FERDI
FONDATION POUR LES ÉTUDES ET RECHERCHES SUR LE DÉVELOPPEMENT INTERNATIONAL

Working Paper 163
September 2016
Development Policies

Review of Theories of Learning for Adopting

ELISABETH SADOULET

Elisabeth SADOULET, Professor at the University of California at Berkeley, Senior Fellow Ferdi.

Overview

The diffusion of a new agricultural technology requires farmers to learn about the existence and the benefits of the technology. What do they have to learn, how do they learn it, and from whom, is the subject of a large literature, both theoretical and empirical. The purpose of this brief is to review the most prominent learning models, briefly assess recent empirical results derived from these theories, and raise a few important remaining issues not explicitly addressed by the theories.

We will focus on the literature that refers to learning from experience, either own or that of others, giving prominence to the network of connections that farmers have. This review is purposefully very selective, with the objective of illustrating concepts and categories of models, rather than providing a genuine literature review.

.../...

Paper prepared for the Ferdi and SPIA workshop ***Learning for adopting: Technology adoption in developing country agriculture***, held in Clermont-Ferrand (France), June 1 and 2, 2016

LA FERDI EST UNE FONDATION RECONNUE D'UTILITÉ PUBLIQUE.
ELLE MET EN ŒUVRE AVEC L'IDDRI L'INITIATIVE POUR LE DÉVELOPPEMENT ET LA GOUVERNANCE MONDIALE (IDGM).
ELLE COORDONNE LE LABEX IDGM+ QUI L'ASSOCIE AU CERDI ET À L'IDDRI.
CETTE PUBLICATION A BÉNÉFICIÉ D'UNE AIDE DE L'ÉTAT FRANCAIS GÉRÉE PAR L'ANR AU TITRE DU PROGRAMME «INVESTISSEMENTS D'AVENIR»
PORTANT LA RÉFÉRENCE «ANR-10-LABX-14-01».

Review of Theories of Learning for Adopting

Elisabeth Sadoulet

July 10, 2016

1 Overview

The diffusion of a new agricultural technology requires farmers to learn about the existence and the benefits of the technology. What do they have to learn, how do they learn it, and from whom, is the subject of a large literature, both theoretical and empirical. The purpose of this brief is to review the most prominent learning models, briefly assess recent empirical results derived from these theories, and raise a few important remaining issues not explicitly addressed by the theories. We will focus on the literature that refers to learning from experience, either own or that of others, giving prominence to the network of connections that farmers have. This review is purposefully very selective, with the objective of illustrating concepts and categories of models, rather than providing a genuine literature review.

Models differ along four main dimensions: (i) what is to be learned, (ii) what is observed or transmitted by the social network, (iii) how do farmers aggregate the information that they receive from different sources, and (iv) what is the assumed network structure.

What is to be learned: In some models, farmers learn from others the simple facts of the existence or the adoption by others of a technology. This factual knowledge is simple: the information is either transmitted or not, and if transmitted, it contains no error or noise. This is the essence of a class of models called 'diffusion models' (reviewed in section 2). Being informed is a binary variable. Adoption is then a function of simply knowing about the technology or knowing people that have adopted it.

In other models, what farmers need to learn before deciding to adopt a technology is an expected profit or yield, or a (stochastic) optimal input to be used. This is substantially more difficult, as one never observes expected values, but only specific realizations of the stochastic variable. These realizations provide 'signals' on the underlying outcome of interest. We refer to these models as 'learning models' in section 3 below. A key assumption of most models is that there is no fundamental heterogeneity among farmers. Expected profit/optimal input are the same for all and the signal is unbiased.

A more realistic view of the world of heterogeneity in agricultural production suggests that what farmers should be learning is a more complex multivariate relationship, $\pi^*(x, z)$ between input and characteristics x, z, respectively, and outcome, here expected profit or yield, π^*. Only such models allow learning from cross-sectional (land quality, input use, farmer ability) or over time (function of weather realizations) heterogeneity. We will review one such model by Schwartzstein in section 3.4 below. The paper focuses on a specific aspect of the challenge facing farmers, called 'selective attention', which is to properly assess the set of covariates that matter in the relationship.

What is transmitted by the social network: In diffusion models, the information that is transmitted is without ambiguity. In more complex learning models, farmers can transmit to each other either the information they have about the technology or their own action regarding adoption (resulting from their net assessment of the information they have). The full information (including where it comes from beyond one's own experience) is more informative, but not as easily transmissible. A recent literature (described in section 5) addresses this question in the context of fully controlled experimental games. Throughout the empirical literature, authors have pointed out cases where farmers resist transmitting their decisions or all the information they have, which of course is only possible when it is not directly observable by others (Conley and Udry, 2010; Cai et al., 2015).

How do people aggregate the information they receive: Let's say that a few farmers have the information on or have adopted the new technology, and that the information starts diffusing in the network. Uninformed people receive the information or signals on the variable of interest, possibly from different sources. A key question is how do they aggregate these different sources of information.

In diffusion models, the information is simple, and once acquired is not reversible: once you are informed or have adopted, this is it. Different models however specify different rules by which information from others translate into being informed or adopting., e.g., a non-adopter will adopt as soon as he is in contact with a threshold number of adopters, or a fraction of his friends have adopted, either with certainty or with a certain probability.

In learning models, people receive signals on the outcome of interest, and use them to update their prior. Key issue is whether people aggregate following the sophisticated Bayesian rule, where different signals are weighted according to their probability of occurrence and precision, or whether people use more heuristic formulae, with some ad'hoc weighting schemes, as suggested by DeGroot. Another issue is whether people can recognize the origin of the information, and for example whether they can properly correct for a unique information that reached them through two different channels (e.g., a signal originating from person A, and transmitted to D by both B and C). While early models simply stated either a Bayesian or a DeGroot rule, recent empirical work described in section 5 use experimental games to study the issue.

What is the structure of the network: Most of the early literature assumed that networks are 'complete' in the sense that all links between the different members of the networks exist, and the network itself is defined by a large population, most often the village. When network are complete no one has any particular position. In contrast, the recent literature has paid attention to the structure of the network, i.e. the links that exist between any two members of the population. In such incomplete networks, different people have different ability to facilitate the diffusion of information, some people are more 'central' than others. Recent work described in section 4 show how the centrality concept is related to the diffusion model.

With learning models, where the information or signal (such as a realized profit) needs to be generated by each member of the network, the nodes of the network acquire some other important characteristics, such as their ability as demonstrators to develop useful information. In addition, links themselves may be of different strength, if for example information provided by more trusted members of the network is more persuasive. There is not much theoretical development on these aspects, but this is an important question with strong policy implications regarding the choice of whom to select as injection points for diffusing a new technology.

The paper proceeds as follows. We first review early models that assume full netwroks (section 2 and 3 for the diffusion and learning models, respectively) and then review models and empirical studies that are anchored in the specific structure of the network (in section 4 and 5 for the diffusion and learning models, respectively). In section 6, we ask the question of how to select the injection points in a network in order to maximize the diffusion of a technology, a policy relevant question addressed in many current empirical works.

