ONLINE APPENDIX for:

The Complexity of Multidimensional Learning in Agriculture

Rachid Laajaj and Karen Macours

Appendix A: Description of the Agronomic Trials

A.1 Researcher-designed and farmer-managed trials

The trials were designed by soil scientists and agronomists of IITA, who provided general oversight while the day-to-day management of the trial plots fell to the farmers, according to the following division of tasks: All inputs were provided by the research team, with the exception of the local maize seed, tested in two out of the six maize subplots (where farmers were asked to apply their own, for comparison with treated subplots). A researcher (local expert agronomist) was present and led planting, gapping and thinning, all fertilizer applications, and harvesting. In these activities, labor was typically provided by the farmers. Planting dates were mostly decided by the researchers to best target the onset of rains, also responding to the farmers' feedback on beginning of rains and availability to schedule the visit for planting. The farmers were in charge of land preparation, weeding and other management, with the researcher providing guidelines on those practices. Farmers were also asked to inform the contact person in case of any pest or disease, in which case the researcher provided the required pesticide or fungicide. Finally, farmers provided the land area, that needed to fit the criteria for a trial, including sufficient space with reasonably low inclination and no shade. This was rarely a binding factor in the selection of farmers. Farmers tended to have many locations that satisfied these criteria, and it was up to them to select which location within their parcels they wanted to dedicate to the trials.

A.2 Trial design and inputs tested

For each farmer, the trials followed a factorial design of 2x3 subplots as presented in figures A1A (for maize) and A1B (for soya). The intercrop trials followed the same factorial design as the soya trials, testing the same combination of soya inputs, but tracking performance on both soya and maize yields. We refer to the "trial plot" for the entire area dedicated to the trial with IITA, and to subplots to refer to each one of the 6 subdivisions where a specific input combination was tested. Each subplot was approximately 4.5 × 5 m and treatments were randomized between the six subplots. Between subplots, a 1 m buffer of sweet potatoes was planted to prevent inter-plot contamination. Plot layout and treatments were maintained for three seasons. All the trials aimed to test some of the most viable solutions that an integrated soil fertility management combination of inputs can bring to the conditions faced by farmers in the region. See Laajaj et al (2020) for details on all inputs and justification for their choices.

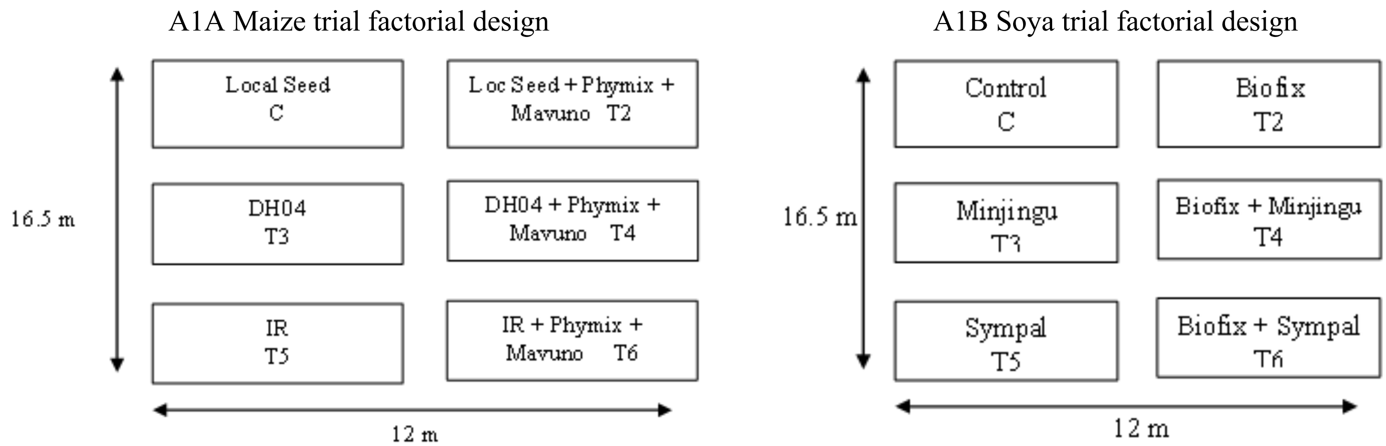


Figure A1: Visual representation of factorial designs of agronomic trials for the cases of maize and soya trials. Each rectangle represents a subplot with a specific set of inputs tested. In maize trials, where DH04 is indicated, in fact half of the maize trials tested KSTP 94 seed and the other half tested DH04 maize seed (randomly allocated at the village level). In soya trials, where Biofix is indicated, in fact half of the soya (or intercrop) trials tested Legumefix inoculant and the other half tested Biofix inoculant (randomly allocated at the village level).

A.2 Main agronomic Findings from the trials

The key findings from the agronomic analysis of the trial data are published in Thuita et al. (2018) and (Laajaj et al 2020). Notably, the maize trials showed that the combination of chemical (Mavuno planting and Mavuno top dressing) and non-chemical fertilizers (Phymix vermicompost) led to large gains in yield and in benefits in all seasons. The yield gains profitability increases over time. There were no significant differences between local seeds and the improved seeds (IR, KSTP94 or DH04). The soya trials in turn showed that : i) the P-source alone (either Sympal or Minjingu) had a significant positive impact on yield and on benefits compared to not using it. ii) One of the P-sources (Sympal) consistently outperformed the other (Minjingu). iii) The tested bio-fertilizer (Biofix or Legumefix, inoculants with a nitrogen-fixing bacteria) is a cost-saving substitute to traditional nitrogen sources (like CAN or Urea); iv) There is a complementarity between the P-source (Sympal or Minjingu) and bio-fertilizer; and v) The results are conditional on management quality and other production practices.

A3 Yield increments observed from the trials

The monitoring data from the agronomic trials include crop cut measures of yield within each subplot of the trials. We use this to compute a "yield increment" of the "best bet" package compared to its control, following Laajaj et al (2020). These yield increments can be interpreted as the potential of the input package observed when tried in the farmer's plot. We use it in Table 2 to estimate its effect on the farmers' decisions to purchase the trial inputs. The measures can be noisy measures of the true potential of the input (because of the small size of each subplot and various sources of seasonal variation), but it reflects well the farmer's realization and observation that allows her to update her prior about the potential of the tested inputs.

Yield increments for maize are calculated using three pairs of treatment–control plots: T2–T1, i.e. subtracting yields in the plot with local seeds and no fertilizer (T1) from yields in the plot with local seeds and the full fertilizer package (Mavuno and Phymix). Similarly, calculations are done for hybrid seed plots with and without fertilizer (T4–T3), and for IR seeds plots with and without fertilizer (T6–T5)..

Yield increments for soybean are calculated using two pairs of treatment–control plots: (a) subtracting yields in the control plot from yields in plots containing a soybean inoculant and Sympal (T6); and (b) subtracting yields in the control plot from yields in plots containing a soybean inoculant and Minjingu (T5). Figure A2 shows a relatively large spread in potential realizations of yield increment in the different trials. It also shows no systematic differences in yield increment between the trials of the LSFs and HSFs for the maize and soya trials, but a more systematic difference in intercrop trials.

