239–268, 2018.

Eric P. Bettinger, Bridget Terry Long, Philip Oreopoulos, and Lisa Sanbonmatsu. The Role of Application Assistance and Information in College Decisions: Results from the H&r Block Fafsa Experiment. *Quarterly Journal of Economics*, 127(3):1205–1242, August 2012.

Gharad Bryan, Shyamal Chowdhury, and Ahmed Mushfiq Mobarak. Underinvestment in a Profitable Technology: The Case of Seasonal Migration in Bangladesh. *Econometrica*, 82 (5):1671–1748, 2014.

Michael Carter, Mary Mathenge, Samuel Bird, Travis Lybbert, Tim Njagi, and Emilia Tjernström. Filling a niche: the impacts of a local seed company on maize productivity in Kenya. Working paper, 2019.

Gary Charness and Angelino Viceisza. Three risk-elicitation methods in the field: Evidence from rural senegal. *Review of Behavioral Economics*, 3(2):145–171, 2016.

Timothy G. Conley and Christopher R. Udry. Learning about a New Technology: Pineapple in Ghana. *The American Economic Review*, 100(1):35–69, 2010.

Adeline Delavande. Probabilistic Expectations in Developing Countries. *Annual Review of Economics*, 6(1):1–20, 2014.

Adeline Delavande, Xavier Giné, and David McKenzie. Eliciting Probabilistic Expectations with Visual Aids in Developing Countries: How Sensitive Are Answers to Variations in Elicitation Design? *Journal of Applied Econometrics*, 26(3):479–497, 2011a.

Adeline Delavande, Xavier Giné, and David McKenzie. Measuring Subjective Expectations in Developing Countries: A Critical Review and New Evidence. *Journal of Development Economics*, 94(2):151–163, March 2011b.

Roberta Dessí and Xiaojian Zhao. Overconfidence, Stability and Investments. *Journal of Economic Behavior & Organization*, 145:474–494, January 2018.

Thomas Dohmen, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G. Wagner. Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences. *Journal of the European Economic Association*, 9(3):522–550, June 2011.

Pascaline Dupas. Do Teenagers Respond to HIV Risk Information? Evidence from a Field Experiment in Kenya. *American Economic Journal: Applied Economics*, 3(1):1–34, January 2011.

Stephen M Fiore, Glenn W Harrison, Charles E Hughes, and E Elisabet Rutström. Virtual experiments and environmental policy. *Journal of Environmental Economics and Management*, 57(1):65–86, 2009.

Andrew D. Foster and Mark R. Rosenzweig. Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture. *Journal of Political Economy*, 103 (6):1176–1209, 1995.

Neil Gordon, Mike Brayshaw, and Simon Grey. Maximising gain for minimal pain: Utilising natural game mechanics. *Innovation in Teaching and Learning in Information and Computer Sciences*, 12(1):27–38, 2013.

Aurélie Harou, Malgosia Madajewicz, Chris Magomba, Hope Michelson, Kevin Tschirhart, Nyambi Amuri, and Cheryl Palm. Can information improve investment? Effects of site-specific soil recommendations on fertilizer demand. Working paper, 2018.

Mitchell Hoffman. How is Information Valued? Evidence from Framed Field Experiments. *The Economic Journal*, 126(595):1884–1911, September 2016.

Wallace E. Huffman, Matthew Rousu, Jason F. Shogren, and Abebayehu Tegene. The Effects of Prior Beliefs and Learning on Consumers' Acceptance of Genetically Modified Foods. *Journal of Economic Behavior & Organization*, 63(1):193–206, May 2007.

B Kelsey Jack. Constraints on the adoption of agricultural technologies in developing countries. *White paper, Agricultural Technology Adoption Initiative, Boston: J-PAL (MIT) and Berkeley: CEGA (UC Berkeley)*, 2011.

Joey J. Lee and Jessica Hammer. Gamification in education: What, how, why bother? *Academic exchange quarterly*, 15(2):146, 2011.

Jeremy R Magruder. An assessment of experimental evidence on agricultural technology adoption in developing countries. *Annual Review of Resource Economics*, 10:299–316, 2018.

Helia Marreiros, Mirco Tonin, Michael Vlassopoulos, and M. C. Schraefel. “Now That You Mention It”: A Survey Experiment on Information, Inattention and Online Privacy. *Journal of Economic Behavior & Organization*, 140:1–17, August 2017.

Don A. Moore and Paul J. Healy. The Trouble with Overconfidence. *Psychological Review*, 115(2):502–517, April 2008.

Kaivan Munshi. Social Learning in a Heterogeneous Population: Technology Diffusion in the Indian Green Revolution. *Journal of Development Economics*, 73(1):185–213, February 2004.

Jorge Simões, Rebeca Díaz Redondo, and Ana Fernández Vilas. A social gamification framework for a K-6 learning platform. *Computers in Human Behavior*, 29(2):345–353, 2013.

P. Tittonell, B. Vanlauwe, M. Corbeels, and K. E. Giller. Yield Gaps, Nutrient Use Efficiencies and Response to Fertilisers by Maize across Heterogeneous Smallholder Farms of Western Kenya. *Plant and Soil*, 313(1-2):19–37, December 2008.

Emilia Tjernström. Learning from Others in Heterogeneous Environments. Working paper, April 2018.

Steffen P. Walz and Sebastian Deterding. *The Gameful World: Approaches, Issues, Applications*. MIT Press, January 2015.

Matthew Wiswall and Basit Zafar. How Do College Students Respond to Public Information about Earnings? *Journal of Human Capital*, 9(2):117–169, 2015.

Appendices

A *MahindiMaster* Design Details

MahindiMaster was built using the crop modeling software DSSAT. DSSAT uses plot-level soil characteristics and historical weather to simulate yields under different fertilizer types and application rates. We calibrated each simulation based on soil samples from farmers' fields (collected in October 2016). An ISO-certified laboratory analyzed these soil samples. We also connect historical weather data from AgMERRA to the field's GPS location to identify low, medium, and high rainfall scenarios.

