Journal of Economic Behavior and Organization 189 (2021) 28–50

Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

journal homepage: www.elsevier.com/locate/jebo

Learning by (virtually) doing: Experimentation and belief updating in smallholder agriculture[☆]

Emilia Tjernström[a,*], Travis J. Lybbert[b], Rachel Frattarola Hernández[c,1], Juan Sebastian Correa[b]

[a] University of Sydney, Australia
[b] University of California, Davis, USA
[c] Office of Management and Budget, USA

ARTICLE INFO

Article history:
Received 12 April 2019
Revised 26 February 2021
Accepted 1 March 2021
Available online 7 July 2021

JEL classification:
D84
O12
O13

Keywords:
Belief updating
Responsiveness to information
Gamification
Technology adoption

ABSTRACT

In much of sub-Saharan Africa, soil quality heterogeneity hampers farmer learning about the returns to different inputs. This may help explain why relatively few farmers in the region use improved inputs. We study how Kenyan farmers respond to an interactive app that enables them to discover the returns to different inputs on a virtual farm that is calibrated to resemble their own. Farmers update both their beliefs and their behaviors after engaging with the virtual learning app. We measure beliefs by eliciting probability distributions and use an incentive-compatible experiment to measure behavior change. The experiment gave participants an input budget that they could allocate across farm inputs. After playing several virtual seasons on the app, they could update these allocations. Farmers revise their input allocations along several dimensions after the virtual learning experience. To support our interpretation that these adjustments stem 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.

☆ We are grateful to Matt Kimball and Tyler Lybbert for their help animating and programming *MahindiMaster* as well as to David Cammarano and Christopher Kucharik who provided generous advice for the underlying crop models. We thank two anonymous referees, Jen Alix-Garcia, Laura Schechter, and seminar audiences at IFPRI, the University of California – Davis, the University of Wisconsin – Madison, Oklahoma State University, Rutgers University, the 2018 AAEA Annual Meeting, the USAID Global Development Lab, and the 2018 ICABR-World Bank Conference 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).

* Corresponding author.
E-mail addresses: emilia.tjernstrom@sydney.edu.au (E. Tjernström), tlybbert@ucdavis.edu (T.J. Lybbert).

1 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.

https://doi.org/10.1016/j.jebo.2021.03.001
0167-2681/

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).

This paper aims to increase our understanding of how farmers update and act upon their beliefs about new and existing technologies. We do so by studying how Kenyan maize farmers respond to an app that lets them experiment with agricultural inputs on virtual plots calibrated to match their own. Many pre-requisites precede a household's decision to adopt a new technology; nevertheless, perhaps the most essential requirement is knowledge—knowledge about the technology's existence as well as its use and profitability. The literature examines a vast array of constraints that prevent households from adopting, even if these pre-conditions are fulfilled, but our study design allows us to bypass many of them in order to focus on information.[2] Researchers have examined many types of information interventions, but we still know relatively little about how farmers form and update their beliefs about the benefits of new products.

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. This suggests that the learning is productive, since farmers in our sample generally believe that yields are lower than observed historical averages. Furthermore, using an incentive-compatible method that allows farmers to allocate researcher-provided funds toward agricultural inputs, we find that farmers who were *ex ante* most likely to benefit from a new, unfamiliar input respond most to information about it. Our enumerators were blind to participants' predicted returns to the new technology, suggesting that this result is not driven by demand characteristics.

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 has the potential to hamper farmer learning in multiple ways: first, regional fertilizer recommendations will rarely be accurate for a given farmer, which reduces the efficacy of information diffusion by central agencies. 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).

We test an interactive information intervention designed specifically to overcome these learning limitations. Drawing on design insights from the gamification literature, we designed an accessible app that gives farmers proxy-observations for various input combinations using animated output from a crop model. Participants use the app, which we call *MahindiMaster* (*Mahindi* means maize in *kiSwahili*), on a researcher-provided tablet.

MahindiMaster simulates yields based on farmers' choices from a menu of input options; importantly, these yields are individually-tailored using crop model outputs that leverage plot-level soil samples and historical weather data. 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, reduces 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 plants from absorbing soil nutrients (including the nutrients in 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.

We study how the app influences farmers' beliefs and behavior. Given the uncertainty surrounding fertilizer returns, we expect farmers to have some beliefs about these returns that can be represented by an assignment of probabilities to potential outcomes (Delavande, 2014). We elicit these subjective 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 then go a step further and analyze whether farmers' interactions with *MahindiMaster* induced them to change their behavior. We achieve this by allowing farmers to allocate a researcher-provided budget across three inputs (DAP, CAN, and lime) during a pre-intervention survey and then giving them a chance to revise their allocation after they interact 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 on changing

[2] 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 can control for risk aversion in our results.

[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.

the information flow that users receive (Walz and Deterding, 2015). We use gamification primarily to provide new information to farmers and to display this information in an interface that we hope eases comprehension. Similarly, virtual reality enables researchers to represent tasks in a less abstract and more naturalistic manner; in other words, a manner that better reflects individuals' decision-making reality. Fiore et al. (2009) find that participants' subjective beliefs more closely align with actual risks in a virtual experiment with virtual reality technology.