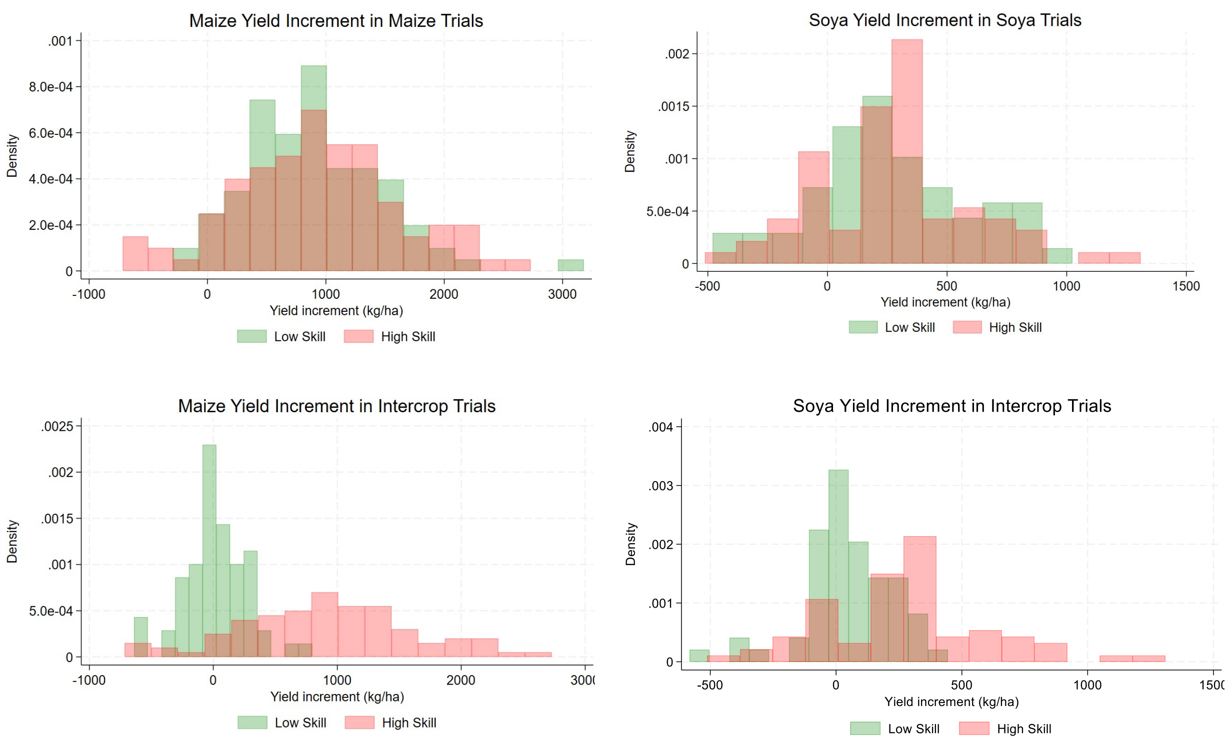


Figure A2 PDFs of yield increments in the agronomic trials

The figure displays the distribution of yield increment. Its calculation is described above. Yield increments in intercrop trials are presented separately since yields in intercropping and monocrop systems are not comparable.

Appendix B: Verification of the Randomization

Table B.1 Balance Test

Variable	Treatment status: Control	Treatment status: Treatment	p-value of difference
Household head is female	0.24 [0.02]	0.25 [0.02]	0.889
Age of respondent	45.77 [0.58]	45.29 [0.61]	0.555
Household size	5.68 [0.11]	5.57[0.12]	0.504
Number of cattle	2.7 [0.19]	2.89 [0.22]	0.654
Land Size (log of ha)	0.27 [0.05]	0.26 [0.07]	0.885
Years of education	5.95 [0.15]	6.16 [0.17]	0.396
Used any inputs of the trials	0.09 [0.01]	0.09 [0.01]	0.952
Grew soya (dummy)	0.04 [0.01]	0.03 [0.01]	0.366
Profit (KES)	5963 [349]	5281 [333]	0.176
Used soil conservation practice	0.29 [0.02]	0.28 [0.02]	0.596
Fertilizer (dummy)	0.74 [0.02]	0.81 [0.02]	0.212
Fertilizer (every year in past 10 years)	0.31 [0.02]	0.3 [0.02]	0.923
Wealth Index	-0.02 [0.04]	0.02 [0.05]	0.647
p-value of joint significance of all variables to explain treatment:			0.6633
Total number of observations	472	472	944

Standard errors of means in brackets

Table B2: Attrition rates by season and treatment status

Season	Attrition rates: Full sample	Treatment	Control	p-val of dif in attrition
0	1.7%	1.7%	1.7%	1
1	2.1%	1.9%	2.3%	0.618
2	2.6%	2.7%	2.5%	0.834
3	4.1%	4.8%	3.3%	0.226
4	4.6%	5.2%	4.0%	0.345
5	3.5%	3.3%	3.8%	0.707
All	**3.1%**	**3.3%**	**2.9%**	**0.637**

Note: attrition is equal to one if the farmer did not answer the main survey in a given season.

Appendix C: Definitions of variables and aggregate indices

The breadth of outcomes considered in this paper is part of its contribution. We aggregate outcomes by type in a set of standardized indices. Below, we provide explanations about each index and references to more detailed sources, starting with an in-depth discussion of the skill measurement, before turning to construction and definitions for all the outcome variables.

C.1 The Skill Index

The sample, data and index calculation used to measure skills are the same as the ones used in Laajaj and Macours 2021, which tests the reliability and validity of the measures and more generally provides guidelines for such measures in rural contexts of low- and middle-income countries. The online appendix of that article provides detailed lists of items and the replication file available online (https://www.openicpsr.org/openicpsr/project/124141/version/V1/view) provides the code for the estimation of the indices. This paper uses the exact same measure (applied to the same population using the same data). The skill index is a non-weighted average of 3 components: cognitive, noncognitive, and technical agronomic skills. Here we provide a brief description of the 3 components.

The cognitive index is obtained using Item Response Theory (two parameter model) with the responses from five cognitive tests: (i) the Raven Colored Progressive matrices, measuring visual processing and analytical reasoning; (ii) the digit span forwards and backwards, measuring short-term memory and executive functioning; (iii) a written and timed test of basic math skills; (iv) an oral nine-item test containing short math puzzles relevant for agriculture; and (v) a reading comprehension test.

The non-cognitive skill measure results from a subset of items from the 44-item Big Five Index (a commonly used instrument for the Big Five personality traits) together with commonly used instruments for lower-order constructs such as locus of control, self-esteem, perceptions about the causes of poverty, attitudes towards change, organization, tenacity, metacognitive ability, optimism, learning orientation and self-control. Most of these subscales are derived from a set of questions asking the respondent the level at which they agree or disagree with general statements about themselves, with answers on a Likert scale from one to five. In addition, we asked a set of locus-of-control questions with visual aids in which people are asked to attribute success to effort and good decisions, luck, or endowments. We also included the CESD, a commonly used depression scale, validated in many developing countries, as it relates to some noncognitive domains captured in other scales (neuroticism and optimism). Factorial analysis revealed that the non-cognitive index is multidimensional, hence we use the average of the non-cognitive factors to obtain an aggregated non-cognitive skills construct.

To build the technical agronomic skills items, we started from different types of questions that can be found in the literature, and worked closely with the team of agronomists from IITA who were overseeing the trials to identify which knowledge could be the most relevant: questions on the timing at which inputs should be used, how to apply the inputs (quantity, location, etc.) and knowledge of both basic and more complex practices (spacing, rotation, composting, conservation). The items cover the most common practices and inputs in Western Kenya. We use a mix of open questions and multiple-choice questions. Some questions allow multiple answers, and a subset of questions had visual aids (for example, pictures of inputs). The set of questions covered a relatively broad spectrum of practices, including a set of questions on maize, soybean, banana, soil fertility practices, composting, and mineral fertilizer. We used Item Response Theory to compute the agronomic technical skill index.

All skill items were obtained from a baseline skill survey specifically designed for this purpose, implemented twice at a three-week interval to evaluate the reliability through test-retest correlations. To reduce measurement error for each index, we take the average of the two indexes obtained from the two rounds of survey. Extensive tests of reliability and consistency were applied to the data (Laajaj and Macours, 2021). The cognitive measure is highly reliable and with limited noise, the technical agronomic skills measure is noisier, but still quite reliable. By contrast, the noncognitive skill measure is noisy and prone to systematic measurement errors (such as acquiescence bias, which we adjusted for, and social desirability bias, harder to correct) and thus requires more caution in its use and interpretation. Despite the limitations, this is arguably a very comprehensive effort to obtain in-depth skills measures.

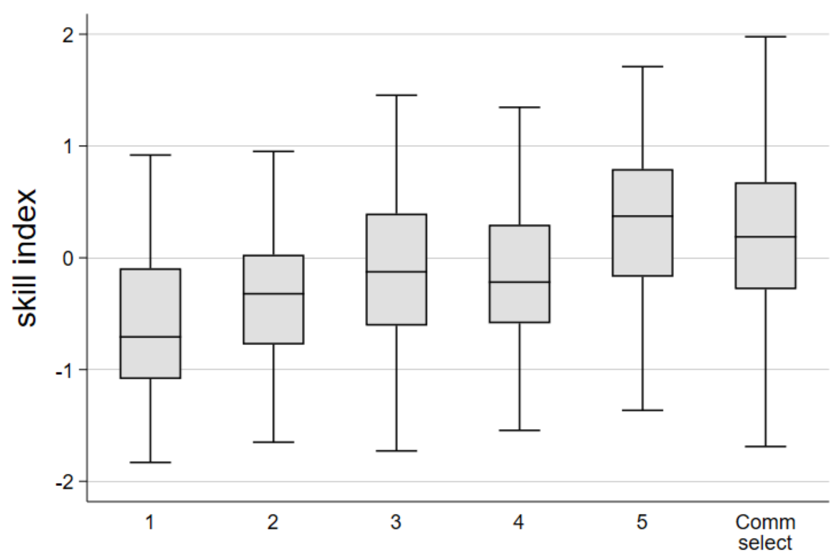


Figure C1 The box plot displays the median (the central line), the 25th percentile (bottom of the box), the 75th percentile (top of the box), as well as the lower adjacent value and upper adjacent value of the skill distribution for each possible quintile in the ex-ante skill index (among randomly selected farmers), and for the community selected farmers (right hand box).