DSSAT requires some soil information that we could not get from our soil samples. We therefore supplement the soil sample data with soil characteristic information from Africa Soil Information Service (AfSIS), which provides estimates of soil characteristics for 250x250m grids across the African continent. Data for calcium, CEC, potassium, phosphorus, and pH come from our soil samples, while bulk density, stones (%), clay (%), silt (%), and nitrogen come from the AfSIS data. We use the median value from 22 WISE profiles from Kenya for the lower limit of plant extractable soil water, drained upper limit, saturated upper limit, albedo, evaporation limit, drainage rate, runoff curve number, mineralization factor, and soil fertility factor. Finally, we set saturated hydraulic conductivity to 0.06.

DSSAT also requires daily rainfall, maximum temperature, minimum temperature, and solar radiation. We categorize rainfall into three categories based on historical data, classifying poor rainfall as the 35th percentile of historical rainfall, medium rainfall as the median historical rainfall, and good rainfall as the 60th percentile of historical rainfall. We use the median daily value for temperature and solar radiation. DSSAT models the effect of nitrogen and phosphorus applications but cannot simulate the effect of potassium or lime. We therefore manually input the lime applications into DSSAT, with each 250kg application corresponding to a .135 increase in soil pH.

We vary application rates for DAP and CAN between 0 to 125kg per acre in steps of

25kg. Lime is applied in much larger quantities, so we allow it to vary between 0 to 2000kg in 250kg steps. We simulate yields under all combinations of DAP, CAN, and lime and the three weather scenarios, resulting in 972 yield simulations per household.

B Do effects vary by confidence and fertilizer experience?

Table B1: Updating and confidence

	%Δ Mean	%Δ Mean	%Δ Mean
Fertilizer Distribution:			
(No. believed correct - No. correct)/ No. believed correct	-0.81		
	(0.82)		
No. questions correct	3.66	4.35	0.38
	(12.57)	(8.18)	(8.45)
(No. believed correct - No. correct)		-17.41**	
		(6.70)	
Overconfident (0/1)			-85.12***
			(26.76)
Intercept	119.76	137.39***	143.59***
	(73.89)	(38.46)	(37.22)
R^2	0.05	0.07	0.08
N	173	173	173
Mean of dep. var:	93.55	93.55	93.55
Fertilizer + Lime Distribution:			
(No. believed correct - No. correct)/ No. believed correct	-0.84		
	(0.99)		
No. questions correct	20.46	22.10**	17.36
	(14.95)	(10.76)	(11.06)
(No. believed correct - No. correct)		-16.62*	
		(9.35)	
Overconfident (0/1)			-86.44**
			(34.12)
Intercept	98.53	109.50**	121.36***
	(86.14)	(47.88)	(43.20)
R^2	0.05	0.07	0.07
N	173	173	173
Mean of dep. var:	120.57	120.57	120.57

Notes: Dependent variable is the percentage change in mean of the subjective yields distribution for fertilizer and fertilizer + lime, where noted.
Standard errors in parentheses
$*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

Table B2: Past fertilizer use and belief updating

	%Δ Mean Fertilizer Dist.	%Δ Mean Fertilizer + Lime Dist.	%Δ Mean Fertilizer Dist.	%Δ Mean Fertilizer + Lime Dis
Short and long rains:				
No. of seasons used fertilizer	-6.36***	-6.39**	0.05	0.03
	(2.10)	(2.53)	(0.09)	(0.08)
Intercept	118.88***	145.69***	0.41	0.44
	(17.24)	(21.81)	(0.54)	(0.52)
R^2	0.04	0.02	0.00	0.00
Long rains only:				
No. of seasons used fertilizer	-12.53***	-12.55**	0.09	0.05
	(4.15)	(4.97)	(0.18)	(0.16)
Intercept	118.55***	145.31***	0.43	0.46
	(17.19)	(21.73)	(0.53)	(0.52)
R^2	0.04	0.02	0.00	0.00
N	175	175	175	175
Mean of dep. var:	92.77	119.49	0.62	0.57

Notes: The dependent variable is the percentage change in the fertilizer and fertilizer-plus-lime distributions: ((post-mea premean)/pre-mean)*100

Standard errors in parentheses

$*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

C Correlations between pH, orders, and in-game behavior

Table C1: Correlations between soil pH and DAP, CAN orders

	Share of final order to DAP	Share of final order to CAN
DAP share (pre)	0.50***	
	(0.13)	
CAN share (pre)		0.24
		(0.17)
$5.5 < pH <= 6$	-1.91	4.11
	(4.41)	(6.41)
$6 < pH <= 6.5$	2.09	12.74**
	(3.97)	(6.34)
$6.5 < pH <= 7$	3.76	11.16
	(5.82)	(7.58)
$pH > 7$	-2.00	20.14**
	(5.73)	(7.76)
Intercept	12.95**	25.60***
	(6.50)	(7.22)
R^2	0.07	0.06
N	167	167

Notes: Dependent variables: (1) Share of the value of the post-game order allocated to DAP, (2) Share of the value of the post-game order allocated to CAN
The omitted pH category is pH < 5.5.
Standard errors in parentheses
$*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

Table C2: Correlations between soil pH and game behavior

	Share DAP game	Share CAN game	Share lime game
5.5 < pH <= 6	-0.13**	-0.19***	-0.15**
	(0.06)	(0.07)	(0.07)
6 < pH <= 6.5	-0.09*	-0.06	-0.24***
	(0.05)	(0.05)	(0.06)
6.5 < pH <= 7	-0.13**	-0.10*	-0.21***
	(0.06)	(0.05)	(0.07)
pH > 7	-0.17***	-0.12*	-0.32***
	(0.06)	(0.06)	(0.07)
Intercept	0.89***	0.90***	0.81***
	(0.04)	(0.04)	(0.05)
R^2	0.04	0.04	0.08
N	175	175	175

Notes: Dependent variables are as follows: (1) Share of rounds played using DAP, (2) Share of rounds played using CAN. (3) Share of rounds played using lime
The omitted pH category is pH < 5.5.
Standard errors in parentheses
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

D Online appendix: Pre-analysis plan and results

For transparency, this appendix (meant for an online appendix) presents the complete set of results that were part of the pre-analysis plan but excluded from the body of the paper. Tables D1-D4 summarize the pre-specified hypotheses. In each table, column (4) shows the expected sign of each hypothesis (if pre-registered), and column (5) reports the actual results. For the hypotheses in Table D4, we did not have clear priors on signs, but wanted to register that these were hypotheses we wanted to explore. Since some hypotheses are reported elsewhere in the paper, the last column of each table notes where the results for each hypothesis can be found, either in the appendix or in the paper.[18]

The full pre-analysis plan for this study can be found at the Registry for International Development Impact Evaluations (RIDIE). The sections below explain the hypotheses in more detail and display the full results for each of the hypotheses.

