American Economic Review 2010, 100:1, 35–69
http://www.aeaweb.org/articles.php?doi=10.1257/aer.100.1.35

Learning about a New Technology: Pineapple in Ghana

By TIMOTHY G. CONLEY AND CHRISTOPHER R. UDRY*

This paper investigates the role of social learning in the diffusion of a new agricultural technology in Ghana. We use unique data on farmers' communication patterns to define each individual's information neighborhood. Conditional on many potentially confounding variables, we find evidence that farmers adjust their inputs to align with those of their information neighbors who were surprisingly successful in previous periods. The relationship of these input adjustments to experience further indicates the presence of social learning. In addition, applying the same method to input choices for another crop, of known technology, correctly indicates an absence of social learning effects. (*JEL* D83, O13, O33, Q16)

The transformation of technology is fundamental to the development process. For a new technology to be adopted by an agent, particularly in agriculture, it must be adapted to the circumstances faced by that agent. Its characteristics usually will not be transparent to the new user (Robert E. Evenson and Larry E. Westphal 1995). Consequently, an investment in learning about the new technology is associated with its adoption. If there are multiple adopters of the new technology in similar circumstances, as is often the case with an innovation in agriculture, then the process of learning about the new technology may be social. New users of the technology may learn its characteristics from each other.

The role of social learning in promoting growth and technology diffusion has been featured in the endogenous growth literature (Paul M. Romer 1986; Robert E. Lucas, Jr., 1988; Philippe Aghion and Peter Howitt 1998; Daron Acemoglu 2009). Social learning that generates knowledge spillovers is a central idea in the large literature on urbanization and growth (e.g. Alfred Marshall 1890; Jane Jacobs 1969; Michael E. Porter 1990; Edward L. Glaeser et al. 1992). These interactions are also an integral part of current practice in agricultural research and extension systems in developing countries. New technologies are introduced either by farmers' own experimentation or through formal sector intervention, and the process of social learning encourages their diffusion (Everett M. Rogers 1995; Vishva Bindlish and Robert E. Evenson 1997). There is a large body of empirical work looking at country- and city-level evidence on the role of knowledge spillovers and

* Conley: Booth School of Business, University of Chicago, 5807 South Woodlawn Street, Chicago, IL 60637 (e-mail: tim.conley@chicagobooth.edu); Udry: Department of Economics, Yale University, 27 Hillhouse Ave., Rm. 33, New Haven, CT 06520-8269 (e-mail: christopher.udry@yale.edu). The authors have benefitted from the advice of Richard Akresh, Muneeza Alaam, Federico Bandi, Alan Bester, Dirk Bergemann, Larry Blume, Adeline Delavande, Steven Durlauf, Ana Fernandez, Garth Frazer, Chris Hansen, Lars Hansen, Ethan Ligon, Charles Manski, Francesca Molinari, Stephen Morris, Melanie Morton, Jeff Russell, Chris Taber, Otto Toivanen, Giorgio Topa, and seminar participants at a number of seminars. We especially thank Yaw Nyarko, Mark Rosenzweig, and anonymous referees for their advice. The data used in this paper were collected by Ernest Appiah, Robert Ernest Afedoe, Patrick Selorm Amihere, Esther Aku Sarquah, Kwabena Moses Agyapong, Esther Nana Yaa Adofo, Michael Kwame Arhin, Margaret Harriet Yeboah, Issac Yaw Omane, Peter Ansong-Manu, Ishmaelina Borde-Koufie, Owusu Frank Abora, and Rita Allotey under the direction of the second author and of Markus Goldstein, who were visiting scholars at the Institute of Statistical, Social and Economic Research at the University of Ghana. This research has received financial support from the National Science Foundation (grants SBR-9617694 and SES-9905720), International Food Policy Research Institute, World Bank Research Committee, Fulbright Program, Social Science Reseach Council, and Pew Charitable Trust. The authors are of course responsible for any errors in this paper.

growth (e.g. Glaeser et al. 1992; James E. Rauch 1993; Robert J. Barro 1991). This paper is an effort to contribute to the recent small, but growing, literature that uses individual-level data to measure the quantitative importance of learning from others. Important examples of this work include Andrew D. Foster and Mark R. Rosenzweig (1995), Oriana Bandiera and Imran Rasul (2006), Kaivan Munshi (2004), Esther Duflo, Michael Kremer, and Jonathan Robinson (2006), Kremer and Edward Miguel (2007), and Patrick J. Bayer, Randi Pintoff, and David Pozen (2009).[1]

In this paper we investigate learning about a new agricultural technology by farmers in the Akwapim South district of Ghana. Over the 1990s, an established system of maize and cassava intercropping for sale to urban consumers began a transformation into intensive production of pineapple for export to European markets (Irene S. Obeng 1994). This transformation of the region's farming system involved the adoption of a set of new technologies, in particular the intensive use of fertilizer and other agricultural chemicals. What was the role of social learning in this process?