As treatment stratification was based on the ex-ante skill proxy (described in section 2.3), Figure C1 displays the spread of the ex-post skill index for the 5-levels of ex-ante classification and, separately for the community-selected farmers. The ex-post skill measure increases with the ex-ante proxy as expected, and the community selected farmers tend to have a skill distribution relatively similar to farmers in the 5th quintile of ex-ante skill proxy.

The share of HSFs (according to the ex-post measure) varies from 16% in the lowest (ex-ante) quintile, to 70% in the fifth quintile. Among the community-selected group 62% are classified as HSF. This confirms that community-selected farmers are closest to the top quintile in the randomly selected sample, and that our ex-ante skill proxy is a well correlated but imperfect proxy for the farmers' skill level. Among the 50% randomly selected farmers, the cutoff of the aggregate skill index is at percentile 62, so that being a HSF is roughly equivalent to being in the top 2 quintiles of the village distribution.

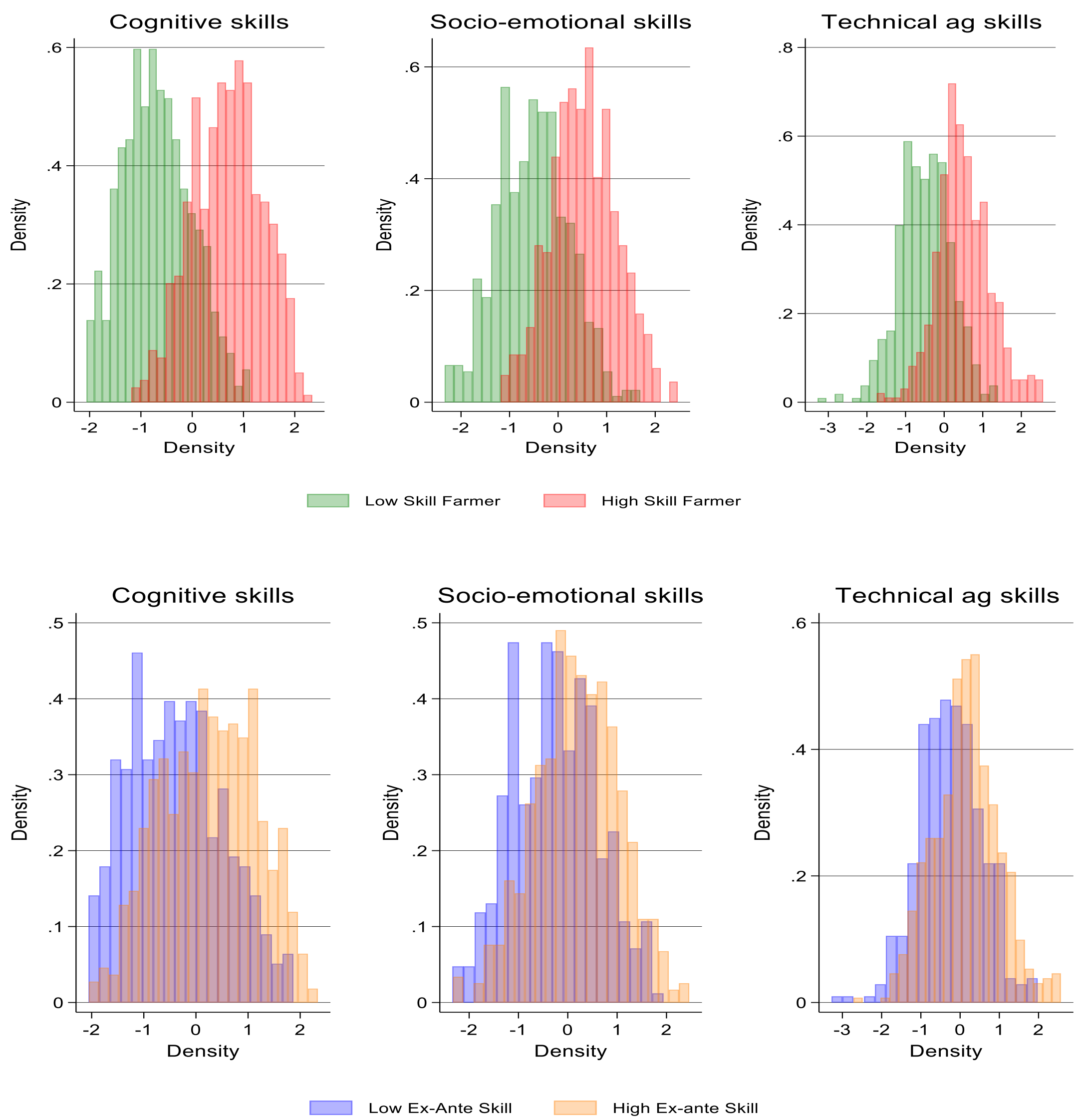


Figure C2: distribution of ex-post (top-panel) and ex-ante (bottom-panel) aggregate skill measure

Figure C2 further helps understand the relationship between the ex-ante proxy and the ex-post, more comprehensive skill measure. The ex-ante proxy captures a reasonable share of the difference in skill in each one of the 3 dimensions (bottom panel). As expected, it differentiates less, however, than the LSF-HSF contrast that was created by splitting the farmers based on the median of the aggregate skill distribution (upper panel).

Table C1 shows clear systematic differences between LSF and HSF in baseline socio-economic variables, as well as agricultural practices and outcomes. Land size, number of cattle and the likelihood to grow soya do not differ significantly. As skills are far from being randomly assigned, we do not attribute differences in treatment effects for LSFs and HSFs to the skills themselves. For example, we cannot rule out that since our HSFs are wealthier on average, they can take more risk, and that this drives some of the differences in treatment effects between the two groups. In a similar way, any other observed or non-observed differences between LSFs and HSFs could contribute to the differences in treatment effects.

Table C1: Comparison of socio-economic characteristics between LSFs and HSFs.

	Skill Level		p-value of
Variable	Low	High	difference
Household head is female	0.33 [0.02]	0.15 [0.02]	0.000
Age of respondent	47.7 [0.62]	43.38 [0.55]	0.000
Household size	5.06 [0.11]	6.19 [0.11]	0.000
Number of cattle	2.71 [0.19]	2.87 [0.22]	0.532
Land Size (log of ha)	0.28 [0.05]	0.26 [0.07]	0.747
Years of education	4.03 [0.14]	8.08 [0.12]	0.000
Used any inputs of the trials	0.07 [0.01]	0.11 [0.01]	0.018
Grew soya (dummy)	0.03 [0.01]	0.03 [0.01]	0.550
Profit (KES)	5031 [326]	6221 [355]	0.025
Used soil conservation practice	0.26 [0.02]	0.31 [0.02]	0.077
Fertilizer (dummy)	0.69 [0.02]	0.86 [0.02]	0.000
Fertilizer (every year in past 10 years)	0.24 [0.02]	0.38 [0.02]	0.000
Wealth Index	-0.09 [0.04]	0.09 [0.05]	0.001
p-value of joint significance of all variables to explain treatment:			0.000
Total number of observations	472	470	942

Standard errors of means in brackets

C.2 Inputs and practices: adoption and exploration

The survey instrument includes a section where farmers enumerate their plots, which crops they grow in each plot, followed by detailed questions on inputs, practices and production. To measure input adoption, we compare the list of commercial inputs that they mention using on their non-trial plots to the list of inputs tested in the trials. We define a dummy equal to one if the farmer used in his own parcels, at least one of the inputs tested in the trials, which includes soya fertilizers (Mavuno and Sympal), biofertilizer (Biofix) or seed (SB19), or the maize fertilizer (Mavuno planting or top dressing), or seed (IR, KSTP94 or DH04).

Regarding practices, we ask about many practices, including intercropping, manure, number of weeding, number of seeds sowed per hole, soil conservation practices, etc, which are part of the instructions provided by the agronomists in the management of the trials. We then construct an index that represents the fraction of these practices adopted by the farmer in at least one plot (not including trials). The 10 practices are dummies presented below, with their average across our sample:

Description of the dummy: equal to one if…	**Adoption rate**
The household used manure in at least one parcel	67%
The household used a soil conservation practice in at least one parcel	43%
The household harrowed twice or more at least one parcel	31%
The household applied weeding twice or more in at least one parcel	59%
The household used gapping in at least one parcel (filling with a plant that grew somewhere else in places where the crop did not germinate)	35%
The household used intercropping in at least one parcel	82%
The household used sweet potato in at least one parcel	12%
The household grew soya in at least one parcel	8%
The household used crop rotation in at least one parcel	57%
The household used only 1 seed per hole in at least one parcel	57%

The practices adoption index is a non-weighted average of the 10 dummy variables, so a 0.1 increase in the index can be interpreted as the adoption of one additional practice out of the 10 practices.