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; however, the type of information, targeting, and the characteristics of those who receive it all matter. Dupas (2011) finds that a national HIV/AIDS curriculum in Kenya had no effect on pregnancy or sexual partners' age, whereas 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 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 farmers to value the information in *MahindiMaster* more if it is relevant to their context. Information about lime will be less relevant for farmers whose soils are not acidic. Therefore, we expect households with acidic soils (whose soils are unproductive without added lime) to react more strongly to the intervention. However, existing empirical evidence shows 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. Similarly, Ambuehl and Li (2018) 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 many 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. This difference arises because overconfidence is less beneficial in a stable environment than 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.

Adding another dimension to the mix, Barham et al. (2018) study how receptiveness to advice and cognitive ability influence US farmers' technology adoption behaviors. They find that the two interact: receptiveness to advice can either

[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.

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 1800 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 Fig. 1 shows the main agricultural seasons in Kenya. The rainfall distribution in most of Kenya is bimodal, with a main season beginning in March and a short season between October and December–January. For both the RCT and 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, enabling participating farmers to apply their chosen inputs in the upcoming main season. The bottom panel of Fig. 1 presents a timeline of the original intervention, the associated household panel data collection, and the app-based study. Throughout the paper, we refer to data coming from the RCT panel baseline survey as "baseline data." We call the data elicited shortly before (after) farmers interacted with the app as pre-game (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 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.

2.3. MahindiMaster

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

[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.

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

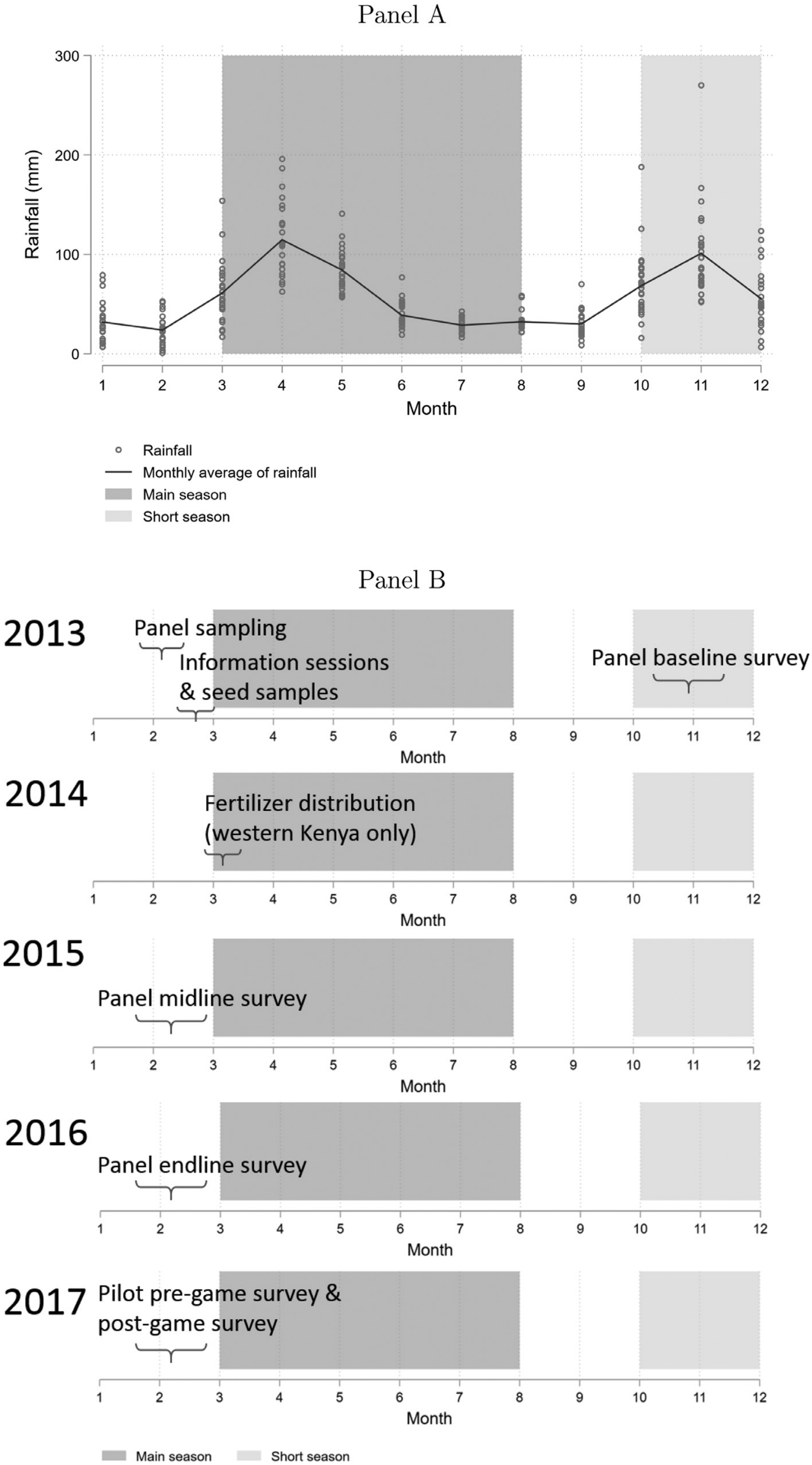


Fig. 1. Rainfall distribution in sample area and data collection timeline.