2 Diffusion models

In diffusion models, people adopt when they come in contact with others who have already adopted. There is no explicit theory of learning, but a dynamic model of transmission of behavior. We follow Young (2009) in presenting the models in the context of a large population with random encounters.

Contagion models: In the simple contagion model, a non-adopter will adopt as soon as he encounters an adopter. Let $\lambda > 0$ be the instantaneous rate at which a current non-adopter 'hears about' the innovation from a previous adopter within the group, and let $\gamma > 0$ be the instantaneous rate at which he hears about it from sources outside of the group. In the absence of heterogeneity, the proportion of adopters in period t, $p(t)$, follows the differential equation:

$$p'(t) = (\lambda p(t) + \gamma)(1 - p(t)),$$

and the solution is

$$p(t) = [1 - e^{-(\lambda+\gamma)t}]/[1 + \frac{\lambda}{\gamma}e^{-(\lambda+\gamma)t}].$$

Individual are characterized by their individual values (λ, γ), λ is a sort of rate of social interaction with the rest of the population, and γ with the external world.

In complex contagion models individuals adopt if they are connected to at least a threshold number of adopters. A recent test of these models in the context of incomplete networks is proposed by Beaman et al. (2014) and discussed in section 4.

Social influence models: In these models, non-adopters are persuaded to adopt when a certain fraction of the population has already adopted, what has been called a 'conformity' motive. Each agent i is characterized by the minimum proportion $r_i \geq 0$ that needs to have adopted before he adopts. The parameter measures a degree of responsiveness to social influence. A key feature of such a model is that adoption depends on the innovation's current popularity rather than on how good or desirable the innovation has proven to be.

You need a group of people that are willing to adopt on their own, even without anyone else having adopted before them. After that, those with the lower level of r_i adopts first, and then on. Let $\lambda > 0$ be the instantaneous rate at which these people convert. Then the adoption process is described by the differential equation:

$$p'(t) = \lambda[F(p(t)) - p(t)],$$

where $F(.)$ is the cumulative distribution function of thresholds r in the population.

Susceptible infected models: In these models, being informed does not automatically translate into adoption, but only makes you 'susceptible' to adopt. Informed non-adopters only adopt with a certain probability, that possibly depends on their own characteristics.

For example, Banerjee et al. (2013) develop a model of information diffusion through a social network that discriminates between information passing (individuals must be aware of the product before they can adopt it, and they can learn from their friends) and endorsement (the decisions of informed individuals to adopt the product might be influenced by their friends' decisions). They apply it to the diffusion of microfinance loans, in a setting where the set of potentially first-informed individuals is known. The underlying model is as follows:

- An informed person transmits the information with probability q^P if he participates in the MFI, and q^N if he does not.

- An informed individual i decides to participate in the MFI with probability p_{it} at time t:

$$log\left(\frac{p_{it}}{1-p_{it}}\right) = X_i'\beta + \lambda F_{it}$$

  where F_{it} is the fraction of his informed network links that participate.

The model allows to estimate separately the information channels (q^P and q^N) and the endorsement ('action') channel (λ). They find no evidence of endorsement effect. And the estimates for the information transmission are $q^P = 0.35 to 0.50$ and $q^N = 0.05$. This suggests that in this context non-adopters have little influence, and transmission of adoption is quite partial.

3 Learning models

In this section we present examples of learning models. Each is built as a specific combination of the outcome to be learned, the information that is transmitted, and the aggregation rule used by the receiving agent to update his prior information. The first two models are the widely used targeted input model with Bayesian updating, and a model that illustrates a DeGroot aggregation mechanism. We then present three models that each focus on an additional aspect of the learning process: (i) a dynamic learning model, in which farmers can strategically adopt the new technology to increase learning, (ii) a model in which what is to be learned is a complex production function, and (iii) a model that points to the difference between the time series information collected by self experience over time and cross sectional information collected from experiences by others.

3.1 The Target Input Model: Bayesian learning based on unbiased signals reduces uncertainty and hence increases $E\pi$

This model is presented in Bardhan and Udry (1999) and used by BenYishay and Mobarak (2015). The production function is known to the producer with certainty, except for one parameter, usually conceptualized as the optimal input:

$$q_{it} = 1 - (k_{it} - h_{it}^*)^2 \tag{1}$$

where q_{it} is output or profit, k_{it} is input used, and h^*_{it} is the optimal ('target') input.

The optimal input level is subject to idiosyncratic variation μ_{it} around a mean value h^*, i.e., $h^*_{it} = h^* + \mu_{it}$, with $\mu_{it} \sim N(0, \sigma^2_{\mu i})$.

If h^* is known, maximization of expected profit leads to choosing $k_{it} = E_t(h^*_{it}) = h^*$ and expected profit is $\pi_{it} = 1 - \sigma^2_{\mu i}$. This variance $\sigma^2_{\mu i}$ is due to the inherent variation in conditions that implies that the optimal input cannot be known at the onset of period t. It can be specific to producer i. There is thus fundamental heterogeneity in expected profitability.

If, however, h^* is unknown, producers rely on beliefs about h^*, further reducing expected profit. Beliefs are modeled as a distribution of potential level for h^*. Say that producer i's belief at the beginning of year t is normally distributed $N(h_{it}, \sigma^2_{uit})$. Maximization of expected profit leads to choosing $k_{it} = E_t(h^*_{it}) = h_{it}$ and expected profit is $\pi_{it} = 1 - \sigma^2_{\mu i} - \sigma^2_{uit}$.

Updating from own experience. At the end of year t, producers can observe q_{it} and hence infer what should have been h^*_{it}. The useful information from that observation is what it tells him about h^*, since μ_{it} is structural. h^*_{it} is thus an unbiased signal of variance $\sigma^2_{\mu i}$ about h^*. Beliefs are updated in year $t+1$ as the posterior distribution of h^*, a normal distribution with mean and variance as follows:

$$
\begin{aligned}
h_{it+1} &= \frac{(1/\sigma^2_{uit})h_{it} + (1/\sigma^2_{\mu i})h^*_{it}}{(1/\sigma^2_{uit}) + (1/\sigma^2_{\mu i})} \\
\frac{1}{\sigma^2_{uit+1}} &= \frac{1}{\sigma^2_{uit}} + \frac{1}{\sigma^2_{\mu i}}
\end{aligned}
\tag{2}
$$

Producers choose to apply $k_{it+1} = h_{it+1}$ and expected profit is $\pi_{it} = 1 - \sigma^2_{\mu i} - \sigma^2_{uit+1}$. Hence as information accumulates, the uncertainty about h^* decreases and expected profit increases until it converges to the expected profit under perfect information about h^*.