Measuring the extent of social learning is difficult for two major reasons. First, the set of neighbors from whom an individual can learn is difficult to define. Second, even with a proper definition of this set, distinguishing learning from other phenomena that may give rise to similar observed outcomes is problematic. In the absence of learning, individuals may still act like their neighbors as a result of interdependent preferences, technologies, or because they are subject to related unobservable shocks.

Direct data on information interconnections is typically unavailable to economists.[2] Consequently, economic investigations of the process of social learning have typically made assumptions that relate observed relationships between individuals—such as geographic proximity—to unobserved flows of information. This set of assumptions is critical for the measurement of the extent of social learning, but can rarely be tested because of data limitations.[3] For example, Foster and Rosenzweig (1995) provide tabulations indicating that "friends and neighbors" are an important source of information about fertilizer use, but must use village aggregates as the relevant information set for social learning.

We have rich data that allow us to address the concerns of neighbor definition more directly. Our approach draws on the classic work by James Coleman, Elihu Katz, and Herbert Menzel (1957) which related adoption of new antibiotics to the network of social interconnections between the doctors. We collected detailed information on whom individuals know and talk to about farming and use this to define information links.[4]

[1] In contrast, there is a long tradition of empirical studies by economists of the adoption of new technologies in agriculture. Zvi Griliches (1957) is the seminal work. For reviews see Gershon Feder, Richard E. Just, and David Zilberman (1985), and Evenson and Westphal (1995). See Tavneet Suri (2006) for an important recent example focused on the adoption and disadoption of fertilizer. This important literature does not, however, isolate the role of learning processes from other determinants of adoption.

[2] Exceptions include Isolde Woittiez and Arie Kapteyn (1998), and Kapteyn (2000), who use individuals' responses to questions about their "social environments" to describe their reference groups. Mattia Romani (2003) uses information on ethnicity and membership in cooperatives in Côte d'Ivoire to infer the probability of information flows. Another exception is Bandiera and Rasul (2006), who have information on the number (though not the identities) of people using a new technology known by particular farmers. James E. Rauch and Alessandra Casella (2001) is a very useful collection of papers that use direct information on social interactions more generally.

[3] In many investigations of learning in developing country agriculture, the reference group is taken to be all farmers in the village (Timothy J. Besley and Anne Case 1994; Foster and Rosenzweig 1995; Futoshi Yamauchi 2007). Munshi and Jacques Myaux (2006) take exceptional care in the construction of reference groups for social learning by using external evidence on communication barriers arising from religion. See Charles F. Manski (1993) for a concise discussion of the importance of reference group designations in identification of endogenous social effects.

[4] Dean Birkhaeuser, Robert E. Evenson, and Feder (1991) and Rogers (1995) provide valuable surveys of research that describes and characterizes the set of neighbors from whom agents learn about new innovations in a wide variety of settings. Christophe Van den Bulte and Gary Lilien (2001) argue that the social contagion effects found by Coleman, Katz, and Menzel (1966) vanish once marketing effort is taken into account.

Once neighborhoods are defined, the identification of learning is still a formidable problem. The classic problem of omitted variables prevents us from inferring that learning effects must be present simply from observations on, say, the diffusion process of a new technology. The fact that a farmer is more likely to adopt a new technology soon after his neighbors have done so might be a consequence of some unobserved variable that is spatially and serially correlated, rather than learning. We believe that correlated unobservables are a general problem in the literature on agrarian technology, and it is apparent that they are important in the sample region (see Sections IIB and IIC). We have collected data to mitigate this problem. Our data contain detailed geographic and soil information as well as information on credit and family relationships, allowing us to control for otherwise confounding factors.

Our identification problem can be thought of as a special case of the general problem of identification in social interactions models studied by Manski (1993, 1997), William A. Brock and Steven N. Durlauf (2001), Robert A. Moffitt (2001), and others. This literature is concerned with the problem of inferring whether an individual's behavior is influenced by the behavior or characteristics of those in his neighborhood or reference group. Our strategy for identifying learning effects relies on using the specific timing of plantings to identify opportunities for information transmission. The staggered plantings in our data naturally provide a sequence of dates where new bits of information may be revealed to each farmer. By conditioning upon measures of growing conditions, we can isolate instances of new information regarding productivity being revealed to the farmer. We then examine whether this new information is associated with innovations in a farmer's input use in a manner consistent with a simple set of assumptions about the nature of learning.[5]

We model farmers' learning about the productivity of inputs. The two key farmer-chosen inputs, fertilizer and labor, are used in essentially fixed, known proportions and farmers need to learn about use of this composite input per pineapple plant. Each harvest opportunity gives the farmer an observation on output for a given composite input, and thus reveals information about the productivity of that input level. We focus on fertilizer usage as a measure of this composite input, since it is the most novel dimension of this new technology and because it is better measured than labor.