To measure exploration, both for inputs and practices (separately), we use an index of the frequency of changes to reflect the extent to which farmers are exploring and trying new things. For this purpose, we generate for each practice or input a dummy equal to 1 if the dummy of use at a given season is different from the one of the preceding seasons (the farmer was not using it on any of their plots in the prior season and started using it on at least one plot, or the farmer was using it in the prior season and then stopped using it), and 0 if it is the same. We then average the index across all inputs and across all practices to obtain an index of changes in input decisions and in index of changes in practice decisions. These indices can be interpreted as the fraction of practices (or inputs) decisions that the farmer changed compared to the prior season.

Section 5.3 uses two continuous measures of adoption. The first one is the sum of the amount spent on trial inputs. For this we simply add across all parcels, input expenditures dedicated to the purchase of any input that belongs to the list of inputs tested in the trials. The second measure is the area dedicated to soya, which we obtain by summing the area of all parcels in which soya was planted (including parcels where soya was intercropped).

C.3 Profits

Profits are computed by deducting the total cost of inputs from the total value of production (using village level prices). These profits incorporate labor costs when hired but do not account for the shadow value of household's labor. Average profits are positive for both LSF and HSF, with seasonal profits being negative for 14% of farmers, and profits over all season only negative for 3%. Farmers using inputs tested in the trial on their own plots have significantly higher profits.

C.4 Agricultural knowledge index (general, maize and soya)

To measure technical know-how with respect to the use of the trial inputs, 15 knowledge questions were asked ranging from recognizing the inputs to questions on where and when to apply the fertilizer or how to store the biofertilizer. These questions (combining open questions and multiple choice) were designed together with the agronomists of IITA, with the intention to capture the key practical decisions that can affect the profitability of the newly tested inputs. All questions were first transformed into binary variables equal to 1 if the farmer provided an answer that the agronomist considered correct. Extensive piloting allowed to avoid "ambiguous" answers that would be correct in some context but not in others. To combine all these items into one index, we use the 2-parameter model of item response theory (IRT). We also generate sub-indices for maize knowledge and soya knowledge by running the 2 parameter IRT model, but in each case limiting the items to the corresponding ones.

C.5 The Willingness to Purchase inputs

One of the major limitations of our index on input adoption is that not all inputs are available at all retail shops (agrodealers) and also that in some cases farmers may wish to adopt but not have the money to do it, which can explain why even though effects on input adoptions are significant, the average values remain modest. We therefore also gathered data on an intermediary outcome: whether farmers would be willing to purchase the trial inputs when we simply simulate the situation where they want to produce maize (or soya) in one of their plots, then ask them to imagine that they enter the shops and find a list of inputs with their real local prices. The list included common inputs sold and used in the region as well as the ones tested in the trials. Which ones would they purchase? This allows us to construct a binary variable, equal to 1 if at least one of the newly tested inputs is chosen in this fictitious scenario. The WTP questions were targeted so that farmers in the treatment were asked about the crop for which they had been randomly selected to receive the agronomic trials, with the equivalent number of randomly selected farmers in the control villages asked for that crop.

We also use this set of questions to compare how much farmers prefer using Sympal versus Minjingu or IR seed compared to another maize seed, without the answers being affected by input availability.

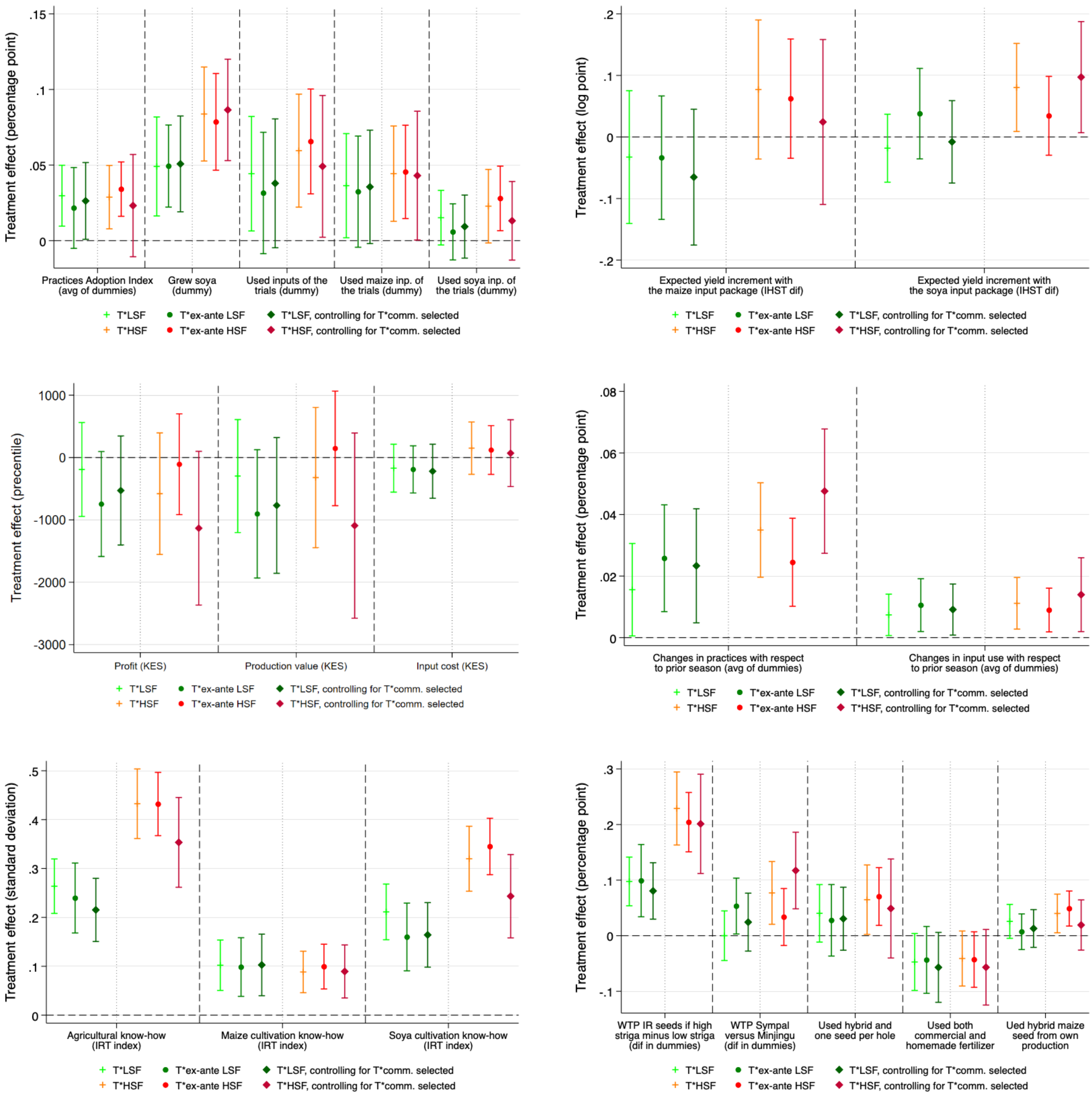
C.6 The Expectation of Yield Increment with the input package (for maize and soya)

Finally, to examine learning about the returns, we estimate impacts on beliefs about expected yield increase when using inputs compared to without inputs in either maize or soya production using a method similar to Carter et al (2021).[1] To do this, after identifying a given plot for each study participant, farmers were asked what production she expected if they used the set of inputs that they had chosen on this parcel (as per the WTP section described in appendix section C4) in (i) a normal year, (ii) a good year, and (iii) a bad year. The same set of questions was asked for a scenario in which they did not use any input in the same parcel and conditions. We then asked the farmer to say, on average, out of 10 years, how many are good years, bad years, and normal years. This set of questions allows us to calculate the expected yield when using the technology package and the expected yield when not using the technology package and accounting for the stochastic nature of yield. We then transform the expected yield values in inverse hyperbolic sine transformations. And then we take the difference between expected yield increase with the input package compared to expected yield increase without the input package. This index can be interpreted as the relative increment in yield that the farmer expects when using the selected input package compared to when not using it. If the treatment allowed farmers to learn that the fertilizer package can allow them to reach higher yield than what they were able to reach before knowing about these inputs, then we would expect a positive effect on this subjective expectation about yield increment. The variables are computed separately for maize and for soya.

[1] Carter et al (2021) estimate the expected yield with the input packaged whereas we estimate the expected yield increment by subtracting expected yield without the input package to the expected yield with the input package, which is arguably closer to what really matters economically for the farmer.