Table D1: Group of Hypotheses A. Impact on farmer's beliefs:

Hypothesis	Outcome Variable	Covariate	Expected sign $\frac{\partial Y}{\partial X}$	Actual sign $\frac{\partial Y}{\partial X}$	Table
A1	Δ mean beliefs	Experience with DAP & CAN	$-$	$-/0$	
	CV prior beliefs	Experience with DAP & CAN	$-$	0	D5
A2	Δ orders	Δ mean beliefs	$\neq$0	0 for DAP 0 for CAN $+/0$ for Lime	D6 D7 D8
A3	Lime budget share post-game	pH	$-$	$-/0$	7
A4a					-
A4b	Δ mean beliefs	Yields and planting sequence	+	$+/0$	D9

$-/0$: main coefficient is ≤ 0 and statistically significant in some models; not distinguishable from zero in others.
$+/0$: main coefficient is ≥ 0 and statistically significant in some models; not distinguishable from zero in others.
CV stands for the coefficient of variation and Δ denotes changes in variables pre- and post- intervention.

[18]In Table D1, hypothesis A4a is greyed out because we were unable to test this hypothesis. The hypothesis described a "sanity check," which we had not clearly defined what it meant in the pre-analysis plan. We therefore do not test it but leave the row in the table for transparency.

Table D2: Group of Hypotheses B. Effect of farmer characteristics on game choices

Hypothesis	Outcome Variable	Covariate	Expected sign $\frac{\partial Y}{\partial X}$	Actual sign $\frac{\partial Y}{\partial X}$	Table
B1	Rounds played in game	Confidence	$-$	$+/0$	D11
B2a	Rounds played, final round	Confidence	$-$	0	D12, D13
B2b*	Rounds played, final round	Confidence	$-$		-
B3	Rounds played	Risk Aversion	$+$	0	D14
B4	Low-rainfall weather scenarios	Risk Aversion	$+$	$-/0$	D15
B5	Share of rounds played w/ CAN, Lime	Risk Aversion	$+$	0	D16, D17
B6	Rounds played, final round	Risk Aversion	$+$	$-/0$	D18

$-/0$: main coefficient is ≤ 0 and statistically significant in some models; not distinguishable from zero in others.
$+/0$: main coefficient is ≥ 0 and statistically significant in some models; not distinguishable from zero in others.
*This hypothesis was defined conditional on the participant having changed the final order at least once. Too few farmers in our sample played more than one final round so we are unable to test this hypothesis.

Table D3: Group of Hypotheses C. Heterogeneity of farmer belief impacts

Hypothesis	Outcome Variable	Covariate	Expected sign $\frac{\partial Y}{\partial X}$	Actual sign $\frac{\partial Y}{\partial X}$	Table
C1	Δ orders	Rounds played	$+$	0	D19
	Δ mean beliefs	Rounds played	$+$	$+/0$	
C2	Δ orders	Correct questions on quiz	$-$	$-$ for DAP, $+/0$ for CAN, $+$ for Lime	**??**
	Δ mean beliefs	Correct questions on quiz	$-$	$+$	9

$-/0$: main coefficient is ≤ 0 and statistically significant in some models; not distinguishable from zero in others.
$+/0$: main coefficient is ≥ 0 and statistically significant in some models; not distinguishable from zero in others.
Δ denotes changes in variables pre- and post- intervention.

Table D4: Group of Hypotheses D. Unclear outcomes

Hypothesis	Outcome Variable	Covariate	Expected sign $\frac{\partial Y}{\partial X}$	Actual sign $\frac{\partial Y}{\partial X}$	Table
D1	Share of rounds CAN, Lime	Confidence	?	0 for CAN, $-$ for Lime	11, 12, 13
D2	Share pre-final rounds	Confidence	?	$+$	D20
D3	Dummy, multiple final rounds played	Risk aversion	?	$-/0$	D21
D4	Share, pre-final rounds played	Risk aversion	?	$+/0$	D22
D5	Δ mean beliefs	Risk	?	0	D23, D24
	Δ mean beliefs	Confidence	?	$-$	B1

$-/0$: main coefficient is ≤ 0 and statistically significant in some models; not distinguishable from zero in others.
$+/0$: main coefficient is ≥ 0 and statistically significant in some models; not distinguishable from zero in others.
Δ denotes changes in variables pre- and post- intervention.

Hypothesis A1

For DAP and CAN, we expect farmers who have less past experience with these two fertilizers (as observed in the RCT the panel data) to have more diffuse prior beliefs than farmers who have used these fertilizer extensively in the past. We expect that farmers with less experience and/or more diffuse priors will update their beliefs more after playing the game.

As a descriptive statistic related to hypothesis A1, we examine the correlations between farmers' past experience with fertilizer (as well as DAP or CAN individually) and their beliefs about returns to fertilizer. The test of the updating hypothesis can be found in Table B2 and how they update their beliefs. We define past fertilizer experience as the number of seasons in which the farmer used inorganic fertilizer in the five years before the baseline survey. Alternately, we also define past fertilizer experience as the number of long rains in which the farmer used inorganic fertilizer in the five years preceding the baseline survey. From the results, we see that households who have more experience with fertilizer in the past have subjective belief distributions with a higher mean but with no difference in the coefficient of variation.