Our primary method to test for social learning is to estimate how farmers' input decisions respond to the actions and outcomes of other farmers in their information network. We know the inputs used and output harvested by each farmer, and thus can infer aspects of the information conveyed by each "experiment" with the new technology by each respondent. We use our data on the spatial relationship between farms to condition on spatially correlated but otherwise unobserved factors that influence both profits and optimal input choices. We use our data on information flow between farmers to trace the impact of the information revealed by each experiment on the future input decisions of other farmers who are in the information neighborhood of the cultivator who conducted the experiment.

We find strong effects of news about input productivity in the information neighborhood of a farmer on his innovations in input use.[6] Specifically, we find for a given farmer: (i) he is more likely to change his fertilizer use after his information neighbors who use similar amounts of fertilizer achieve lower than expected profits; (ii) he increases (decreases) his use of fertilizer after his information neighbors achieve unexpectedly high profits when using more (less) fertilizer than he did; (iii) his responsiveness to news about the productivity of fertilizer in his information neighborhood is much greater if he has only recently begun cultivating pineapple; and (4) he

[5] Duflo, Kremer, and Robinson (2006) use a randomized intervention in Western Kenya to implement the same strategy. They gather data on social connections between farmers and then provide information regarding the profitability of fertilizer to a random subset of these farmers. This permits them to identify the importance of learning from the experience of others in their environment. In Western Kenya, however, it turns out that information, either from neighbors or from one's own experience, plays a very limited role in decisions about fertilizer use.

[6] We use the male pronoun to refer to farmers because the large majority of pineapple farmers in our data are men.

responds more to news about the productivity of fertilizer on plots cultivated by veteran farmers and farmers with wealth similar to his. These conclusions hold when conditioning on the changes in fertilizer use of farmers who are physically nearby and who therefore experience unobserved changes in growing conditions that are highly correlated with his. In addition, they are robust to a variety of different definitions of information flow between farmers, and conditional on the fertilizer use of farmers with whom he has financial ties. Finally, we apply our methods to a traditional maize-cassava mixture and they (correctly) indicate no evidence of learning about this established technology.

I. A Learning Model

This section describes a simple model of learning about a new technology that we use to guide our empirical work. We consider risk-neutral farmers, each with a single plot, who are concerned with maximizing current expected profits. At time period t farmer i chooses a discrete-valued input $x_{i,t}$. We mark time with the six-week intervals of our survey rounds. On this time scale, pineapple output is realized five periods after inputs are applied via the production function

$$y_{i,t+5} = w_{i,t}(f(x_{i,t}) + \varepsilon_{i,t+5}), \tag{1}$$

where $\varepsilon_{i,t+5}$ is an expectation zero productivity shock that is i.i.d. across farmers and time and not observed by either farmers or the econometrician. The variable $w_{i,t}$ is a positive, exogenous growing conditions variable that influences the marginal product of $x_{i,t}$ and is correlated across farmers and time.[7] This is motivated by the fact that agricultural production is often affected by spatially and serially correlated shocks to the marginal product of inputs (examples include variation in soil moisture, weeds, or pests). We assume the $w_{i,t}$ are observable to farmers but not to the econometrician. The price of the input x is a constant which we normalize to one.[8] Profits are therefore $\pi_{i,t+5} = w_{i,t}(f(x) + \varepsilon_{i,t+5}) - x$. Farmers do not know the function f; it is the object of learning. The information set available to each farmer is that generated by all current and past growing conditions and observation of inputs and profits for all previous plantings conducted by his information neighbors,[9] as well as his own previous plantings.

The farmer's beliefs are conveniently summarized by his subjective expectations. We use the notation $f_{i,t}(x)$ and $E_{i,t}(\pi_{i,t+5})$ to refer to farmer i's time t subjective expectations of $f(x)$ and of time $t + 5$ profits, respectively. The farmer's time t problem is to choose inputs to maximize (subjective) expected profits. This is nothing more than choosing input level $x_{i,t}$ so that

$$E_{i,t}(\pi_{i,t+5}(x_{i,t})) = w_{i,t}f_{i,t}(x_{i,t}) - x_{i,t} \geq w_{i,t}f_{i,t}(\tilde{x}) - \tilde{x}, \text{ all } \tilde{x}. \tag{2}$$

This simple model is illustrated in Figure 1 for a case where the input can take on three values: zero (Z), low (L), and high (H). This figure plots expected profits as a function of growing conditions w. Expected profit for any given input level x is a line with slope $f_{i,t}(x)$. As drawn, the lines reflect a situation where beliefs are such that none of the three input levels is dominated. For small w the zero input choice is optimal, for intermediate w the low level is optimal, and for sufficiently large w the high level is best. The upper envelope of these lines characterizes optimal input choice as a function of w. Learning will consist of updating beliefs and subjective

[7] The variable $w_{i,t}$ could include a forward looking component, e.g., a rain forecast.

[8] In our study area, fertilizer prices and wages are common across farmers within villages and essentially constant throughout the sample time span.

[9] See Section IIA for operational definitions of information neighbors.