Updating from others' experience. Suppose all farmers in the village have the same production technology, and that the optimal inputs are drawn from the same distribution, i.e., $h^*_{it} \sim N(h^*, \sigma^2_{\mu})$. When producer i observes the production of N_t farmers, he receives a signal $\bar{h}^*_t$, with variance σ^2_{μ}/N_t. The updating equations for his beliefs are thus:

$$
\begin{aligned}
h_{it+1} &= \frac{(1/\sigma^2_{ut})h_{it} + (N_t/\sigma^2_{\mu})\bar{h}^*_t}{(1/\sigma^2_{ut}) + (N_t/\sigma^2_{\mu})} \\
\frac{1}{\sigma^2_{ut+1}} &= \frac{1}{\sigma^2_{ut}} + \frac{N_t}{\sigma^2_{\mu}}
\end{aligned}
\tag{3}
$$

Adoption. Suppose producers had access to a perfectly known traditional technology with constant profit, and choose to cultivate with either the traditional or the modern technology. First note that once a producer switches to the new technology, he never returns to the older one. This is because in this model expected profit can only improve with more experience. If farmers are myopic, then they will switch to the new technology whenever they have learned enough (from the others) so that the expected profit of the new technology is higher than the profit of the traditional technology. If however they are forward looking, they will include the benefits of experimenting to acquire information and may adopt even when there is expected current loss, if it is lower than the discounted gain in future profitability (see a dynamic learning model in section 3.3 below). If producers are learning from each others, then who decides to experiment and who decides to wait for the others to experiment depends on the structure of their interactions. This model is formalized in Bandiera and Rasul (2006).

Adaptation in BenYishay and Mobarak (2015). Assume that the production function is the same for all farmers, $q_i = 1-(k_i-h^*)^2$. There is a common prior belief regarding the optimal input for the new technology which is distributed $N(0,\sigma^2)$. If a farmer uses the technology with $k=0$, the corresponding expected profit is then $q = 1-\sigma^2$.

Farmers are selected by the experiment to try the new technology. [If h^* was not stochastic, the experimenter would immediately learn about the true value h^*.] These selected farmers x then choose whether to communicate or not the information gained from their experience to others. Since all farmers have the same production function, the signal is unbiased. However, the precision of the signal received by another farmer θ has two components: (i) one, ρ is related to the cost $c(\rho)$ incurred by the sender, and (ii) a second related to the distance $|x-\theta|$ between the two farmers. Using the same notation as above, the signal received by θ on h^* is:

$$s_{x\theta} = h^* + \frac{|x-\theta|}{\rho}\epsilon_\theta$$

where $\epsilon_\theta \sim N(0,1)$. Farmer θ then updates his prior about h^* as follows:

$$\begin{aligned} E[h^*|s_{x\theta},\rho] &= \frac{(\rho^2/(x-\theta)^2)s_{x\theta}}{1/\sigma^2+\rho^2/(x-\theta)^2} \\ \frac{1}{var[h^*|s_{x\theta},\rho]} &= \frac{1}{\sigma^2} + \frac{\rho^2}{(x-\theta)^2} \end{aligned} \tag{4}$$

With this model, the distance between farmers produces an increase in noise (not in bias) of the signal. This noise, in turn, induces a reduction in expected profitability. Note that the precision on the prior σ could be farmer specific σ_θ, indicating farmer θ's own 'ability' for example.

3.2 Munshi (2004): DeGroot updating from observation of the network's average decision and outcome

Farmers have the choice between two technologies, a traditional technology, with a certain yield y_{TV} identical for all farmers, and a modern technology with higher but risky return. The risky yield y_{it} is written:

$$y_{it} = y(Z_i) + \eta_{it} \tag{5}$$

where the expected yield $y(Z_i)$ is function of the farmer's characteristic Z_i and the stochastic term η_{it} is of mean 0 and variance λ_i^2. Note that both expected value and variance of yield are farmer specific.

If the farmer had perfect information, he would choose to allocate its land between the two crops, maximizing utility over the return. Under standard hypotheses, this would lead to land allocated to the new crop to be increasing in expected return and decreasing in the variance of return, i.e.:

$$A_i^* = A(y(Z_i) - y_{TV}, \lambda_i) \tag{6}$$

If the farmer does not know $y(Z_i)$, he uses an estimate $\hat{y}_{it}$ with variance σ_{it}^2, and the land allocation is:

$$A_{it} = A(\hat{y}_{it} - y_{TV}, \lambda_i, \sigma_{it}) \tag{7}$$

At the end of the season, the farmer obtains a realized yield y_{it}.

Timing of decisions and information flows are as follows: Farmers receive private signals. Based on these they update their own yield expectation ($\hat{y}_{it}$) and decide how much to plant, A_{it}. Yields are then realized.

How are these yield estimates $\hat{y}_{it}$ formed?

Social learning when conditions are identical across farmers

Expected yield is the same for all farmers, and information from neighbors are just as good as information from one's own field. Each farmer transmits two pieces of information: Planting decision, which reveals the private signal he received, and realized yield which provides another signal on expected yield. We assume that farmers share a common knowledge $\hat{y}_t$ which they each combine with the personal signal u_{it}. The updating of the common knowledge is based on the new information received by the village, i.e., the average of all signals received by individuals and their realized yields. This gives:

$$\hat{y}_{it} = \alpha \hat{y}_t + (1 - \alpha) u_{it} \tag{8}$$

$$\hat{y}_t = (1 - \beta - \gamma)\hat{y}_{t-1} + \beta \bar{u}_{t-1} + \gamma \bar{y}_{t-1} \tag{9}$$

Using a linear function for (7), the law of motion of land allocation thus becomes:

$$A_{it} = \pi_0 + \pi_1 \hat{y}_{it} + g(X_i, \sigma_{it}) \tag{10}$$
$$= \pi_0 + \pi_1 \alpha(1 - \beta - \gamma)\hat{y}_{t-1} + \pi_1 \alpha\beta \bar{u}_{t-1} + \pi_1 \alpha\gamma \bar{y}_{t-1} + \pi_1(1-\alpha)u_{it} + g(X_i, \sigma_{it}) \tag{11}$$

$\bar{u}_{t-1}$ and $\hat{y}_{t-1}$ are not observable to the farmer, but can be shown to be function of $\bar{A}_{t-1}$ and A_{it-1}, so that :

$$A_{it} = \eta_0 + \eta_1 A_{it-1} + \eta_2 \bar{A}_{t-1} + \eta_3 \bar{y}_{t-1} + \epsilon_{it} \tag{12}$$

A_{it-1} contains all the information about the expected yield that was available at the beginning of period $t-1$. Conditional on A_{it-1}, $\bar{A}_{t-1}$ represents the new information that was received by the village in period $t-1$ through the exogenous signals. Similarly, $\bar{y}_{t-1}$ represents the information that was obtained from the yield realizations in that period.

In the language of the current network theory (section 5), this model is about the aggregation function. The network is implicitly defined as being the whole village sufficiently connected that everyone knows what everyone else does. There is transmission of both the action (area planted based on the signal received) and the information (the obtained yield).