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.

MahindiMaster is a Unity-based Android application that communicates the simulation results to farmers. Once a user chooses a farmer ID and how much of the available inputs to apply, the app retrieves the simulation results and displays the relevant animations. 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 the type of rainfall scenario they wish to see. 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 that at first limited farmers' choices to how much of to apply of the most familiar input, DAP. 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 25 kg-intervals ranging from 0 to 125 kg. Recommended lime application rates are much larger, so farmers selected lime in 250 kg intervals ranging from 0 to 2000 kg. Farmers had to play a minimum of nine fictional 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 a "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.

Fig. 2 shows screenshots from *MahindiMaster*. Fig. 2 a shows the selection screen when only DAP is available, Fig. 2 b shows the animated hand applying fertilizer (which gets deposited into a hole, made with a stick on the previous screen), Fig. 2 c shows the same hand depositing the seeds into the hole, and Fig. 2 d 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 90 kg bag of maize (a common unit for measuring harvests).

2.4. Data sources

2.4.1. 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.

2.4.2. 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]

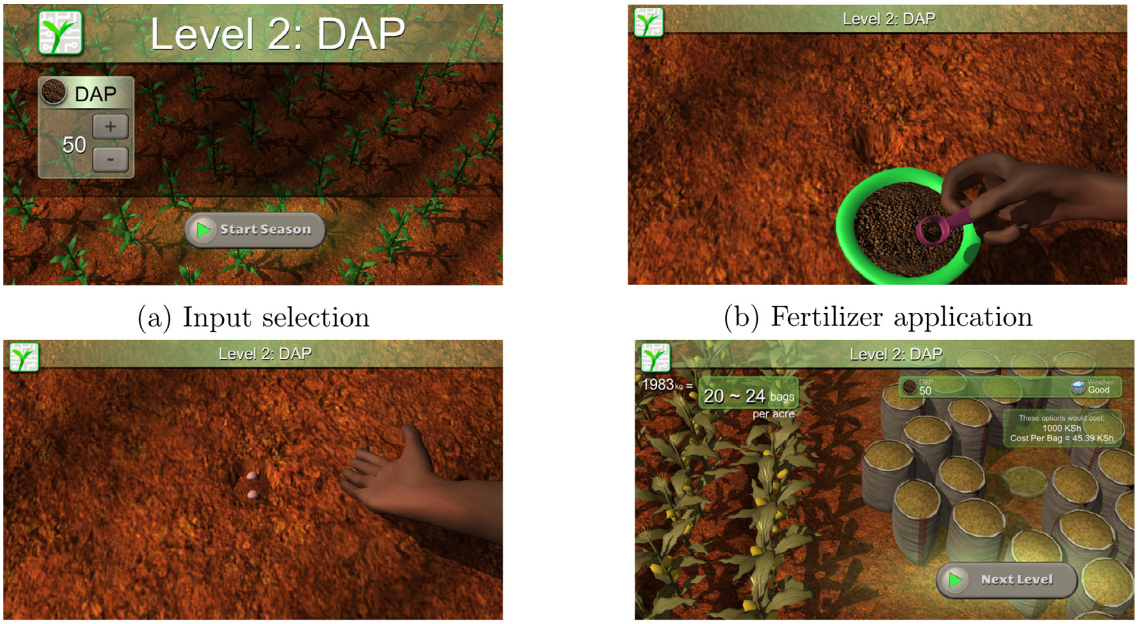
2.4.3. Pre- and post-game data

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

[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 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,b), and Delavande (2014) for an overview of eliciting subjective expectations in developing countries.



(a) Input selection

(b) Fertilizer application

(c) Seed planting

(d) Harvest summary screen

Fig. 2. Screenshots from *MahindiMaster* gameplay.

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 fertilizer-alone distribution.

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 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

[10] For respondents who were unable to state their subjective expectations about yields under fertilizer-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's (2008)) "overestimation" definition of confidence.

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 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]

2.4.4. 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 to plant an experimental plot of 10$\times$10 m.[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, 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.

Fig. 3 shows that many participants played more rounds than was required. 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

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

[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.

Table 1
Baseline characteristics and game interactions.