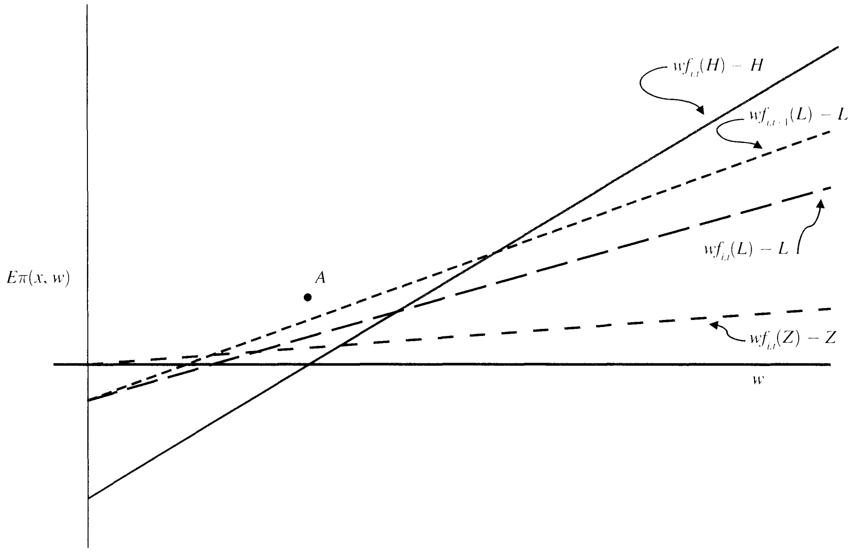

FIGURE 1. UPDATING PRODUCTIVITY KNOWLEDGE GRAPH

expectations $f_{i,t}(\cdot)$ in response to new pieces of information. As the farmer learns, the slopes of the lines in Figure 1 change and of course this can influence farmer's input choices for a given realization of growing conditions.[10]

Farmers update their beliefs about f in response to observations of inputs, growing conditions, and outputs. Suppose that plot j of farmer j is observed by farmer i.[11] This means that at time t farmer i observes profit, $\pi_{j,t}$, the inputs used, $x_{j,t-5}$, and the relevant growing conditions, $w_{j,t-5}$. This new information allows farmer i to calculate $f(x_{j,t-5}) + \varepsilon_{j,t} = (\pi_{j,t} - x_{j,t-5})/(w_{j,t-5})$ and this informs his beliefs about the productivity of input level $x_{j,t-5}$, leading him to update his previously held expectation (dated at some time t_p prior to t): $f_{i,t_p}(x_{j,t-5})$ to $f_{i,t}(x_{j,t-5})$. Note that farmers know growing conditions, w, so these expectations reflect only uncertainty due to ε and imperfect knowledge of f.

Rather than focus on a specific mechanism or type of learning, we consider the empirical implications of a set of three assumptions about the way farmers learn that correspond to farmers' descriptions of their own learning process.[12] We state our assumptions in terms of farmer i's update $\Delta f_{i,t}(x_{j,t-5}) \equiv f_{i,t}(x_{j,t-5}) - f_{i,t_p}(x_{j,t-5})$ in response to observing the event $(\pi_{j,t}, x_{j,t-5}, w_{j,t-5})$:

(i) $\Delta f_{i,t}(\mathrm{x}_{j,t-5})$ has the same sign as $\pi_{j,t}(x_{j,t-5}) - E_{i,t_p}(\pi_{j,t}(x_{j,t-5}))$ and it increases without bound as $\pi_{j,t}(x_{j,t-5})$ exceeds $E_{i,t_p}(\pi_{j,t}(x_{j,t-5}))$.

(ii) We assume that $\Delta f_{i,t}(\mathrm{x})$ attenuates as experience at input level x increases. An increase in i's experience reduces the absolute value of $\Delta f_{i,t}(x)$ in response to a given piece of new information $(\pi_{j,t}, x_{j,t-5}, w_{j,t-5})$.

[10] Figure 1 is drawn for the production function (1). In a more general model in which $y_{t+5} = F(w_{i,t}, x_{i,t}, \varepsilon_{i,t+5})$, the curves need not be linear. In this case there could be disjoint regions of w in which a given level of input application is optimal.

[11] The analysis applies as well for farmer i learning from his own experience.

[12] There are many specific models of learning consistent with these assumptions. In particular, they are consistent with independent Bayesian learning about the elements of the support of x, or with standard models of reinforcement learning (Leslie P. Kaelbling, Michael L. Littman, and Andrew W. Moore 1996; Nick Feltovich 2000).