Social learning when conditions vary across farmers

Without a formal analysis, Munshi's assessment of how the model applies when there is heterogeneity is as follows: "The grower could condition for differences between his own and his neighbors' observed characteristics when learning from them. But the prospects for social learning decline immediately once we allow for the possibility that some of these characteristics may be unobserved, or imperfectly observed. Mistakes that arise because he is unable to control for differences between his own and his neighbors' characteristics when learning from their yields are persistent, and therefore more serious. Take the case where all the neighbors' characteristics are unobserved by the grower. He now has two choices. He could rely on his own information signals and yield realizations, ignoring information from his neighbors. Consistent but inefficient estimates of the expected yield would be obtained with such individual learning. Alternatively, he could continue to utilize information from his neighbors, measured by the mean acreage and the mean yield, as before. The efficiency of his estimates increases with social learning since more information is being utilized, but some bias will inevitably be introduced since the grower cannot control for variation in the underlying determinants of the yield when learning from his neighbors. The grower will ultimately choose between individual learning and social learning on the basis of the trade-off between bias and efficiency.”

3.3 Dynamic learning model, with strategic adoption to increase learning

This model is presented in Besley et al. (1994). The authors develop a dynamic model of learning, where individuals are forward looking and Bayesian. The returns to technology adoption are twofold: it could affect current profits; and it could also induce learning about the value of this technology (information), which is a public good and will pay off when future decisions are made.

In this framework, uncertainty about a new technology can be represented by a state variable α, which can be perceived as the increased profitability from adoption. There are M farmers indexed by i. Each farmer has N_i fields, and he has to choose how many to sow to the new variety (the new technology) at each date t. His current expected payoff from sowing n_{it} fields to the new technology is $f^i(n_{it}, \alpha_t)$, where α_t is the the belief about α at time t. $f^i(n_{it}, \alpha_t)$ is assumed to be increasing, twice differentiable with $\partial^2 f/\partial n_{it}\partial\alpha_t > 0$. Uncertainty in this model comes from the fact that people could not precisely estimate the effect of the technology, but instead only evolve a belief based on past experience, which is represented by a conditional distribution function $H^t(\alpha_{t+1}|\alpha_t, \sum_i n_{it})$.

Given the setup, a farmer's sowing decision can be described by:

$$W_t^i(\alpha_t, \sum_{j\neq i} n_{jt}) \equiv \max_{n_{it}}\{f^i(n_{it}, \alpha_t) + \beta \int V_{t+1}^i(\alpha_{t+1}) dH^t(\alpha_{t+1}|\alpha_t, \sum_{j=1}^{M} n_{jt})|n_{jt} \leq N_i\} \tag{13}$$

where β is the discount factor and $V_t^i(x)$ is the value function, defined as the value of entering period t with state variable x:

$$V_t^i(x) \equiv W_t^i(x, \sum_{j\neq i} n_{jt}(x)) \tag{14}$$

Farmers are assumed to be risk-neutral, and there exists a trade-off between current profitability and the value of learning that arises through the dependence of future beliefs on sowing decisions. Information is a public good and there exists externalities of technology adoption (every farmer's decision affects the conditional distribution function of beliefs about technology), so every farmer's sowing decision should be conditioned on that of all other farmers. The Nash equilibrium is a vector of sowing decisions: $\{n_{1t}^*(\alpha), ..., n_{Mt}^*(\alpha)\}$. As the state variable α_t evolves over time, the farmers reach a succession of Nash equilibria conditioned on the value of α_t in each period t. Therefore, this sequence gives us a Markov

Perfect Equilibrium.

For comparison, the authors also consider two further cases: when learning is undertaken cooperatively by the farmers and when farmers are myopic. In the first case, farmers maximize joint profit, so the problem is no longer M farmers choosing M variables, now only one decision is made, in which the total number of fields sown to the new technology, n_t, is chosen. The model could be further extended such that the planner in this cooperative case may also choose how to allocate the sowing decisions across farmers, making use of side payments to bring about the planning allocation. In the second case, the farmers are assumed to be myopic, so their decisions are based solely on current profitability. This corresponds to $\beta = 0$ in the model. If this is the case, then whether farmers are cooperative or non-cooperative no longer makes a difference, since coordination behavior affects only expected future payoffs.

Bandiera and Rasul (2006) also present a version of this strategic dynamic model. It is based on the targeted input model. Their empirical analysis studies the adoption of sunflower by farmers in the Zambezia region of Northern Mozambique. It is based on cross-sectional data on 198 household heads from 9 villages. Each farmer was asked how many of the people they know have adopted, and how many of those are family or friend. They find the relationship between farmers' decisions to adopt and the adoption choices of their network of family and friends to be inverse-U shaped, suggesting social effects are positive when there are few adopters in the network, and negative when there are many. They also find that (i) adoption decisions of farmers who have better information about the new crop are less sensitive to the adoption choices of others., and (ii) adoption decisions are more correlated within groups of family and friends than in religion-based networks, and uncorrelated among individuals of different religions. Note however that all these results are correlations as there is no identification strategy. Their argument is that contextual effects and mimicry would create positive correlation. So finding an inverse U-shape suggests this is not all.

3.4 Learning a complex relationship: the problem with selective attention

Schwartzstein (2014) explicitly considers the case where what is to be learned is a multivariate relationship between inputs and output. The angle that the author considers is the issue of the bias and distortion that occur in the learning process if the farmer fails to recognize all the necessary dimensions of the production function. The paper presents a model of selective attention: an agent learns to make forecasts based on past information, but is selective as to which information he pays attention to.

Specifically, the agent wants to accurately forecast y given (x, z), where y is a binary variable and x and z are finite random variables. In each period t, the agent observes a random draw of (x, z), (x_t, z_t), from a fixed distribution $g(x, z)$; then he gives his prediction of y, $\hat{y}_t$, to maximize $-(\hat{y}_t - y_t)^2$; then he learns the true y_t. The agent knows that given (x, z), y is independently drawn from a Bernoulli distribution with fixed but unknown success probability $\theta_0(x, z)$ in each period: $p_{\theta_0}(y = 1|x, z) = \theta_0(x, z)$. He also knows the joint distribution $g(x, z)$, which is positive for all (x, z).