	Mean	Std deviation	Min	Max
Panel A: Farmer and plot characteristics				
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=do not know	4.87	2.04	0	9
Overconfident (0/1)	0.59	0.49	0	1
Panel B: Game play				
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[a]	0.80	0.26	0	1
Share of rounds with lime[a]	0.60	0.31	0	1
Share of rounds with no fertilizer	0.046	0.062	0	0.25
Random rainfall scenario[b]	2.01	0.30	1.25	2.80
Chosen rainfall scenario[b]	2.70	0.44	1	3
Rounds played	10.6	2.93	1	19
Observations	175			

[a] Conditional on the input being available in a given round.
[b] Rainfall scenarios: 1=poor, 2=normal, 3=good.

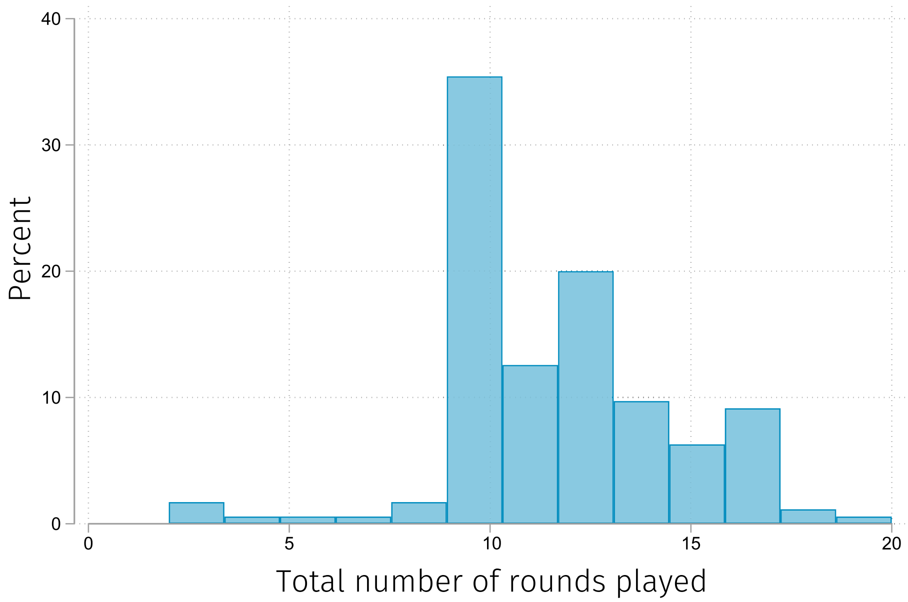


Fig. 3. Number of rounds that experimental participants played.

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

Table 2
Descriptive statistics on yield beliefs.

	Observations	Mean	s.d.	Min	Max
Panel A: Bounds on belief distribution[a]					
Pre-game					
No fertilizer					
Lower bound	175	125.5	129.5	0	720
Upper bound	175	408.0	283.7	9	1440
Fertilizer					
Lower bound	121	269.4	230.3	0	1200
Upper bound	121	788.1	446.3	162	2185.7
Fertilizer and lime					
Lower bound	4	626.3	451.1	135	1200
Upper bound	4	1605	942.5	990	3000
Post-game					
Fertilizer					
Lower bound	175	309.7	285.3	0	1440
Upper bound	175	1078.6	681.1	144	3420
Fertilizer and lime					
Lower bound	60	435.5	348.1	0	1500
Upper bound	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 [b]					
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

[a] All expressed in kg/acre.
[b] Out of 5 bins.

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

inputs, and that many farmers played beyond the required rounds, suggesting interest in the game and in the information presented.

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 these readers that our 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 4 shows how yield beliefs correlate with soil characteristics and past yield reports. Columns (1) and (2) show correlates of no-fertilizer yields, (3) and (4) examine correlates of with-fertilizer subjective yield distributions, and the last

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)
Share of seasons w/ 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 $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

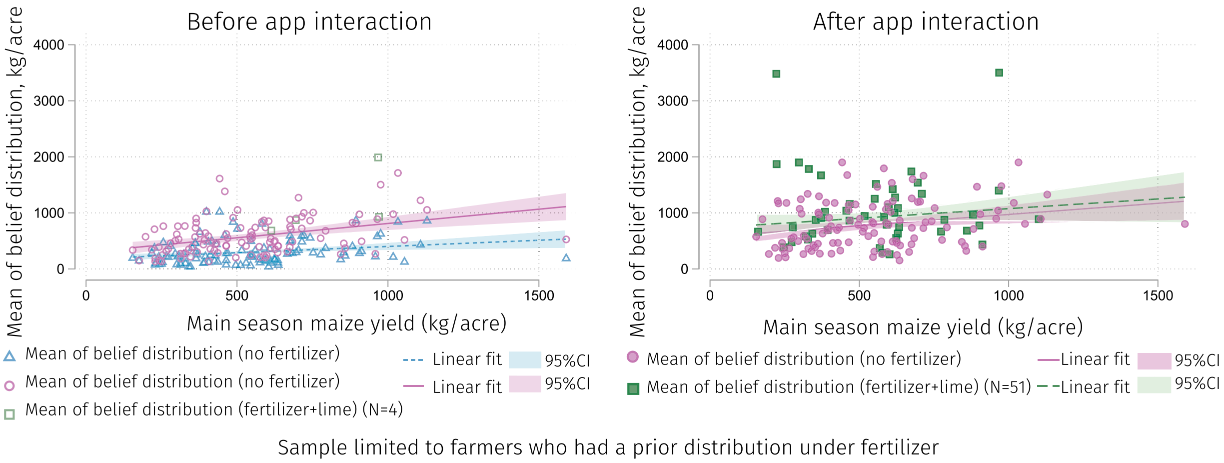


Fig. 4. Subjective yield beliefs vs. main season yields.

four columns follow the same pattern 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.