Appendix D: Robustness Checks and Additional Results

Figure D.1 Pooling seasons since treatment started and with robustness checks



Note: Figure D1 presents the results of the estimation of equation 1, where the treatment effects are separated by LSF versus HSF but pooling all seasons together, with 95% confidence intervals in brackets. It then presents the same specification but using the ex-ante proxy (and level of treatment stratification) to classify farmers into high skill or low skill. And the third specification maintains the ex-post skill measure of the main specification, but controls for the treatment interacted by whether the farmer comes from the community selected process (rather than the random selection). The outcomes presented in the figures include the ones used throughout the article, slightly re-organized, to pool together the ones that share the same unit.

Table D1: Regressions pooling seasons and LSF with HSF, with Treatment interacted with skill index

VARIABLES	(1) Practices Adoption Index (avg of dummies)	(2) Used any inputs of the trials (dummy)	(3) Grew soya (dummy)	(4) Profit (KES)	(5) Production Value (KES)	(6) Input Cost (KES)	(7) Agricultural know-how (IRT index)
Treatment	0.028***	0.050***	0.066***	-431.550	-385.745	-32.338	0.340***
	(0.007)	(0.014)	(0.012)	(309.061)	(347.340)	(139.402)	(0.024)
Treatment * skill index	0.005	0.009	0.029**	-34.495	299.657	329.636*	0.116***
	(0.010)	(0.016)	(0.014)	(415.708)	(503.328)	(191.054)	(0.032)
Observations	4,541	4,655	4,541	4,534	4,518	4,531	4,548
Baseline value control	YES	YES	YES	YES	YES	YES	NO

VARIABLES	(8) Discussed ag with other vlg member in last 3 months (avg of dummies)	(9) Expected yield increment with the maize input package (IHST dif)	(10) Expected yield increment with the soya input package (IHST dif)	(11) Changes in practices with respect to prior season (avg of dummies)	(12) Changes in input use with respect to prior season (avg of dummies)	(13) Willingness to purchase IR seeds if high striga minus low striga (dif in dummies)	(14) Willingness to purchase Sympal versus Minjingu (dif in dummies)
Treatment	0.083***	0.018	0.031	0.025***	0.009***	0.158***	0.037*
	(0.016)	(0.038)	(0.024)	(0.005)	(0.002)	(0.021)	(0.019)
Treatment * skill index	0.019	0.055	0.018	0.015*	-0.000	0.071***	0.051**
	(0.020)	(0.057)	(0.032)	(0.008)	(0.004)	(0.025)	(0.022)
Observations	4,655	3,689	3,267	4,529	4,529	3,742	2,893
Baseline value control	YES	NO	NO	NO	NO	NO	NO

VARIABLES	(15) Used hybrid and one seed per hole	(16) used both commercial and homemade fertilizer on at least one plot	(17) =1 if reports using hybrid maize seed from own production	(18) Used any maize inputs of the trials (dummy)	(19) Used any soya inputs of the trials (dummy)	(20) Maize cultivation know-how (IRT index)	(21) Soya cultivation know-how (IRT index)
Treatment	0.047**	-0.046**	0.032***	0.039***	0.019**	0.091***	0.259***
	(0.020)	(0.018)	(0.010)	(0.013)	(0.008)	(0.018)	(0.024)
Treatment * skill index	0.009	0.021	-0.001	-0.000	0.007	-0.036	0.093***
	(0.024)	(0.027)	(0.017)	(0.015)	(0.010)	(0.026)	(0.029)
Observations	4,431	4,552	4,431	4,655	4,655	4,548	4,548
Baseline value control	NO	NO	NO	YES	YES	NO	NO

Note: Table reports the effects of Treatment alone, pooling all seasons and without separating LSF and HSF. It also includes the Treatment interacted with the continuous skill index to assess heterogenous effects by skill. The estimation is ran for the outcomes that appear throughout the article. Because the skill index has a 0 mean, the coefficient of the Treatment dummy can be interpreted as the average treatment effect in the entire sample, over the different periods. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

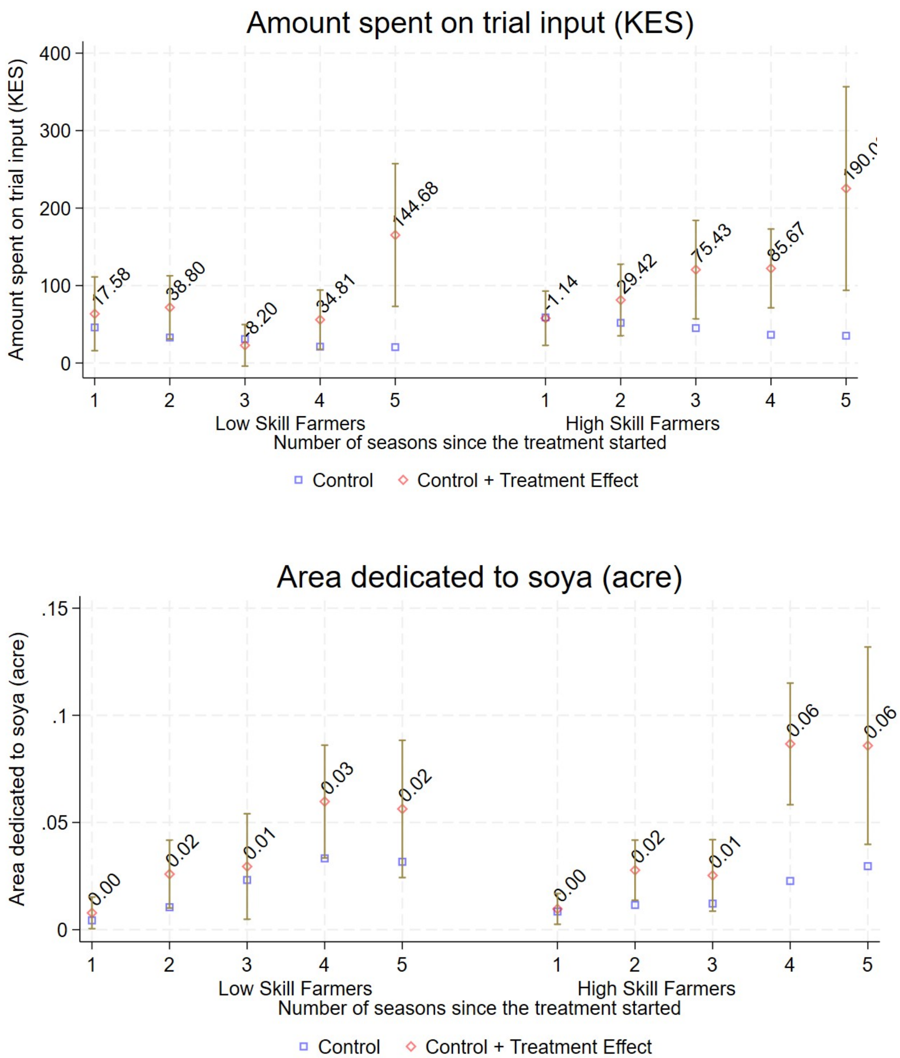


Figure D2 Adoption of trial inputs and soya cultivation on the intensive margin
Treatment effects by skills and number of seasons since treated on continuous measures of adoption. The figure replicates the methodology of Figure 4, displaying the results from the estimation of equation (2). Blue squares represent the average among control farmers in the corresponding season and skill level. Red diamonds represent the expected value in the corresponding treatment group (average value in control group + treatment effect), and the value that appears above each red diamond is the corresponding treatment effect.