The author assumes that z is important to predicting y, while x is important to predicting y in the absence of conditioning on z (there could be cases where x is no longer predictive once we control for z). The agent does not know the functional form of the success probability θ_0. To estimate this function, he needs to (i) choose the model (i.e., decide whether x and/or z are important) and (ii) estimate the parameters that he thinks are important using a standard Bayesian approach. Let $M_{i,j}$ where $i \in \{X, \neg X\}$, $j \in \{Z, \neg Z\}$ designate the four potential models. And let $\pi_X(\pi_Z) \in (0, 1]$, be the subjective prior probability that $x(z)$ is important to predicting y. The learning process is a standard Bayesian one. The history through period t is denoted by:

$$h^t = ((y_{t-1}, x_{t-1}, z_{t-1}), (y_{t-2}, x_{t-2}, z_{t-2}), ..., (y_1, x_1, z_1)) \tag{15}$$

So the agent updates his beliefs about the model and about the parameters based on history, and uses the updated belief to forecast. In period t, his forecast of y given x and z can be written as:

$$E[y|x, z, h^t] = E[\theta(x, z)|h^t] = \sum_{i,j} \pi_{i,j}^t E[\theta(x, z)|h^t, M_{i,j}] \tag{16}$$

where $\pi_{i,j}^t \equiv Pr(M_{i,j}|h^t)$ equals the posterior probability placed on model $M_{i,j}$.

It follows that if the agent is Bayesian and has access to full history h^t at each date, then he should make asymptotically accurate forecasts, and he should learn the true model. Therefore, in this setting, any deviations from such perfect learning must stem from selective attention (the agent fails to pay attention to a variable in certain periods, so could not recall it later).

Standard Bayesian approaches assume that the agent perfectly encodes (y_k, x_k, z_k) for all $k < t$. But if the individual is "cognitively busy" in a given period k, he may not attend to and encode all components of (y_k, x_k, z_k) due to selective attention. Intuitively, this can be thought of as the agent sorting into his memory, and only remembering the elements that were perceived to be important. Therefore, at date t, the agent may only have access to an incomplete mental representation of true history h^t, which is denoted by $\hat{h}^t$.

The author makes several assumptions to put structure on $\hat{h}^t$. Basically, the agent always encodes x and y, so selective attention only applies to z. And the likelihood that the agent attends to z is increasing in the current probability that he thinks z is predictive for y. In addition, the author assumes that the agents are naive: when they cannot recall the z in the past history, they recall such missing information as a fixed but distinct non-missing value. This assumption is important in generating the main results of the paper.

One of the main proposition derived from the model is that if the agent settles on encoding z, he learns the true model almost surely; if the agent settles on not encoding z, he does not learn the true model. The intuition is that when he encodes z, this is identical to standard Bayesian process, so he learns the true model. But if he does not encode, he believes that x is important to predicting y (by assumption, x predicts y when not conditioning on z), and fails to realize that this is driven by his ignorance of z due to the naivete assumption (the agent treats missing values of z as non-missing distinct values). This result means that in some cases, the agent interprets correlative relationships as causal, and he makes such errors persistently because he has selective attention and could not recall the complete history.

Using this framework, Hanna et al. (2014) suggest that failing to notice a gap between knowledge and actual practice, and not the information set itself, may be a key barrier to learning. They show that seaweed farmers in Kenya acted on the information received only when it included descriptions of the relationship between yield and pod size from their own plot.

The strong effect of *field visits* in inducing demand for the new rice variety shown in Emerick et al. (2016) may be interpreted as an opportunity to point to some of the benefits of the new technology and/or how to use it. In which case it would help counteract these failure to notice by making them salient.

3.5 Private learning from time series vs. social learning from cross sectional observations

An interesting point made by Wang et al. (2013) is that private learning proceed from the observation of time series of realized events, while learning from others is based on cross-sectional observations of stochastic events. The authors build a model where farmers consider an investment project, whose value function follows a geometric Brownian motion (a continuous-time stochastic process widely used in finance). Departing from the standard learning framework, the authors assume that a key parameter (the drift rate) of the Brownian motion is unobservable to the farmer (parameter uncertainty). Therefore, the farmer is imperfectly informed about the expected rate of return, which he has to figure out in order to decide the optimal timing of investment. Learning then happens in two ways: (1) private learning, by extracting information on the true drift from a continuous observation of past realized returns on the project value. (2) Social learning, by obtaining discrete noisy signals of the true drift (learning from early adopters in the farmer's social network).

Unfortunately, the authors do not further elaborate on the distinction between the two types of variability. Thinking of agricultural production, it is quite clear that the cross sectional variability of yield across farmers is not at all the same concept as the variability over time. This issue is common to all the models that compare or combine learning from one self and learning from others.

The empirical analysis seems to have lost the interesting distinction between the two learning processes. It is a simple censored tobit model for the time it takes to adopt. They conclude that social learning has a significant positive impact on greenhouse adoption: 10 more adopters in the farmer's social network increase the probability of adoption by 32%, which is an economically significant effect. Moreover, results from the duration analysis confirm this finding with social learning reducing the waiting time significantly in greenhouse adoption.

4 Diffusion models in incomplete networks: The key role of injection points

While the basic diffusion and network models reviewed in section 2 refer to the social network as an important source of information, these networks are relatively unspecified: They are generically referred to as the village population, and assumption is that everyone is equally connected to everyone in the network. In the real world however, networks have structure (or topology). They consist in the set of links that exist between the members of a given population. In these 'incomplete' webs of relationships, the diffusion process depends on where the entry points for the information/adoption are in the network.

With diffusion depending on the diffusion model, the definition of links in the network, and the entry points in a non-separable way, testing for the diffusion model becomes intrinsically linked to the choice of injection points. Beaman et al. (2014) addresses exactly this issue. The different diffusion models of agricultural technology they consider are: 1) simple contagion model with network links defined from a survey; 2) complex contagion model with network links defined from a survey; and 3) complex contagion model with network links define by geographical proximity. For each of these three cases, optimal injection points are selected based on network simulation. The control arm is defined by the status quo, i.e., entry points are the extension workers' choice. The authors then compare rates of diffusion, and find: a) farmers chosen by network theory yield greater adoption rates over three years; b) the learning environment is more consistent with the complex contagion model where farmers need to know more than one person with the new technology to decide to adopt. That is, "the complex contagion model with optimal entry points" performs better than any other model with its associated optimal injection points.

Banerjee et al. (2013) start from a different diffusion model, the susceptible infected model described in section 2, where information is transmitted though active links, one leg per period of time, and informed people decide whether to adopt based on their own characteristics and the adoption rate among their informed network neighbors. After having estimated the model parameters, they can compute a measure of communication centrality for each leader (injection points). This is defined as the fraction of households who would eventually participate if this leader were the only one initially informed. To compute this fraction, they simulate the model with information passing and participation decisions being governed by the estimated values of q^N, q^P, β. Finally, they develop an easily computed proxy for communication centrality, which they call diffusion centrality.

5 Learning models in incomplete networks: Diffusion and aggregation of information

Learning includes three elements: what <u>information</u> is transmitted from one person to the next (either the belief, typically a probability, or the action taken based on that belief), the <u>diffusion</u> of information within the network, and the <u>aggregation</u> of received signals.

As seen above, diffusion models assume that what is transmitted is the action ('adoption'), and that it is passed along one link. The diffusion models then specify different aggregation functions: The contagion models specify that adoption will take place if at least a threshold number of network neighbors have adopted; the social influence model is a certain fraction of the network neighbors that have adopted.