Table D5: Past Fertilizer Use and Fertilizer Beliefs

	Mean Fertilizer Dist	Mean Fert + Lime Dist	CV Fert Dist	CV Fert + Lime Dist
Short and Long Rains:				
No. of Seasons Used Fertilizer	0.26***	0.29***	-0.001	-0.001
	(0.06)	(0.07)	(0.00)	(0.00)
Intercept	4.38***	4.34***	0.64***	0.64***
	(0.28)	(0.28)	(0.00)	(0.00)
R^2	0.11	0.11	0.01	0.01
Test H_0: coef <= 0 (p-value)			0.85	0.82
Long Rains Only:				
No. of Seasons Used Fertilizer	0.52***	0.56***	-0.001	-0.001
	(0.12)	(0.13)	(0.00)	(0.00)
Intercept	4.40***	4.36***	0.64***	0.64***
	(0.28)	(0.28)	(0.00)	(0.00)
R^2	0.11	0.11	0.00	0.01
Test H_0: coef <= 0 (p-value)			0.86	0.83
N	175	175	175	175
Mean of Dependent Variable:	5.47	5.51	0.64	0.64

Notes: The dependent variables are the mean and coefficient of variation for the fertilizer and fertilizer + lime distributions measured before farmers' interactions with Mahindi Master. The units are bags/acre.
$p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.
The results for belief updating are presented in Table B2

Hypothesis A2

We expect that farmers who update their beliefs about the returns to fertilizer and/or lime after playing the game will be more likely to change their fertilizer "order" after playing the game.

We test how updating beliefs is associated with farmers' fertilizer orders. To measure updating, we use the percent change of the mean and coefficient of variation of the subjective yields distribution for fertilizer and for fertilizer and lime. We find that updating beliefs decreases the value of the DAP order, but it increases for farmers that are currently using DAP or CAN. Updating beliefs is not correlated with the value of CAN orders. This order is only affected negatively if farmers are currently using DAP or CAN for updates in the subjective yield distribution for fertilizer. Lime orders increase as a result of updating the subjective yields distribution for fertilizer and lime.

Table D6: DAP Order Change and Fertilizer Use

	(1)	(2)	(3)	(4)	(5)	(6)
Uses DAP/CAN	-1.86	8.90	2.45	8.52	-2.41	1.74
	(10.44)	(8.05)	(10.47)	(8.09)	(10.41)	(10.46)
% change, mean (fertilizer)	-0.07**				-0.07**	
	(0.03)				(0.03)	
Uses DAP/CAN × % change, mean (fert.)	0.10**				0.10**	
	(0.04)				(0.04)	
% change, CV (fertilizer)		-1.94*			-1.94**	
		(1.02)			(0.93)	
Uses DAP/CAN × % change, CV (fert.)		1.78			1.81	
		(1.17)			(1.10)	
% change, mean (fertilizer + lime)			-0.03			-0.03
			(0.04)			(0.04)
Uses DAP/CAN × % change, mean (fert. + lime)			0.05			0.05
			(0.04)			(0.04)
% change, CV (fertilizer + lime)				-1.71		-1.69*
				(1.07)		(1.01)
Uses DAP/CAN × % change, CV (fert. + lime)				1.83		1.83
				(1.28)		(1.25)
Intercept	-24.81***	-33.85***	-28.97***	-33.70***	-24.14**	-28.37***
	(9.50)	(6.98)	(9.50)	(7.03)	(9.41)	(9.42)
R^2	0.04	0.04	0.02	0.03	0.06	0.04
N	166	166	166	166	166	166

Notes: Dependent variable is the percent change in the value of the DAP order
Standard errors in parentheses
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table D7: CAN Order Change and Fertilizer Use

	(1)	(2)	(3)	(4)	(5)	(6)
Uses DAP/CAN	1.99	-7.61	-1.64	-8.04	3.77	-0.60
	(19.27)	(15.50)	(19.18)	(15.51)	(19.43)	(19.30)
% change, mean (fertilizer)	0.05				0.05	
	(0.06)				(0.06)	
Uses DAP/CAN × % change, mean (fert.)	-0.12*				-0.12*	
	(0.07)				(0.07)	
% change, CV (fertilizer)		2.60			2.60	
		(2.17)			(2.12)	
Uses DAP/CAN × % change, CV (fert.)		-3.31			-3.39	
		(2.50)			(2.47)	
% change, mean (fertilizer + lime)			0.01			0.01
			(0.05)			(0.05)
Uses DAP/CAN × % change, mean (fert. + lime)			-0.08			-0.07
			(0.05)			(0.05)
% change, CV (fertilizer + lime)				2.07		2.06
				(2.27)		(2.26)
Uses DAP/CAN × % change, CV (fert. + lime)				-2.23		-2.32
				(2.77)		(2.76)
Intercept	8.84	15.22	13.73	15.10	7.86	12.90
	(15.63)	(11.28)	(15.28)	(11.33)	(15.58)	(15.23)
R^2	0.01	0.01	0.01	0.01	0.02	0.02
N	163	163	163	163	163	163

Notes: Dependent variable is the percent change in the value of the CAN order
Standard errors in parentheses
$^*p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table D8: Lime Order Change and Fertilizer Use

	(1)	(2)	(3)	(4)	(5)	(6)
Uses DAP/CAN	509.66***	445.77***	507.44***	465.08***	516.98***	533.25***
	(172.00)	(143.77)	(166.77)	(143.45)	(171.75)	(166.57)
% change, mean (fertilizer)	0.58				0.58	
	(0.48)				(0.48)	
Uses DAP/CAN × % change, mean (fert.)	-0.34				-0.34	
	(0.96)				(0.96)	
% change, CV (fertilizer)		7.63			7.68	
		(17.34)			(17.81)	
Uses DAP/CAN × % change, CV (fert.)		-13.06			-12.88	
		(26.34)			(26.75)	
% change, mean (fertilizer + lime)			0.73*			0.73*
			(0.42)			(0.40)
Uses DAP/CAN × % change, mean (fert. + lime)			-0.07			-0.08
			(0.49)			(0.47)
% change, CV (fertilizer + lime)				11.67		11.05
				(16.95)		(17.49)
Uses DAP/CAN × % change, CV (fert. + lime)				-42.63		-41.00
				(28.42)		(28.42)
Intercept	-58.46	22.57	-97.97	19.84	-60.72	-101.32
	(123.51)	(97.87)	(121.15)	(96.62)	(123.44)	(120.58)
R^2	0.05	0.05	0.07	0.06	0.05	0.08
N	175	175	175	175	175	175