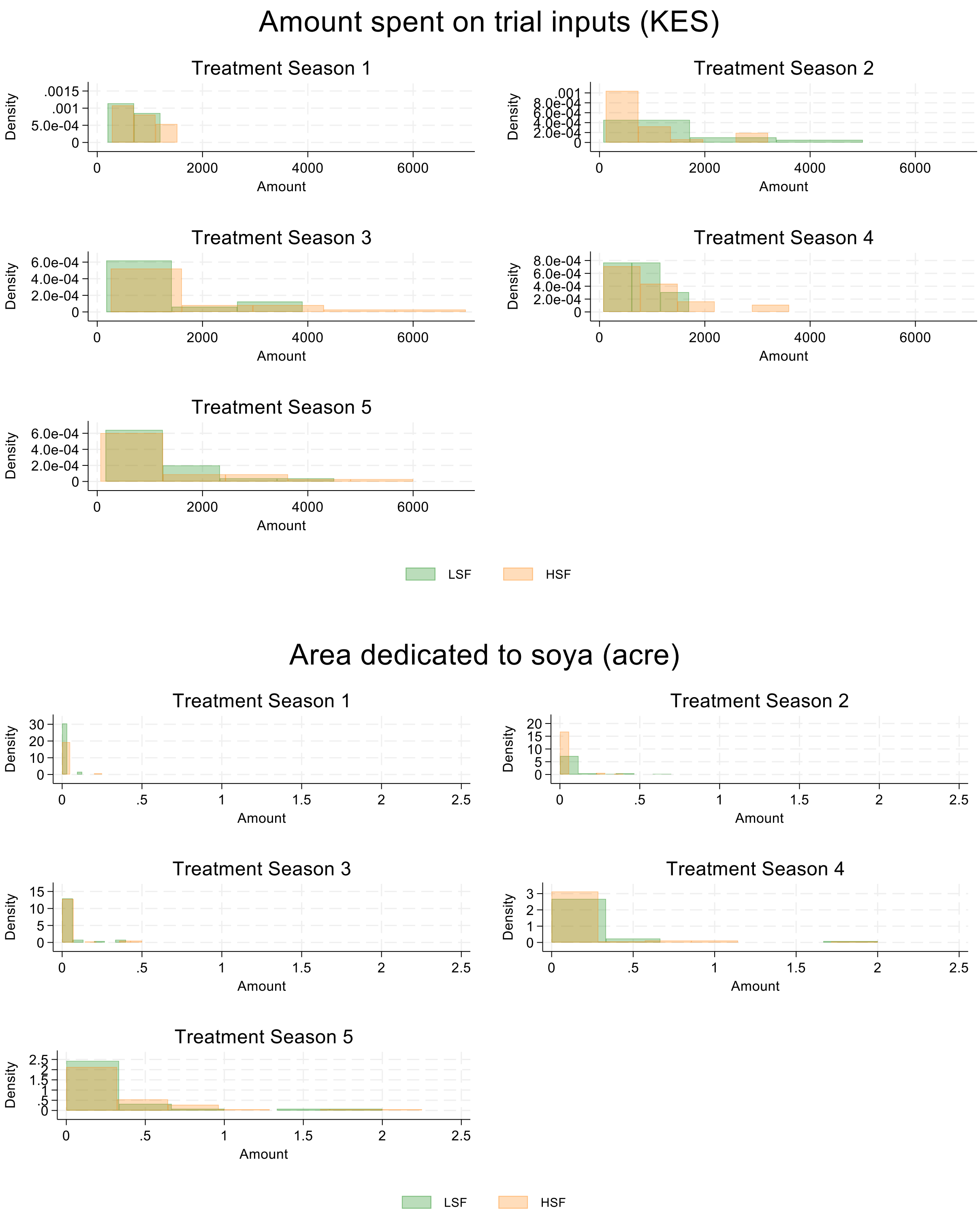


Figure D3: Distribution of continuous measures of technology adoption.
To graph this distribution, all 0 were transformed into missing values, so that it should be interpreted as a descriptive representation of the evolution of the distribution at the intensive margin. The sample is restricted to the treatment group.

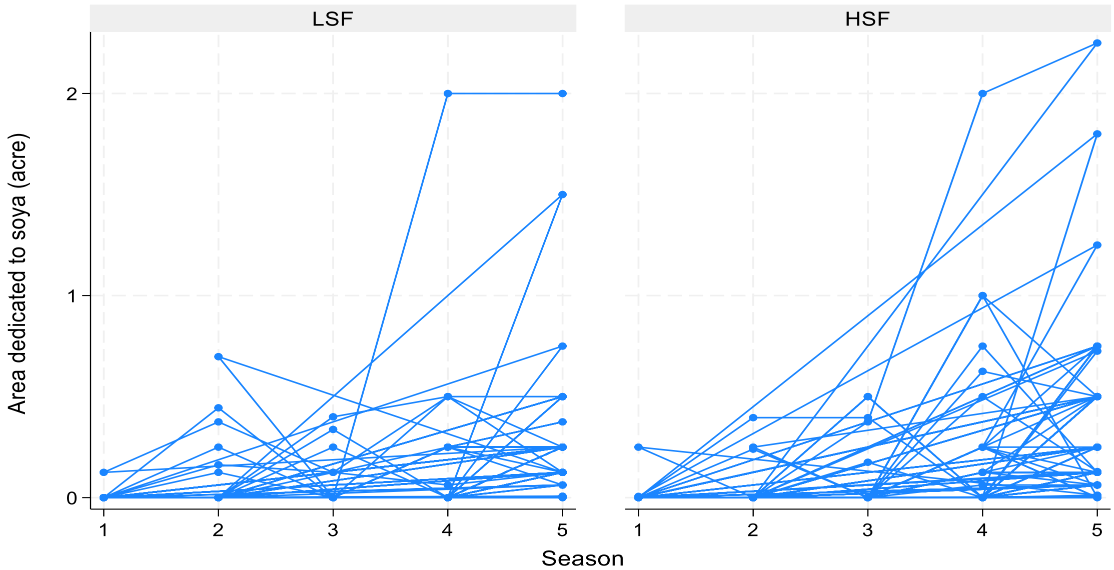


Figure D4: Household trajectories in area dedicated to soya among adopters

Each blue line connects a given household across the 5 seasons of the study. Only treated farmers are included (and only the ones with a strictly positive value of soya area in the last 2 seasons), in order to observe variations in the intensive margin among adopters.

Table D2: Regressions of treatment effect on key outcomes in a spillover sample

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	know-how (IRT)	Practices Adoption Index (avg of dummies)	Used any inputs of the trials (dummy)	WTP maize prog inputs	WTP soya prog inputs	Can identify input names
Spillover * LSF (ex ante proxy)	0.021	0.010	0.019	0.001	-0.000	0.031**
	(0.046)	(0.012)	(0.015)	(0.006)	(0.007)	(0.016)
Spillover * HSF (ex ante proxy)	0.097	0.007	0.018	0.020*	-0.005	0.003
	(0.091)	(0.030)	(0.028)	(0.011)	(0.012)	(0.047)
Observations	942	942	956	942	932	942
Avg outcome in LSF control	-0.129	0.399	0.0234	0.0505	0.0667	0.0962
Avg outcome in HSF control	-0.113	0.428	0.0319	0.0370	0.0724	0.205
P-val of Treat. (LSF & HSF)	0.202	0.595	0.216	0.0896	0.738	0.489
P-val of T * LSF = T * HSF	0.500	0.931	0.975	0.125	0.717	0.573

Estimates with spillover sample, consisting of an additional 480 farmers, i.e. 5 randomly selected farmers within each village, (after excluding the 10 farmers already selected to take part of the study). We administrated follow-up surveys to this additional sample during the last 2 rounds of the study. “Spillover” refers to belonging to a treatment village (where another 10 farmers were assigned to participate to the agronomic trials). The LSF and HSF categorization of these farmers uses the ex-ante proxy measure of skill described in the main text since we do not have detailed skills measures on this sample. These farmers were not actively involved in the trials, nor in discussions or field days throughout the study.

D.4 Treatment effects on soil quality

A possible explanation for the continued adoption, despite negative effects on profits could be that farmers perceived that the trial inputs contribute to improvements in soil quality which will be reflected in future profits. To test this, we assess whether the treatment affected soil quality. In every season, the farmers were asked the soil quality of each parcel on a 5 step scale (1 being very poor and 5 being very fertile). We take the average of the response over all parcels to create an index of the farmers perception of soil quality. Figure D4 displays the effects of the treatment on this outcome variable, separating the treatment effects by skill type and number of seasons since the beginning of the treatment and shows no significant effect. Because some coefficients appear to be marginally significant, we show in table D4, the treatment effects when pooling all seasons together and conclude that there is no significant effect of the treatment on the farmers' perception of the soil quality of their parcels.

We must note, however that the soil quality variable suffers from a high number of missing observations: 20.9% of observations were missing in the control group and 22.7% in the treatment group (p-value of the difference 0.097). This is due to many farmers answering that they don't know about the quality of their parcels. Hence the results are slightly prone to a selection bias if those more likely to answer tend to differ between the treatment and control group, and they are only indicative of no major change in the perception about soil quality that occurred with the treatment.

Finally, one may be concerned that the perception of soil quality may not reflect well actual soil quality. Since we are trying to understand what can explain the increase in adoption decisions despite the negative profits, however, we are interested in the farmers' perception of soil quality rather than the soil quality per se, hence this is not a primary issue.