Fig. 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.

Fig. 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 distinguishable at higher pH levels.

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 *MahindiMaster* 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

[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).

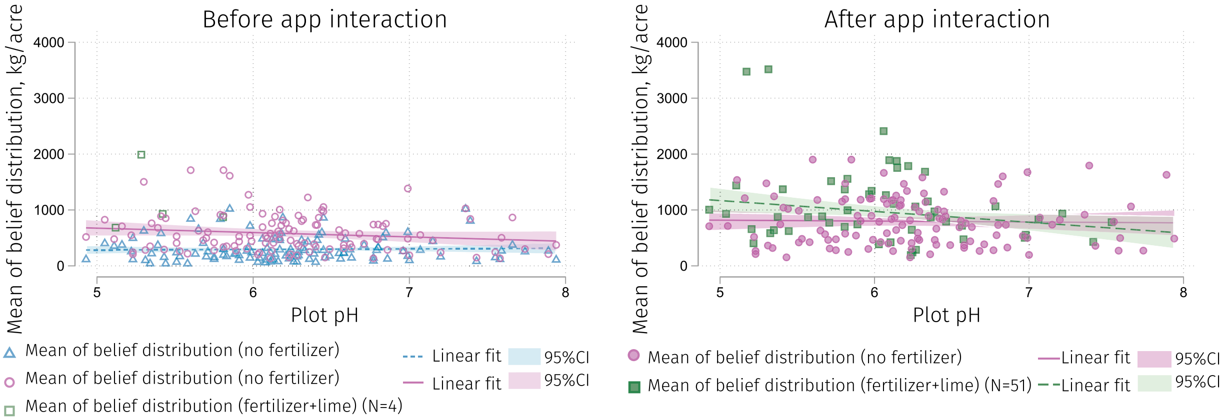


Fig. 5. Subjective yield beliefs vs. maize plot soil pH.

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 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.

Table 5
Fertilizer orders.

	Mean (pre)	Mean (post)	t-statistic[a]
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

[a] Test statistic based on standard errors that are clustered at the village level.

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 believe that our interpretation of the results is defensible.

To evaluate the effect of the intervention, our baseline approach is to regress post-treatment outcomes on pre-treatment values of the outcome variable, controlling for farmer traits. Specifically, we estimate the following equation using Ordinary Least Squares (OLS):

$$\text{Post}_i = \alpha_1 + \beta_1 \text{Pre}_i + \gamma_1 \text{Trait}_i + \varepsilon_i \tag{1}$$

In addition, given that the returns to lime differ across pH levels, we consider a quasi-experiment where we treat farmers with low soil pH as the treatment group within a difference-in-differences framework. This allows us to analyze the effect of the treatment on lime orders for low-pH farmers, using higher-pH farmers as a control group. 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 specifications to analyze the effect of soil pH. In the first (Eq. (2)), we assume that the outcome variable has a quadratic relationship with soil pH. The second specification (Eq. (3)) is more flexible: we include indicator variables for different ranges of pH.

$$\text{Post}_i = \alpha_2 + \beta_2 \text{Pre}_i + \delta_1 \text{pH}_i + \delta_2 \text{pH}_i^2 + u_i \tag{2}$$

$$\text{Post}_i = \alpha_3 + \beta_3 \text{Pre}_i + \phi_k \sum_{k=1}^{5} \text{pH}_i^k + \epsilon_i \tag{3}$$

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

We further analyze whether changes in beliefs (post-game less pre-game) about returns to fertilizer use affect participants' subsequent decisions of how to allocate their budget across the different inputs. We analyze the effect of belief updating as a consequence of playing the game using the following model (and a few variations thereof):

$$(\text{Post}_i - \text{Pre}_i) = \alpha_4 + \beta_4 \text{Fert}_i + \zeta_1 \Delta\text{Belief}_i + \zeta_2 \text{Fert} \times \Delta\text{Belief}_i + \varepsilon_i \tag{4}$$

where Fert_i is a dummy that indicates whether farmer i has used either DAP or CAN in the past and ΔBelief_i is the percentage change in farmer i's beliefs, measured either by the mean or the coefficient of variation.[17]

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 5 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. Some farmers ordered lime before interacting with the game—perhaps in order to experiment with it. The fact that

[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.

[17] We thank an anonymous referee for encouraging us to include this analysis.

Table 6
Subjective expectations pre- and post-game.

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

[a] Test statistic based on standard errors that are clustered at the village level.