Notes: Dependent variable is the percent change in the value of lime ordered (post minus pre).
Standard errors in parentheses
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Hypothesis A3

We expect farmers whose ex ante predicted returns to lime (low pH on their field) are high will allocate a greater share of their budget to lime after playing the game than those with low expected returns to lime.

The result for this hypothesis is presented in Table 7

Hypothesis A4a

We expect farmers for whom the sanity check fails to update their beliefs (and orders) less than those whose sanity check is accurate. We were unable to create the necessary variables to test this hypothesis.

Hypothesis A4b

We expect farmers who do not feel that the game reflects their reality to be less likely to update their beliefs.

Yields and planting sequences that are very different from reality does not have a significant effect on updating of the fertilizer distribution. Farmers who believe that yields in the game were much higher or higher than in real life update beliefs more than those who believe that yields in the game were the same as real life. Larger differences between the planting sequence in the game and real life are also associated with more updating. These results hold for updating of the fertilizer and lime distribution with the exception of yield perceptions, which are not longer significant.

Table D9: Updating of fertilizer mean beliefs

	(1)	(2)	(3)	(4)
Different yields (0/1)		22.50		
		(22.51)		
Planting sequence very different (0/1)				9.06
				(26.41)
Yield perceptions:				
Much higher or higher	42.67**			
	(18.43)			
Lower or much lower	34.68			
	(38.30)			
Planting sequence perceptions:				
Slightly different			60.77**	
			(27.14)	
Very different			59.61*	
			(33.16)	
Intercept	67.09***	90.09***	44.99*	95.54***
	(10.74)	(14.59)	(23.41)	(12.22)
R^2	0.01	0.01	0.02	0.00
N	175	175	175	175

Notes: Dependent variable is the absolute value of the percent change in the mean for the fertilizer distribution. The omitted category for yield perceptions is the same, and the omitted category for planting sequence perceptions is similar or somewhat similar.
Standard errors in parentheses
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table D10: Updating of fertilizer + lime mean beliefs

	(1)	(2)	(3)	(4)
Different yields (0/1)		24.47		
		(28.29)		
Planting sequence very different (0/1)				11.90
				(36.71)
Yield perceptions:				
Much higher or higher	31.61			
	(26.69)			
Lower or much lower	26.36			
	(61.19)			
Planting sequence perceptions:				
Slightly different			80.42***	
			(29.31)	
Very different			78.81*	
			(41.03)	
Intercept	100.02***	114.27***	52.64**	119.55***
	(20.39)	(20.65)	(23.65)	(15.15)
R^2	0.00	0.00	0.02	0.00
N	175	175	175	175

Notes: Dependent variable is the absolute value of the percent change in the mean for the fertilizer and lime distribution. The omitted category for yield perceptions is the same, and the omitted category for planting sequence perceptions is similar or somewhat similar.
Standard errors in parentheses
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Hypothesis B1

The number of rounds that farmers play in the game (pre-final round) is expected to be decreasing in the participant's measured confidence. We use a farming quiz to create a measure of confidence for farmers. Farmers were asked a series of ten questions related to general maize farming knowledge. After answering the questions, farmers were asked how many questions they believed that they answered correctly. We compare the number of questions that farmers answered correctly to the number of questions farmers believed that they answered correctly. If farmers got the same number of quiz questions correct as they reported believing they got correct within one question (i.e. if the farmer thought he answered 6 correct but only answered 5 correct), we consider the farmer appropriately confident. If the farmer believes he answered more questions correctly than he did and the difference is greater than one, we consider the farmer overconfident. If the farmer believes that he answered fewer questions correctly than he did and the difference is greater than one, we consider the farmer underconfident. We only have two farmers that are classified as underconfident in the sample. We create a dummy variable that equals one if the farmer is overconfident and zero otherwise.

We see that overconfident farmers play more rounds in the game (pre-final round). This result is only marginally significant.

Table D11: Number of Pre-Final Rounds and Confidence

	(1)	(2)	(3)	(4)
Percentage difference between No. quiz questions believed correct & actually correct	0.01*			
	(0.00)			
No. questions correct	0.07	0.04		
	(0.24)	(0.23)		
No. believed correct - No. correct		0.07		
		(0.11)		
Overconfident (0/1)			0.95*	
			(0.50)	
Farming self-doubt:				
Almost always / Very often				0.50
				(1.35)
Sometimes				0.08
				(0.88)
Rarely				0.51
				(0.94)
Intercept	8.65***	8.87***	8.63***	9.00***
	(0.86)	(0.92)	(0.37)	(0.81)
R^2	0.01	0.00	0.02	0.00
N	173	175	175	166

Notes: Dependent variable is the total number of rounds played in the game, excluding the sanity check round and the final founds. The omitted category for farming self-doubt is 'Never.'
Standard errors in parentheses
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Hypothesis B2a

We expect more confident farmers to be less likely to want to tweak their order (i.e. more likely to only play one final round).

From the results, we see that overconfidence is not correlated with the number of final rounds played. Farmer self-doubt is positively correlated with the number of final rounds played; indeed, having more frequent self-doubt increases the probability of playing multiple final rounds.