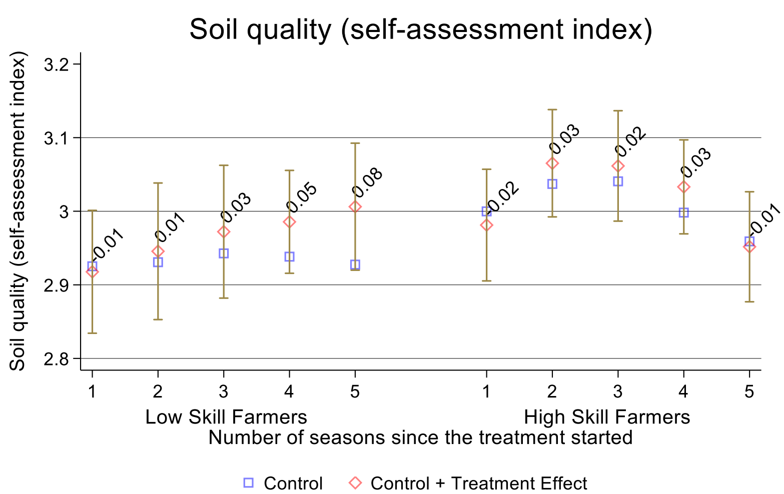


Figure D5 Treatment effects by skills and number of seasons on the perception of soil quality.
This figure displays the results from the estimation of equation (2). Blue squares represent the average among control farmers in the corresponding season and skill level. Red diamonds represent the expected value in the corresponding treatment group

(average value in control group + treatment effect), and the value that appears above each red diamond is the corresponding treatment effect.

Table D3: Regressions of treatment effect on Soil Quality

	(29)
VARIABLES	Soil quality (self-assessment index)
Treatment * low skill	0.017
	(0.026)
Treatment * high skill	0.044
	(0.087)
Observations	4,470
p-val average skill	0.508
p-val low skill treat=high skill treat	0.770

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Appendix E: Why would HSF learn with higher precision? A behavioral inattention explanation

This section provides a minor extension to the theoretical model to provide behavioral foundations for why HSFs would learn faster than LSFs in the form of observing realizations with more precision and thus less noise. However, this noise is not due to additional randomness (or risk exposures) in their realizations. Any farmer faces the same profit function:

$$\pi = f(X) + \epsilon_{it}$$

We assume that the randomness of profit can be modeled as k random shocks such as weather, bug infestation or weeds:

$$\epsilon_{it} = \eta_{1t} + \eta_{2t} + \cdots + \eta_{kt}$$

These random shocks can be observed, but this requires attention, which itself requires the allocation of scarce cognitive resources. We assume that farmers have a limited bandwidth, which limits the number of random shocks that they are able to observe. Importantly, a farmers' skill determines the number of random shocks that she can observe, hence HSFs can observe more shocks than LSFs. When a shock is observed then the noise of that shock does not affect the precision of the Bayesian update, only the variance of unobserved shocks enters in the calculation of the precision (which is given by the inverse of the total variance). Since farmers are rational, they will observe the shocks with the highest variance until they reach the maximum number of shocks that they can observe. Because the HSFs are left with a smaller number of unobserved shocks, they face less noise in their Bayesian update and can update their beliefs with more precision at each round. This is one possible model where behavioral inattention and differences in attention capacity can drive the difference in learning capacity. Among the possible alternatives, the LSFs could observe the decision variables X or the outcome variable Y with some noise (larger than the one of the HSFs), because they are less able to pay attention. This would also result in differences in precision of the signals between LSFs and HSFs.

Appendix F: Propositions and their proofs

<u>Proposition 1</u>: *With T high enough, all farmers converge to a stationary point that is a local maximum.* First, because the distribution of the beliefs at any X is normal, its expected value and variance of the beliefs ($f_t^b(X)$ and $V_t^b(X)$) are sufficient statistics to describe the distribution at each point. Hence, we can re-write equation 5 as:

$$AF_t\left(f_t^b, V_t^b\right) = \underbrace{Eu\pi\left(f_t^b, V_t^b\right)}_{Exploitation\ gain} + \underbrace{\alpha_{Ft}\left[f_t^b + \lambda V_t^b\right]}_{Exploration\ gain}$$

where $Eu\pi\left(f_t^b, V_t^b\right)$ is a function that describes the expected utility derived from the profit of a random variable that follows a normal distribution $\aleph\left(f_t^b, V_t^b\right)$. Even though f_t^b and V_t^b are functions of X, here we omit the X to focus on the effects of changes in these variables (as if they were exogenous), before looking at how they vary depending on X.

Notice that f_t^b increases the value of both exploitation and exploration gains. By contrast V_t^b only increases the exploration gain but reduces the exploitation gain. This tradeoff in the variance will play a key role in the farmer's arbitrage. Hence, we describe the behavior of the function $AF_t\left(\overline{f_t^b(X)}, V_t^b(X)\right)$, i.e. when keeping $f_t^b(X)$ fixed and varying only $V_t^b(X)$.

$$\frac{dAF_t\left(\overline{f_t^b}, V_t^b\right)}{dV_t^b} = \frac{dEu\pi\left(\overline{f_t^b}, V_t^b\right)}{dV_t^b} + \alpha_{Ft}\lambda$$

Given that the utility function exerts increasing absolute risk aversion, we know that:

1. $\frac{dEu\pi\left(\overline{f_t^b}, V_t^b\right)}{dV_t^b}$ is negative and concave (in other terms the absolute utility cost of variance is convex)
2. $\frac{dEu\pi\left(\overline{f_t^b}, 0\right)}{dV_t^b} = 0$
3. $\lim\limits_{V_t^b \to \infty} \frac{dEu\pi\left(\overline{f_t^b}, V_t^b\right)}{dV_t^b} = -\infty$

Taking these elements together with the fact that $\alpha_t\lambda$ is a positive constant, we know that $AF_t\left(\overline{f_t^b(X)}, V_t^b(X)\right)$ is single-peaked, first increasing $\left(\frac{dAF_t\left(\overline{f_t^b}, V_t^b\right)}{dV_t^b} > 0\right)$ and then decreasing $\left(\frac{dAF_t\left(\overline{f_t^b}, V_t^b\right)}{dV_t^b} < 0\right)$. If $f_t^b(X)$ was constant, the X at which this maximum is located is the location of the best next step (that maximizes the AF_t). When $f_t^b(X)$ is increasing, this optimal step will be located further away from $X_{t_{-1}}$ where the loss due to the increase in the variance exactly compensates the gain in AF due to the increase in $f_t^b(X)$.

We first show that any point $\acute{X}$ that is not a local maximum cannot be a stationary point. Assume that $\acute{X}$ has a strictly positive expected gradient when varying X in a given direction ΔX^{PG}, then $\frac{dAF_t\left(f_t^b,\overline{V_t^b}\right)}{df_t^b} > 0$ and, as shown in the preambule, if the variance is sufficiently low at the vicinity of $\acute{X}$ then $\frac{dAF_t\left(\overline{f_t^b},V_t^b\right)}{dV_t^b} > 0$, and if it is not sufficiently low, then being a stationary point would require the farmer to remain at this position, but as the farmer does so, the variance would be reduced until the point where $\frac{dAF_t\left(\overline{f_t^b},V_t^b\right)}{dV_t^b} > 0$. Once this condition satisfied, then if ϵ is small enough, a step $\epsilon\, \Delta X^{PG}$ can only lead to an increase in AF_t and thus will always be preferred to remaining at $\acute{X}$, hence $\acute{X}$ cannot be a stationary point. As a result, with enough time, any farmer will continue exploring until a local maximum is reached. A local maximum is not necessarily a stationary point, but if it not a stationary point, then the exploration continues until a local maximum that is high enough to be a stationary point (because it provides a higher AF_t than jumping into the wild or any alternative previously explored), hence with T high enough any farmer converges to a stationary point that is a local maximum.

Proposition 2: Under the same conditions, HSFs are expected to converge to a (weakly) higher local maximum than LSFs

Assume a local maximum $\hat{X}$. For this local maximum to be a stationary point, $AF_t(\,\hat{X})$, it needs to exceed the AF_t associated to jumping into the wild. The expected benefit of exploiting $\hat{X}$ is given by $E[u(\pi(X))]$ (the exploitation gain in equation 5, in the case where the variance is only driven by the randomness of the profit function, but not variance in the beliefs) and is independent of the skill level of the farmer. By contrast the AF_t of jumping into the wild is strictly higher for HSFs than LSFs since $\alpha_{HSF\ t} > \alpha_{LSF\ t}$. Hence the minimum value of $f(\hat{X})$ required for $\hat{X}$ to be a stationary point is strictly higher for HSFs than LSFs.

Proposition 3: The curse of dimensionality results from complementarities and substitutabilities between inputs.

The curse of dimensionality refers to the fact that the area to explore increases exponentially with the number of dimensions of X. To simplify it, assume that each dimension of X is defined over a finite segment of size s. The Hilbert space over which $f(X)$ is define is then a hypercube with an n-dimensional volume of s^n. Hence the area to be explored grows exponentially with n (the number of dimensions of X).

In the absence of complementarities (and substituabilities) $\frac{d^2 f(X)}{dx_i dx_j} = 0\ \ \forall i,j\ i,j\ \in\ 1,\ldots,n$ and $i \neq j$. As a result, any lesson about $\frac{df(x_i,x_{-i})}{dx_i}$ would be relevant independently of x_{-i} and each dimension can be explored separately. In this setting, each additional n^{th} dimension adds s to the area that needs to be explored but is not multiplied by the previous $n-1$ hypercube.