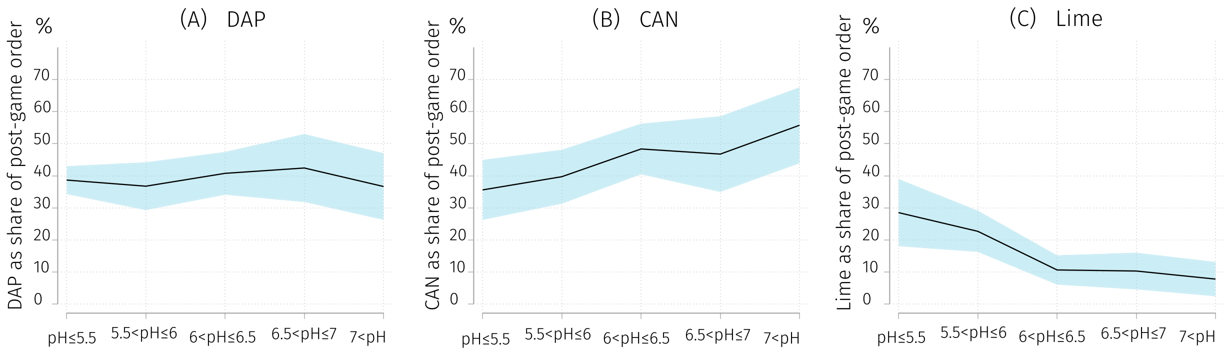


Fig. 6. Average share of post-game order allocated to inputs, by soil pH.

some of them revised their choices and removed lime from their order after interacting with the app is consistent with the app encouraging learning.

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 6). 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.

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 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.

Fig. 6 plots the predictive margins (marginal means) of the share of post-orders allocated to the different inputs for different bins of soil pH.[18] 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 particular 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

[18] 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.

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). We report q-values in square brackets. 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. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 8
Lime order change and fertilizer use.

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

Notes: Dependent variable: the change in the amount of lime that a participant allocates to lime (post less pre). Standard errors in parentheses $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 9
Orders and farming ability.

	DAP (post-pre)	CAN (post-pre)	Lime (post-pre)	DAP (post-pre)	CAN (post-pre)	Lime (post-pre)
No. correct questions	−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=do not 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 $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 10
Belief revisions.

	Fertilizer			Fertilizer + lime		
	%Δ Mean	%Δ Mean	%Δ Mean	%Δ Mean	%Δ Mean	%Δ Mean
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 = Do not 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. "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)			
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

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 11
Effect of experimenting with lime on belief updating.

First stage		
Rounds w/ poor rain[a]	0.870***	0.870***
	(7.23)	(7.23)
Rounds w/ fair rain[a]	0.771***	0.771***
	(7.72)	(7.72)
Rounds w/ good rain[a]	0.829***	0.829***
	(7.93)	(7.93)
pH	−0.0972***	−0.0972***
	(−3.75)	(−3.75)
Intercept	0.343*	0.343*
	(1.81)	(1.81)
Observations	175	175
Second stage		
Rounds w/ lime[a]	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
F-stat[b]	35.60	35.60

[a] The share of rounds with a given rainfall scenario.
[b] Kleibergen–Paap F-statistic. 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, fair, and bad rainfall draws (during rounds when they could not choose the rainfall scenario). Standard errors in parentheses $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

necessarily expect the same kind of gradient for DAP or CAN, and panels A and B of Fig. 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 after the intervention slightly increases with pH (panel B). This latter gradient may suggest that low-pH farmers substitute away from CAN to lime.

Table 8 shows estimates of the correlations between farmers' belief-updating and their lime orders. The dependent variable in all columns is the value of the lime order post-game minus the pre-game order. The results suggest that participants who update their beliefs about the returns to lime to a greater extent—as measured by changes in the first moment of their belief distribution—increase the amount of lime that they order by a larger amount. Columns (4) and (6) show that this result only holds for updates to farmers' beliefs about the returns to fertilizer *and* lime, not for fertilizer alone.[19]. These results seem reasonable, but the estimated coefficients on belief changes is relatively small; therefore, a large change in beliefs may be needed to substantially change behavior.

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 in a general sense, but lack plot-specific information on returns, then we might expect them to be better equipped to update in response to the information.

Table 9 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 10 shows the results from this analysis, with the dependent variable being either the per-

[19] We report the results for DAP and CAN as outcome variables in Appendix D.