In a series of recent articles (Chandrasekhar et al., 2012; Grimm and Mengel, 2014; Mobius et al., 2015), researchers have resorted to lab experiments to better understand the diffusion and aggregation of information in networks. These experiments are about the discovery of one truth (among several options), and whether the learning process converges to the truth. So for example in Chandrasekhar et al. (2012), the world (a bag containing 7 balls) is either blue or yellow. In the blue bag there are 5 blue balls and 2 yellow balls, with the reverse in the yellow bag. There are 7 participants. Each participant receives a signal (blue or yellow), with a 5/7 probability that the signal is correct. Each participant only receives one signal and then relies on additional information from its limited network. Each individual's initial best guess of the color of the world is 1/2 (since the bag was randomly selected). After receiving the signal and collecting information from his network, each individual provides a new assessment of the color of the world. These second round guesses are transmitted through the networks, leading to a third round set of assessments and guesses, etc.

5.1 Bayesian vs. DeGroot aggregation of information

In Chandrasekhar et al. (2012), the network transmits the best guess of each individual (and not the information that served to establish it, nor the mechanism by which the person aggregated this information): This is an "action" model. The diffusion along the network is perfect, as it is done by the experiment itself. What varies is the structure of the network. What the paper is trying to uncover is the aggregation rule used by the subjects in the experiment. Specifically, are they Bayesian (the aggregation rule is a Bayesian updating of their belief) or DeGroot (aggregation is some weighted average of their own and their network's past actions, with ad'hoc weights). The way this is done is by simulating the outcomes that we should observe under a number of scenarios: all are Bayesian, all are DeGroot, a certain fraction are Bayesian and this is common knowledge, all are Bayesian but they don't know what others are, etc. The authors find that it is the "all DeGroot" model that comes closest to what is observed.

Why is this important? Any model but a correct Bayesian model can lead to some cluster of participants being stuck in error (because they initially received some wrong signals, which are never properly reassessed with correct (Bayesian) weights).

Grimm and Mengel (2014) present a horserace between the Bayesian and naive DeGroot models of learning in an experimental game similar to Chandrasekhar et al. (2012). All games are with 7 players, with 3 different network structures (circle, star, and kite), 2 different initial signal distributions (more or less clustered), and 3 degrees of information given to participants on the network structure. They ask whether agents reach a consensus, and if so whether they agree on the correct truth, and how long it takes them. They can predict the outcome under perfect information with each of these two rules, and find that the naive model is a better predictor of *individual* decisions than the Bayesian model. However this model fails to predict the overall network performance (in terms of convergence, convergence to the correct answer, and speed of convergence), so it seems that the equal weights of the pure naive model do not represent what people use.

An interesting part of the paper estimates the empirical aggregation rule, i.e., the weights $\lambda_{ij}(t)$ given by each participant i to each of his network member j over time in the following model:

$$g_i^t = 0 \qquad \text{if and only if} \qquad \lambda_{ii}(t) g_i^{t-1} + \sum_{j \in N_i} \lambda_{ij}(t) g_j^{t-1} < \frac{1}{2} \tag{17}$$

where $g_j^t \in \{0, 1\}$ is the guess by player j at time t about the correct urn.

This of course requires observing interactions between the same players in multiple games. Further, they analyze the estimated weights λ_{ii} that players give themselves as function of their network position, etc. They find that people put more weight on themselves than the pure naive model. This leads them to define some alternative model for the weights that depends on the position in the network and the degree of clustering in the network, etc., and that nests the naive rule and can approximate the Bayesian rule. The model is then estimated in more complex networks (rectangle and pentagon). Note that the paper offers a good literature review on experimental papers testing these network learning models.

5.2 Diffusion of information

In Mobius et al. (2015), participants again have to discover a truth (the answer to three binary-choice questions). The pool of players is a group of 800 students from which the network of up to 10 best Facebook friends was elicited. The game is played on line. Participants receive an initial signal with three suggested answers, and are told that all participants received independent signals, correct in 60% of the case. They make a first choice. They are then encouraged and incentivized to talk to as many people as they want from the group of people playing the game (they can search for participants). They can update their choice as often as they want. Whenever they submit a choice, they also have to record the name of all the people they talked to since the last submission. The experiment thus provides the full endogenous network of conversation links with a time stamp. The paper addresses the two questions of <u>diffusion</u> and <u>aggregation</u>. On diffusion, the authors show that the information does not travel beyond two nodes. In terms of aggregation, they examine whether people are aware that there is some double counting in the signal received. This is for example the case if you get information from two people (A and B) who both had talked to a common third (C) person. This is done by estimating the weight given to information that itself contains different information. For example they find that the weight given to a direct contact is not influenced by the number of paths to it (either several conversations with the same person or an indirect link in addition to the direct link), but that weights given to an indirect partner (C in the example above) does depend on how many paths you have to this partner (two in the case described). This is very plausible if your direct contacts did not tell you whom they themselves were influenced by. Similarly, in Cai et al. (2015), people who had a direct experience of insurance (either receiving or not a payout) are not influenced by the others.

The paper discusses the 'tagged' model which is when there is transmission not only of the signals but also of where it comes from, which allows the recipient to properly avoid double counting, for example if the same information reaches you through two channels.

5.3 What do networks transmit? Information vs. action

In Tjernström (2015), the author conducts a RCT in rural Kenya which explicitly elicits farmers' experiences with the technology to examine the influence of social networks on knowledge about and adoption of a new agricultural technology. Specifically, they randomly select treated villages, in which some farmers (directly treated) receive small packs of a new maize variety and conduct on-farm trials with the seeds; their fellow villagers (indirectly treated) only have access to information about the seeds through their social networks. No intervention is conducted in the control villages. In the treated villages, the author obtains the directly treated farmers' evaluation of how well the on-farm experiment went, and assumes that the signal that a given farmer receives about the new technology is a function of the distribution of these evaluations in his information network. This design could help separate two competing theories in the network and adoption literature: if social pressure is the main reason for adoption, then the number of treated links should largely explain the adoption decisions; if learning actually causes adoption, then farmers should respond to the actual evaluation of the new seeds by their network.

The author finds that networks transmit information (as opposed to 'action') and affect respondents' willingness to pay for the variety: the indirectly treated farmers respond strongly to the signals available in the network, above and beyond the impact of the number of treated links in their network. She also finds that the observed social network effects are weaker in villagers where soil quality is more varied (greater heterogeneity), which illustrates how heterogeneity in returns can handicap network effects. This further confirms that the observed network effects come from learning rather than imitation.

In Miller and Mobarak (forthcoming), the authors design a two-stage RCT to study the adoption of non-traditional stoves in Bangladesh. They promote two types of stoves: "efficiency" stoves whose effects are less observable ex-ante; and "chimney" stoves whose effects are more observable ex-ante. Based on ex-post feedback, "efficiency" stoves are not useful, while "chimney" stoves are useful. These two types of stoves therefore also allow them to study the heterogeneous learning effects caused by positive and negative information.