(iii) $\Delta f_{i,t}(x) = 0$ for all x other than $x_{j,t-5}$. Learning is *local*, so that this new information changes i's beliefs only about the productivity of input level $x_{j,t-5}$. Beliefs regarding other input levels are unchanged.[13]

The counterpart of these assumptions in Figure 1 is that only one line will have its slope changed in response to each new piece of information, this slope will rise (fall) if profit is higher (lower) than expected, and for a given piece of new information the change in slope will become less pronounced as farmer experience grows. So in Figure 1, if farmer i observes at time $t+1$ that farmer j using, say, fertilizer level L has achieved a higher profit than expected given j's growing conditions (at, say, point A), then i updates his beliefs about the productivity of L from $f_{i,t}(L)$ to $f_{i,t+1}(L)$.

Assumptions 1, 2, and 3 have the following main empirical implications:

Implication 1: *Farmers tend to adjust input use toward surprisingly successful input levels, and higher than expected profits at the currently utilized input level will make farmers less likely to change from that level.* When observed profit is higher than expected, $\Delta f_{i,t}(x_{j,t-5})$ is positive. When $\Delta f_{i,t}(x_{j,t-5})$ is positive, the set of w for which $x_{j,t-5}$ is the optimal input increases; hence, this input level becomes more likely to be chosen. This is easy to see in Figure 1, where a higher than expected output for a given input level will raise the slope of the corresponding line, leaving the other lines unchanged. This increases the set of w for which this given input level is optimal and, hence, chosen.

Implication 2: *Farmers tend to adjust input use away from an input level that was less profitable than expected.* When observed profit is lower than expected, $\Delta f_{i,t}(x_{j,t-5})$ is negative. Negative $\Delta f_{i,t}(x_{j,t-5})$ decreases the set of w for which $x_{j,t-5}$ is the optimal input and so it is less likely to be chosen. Again in Figure 1, one line will rotate down, decreasing the set of w where its corresponding input is optimal, perhaps even rendering an input level dominated.

Implication 3: *An observation of profit sufficiently above expectations will induce a farmer to switch to that level of input use.* If $\pi_{j,t}(x_{j,t-5}) - E_{i,t_p}(\pi_{j,t}(x_{j,t-5}))$ is sufficiently large, $\Delta f_{i,t}(x_{j,t-5})$ will be large enough to dominate expected profits under alternate input levels, for any given w. Therefore, farmer i will switch from x_{i,t_p} to $x_{j,t-5}$ in period t if i observes an outcome of j's choice of $x_{j,t-5}$ that is sufficiently good. Defining $\Delta x_{i,t}$ as $x_{i,t} - x_{i,t_p}$, we summarize this implication as

$$(3) \qquad \Delta x_{i,t} = 1\{\pi_{j,t}(x_{j,t-5}) - E_{i,t_p}(\pi_{j,t}(x_{j,t-5})) > c_{i,t}(x_{j,t-5})\}(x_{j,t-5} - x_{i,t_p}),$$

where $1\{\cdot\}$ is an indicator function equal to one if its argument is true. The threshold value $c_{i,t}(x)$ at any fertilizer level x depends upon growing conditions $w_{i,t}$, farmer i's beliefs prior to this information revelation $f_{i,t-1}(\cdot)$, and the characteristics of farmers i and j, particularly i's experience.

Implication 4: *The probability of changing input levels in response to a given piece of information is decreasing in a farmer's experience.* This is a direct implication of our assumption that experience attenuates changes in beliefs induced by new information.

[13] The implications of our model do not hinge upon inputs being discrete. We take them as discrete for ease of exposition. We could allow agents update only within one of a set of input ranges or only within a bandwidth of observed input use via a local-average (kernel) regression. This would allow us to obtain identical implications with continuous inputs, at the cost of complicating notation and our definition of local learning.

Implication 5: *Correlations in growing conditions across space and time can look like social learning to the econometrician who does not observe w.* In this environment, growing conditions $w_{i,t}$ are positively spatially and serially correlated. If there is any positive association in beliefs across farmers, i.e., similar rankings of the productivity of different input levels, then positive correlations in $w_{i,t}$ will result in positive correlations in optimal input choices. In other words, positive spatial and temporal correlations in $w_{i,t}$ mean that proximate (close in space and time) farmers will have similar w values, and thus if their beliefs about productivity are similar, they will tend to choose similar input levels. This is true regardless of whether farmers learn about input productivity.

Moreover, when plantings by different farmers are staggered in time, as in our application, positive dynamic correlations between $w_{i,t}$ and $w_{j,t+k}$ for spatially close i,j can easily lead to *innovations* in choices being positively correlated with lagged neighbors' choices, even if there is no learning at all. To see this, consider the example in Figure 1 with input levels zero, low, and high, and suppose farmers know the technology (so $f_{i,t}(\cdot) \equiv f(\cdot)$ and there is no learning). Take farmers i and j who are physical neighbors, who are likely to experience similar w at similar times, and who thus are likely to choose similar input levels. Suppose that i and j both plant at period t and that they choose $x_{i,t} = x_{j,t} = L$. Suppose, as is common in our setting, that farmer j plants another plot in, say, period $t + 3$, and that farmer i plants another plot in $t + 8$. At time $t + 3$, let j experience a positive growing conditions shock ($w_{j,t+3} > w_{j,t}$) sufficiently large to induce him to change to $x_{j,t+3} = H$. Positive spatial and serial dependence in w is consistent with this shock to j, making it more likely that i, also, will experience a sufficiently large innovation in w so that at $t + 8$ he will change to $x_{i,t+8} = H$. High values of w will of course tend to result in higher than usual values of yields and profits. The econometrician observes farmer j choosing H and receiving higher than usual yields/profits, followed by his neighbor i changing inputs to H. This is of course a sequence of events that would also be predicted if there were no change in growing conditions, but farmer i learned about the profitability of input level H from observing farmer j get a good outcome after choosing H at time $t + 3$. Thus, it will be very important for us to disentangle the effects of learning from the reactions to growing conditions.