This can be understood more easily if, for illustration, we assume that each dimension x_i is discrete (of size $s \geq 2$) rather than continuous. In the presence of complementarities, the total number of combinations of inputs and thus parameters to learn about is s^n. In the absence of complementarities, however, it is sufficient to know the effect of each possible variation in x_i (for any other inputs x_{-i}) to figure out the entire distribution, (because one can perfectly extrapolate to other combinations of x_{-i}, hence the number of parameters to learn about is sn.)

The next set of propositions describe the reaction to an unexpected source of information, coming from the observation of an exogenous realization of $\pi(X)$. This mimics the effect of exposure to the trials, or the effect of observing the input decision and profit of other farmers. For simplicity, our model considers such info as exogenous and does not incorporate strategic actions such as the delaying of one's exploration to wait for information from others (farmers or external intervention).

Proposition 4: *A new signal can trigger a "jump in the wild" if its precision is sufficiently high and if it reveals profits that are sufficiently above the prior at that point.*

We model the effect of a new information such as the one coming from the trials as a new signal that occurs at an unexplored part of the production function. Let $\dot{X}$ be the location of the new signal and $\pi_{t+1}(\dot{X})$ be the new signal observed with precision $\dot{\rho}$. The fact that $\hat{X}$ is a stationary point before the signal implies that:

$$AF_t(\hat{X}) \geq AF_t(\dot{X})$$

And using our re-writing:

$$AF_t\left(f_t^b(\hat{X}), V_t^b(\hat{X})\right) \geq AF_t\left(f_t^b(\dot{X}), V_t^b(\dot{X})\right)$$

We first describe how the signal is expected to affect $AF_t(\dot{X})$ through $f_t^b(\dot{X})$ and then describe how it affects $AF_t(\dot{X})$ through $V_t^b(\dot{X})$. The proposition requires that the realization of profit at the newly explored point exceeds the prior: $\pi_{t+1}(\dot{X}) > f_t^b(\dot{X})$. In Bayesian update the posterior is a weighted average of the prior and the new signal, where the weights are proportional to the precisions of the signals. Hence the adjustment of $f_t^b(\dot{X})$ is upward and increasing in $\dot{\rho}$. We also know that $AF_t(\dot{X})$ is increasing in $f_t^b(\dot{X})$, hence more precision increases $AF_t(\dot{X})$ through its positive effect on $f_t^b(\dot{X})$.

We now look at how $\dot{\rho}$ affects $AF_t(\dot{X})$ through $V_t^b(\dot{X})$. In Bayesian update, the variance of the posterior is decreasing in the precision of the signal, hence $\frac{dV_t^b(\dot{X})}{\dot{\rho}} > 0$. Additionally, because $\dot{X}$ is located in the unknown area, it is in the area with high variance and a negative effect of variance on AF_t (because of the dominating effect of risk reduction – see proof of proposition 1), hence $\frac{dAF_t\left(f_t^b(\dot{X}),V_t^b(\dot{X})\right)}{dV_t^b(\dot{X})} < 0$. Putting these 2 elements together, the higher the precision, the higher the reduction of variance, which contributes to more increase in $AF_t(\dot{X})$.

The last two paragraph can be summarized by the following equation:

$$\frac{dAF_{t+1}\left(f_{t+1}^b(\dot{X}),V_{t+1}^b(\dot{X})\right)}{d\dot{\rho}} = \underbrace{\frac{dAF_{t+1}\left(f_{t+1}^b(\dot{X}),V_{t+1}^b(\dot{X})\right)}{df_{t+1}^b(\dot{X})}}_{>0}\underbrace{\frac{df_{t+1}^b(\dot{X})}{\dot{\rho}}}_{>0} + \underbrace{\frac{dAF_{t+1}\left(f_{t+1}^b(\dot{X}),V_{t+1}^b(\dot{X})\right)}{dV_{t+1}^b(\dot{X})}}_{<0}\underbrace{\frac{dV_{t+1}^b(\dot{X})}{\dot{\rho}}}_{<0}$$

Conditional on $\pi_{t+1}(\dot{X}) > f_t^b(\dot{X})$ an increase in $\dot{\rho}$ increases $f_t^b(\dot{X})$ and reduces $V_t^b(\dot{X})$ and both of these effects increase $AF_t(\dot{X})$. Hence $AF_t(\dot{X})$ is increasing in $\dot{\rho}$. Finally, conditional on the realization of $\pi_{t+1}(\dot{X})$, being high enough, when $\dot{\rho}$ is high enough then $AF_t(\hat{X}) < AF_t(\dot{X})$ and the farmer switches from exploiting $\hat{X}$ to exploring $\dot{X}$.

Proposition 5: *When jumping into the wild, short-term losses in expected utility can be tolerated, and even more so for a HSF than a LSF.*

If a farmer decides to jump from exploiting $\hat{X}$ to exploring $\dot{X}$, it must be the case that $AF_t(\hat{X}) < AF_t(\dot{X})$. Since $\hat{X}$ is a stationary point its gain does not include any gain from information. Hence we have:

$$AF_t(\dot{X}) = \underbrace{E\left[u(\pi(\dot{X}))\right]}_{Exploitation\ gain} + \underbrace{\alpha_{Ft}\left[f_t^b(\dot{X}) + \lambda\, V_t^b(\dot{X})\right]}_{Exploration\ gain}$$

$$AF_t(\hat{X}) = \underbrace{E\left[u(\pi(\hat{X}))\right]}_{Exploitation\ gain}$$

And thus:

$$\underbrace{E\left[u(\pi(\dot{X}))\right]}_{Exploitation\ gain\ at\ \dot{X}} + \underbrace{\alpha_{Ft}\left[f_t^b(\dot{X}) + \lambda\, V_t^b(\dot{X})\right]}_{Exploration\ gain\ at\ \dot{X}} > \underbrace{E\left[u(\pi(\hat{X}))\right]}_{Exploitation\ gain\ at\ \hat{X}}$$

Which is totally consistent with

$$\underbrace{E\left[u(\pi(\dot{X}))\right]}_{Exploitation\ gain\ at\ \dot{X}} < \underbrace{E\left[u(\pi(\hat{X}))\right]}_{Exploitation\ gain\ at\ \hat{X}}$$

To the extent that:

$$\underbrace{\alpha_{Ft}\left[f_t^b(\dot{X}) + \lambda V_t^b(\dot{X})\right]}_{Exploration\ gain\ at\ \dot{X}} > \underbrace{E\left[u(\pi(\hat{X}))\right]}_{Exploitation\ gain\ at\ \hat{X}} - \underbrace{E\left[u(\pi(\dot{X}))\right]}_{Exploitation\ gain\ at\ \dot{X}}$$

In plain English, a farmer is willing to accept a drop in her profit to the extent that the loss in expected utility is at least compensated by the gain in utility from the information acquired.

Finally, since $\alpha_{HSF\ t} > \alpha_{LSF\ t}$ the first term is greater for a HSF than a LSF and thus the acceptable loss in expected utility is greater for a HSF than a LSF. In simple terms, since a HSF acquires more information when exploring, her acceptable loss is higher.

Proposition 6: *When exploring a new area, the belief about profit with the new package can, in expectation, move in a non-monotonic way*.

We model the adoption a new input package as a switch from $\hat{X}$ to $\dot{X}$ and the vicinity of $\dot{X}$. It is possible that as the farmer explores the vicinity of $\dot{X}$ it reaches a new local maximum $\ddot{X}$. In the case where $\dot{X} < \hat{X} < \ddot{X}$, then the beliefs would in expectation move first down and then up, and the farmer can continue exploring the new area despite $\dot{X} < \hat{X}$ because of proposition 5.

To put it in words, assume that a farmer starts using a new input package. The immediate return may be relatively low, because the farmer still needs to explore further the profit function to reach the global max. Hence the adjustment of expectations follows a moving target. In this case, the expected beliefs about the returns with the new inputs first decrease and then increase.

The proposition is relatively weak in the sense that it simply states that it is a possible scenario (as opposed to alternative standard models).

Proposition 7: *A LSF can obtain more valuable information from observing a HSF than a LSF.*

Again, the proposition is relatively weak because it simply states a possibility, making its proof very simple in such a broad model where a wide range of scenarios are possible. Homogamy among LSFs has 2 consequences: first similar farmers tend to have values of the input vector X that are more similar (including perhaps some values that are not malleable because the ability to execute them depends on skills). Second, because this other farmer is also a LSF, this other farmer would explore new areas less than a HSF (proposition 5). Which effect dominates is an empirical question. The more LSFs value "finding new equilibria" over exploring in the vicinity of their current practice, the more they are likely to value observing HSFs than LSFs.