In the first period, the authors randomly publicize whether or not the local "opinion leaders" choose to order the non-traditional stoves, and look at the effects of this information on the adoption decisions of other households. They find that villagers' adoption decisions are affected by the decisions of opinion leaders, and the effects are stronger for the less observable "efficiency" stoves. Also, the results are more salient for negative information as compared to positive information.

In the second period, the authors study how the first period adoption decisions would affect the decisions of other households in the same network. The difficulty of studying this questions is that it is hard to distinguish social learning from common unobservable shocks faced by network members. To address this issue, the authors randomly assign subsidies to induce stove adoption in the first period, which creates exogenous variation in stove adoption and allows them to study whether the presence of network members who are stove owners (causally) affects other households' subsequent propensity to purchase stoves. They find that for both stove types, social ties to first-round participants reduce the likelihood that second-round participants purchase any stoves, suggesting that all villagers were overly optimistic about the effect initially. This negative social network effect is much larger for the "efficiency" stoves, which have been proved to be not useful.

5.4 The difference between lab experiment and real world network diffusion

While the lab experiments provide rigorous tests on how networks function in their own context, it is unclear how much their insights can be extended to a real world situation. This is because there are at least two fundamental differences between lab experiments and the real world that are more than a question of degree or simplification.

- In experimental games, players seek the information (they know that they need it and know where it is available). In the real world of agricultural extension, agricultural officers or selected experimentators know that the information exists but have no incentive to push it onto others (BenYishay and Mobarak, 2015), while farmers know their problems, but they don't know whether a solution exists for any particular one, and as a consequence are not likely to be seeking it.

  Do we expect the info on agricultural technology to be pulled or pushed?
  Can/should we facilitate this communication, e.g., through farmer field days as in Emerick et al. (2016))?

- In real world settings, the value of the information may depend on the person that transmits it (quality of signal, trustworthiness). Hence weights correspond to some underlying relationship between people.

5.5 Network learning when there are heterogeneous benefits

In Magnan et al. (2015), the authors study how social learning influences demand for a resource-conserving technology (Laser Land Leveling or LLL) in India. They design a RCT with two components: (1) a pair of binding experimental auctions for LLL custom service hire held one year apart, and (2) a lottery to determine who among the winners of the first auction would actually adopt the technology. The auctions capture demand for LLL before and after its introduction, allowing the authors to compare the benefits of LLL that farmers perceive to the actual benefits before and after any social learning takes place. The lottery generates exogenous variation in the number of adopters in each farmer's network, allowing them to estimate network effects. This randomization also allows them to estimate the benefits of the technology within the sample. The main point of the paper is to show the effect of heterogeneity in benefits in the diffusion of the technology.

Their results demonstrate some important nuances in how social networks drive technology adoption. On average, LLL reduced water use by 25% within the sample and appears to be profitable for 43 to 59% of farmers at the likely market price. However, in the first auction only two percent of farmers bid at or above this price, indicating that although the technology would benefit many farmers, potential benefits were initially not widely-appreciated by farmers. The authors find strong evidence that farmers learned about LLL benefits over the course of the study, and their demand in the second auction reflects this. Having a benefiting in-network adopter increased WTP by over 50%, equivalent to a 32% subsidy of the likely market price. Adjusting initial demand for LLL by this mean network effect indicates that for 39% of farmers network effects could incite adoption. However, not all farmers receive this network effect because networks are sparse and the technology is not profitable for all farmers. Consequently, the authors calculate that in a village where 12% of farming households initially adopt LLL at a discounted price, network effects would increase adoption by 9% the following year.

6 Learning models in incomplete networks: What are the optimal injection points?

The choice of injection points must depend on two factors: what information is to be transmitted and what the diffusion process is.

1. If what is being transmitted is simple information (such as the existence of a product), all that matters is the injection point's network position in terms of the diffusion model.

For example Beaman et al. (2014) defined by simulation the optimal entry points for either a simple or complex contagion model, based on a complete map of network links.

2. If we are interested in the transmission of adoption decisions rather than information, it is likely to be quite imperfect, in that an informed person may only adopt with a certain probability. If the probability is constant (as in the standard Susceptible-Infected model), then transmission is lower but the choice of entry points is unaffected. If however different people have different propensities to convert information into adoption, then the choice of entry points would need to take into account the network structure in terms of both links and adoption propensities.

3. More generally, if what needs to be transmitted is information about the net benefit of a technology, the entry points need to be both good "demonstrators" and good "communicators" for the first round of information transmission to begin. The proper balance between these qualities obviously depends on the product.

4. Finally, the benefits may be heterogeneous in the population. One would need a characterization of the source of heterogeneity to model the diffusion process and the corresponding optimal choice of entry points

Beaman et al. (2014) and Banerjee et al. (2013) simulated models give example on how to deal with cases 1 and 2 above. The selection of best demonstrators could be addressed with the use of 'selective trials', as proposed by Chassang et al. (2012) and Jack (2013). It consists in using a bidding mechanism (uniform price, sealed bid procurement auction), to select the recipients with the highest expected return to what is offered.

In term of empirical work, several papers report on RCTs, where the new technology was introduced in a village through varying entry points. BenYishay and Mobarak (2015) test the influence of three different communicators: 1) extension agents, 2) lead farmers who are more educated and can afford the new technology; and 3) peer farmers who represent the general population of target farmers. They find that peer farmers with small performance-based incentives are most effective in promoting new agricultural technology. Without incentives, peer farmers do not learn about the new technology or put effort in dissemination. This is a point made in Kondylis et al. (2014) where they implemented a randomized training of the "contact" farmers (CF) to study the impact on diffusion of the new technology. They find that directly training CFs leads to a large, significant increase in CF adoption, with no immediate increase in knowledge. Higher levels of CF adoption have limited impact on the behavior of other farmers.

Emerick et al. (2016) study the diffusion of new rice varieties, attempting to contrast injection points selected in three different ways: (i) by the village official/elite, (ii) in a village meeting, and (iii) by the women Self-Help Group. While these different injection points are notably different in their observable characteristics, they find no difference in diffusion one year later.

There are a number of less well identified analyses that address the same issue:

- Maertens (2015) looks at the diffusion of Bt cotton in India, based on recall data on when each farmer started to use Bt cotton. She finds that farmers appear to be exclusively learning from, and free-riding on the experimentation of, a small set of "progressive" farmers in the village.

- Genius et al. (2014) find that extension services and learning from peers are complement.

- Feder and Savastano (2006) review the literature on the characteristics and impact of opinion leaders on the diffusion of new knowledge, concluding that there is no clear evidence on whether opinion leaders are more effective if they are similar in socio-economic attributes to the other farmers rather than superior to would-be followers. A multivariate analysis of the changes in integrated pest management knowledge in Indonesia among follower farmers over the period 1991-98 indicates that opinion leaders who are superior to followers, but not excessively so, are more effective in transmitting knowledge. Excessive socio-economic distance is shown to reduce the effectiveness of diffusion.