Our model does not capture some aspects of learning about a new technology that are important in other applications. These aspects are largely absent, or concerns about them strongly mitigated by the specific circumstances of our context. First, there are multiple inputs to farming pineapple. The main inputs, fertilizer and labor, are used in essentially fixed proportions so a single, composite input reflects reality well. Measurements of fertilizer provide our measure of this composite input, since it is better measured than labor inputs. Second, we abstract from strategic behavior on the part of farmers. Strategic considerations in information transmission are less salient in this environment than in others in the literature. Our surveyed farmers are operating in a competitive environment in both output and input markets. The fertilizer choices of any farmer will have no impact on the prices of pineapple, or on the costs of farm inputs. Third, we model farmers' information flow from neighbors as being generated just from observation of their inputs and profit outcomes. This implicitly limits the extent of communication between farmers and the ability/willingness of a farmer to model others' behavior. For example, it rules out learning via conversations about third parties' activities and farmers making inferences about a third party's outcomes from an information neighbor's action. Ruling out such aspects of learning is well motivated in the villages we study, as it is considered inappropriate gossip for a pair of farmers to discuss other farmers' activities. Furthermore, farmers have little to no knowledge of others' information connections. As a consequence, second-order inference from the input choices of other farmers has little power. Fourth, we model farmers as focused on short-run profits, so there is no scope for experimentation. Our main implications can be shown to survive the introduction of forward looking farmers who take into account the value of experimentation for future profits (see Conley and Udry 2005, Appendix 1), so this is not a major concern. Furthermore, surveyed farmers' descriptions of the reasoning behind their input choices indicate that their overwhelmingly dominant concerns are returns from current rather than future plantings.

Our choice to model learning as local is also specific to our setting. It contrasts with some models of learning in other contexts in which information about the production function can be deduced regardless of the portion of the production function used (Edward C. Prescott 1972; Foster and Rosenzweig 1995; Boyan Jovanovic and Yaw Nyarko 1996). The empirical implications of our local learning model are different from those of a model of global learning. In the latter class of models, there can be no general implication that farmers adjust inputs toward (away from) surprisingly successful (unsuccessful) levels of input use. For us, local learning is motivated by both surveyed farmers' own descriptions of what they learned from past experiences with using fertilizer, and a substantial descriptive literature from Africa (Paul Richards 1985; Kojo S. Amanor 1994). This evidence strongly suggests that an appropriate model should have the feature that farmers must use or observe inputs in a given range in order to learn about the corresponding part of the production function.

Basic Empirical Approach.—We will investigate these five implications by estimating models, using the full sample, predicting the occurrence of a change in input use, $\Delta x \neq 0$, and for the change in inputs itself, Δx. Here, we preview our baseline models and heuristically describe the key regressors whose construction is detailed in the following section. We estimate a logistic model of the probability that inputs change, $\Delta x \neq 0$. The regressors of interest reflect whether events observed by the farmer in between his planting opportunities had profits above or below expectations, which we refer to as good or bad news, respectively. Let t_p refer to the farmer's previous planting opportunity. We use the notation $s(good, x = x_{i,t_p})$ for the share of time t total observed events in farmer i's information neighborhood (from the beginning of the survey until time t) that are good news events at input level x_{i,t_p} and occur between periods t_p and t. Analogous notation is used for the shares of other combinations of good/bad news and alternative inputs. These regressors are expressed as a share of time t total observed events so that they attenuate with increases in experience. They equal zero when farmer i does not observe any planting at a given input level between t_p and t. We estimate

$$(4)\quad \Pr\{\Delta x_{i,t} \neq 0\} = \Lambda\left[\begin{array}{c} \alpha_1 s(good, x = x_{i,t_p}) + \alpha_2 s(good, x \neq x_{i,t_p}) \\ + \alpha_3 s(bad, x = x_{i,t_p}) + \alpha_4 s(bad, x \neq x_{i,t_p}) \\ + \alpha_5(\Delta \text{ growing conditions}) + (\text{experience and other controls})'\alpha_6 \end{array}\right],$$

where $\Lambda(\cdot)$ denotes the logistic function. The empirical counterpart of Implications 1 and 2 is that α_1, α_4 are negative and α_2, α_3 are positive.