Table D12: Multiple Final Rounds and Confidence

	(1)	(2)	(3)	(4)
Percentage difference between No. quiz questions believed correct & actually correct	-0.00***			
	(0.00)			
No. questions correct	-0.09***	-0.07***		
	(0.03)	(0.03)		
No. believed correct - No. correct		-0.02*		
		(0.01)		
Overconfident (0/1)			-0.09	
			(0.06)	
Farming self-doubt:				
Almost always / Very often				0.17
				(0.11)
Sometimes				0.27***
				(0.05)
Rarely				0.06*
				(0.04)
Intercept	0.53***	0.44***	0.21***	-0.00
	(0.11)	(0.11)	(0.05)	(0.00)
R^2	0.07	0.05	0.01	0.08
N	173	175	175	166

Notes: Dependent variable is a dummy that equals one if the participant played multiple final rounds and zero otherwise. Regressions were run for the full sample. The omitted category for farming self-doubt is 'Never'.
Standard errors in parentheses
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table D13: Multiple Final Rounds and Confidence

	(1)	(2)	(3)	(4)
Percentage difference between No. quiz questions believed correct & actually correct	-0.00			
	(0.00)			
No. questions correct	-0.20***	-0.17***		
	(0.04)	(0.04)		
No. believed correct - No. correct		-0.01		
		(0.02)		
Overconfident (0/1)			0.00	
			(0.08)	
Farming self-doubt:				
Almost always / Very often				0.29
				(0.17)
Sometimes				0.35***
				(0.06)
Rarely				0.11*
				(0.06)
Intercept	0.94***	0.81***	0.23***	-0.00
	(0.16)	(0.15)	(0.05)	(.)
R^2	0.20	0.16	0.00	0.10
N	119	121	121	116

Notes: Dependent variable is a dummy that equals one if the participant played multiple final rounds and zero otherwise. Regressions were run for the sample of farmers that played any number of final rounds. The omitted category for farming self-doubt is Never.
Standard errors in parentheses
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Hypothesis B2b

Conditional on tweaking their "final order" once, we expect the number final rounds played (i.e. the number of tweaks to the final order) to be decreasing in measured farmer confidence.

There are too few farmers that played multiple final rounds to conduct this analysis.

Hypothesis B3

The number of rounds played by farmers in the game (pre-final round) is increasing in risk aversion.

Risk aversion does not appear to significantly affect the number of pre-final rounds played. The point estimates on risk aversion are positive as expected but not significant. Farmers who are risk loving play fewer rounds.

Table D14: Number of Pre-Final Rounds and Risk

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Risk averse (Lottery)	-0.18							
	(0.62)							
Chose safe option for Q1		0.33						
		(0.54)						
Risk averse (General)			1.15					
			(1.04)					
Risk averse (Farming)				1.50				
				(1.06)				
Risk loving (General)					-0.42			
					(0.45)			
Risk loving (Farming)						-0.77*		
						(0.46)		
Risk measure with Q1=2							-0.38	
							(0.72)	
Risk measure with Q1=3							-0.63	
							(0.85)	
Risk measure with Q1=4							-0.36	
							(0.72)	
Risk measure with Q1=5							0.11	
							(0.70)	
Risk measure=2								-0.25
								(0.91)
Risk measure=3								0.02
								(0.78)
Risk measure=4								0.49
								(0.76)
Intercept	9.25***	9.12***	9.09***	9.06***	9.30***	9.40***	9.44***	9.07***
	(0.28)	(0.30)	(0.26)	(0.26)	(0.31)	(0.32)	(0.46)	(0.55)
R^2	0.00	0.00	0.01	0.02	0.00	0.01	0.01	0.01
N	175	175	175	175	175	175	175	130

Notes: Dependent variable is the number of pre-final rounds played.
Standard errors in parentheses
$^*p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Hypothesis B4

The number of low-rainfall weather scenarios chosen is increasing in risk aversion.

Risk aversion does not appear to affect the number of low rainfall scenarios chosen. Risk aversion, as measured by the survey questions, is associated with fewer low rainfall scenarios chosen, but the point estimate is -.05 rounds, which is very close to zero.

Table D15: Low-Rainfall Scenarios and Risk

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Risk averse (Lottery)	-0.02							
	(0.03)							
Rounds with Random Poor Rain	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Chose safe option for Q1		0.00						
		(0.04)						
Risk averse (General)			-0.05***					
			(0.02)					
Risk averse (Farming)				-0.05***				
				(0.02)				
Risk loving (General)					0.01			
					(0.04)			
Risk loving (Farming)						0.00		
						(0.03)		
Risk measure with Q1=2							-0.02	
							(0.04)	
Risk measure with Q1=3							-0.01	
							(0.05)	
Risk measure with Q1=4							0.00	
							(0.05)	
Risk measure with Q1=5							0.02	
							(0.06)	
Risk measure=2								0.01
								(0.04)
Risk measure=3								0.02
								(0.05)
Risk measure=4								0.05
								(0.05)
Intercept	0.03	0.02	0.03	0.03	0.02	0.02	0.03	0.00
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)
R^2	0.01	0.00	0.01	0.01	0.00	0.00	0.01	0.01
N	175	175	175	175	175	175	175	130

Notes: Dependent variable is the number of rounds farmers choose to play with low rainfall.
Standard errors in parentheses
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Hypothesis B5

The share of rounds played with unfamiliar inputs is increasing in risk aversion.

We test the effect of risk aversion on the share of rounds played with unfamiliar inputs. Most point estimates are positive but not significant. Choosing the safe option for the practice question in the lottery is associated with an increase of .10 in the share of rounds played with CAN, and being risk loving is associated with an increase of .12 in the share of rounds with CAN. We see larger increases for risk loving farmers for the share of rounds played with lime.