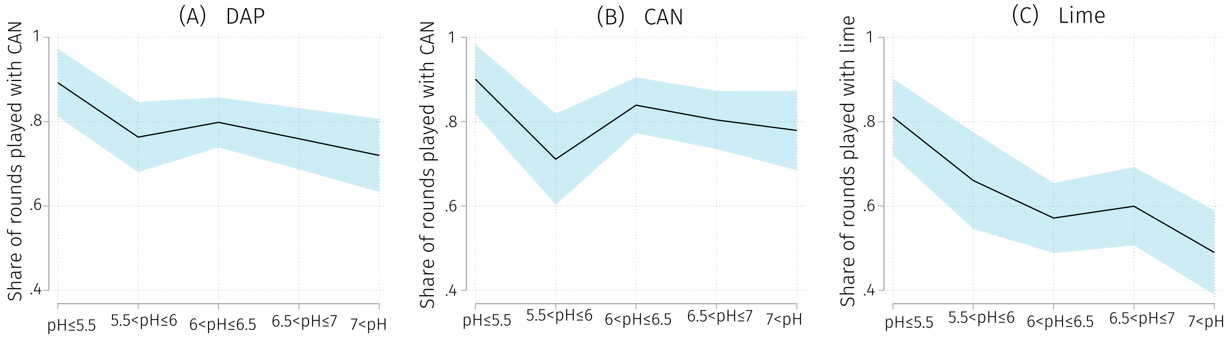


Fig. 7. Average share of rounds played with non-zero input amounts, by soil pH.

Table 12
Share of rounds with DAP.

	Share	Share	Share
(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 $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

centage change in the mean subjective yields for fertilizer (columns 1 and 3) or for fertilizer-plus-lime. Consistent with the results from Table 9, 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 for which farmers were aware they did not know the answer to (columns 3 and 4).

Appendix B (Tables B1 and B2) 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. An indicator variable for overconfidence 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 (reported in Table 11), we use two-stage least squares to examine the effect of in-game experimentation on belief updating. We exploit the fact that rainfall simulations were random for the first seven “seasons” in the app. We use this random variation in rainfall scenarios as an instrument for the share of rounds that a farmer experiments with lime in the app. Random weather correlates quite strongly with participants’ choices in the game, suggesting that there is a decent first-stage relationship. However, the analysis 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.

The 2SLS results in Table 11 suggest that the proportion of rounds that a farmer plays with lime (instrumented for using random weather draws) influences 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 Fig. 7 that the pH of a farmer’s field is not strongly associated with

Table 13
Share of rounds with CAN.

	Share	Share	Share
(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 14
Share of rounds with lime.

	Share	Share	Share
(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 $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

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. Appendix C (Tables C1 and C2) shows the regressions that underlie Fig. 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 12-14 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.

6. Conclusion

Of the many constraints and impediments to agricultural technology adoption that poor farmers face, one of the most fundamental is learning about the properties and performance of a new technology. This learning challenge stems from pronounced heterogeneity in the returns to agricultural technologies and the relatively small number of seasons that a given farmer can observe. 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. The limitations to farmers' ability to proactively experiment in ways that generate useful knowledge about the returns to the technology in their specific case exacerbates this effect. 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 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 show 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 *ex ante* best positioned to benefit from a new, unfamiliar input—discover this opportunity by 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 only 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.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix 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 250×250 m 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 0.135 increase in soil pH.

We vary application rates for DAP and CAN between 0 to 125 kg per acre in steps of 25 kg. Lime is applied in much larger quantities, so we allow it to vary between 0 to 2000 kg in 250 kg steps. We simulate yields under all combinations of DAP, CAN, and lime and the three weather scenarios, resulting in 972 yield simulations per household.

Appendix 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 Fert. dist.	%Δ Mean Fert. + Lime dist.	%Δ Mean Fert. dist.	%Δ Mean Fert. + Lime dist.
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-mean − premean)/pre-mean)*100 Standard errors in parentheses $^*p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Appendix 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 $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jebo.2021.03.001.

References

Ajayi, K.F., Friedman, W.H., Lucas, A.M., 2017. The importance of information targeting for school choice. Am. Econ. Rev. 107 (5), 638–643.
Aker, J.C., Ksoll, C., Lybbert, T.J., 2012. Can mobile phones improve learning? Evidence from a field experiment in Niger. Am. Econ. J.: Appl. Econ. 4 (4), 94–120.
Ambuehl, S., Li, S., 2018. Belief updating and the demand for information. Games Econ. Behav. 109, 21–39.
Anderson, M.L., Magruder, J., 2017. Split-Sample Strategies for Avoiding False Discoveries. Technical Report. National Bureau of Economic Research.
Ashraf, N., Jack, B.K., Kamenica, E., 2013. Information and subsidies: complements or substitutes? J. Econ. Behav. Organ. 88, 133–139.
Attanasio, O.P., Kaufmann, K.M., 2017. Education choices and returns on the labor and marriage markets: evidence from data on subjective expectations. J. Econ. Behav. Organ. 140, 35–55.
Barham, B.L., Chavas, J.-P., Fitz, D., Schechter, L., 2018. Receptiveness to advice, cognitive ability, and technology adoption. J. Econ. Behav. Organ. 149, 239–268.