We investigate Implications 3 and 4 using a regression model for changes in x with the following form:

$$(5)\quad \Delta x_{i,t} = \beta_1 M_{i,t} + \beta_2(\Delta \text{ growing conditions}) + (\text{experience and other controls})'\beta_3 + v_{i,t}.$$

The regressor of interest M is an empirical analog of the right side of equation (3). M is constructed so that if an inexperienced farmer i observes good news using inputs well above (below) his previous input level, x_{i,t_p}, this index will be positive (negative) and large. If the farmer observes good news close to x_{i,t_p}, or in the absence of good news, M will be near/at zero. Motivated by Implication 4, M is constructed so that its absolute magnitude is inversely proportional to farmer i's experience. Conditional on growing conditions, this index should be a good predictor of variation in Δx induced by observations of good news, and its coefficient should be positive. Implication 2 is not informative about the direction or magnitude of changes in response to bad news; hence, measures of bad news events do not appear in (5).

Our goal is to identify possible learning interactions via what Moffitt (2001) describes as a type of policy intervention that "changes the fundamentals for a subset of the population in a group in an attempt to influence the outcomes of the others in the group." Our "intervention" is the realization of surprising (given growing conditions) profits by another farmer in one's information network, which is reflected in M. Implication 5 focuses our attention on the challenge of disentangling the effects of social learning from the confounding influences of growing conditions. We do so by using the high spatial and serial correlation in $w_{i,t}$ to construct a variable that permits us to control for changes in growing conditions (see Section IIC) .

Our identification assumption is that conditional on our measures of changes in growing conditions and other farmer-level characteristics, our information measure, M, is uncorrelated with unobserved determinants of changes in growing conditions. Evidence in support of this assumption is provided in Section III and is the focus of the specification test in Section IVB.

There are at least two important concerns about this empirical strategy. First, individuals clearly choose their information neighbors. This presents the possibility that neighborhood sorting or selection effects could influence our results. As we discuss in Section IVA, details of our empirical results permit us to rule out the most plausible models of endogenous sorting that might lead to spurious results with respect to learning. Section IVA also presents results using an arguably exogenous definition of information neighborhoods based on a prediction of information links between farmers, given the deeper social relationships between individuals.

Second, those who farm pineapple are a selected group of farmers. Thus, our estimates of β_1 may not be representative of potential learning effects that nonadopters might face if or when they adopt pineapple. If there is heterogeneity in the extent to which farmers learn from others, it is possible that those who adopt first are those most responsive to information from others. If so, our sample of adopters would be selected to be responsive, and β_1 would overstate the importance of social learning in the overall population. As we have information on learning only for adopters, we cannot estimate learning effects for the nonadopting subpopulation, and this second concern must remain a caveat. However, in Section III we investigate the extent of heterogeneity in ability to learn within our sampled subpopulation of adopters.

II. Data

This section describes our data and the construction of variables. First, we discuss the basic features of our estimation sample. We then discuss the measures we use to define information neighborhoods. We then describe the data on farmers' inputs and outputs, our methods to control for growing conditions, and our methods for approximating farmers' subjective expectations and innovations in information.

Our main data source is a two-year survey (1996–1998) of 180 households, drawn from a population of 550 households in three villages in southern Ghana.[14] Our study region is in the center of a recently growing area of intensive pineapple cultivation. Two enumerators lived in

[14] These three villages are a subset of four total villages in which surveys were administered. Pineapple farming was not present in the fourth village. A detailed description of survey procedures, copies of the survey instruments, and the data archive can be found at http://www.econ.yale.edu/~cru2/ghanadata.html.