Table D16: Share of Rounds with CAN

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Risk averse (Lottery)	-0.05							
	(0.05)							
Uses CAN	0.07*	0.07*	0.07*	0.07*	0.06	0.06	0.08*	0.10*
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)
Chose safe option for Q1		0.10***						
		(0.04)						
Risk averse (General)			0.05					
			(0.07)					
Risk averse (Farming)				0.08				
				(0.06)				
Risk loving (General)					0.12***			
					(0.03)			
Risk loving (Farming)						0.06		
						(0.04)		
Risk measure with Q1=2							-0.12**	
							(0.05)	
Risk measure with Q1=3							-0.09	
							(0.06)	
Risk measure with Q1=4							-0.01	
							(0.06)	
Risk measure with Q1=5							-0.19***	
							(0.05)	
Risk measure=2								0.03
								(0.07)
Risk measure=3								0.11
								(0.07)
Risk measure=4								-0.07
								(0.06)
Intercept	0.79***	0.75***	0.77***	0.77***	0.75***	0.77***	0.85***	0.73***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.06)
r2	0.02	0.05	0.02	0.02	0.05	0.02	0.08	0.07
N	175	175	175	175	175	175	175	130

Notes: Dependent variable is the number of rounds played with a positive amount of CAN, where CAN was a choice. $p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table D17: Share of Rounds with Lime

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Risk averse (Lottery)	-0.00							
	(0.06)							
Chose safe option for Q1		0.02						
		(0.05)						
Risk averse (General)			-0.12					
			(0.09)					
Risk averse (Farming)				0.02				
				(0.09)				
Risk loving (General)					0.18***			
					(0.05)			
Risk loving (Farming)						0.16***		
						(0.05)		
Risk measure with Q1=2							-0.02	
							(0.07)	
Risk measure with Q1=3							-0.10	
							(0.07)	
Risk measure with Q1=4							-0.05	
							(0.07)	
Risk measure with Q1=5							0.09	
							(0.07)	
Risk measure=2								-0.08
								(0.08)
Risk measure=3								-0.03
								(0.08)
Risk measure=4								0.11
								(0.07)
Intercept	0.60***	0.59***	0.61***	0.59***	0.55***	0.55***	0.61***	0.59***
	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)	(0.05)	(0.05)
R^2	0.00	0.00	0.01	0.00	0.06	0.05	0.04	0.05
N	175	175	175	175	175	175	175	130

Notes: Dependent variable is the share of rounds played with a positive amount of lime, where lime was a choice.
Standard errors in parentheses
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Hypothesis B6

The number of times the final fertilizer order is changed is increasing in risk aversion.

The results suggest that the number of times the final fertilizer order is changed is decreasing in risk aversion.

Table D18: Number of Final Rounds

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Risk averse (Lottery)	0.05							
	(0.30)							
Chose safe option for Q1		-0.46***						
		(0.17)						
Risk averse (General)			-0.59***					
			(0.19)					
Risk averse (Farming)				-0.75***				
				(0.17)				
Risk loving (General)					-0.19			
					(0.19)			
Risk loving (Farming)						-0.31*		
						(0.18)		
Risk measure with Q1=2							0.38	
							(0.30)	
Risk measure with Q1=3							0.75***	
							(0.28)	
Risk measure with Q1=4							0.31	
							(0.19)	
Risk measure with Q1=5							0.39	
							(0.32)	
Risk measure=2								0.38
								(0.38)
Risk measure=3								-0.07
								(0.32)
Risk measure=4								0.02
								(0.41)
Intercept	1.11***	1.24***	1.18***	1.19***	1.16***	1.20***	0.78***	1.16***
	(0.11)	(0.14)	(0.12)	(0.11)	(0.13)	(0.14)	(0.10)	(0.28)
R^2	0.00	0.02	0.02	0.02	0.00	0.01	0.03	0.01
N	175	175	175	175	175	175	175	130

Notes: Dependent variable is the number of final rounds played for the full sample of participants.
$p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Hypothesis C1

We expect farmers who play fewer rounds of the game to be less likely to update their beliefs and fertilizer orders.

The total number of rounds played does not have a significant effect on the updating of the subjective yields distribution. The results show that the total number of rounds played increases the change in the order, in terms of value, for DAP and CAN and decreases the share of the order allocated to DAP.

Table D19: Total Rounds and Updating

	(1)	(2)	(3)	(4)	(5)	(6)
Total number of rounds	-3.94	-5.35	3.68***	5.73**	17.16	-2.53***
	(3.12)	(4.32)	(0.86)	(2.48)	(18.47)	(0.61)
Share allocated to DAP (pre)						0.51***
						(0.12)
Intercept	140.25***	180.07***	4.61	1.10	461.07**	40.16***
	(38.69)	(53.70)	(8.96)	(23.01)	(203.04)	(8.18)
R^2	0.01	0.01	0.09	0.05	0.00	0.14
N	175	175	166	163	175	169

Notes: Dependent variable for each column is as follows: (1) absolute value of the percentage change in the mean of the subjective yields distribution with fertilizer, (2) absolute value of the percentage change in the mean of the subjective yields distribution with fertilizer and lime, (3) absolute value of the percentage change in DAP value, (4) absolute value of the percentage change in CAN value, (5) absolute value of the change in lime value, and (6) DAP Share (Post).
Standard errors in parentheses
$^*p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Hypothesis C2

We expect farmers with a higher stated farming ability to be less likely to update their beliefs and fertilizer orders because they will be less responsive to information.

The results for these hypotheses are presented in Tables **??** and 9.

Hypothesis D1

It is not clear how confidence affects the share of rounds played with unfamiliar inputs.

The results for this hypothesis is presented in Tables 11-13

Hypothesis D2

It is not clear how confidence affects the proportion of rounds spent in the regular game versus the final round.

Overconfident farmers spend more rounds on the regular game relative to the final round.

Table D20: Share of Pre-Final Rounds

	(1)	(2)	(3)	(4)
Percentage difference between No. quiz questions believed correct & actually correct	0.00*			
	(0.00)			
No. questions correct	0.01	0.01		
	(0.01)	(0.01)		
No. believed correct - No. correct		0.01**		
		(0.01)		
Overconfident (0/1)			0.07***	
			(0.03)	
Farming self-doubt:				
Almost always / Very often				-0.08
				(0.06)
Sometimes				-0.10***
				(0.02)
Rarely				-0.02
				(0.03)
Intercept	0.82***	0.80***	0.83***	0.93***
	(0.05)	(0.05)	(0.02)	(0.01)
R^2	0.01	0.02	0.04	0.05
N	173	175	175	166

Notes: Dependent variable is the share of pre-final rounds of the total number of rounds played. The omitted category for farming self-doubt is 'Never.'
Standard errors in parentheses
$*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

Hypothesis D3

It is unclear how risk aversion affect the probability of changing the final fertilizer order.