7 Conclusion

We conclude with mention of a few salient unresolved issues.

What is to be learned may vary from simple information on the existence of a technology or its adoption by others, to more complex information such as the expected benefit of the technology, or even the relationship between key characteristics and inputs and expected profits. In the more complex cases, the information that is transmitted in the network is only a signal on the outcome of interest, which is then used to update priors. A key unresolved issue is how to deal with heterogeneity in benefits. What value does a signal have if it is biased, and the extent of the bias is unknown? Are some outcomes less heterogeneous than others, and hence more amenable to be usefully transmitted in networks? Are there ways by which the heterogeneity of outcomes can be made more transparent and transmissible? Can the transmission of a relationship characterize this heterogeneity?

How does the information circulate and is aggregated through the network? We have seen the first couple of tests and estimations of diffusion models. We are still far from specifying and testing models of diffusion and aggregation of more complex information, such as signals on expected benefits or the distribution of benefits, or on conditional expected benefits.

Who should generate the information/signals on what is to be learned? This is the key question of the choice of optimal injection points in social networks that would maximize the diffusion of the new technology. In diffusion models, people are solely characterized by their position in the network, and the objective is to define the most 'central' person in relation to the specific diffusion model. However, when signals have to be generated, like in most learning models, the quality of the person as experimenter also matters. The best experimenters are those that generate the most useful information for the others. Are they the best farmers, the median farmers, etc.? There is also potential tension between who are the best "diffusers" (in terms of centrality for the diffusion) and who are the best "demonstrators".

References

Bandiera, Oriana and Imran Rasul, "Social Networks and Technology Adoption in Northern Mozambique," *The Economic Journal*, 2006, *116* (514), 869–902.

Banerjee, Abhijit, Arun G Chandrasekhar, Esther Duflo, and Matthew O Jackson, "The diffusion of microfinance," *Science*, 2013, *341* (6144), 1236498.

Bardhan, Pranab and Christopher Udry, *Development microeconomics*, Oxford University Press, USA, 1999.

Beaman, Lori, Ariel BenYishay, A. Mushfiq Mobarak, and Jeremy Magruder, "Can Network Theory Based Targeting Increase Technology Adoption?," 2014.

BenYishay, Ariel and A Mushfiq Mobarak, "Social Learning and Incentives for Experimentation and Communication," 2015. Unpublished, Yale University.

Besley, Timothy, Anne Case et al., "Diffusion as a learning process: Evidence from HYV cotton," Technical Report, Princeton University 1994.

Cai, J, A de Janvry, and E Sadoulet, "Social Networks and the Decision to Insure," *American Economic Journal: Applied Economics*, 2015, *7* (2), 81–108.

Chandrasekhar, Arun G, Horacio Larreguy, and Juan Pablo Xandri, "Testing models of social learning on networks: Evidence from a framed field experiment," *Work. Pap., Mass. Inst. Technol., Cambridge, MA*, 2012.

Chassang, Sylvain, Padró I Miquel, and Erik Snowberg, "Selective Trials: A Principal-Agent Approach to Randomized Controlled Experiments," *American Economic Review*, 2012, *102* (4), 1279–1309.

Conley, Timothy G and Christopher R Udry, "Learning about a new technology: Pineapple in Ghana," *The American Economic Review*, 2010, pp. 35–69.

Emerick, Kyle, Alain de Janvry, Elisabeth Sadoulet, and Manzoor H Dar, "Enhancing the diffusion of information about agricultural technology," 2016. Unpublished.

Feder, Gershon and Sara Savastano, "The role of opinion leaders in the diffusion of new knowledge: The case of integrated pest management," *World Development*, 2006, *34* (7), 1287–1300.

Genius, Margarita, Phoebe Koundouri, Céline Nauges, and Vangelis Tzouvelekas, "Information transmission in irrigation technology adoption and diffusion: social learning, extension services, and spatial effects," *American Journal of Agricultural Economics*, 2014, *96* (1), 328–344.

Grimm, Veronika and Friederike Mengel, "An experiment on belief formation in networks," *Available at SSRN 2361007*, 2014.

Hanna, Rema, Sendhil Mullainathan, and Joshua Schwartzstein, "Learning through noticing: Theory and evidence from a field experiment," *The Quarterly Journal of Economics*, 2014, *129* (3), 1311–1353.

Jack, B Kelsey, "Private information and the allocation of land use subsidies in Malawi," *American Economic Journal: Applied Economics*, 2013, *5* (3), 113–35.

Kondylis, Florence, Valerie Mueller, and Siyao Jessica Zhu, "Seeing is believing? Evidence from an extension network experiment," *World Bank Policy Research Working Paper 7000*, 2014.

Maertens, Annemie, "Who cares what others think (or do)? Social learning, social pressures and imitation in cotton farming in India," 2015. Unpublished, University of Sussex.

Magnan, Nicholas, David J Spielman, Travis J Lybbert, and Kajal Gulati, "Leveling with friends: Social networks and Indian farmers' demand for a technology with heterogeneous benefits," *Journal of Development Economics*, 2015, *116*, 223–251.

Miller, Grant and A Mushfiq Mobarak, "Learning about new technologies through opinion leaders and social networks: experimental evidence on non-traditional stoves in rural Bangladesh," *Marketing Science*, forthcoming.

Mobius, Markus, Tuan Phan, and Adam Szeidl, "Treasure hunt: Social learning in the field," Technical Report, National Bureau of Economic Research 2015.

Munshi, Kaivan, "Social learning in a heterogeneous population: technology diffusion in the Indian Green Revolution," *Journal of Development Economics*, 2004, *73* (1), 185–213.

Schwartzstein, Joshua, "Selective attention and learning," *Journal of the European Economic Association*, 2014, *12* (6), 1423–1452.

Tjernström, Emilia, "Signals, Similarity and Seeds: Social Learning in the Presence of Imperfect Information and Heterogeneity," 2015. Unpublished, University of California, Davis.

Wang, Honglin, Fan Yu, Thomas Reardon, Jikun Huang, and Scott Rozelle, "Social learning and parameter uncertainty in irreversible investments: Evidence from greenhouse adoption in northern China," *China Economic Review*, 2013, *27*, 104–120.

Young, H Peyton, "Innovation diffusion in heterogeneous populations: Contagion, social influence, and social learning," *The American Economic Review*, 2009, *99* (5), 1899–1924.

"Sur quoi la fondera-t-il l'économie du monde qu'il veut gouverner? Sera-ce sur le caprice de chaque particulier? Quelle confusion! Sera-ce sur la justice? Il l'ignore."

Pascal

FERDi

Created in 2003 , the **Fondation pour les études et recherches sur le développement international** aims to promote a fuller understanding of international economic development and the factors that influence it.

Contact
www.ferdi.fr
contact@ferdi.fr
+33 (0)4 73 17 75 30