Bettinger, E.P., Long, B.T., Oreopoulos, P., Sanbonmatsu, L., 2012. The role of application assistance and information in college decisions: results from the H&R block fafsa experiment. Q. J. Econ. 127 (3), 1205–1242.
Bryan, G., Chowdhury, S., Mobarak, A.M., 2014. Underinvestment in a profitable technology: the case of seasonal migration in bangladesh. Econometrica 82 (5), 1671–1748.
Carter, M., Mathenge, M., Bird, S., Lybbert, T., Njagi, T., Tjernström, E., 2019. Filling a Niche: the Impacts of a Local Seed Company on Maize Productivity in Kenya. Working paper.
Charness, G., Viceisza, A., 2016. Three risk-elicitation methods in the field: evidence from rural senegal. Rev. Behav. Econ. 3 (2), 145–171.
Conley, T.G., Udry, C.R., 2010. Learning about a new technology: pineapple in ghana. Am. Econ. Rev. 100 (1), 35–69.
Delavande, A., 2014. Probabilistic expectations in developing countries. Annu. Rev. Econom. 6 (1), 1–20.
Delavande, A., Giné, X., McKenzie, D., 2011a. Eliciting probabilistic expectations with visual aids in developing countries: how sensitive are answers to variations in elicitation design? J. Appl. Econom. 26 (3), 479–497.
Delavande, A., Giné, X., McKenzie, D., 2011b. Measuring subjective expectations in developing countries: a critical review and new evidence. J. Dev. Econ. 94 (2), 151–163.
Dessí, R., Zhao, X., 2018. Overconfidence, stability and investments. J. Econ. Behav. Organ. 145, 474–494.
Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., Wagner, G.G., 2011. Individual risk attitudes: measurement, determinants, and behavioral consequences. J. Eur. Econ. Assoc. 9 (3), 522–550.
Dominitz, Jeff, Charles, F. Manski, 1997. Using Expectations Data to Study Subjective Income Expectations. Journal of the American Statistical Association 92 (439), 855–867.
Dupas, P., 2011. Do teenagers respond to HIV risk information? Evidence from a field experiment in Kenya. Am. Econ. J.: Appl. Econ. 3 (1), 1–34.
Fiore, S.M., Harrison, G.W., Hughes, C.E., Rutström, E.E., 2009. Virtual experiments and environmental policy. J. Environ. Econ. Manag. 57 (1), 65–86.
Foster, A.D., Rosenzweig, M.R., 1995. Learning by doing and learning from others: human capital and technical change in agriculture. J. Polit. Econ. 103 (6), 1176–1209.
Gordon, N., Brayshaw, M., Grey, S., 2013. Maximising gain for minimal pain: utilising natural game mechanics. Innov. Teach. Learn. Inf. Comput. Sci. 12 (1), 27–38.
Harou, A., Madajewicz, M., Magomba, C., Michelson, H., Tschirhart, K., Amuri, N., Palm, C., 2018. Can Information Improve Investment? Effects of Site-Specific Soil Recommendations on Fertilizer Demand. Working paper.
Hoffman, M., 2016. How is information valued? Evidence from framed field experiments. Econ. J. 126 (595), 1884–1911.
Huffman, W.E., Rousu, M., Shogren, J.F., Tegene, A., 2007. The effects of prior beliefs and learning on consumers' acceptance of genetically modified foods. J. Econ. Behav. Organ. 63 (1), 193–206.
Jack, B.K., 2011. Constraints on the Adoption of Agricultural Technologies in Developing Countries. Agricultural Technology Adoption Initiative, Boston: J-PAL (MIT) and Berkeley: CEGA (UC Berkeley).
Lee, J.J., Hammer, J., 2011. Gamification in education: what, how, why bother? Acad. Exch. Q. 15 (2), 146.
Levitt, Steven D., John, A., 2009. Field Experiments in Economics: The Past, the Present, and the Future. European Economic Review 53 (1), 1–18.
Magruder, J.R., 2018. An assessment of experimental evidence on agricultural technology adoption in developing countries. Annu. Rev. Resour. Econ. 10, 299–316.
Marreiros, H., Tonin, M., Vlassopoulos, M., Schraefel, M.C., 2017. "Now that you mention it": a survey experiment on information, inattention and online privacy. J. Econ. Behav. Organ. 140, 1–17.
Moore, D.A., Healy, P.J., 2008. The trouble with overconfidence. Psychol. Rev. 115 (2), 502–517.
Munshi, K., 2004. Social learning in a heterogeneous population: technology diffusion in the indian green revolution. J. Dev. Econ. 73 (1), 185–213.
Simões, J., Redondo, R.D., Vilas, A.F., 2013. A social gamification framework for a k-6 learning platform. Comput. Hum. Behav. 29 (2), 345–353.
Tittonell, P., Vanlauwe, B., Corbeels, M., Giller, K.E., 2008. Yield gaps, nutrient use efficiencies and response to fertilisers by maize across heterogeneous smallholder farms of Western Kenya. Plant Soil 313 (1–2), 19–37.
Tjernström, E., 2018. Learning from Others in Heterogeneous Environments. Working paper.
Walz, S.P., Deterding, S., 2015. The Gameful World: Approaches, Issues, Applications. MIT Press.
Wiswall, M., Zafar, B., 2015. How do college students respond to public information about earnings? J. Hum. Cap. 9 (2), 117–169.