The probability of changing the final order appears to be decreasing in risk aversion and is negatively correlated with being risk loving as measured through the survey questions.

Table D21: Multiple Final Rounds

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Risk averse (Lottery)	-0.04							
	(0.06)							
Chose safe option for Q1		-0.10*						
		(0.05)						
Risk averse (General)			-0.11*					
			(0.06)					
Risk averse (Farming)				-0.18***				
				(0.03)				
Risk loving (General)					-0.08			
					(0.06)			
Risk loving (Farming)						-0.10*		
						(0.05)		
Risk measure with Q1=2							0.04	
							(0.07)	
Risk measure with Q1=3							0.16*	
							(0.09)	
Risk measure with Q1=4							0.04	
							(0.08)	
Risk measure with Q1=5							0.15*	
							(0.09)	
Risk measure=2								0.12
								(0.09)
Risk measure=3								-0.01
								(0.09)
Risk measure=4								0.11
								(0.10)
Intercept	0.17***	0.18***	0.17***	0.18***	0.18***	0.19***	0.09**	0.13**
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.05)
R^2	0.00	0.01	0.01	0.02	0.01	0.01	0.03	0.02
N	175	175	175	175	175	175	175	130

Notes: Dependent variable is a dummy variable that equals one if multiple final rounds were played and zero otherwise.
Standard errors in parentheses
$*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

Hypothesis D4

It is unclear how risk aversion affects the proportion of rounds spent in the regular game versus the final round.

The proportion of rounds spent in the regular game appears to be increasing in risk aversion.

Table D22: Share of Pre-Final Rounds

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Risk averse (Lottery)	0.00							
	(0.03)							
Chose safe option for Q1		0.05**						
		(0.02)						
Risk averse (General)			0.08***					
			(0.02)					
Risk averse (Farming)				0.09***				
				(0.02)				
Risk loving (General)					0.03			
					(0.02)			
Risk loving (Farming)						0.04*		
						(0.02)		
Risk measure with Q1=2							-0.04	
							(0.03)	
Risk measure with Q1=3							-0.10**	
							(0.04)	
Risk measure with Q1=4							-0.04	
							(0.03)	
Risk measure with Q1=5							-0.03	
							(0.04)	
Risk measure=2								-0.06
								(0.05)
Risk measure=3								-0.00
								(0.04)
Risk measure=4								0.00
								(0.05)
Intercept	0.87***	0.86***	0.86***	0.86***	0.86***	0.86***	0.91***	0.87***
	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.03)
R^2	0.00	0.02	0.02	0.02	0.01	0.01	0.03	0.02
N	175	175	175	175	175	175	175	130

Notes: Dependent variable is the share of pre-final rounds of the total number of rounds played.
Standard errors in parentheses
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Hypothesis D5

We will also evaluate if farmers update beliefs differently along the following dimensions: risk preferences, confidence, subjective confidence.

Risk does not have a significant effect on the updating of beliefs.

Table B1 shows the results for confidence

Table D23: Updating and Risk

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Risk averse (Lottery)	8.93							
	(29.18)							
Chose safe option for Q1		13.38						
		(26.95)						
Risk measure with Q1=2			-3.31					
			(36.05)					
Risk measure with Q1=3			12.64					
			(32.40)					
Risk measure with Q1=4			-33.91					
			(35.90)					
Risk measure with Q1=5			-40.73					
			(30.85)					
Risk measure=2				15.95				
				(34.61)				
Risk measure=3				-30.60				
				(37.91)				
Risk measure=4				-37.42				
				(33.16)				
Risk averse (General)					10.21			
					(35.41)			
Risk averse (Farming)						56.18		
						(60.83)		
Risk loving (General)							-21.68	
							(24.08)	
Risk loving (Farming)								-17.25
								(23.40)
Intercept	96.15***	95.00***	108.38***	105.07***	97.45***	93.31***	103.52***	102.98***
	(11.91)	(12.60)	(24.03)	(26.91)	(11.87)	(10.66)	(13.23)	(13.56)
R^2	0.00	0.00	0.02	0.02	0.00	0.01	0.00	0.00
N	175	175	175	130	175	175	175	175

Notes: Dependent variable is the absolute value of the percent change in mean of the subjective yields distribution for fertilizer
Standard errors in parentheses
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table D24: Updating and Risk

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8
Risk averse (Lottery)	35.87							
	(43.27)							
Chose safe option for Q1		3.81						
		(33.31)						
Risk measure with Q1=2			23.82					
			(50.25)					
Risk measure with Q1=3			26.16					
			(40.77)					
Risk measure with Q1=4			-37.31					
			(39.72)					
Risk measure with Q1=5			-52.00					
			(34.42)					
Risk measure=2				2.34				
				(50.53)				
Risk measure=3				-61.13				
				(49.69)				
Risk measure=4				-75.82*				
				(45.55)				
Risk averse (General)					33.53			
					(51.83)			
Risk averse (Farming)						85.33		
						(99.33)		
Risk loving (General)							-20.49	
							(30.87)	
Risk loving (Farming)								-21.
								(29.
Intercept	114.13***	122.38***	126.18***	150.00***	120.10***	115.56***	128.16***	129.03
	(14.03)	(17.42)	(28.64)	(41.34)	(15.55)	(12.85)	(17.78)	(18.
R^2	0.01	0.00	0.02	0.03	0.00	0.02	0.00	0.0
N	175	175	175	130	175	175	175	17

Notes: Dependent variable is the absolute value of the percent change in mean of the subjective yields distribution for fertilize
lime.
Standard errors in parentheses
$*p < 0.10$, $**p < 0.05$, $***p < 0.01$.