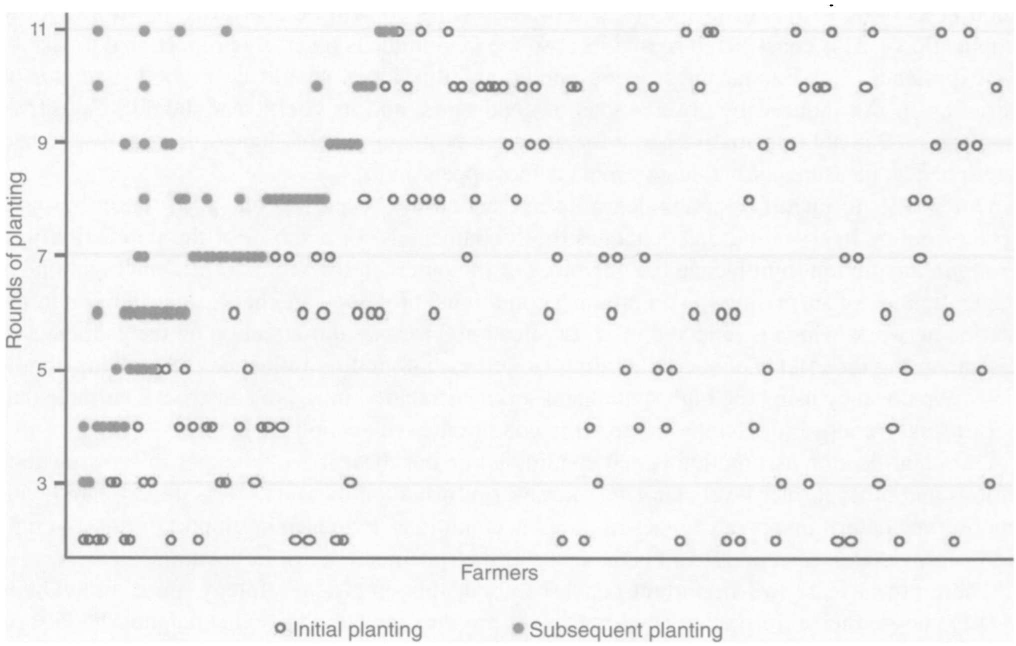

Figure 2. Distribution of Observations across Farmers and Survey Rounds

or near each village and interviewed respondents in 15 rounds at intervals of approximately six weeks. In addition to survey-based information, all plots were mapped using global positioning system equipment. This procedure yields accurate measures of plot size and location, data that are seldom available for developing countries.

Our main estimation sample is constructed as follows. We begin with information on pineapple being grown on 406 plots by 132 farmers. Of these plots, 288 were planted during our survey. Plot input data are missing on 3 of these plots, leaving 285; 77 of these were planted too late in our survey for fertilizer application to be completed before the end of fieldwork, leaving 208 plantings. We are missing data for some rounds on 8 of these, leaving 200 plantings; 87 of these are the first planting in our survey by particular farmers, leaving 113 observed changes in fertilizer use. GIS information is missing on 6 of these plots, leaving information on 107 changes in fertilizer use by 47 farmers. Figure 2 depicts the pineapple plots for which we have GIS and input/output information. Farmers identification numbers are the horizontal coordinates and the vertical coordinate for each point is the round in which the planting began. The first planting by each farmer is marked with an open circle; second and later plantings are denoted by closed circles. The 47 farmers who contribute observations on input changes are arranged at the left to make reading the graph easier.[15]

[15] There are 81 farmers who cultivate plots for which we have GIS and input/output information: 11 live in village 1, 32 in village 2, and 38 in village 3. Of the 47 farmers who provide data on input changes, 8 live in village 1, 14 in village 2, and 25 in village 3.

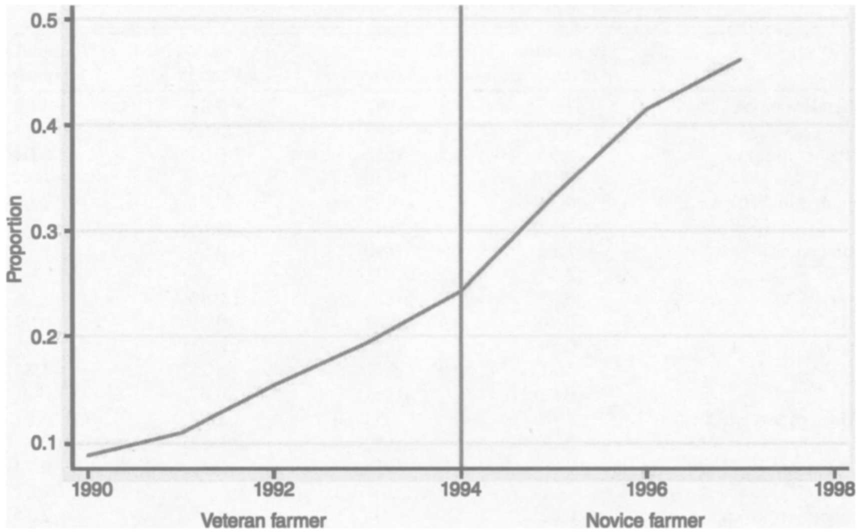

FIGURE 3. SAMPLE PROPORTION OF FARMERS CULTIVATING PINEAPPLE BY YEAR
(*Retrospective data from authors' survey*)

Figure 3 shows the pattern of adoption of pineapple in our sample villages: from less than 10 percent in 1990, pineapple spread very rapidly until more than 46 percent of farmers were cultivating pineapple in 1997. We utilize two measures of experience. $Experience_{i,t}$ is the total number of pineapple plants planted by farmer i from the start of our survey until time t. We also utilize data on years of farming pineapple to define a binary indicator, dividing pineapple farmers into two groups: veteran farmers who adopted pineapple before 1994, and novice farmers who adopted in 1994 or after.

Table 1 reports summary statistics for our data. We report statistics by novice/veteran status as well. Wealth is defined as the value of the nonland assets held by the farmer at the start of the survey period. Veteran farmers are far wealthier. In addition, pineapple farmers as a class are much wealthier than nonpineapple farmers in the area (not shown in table). Those who do not cultivate pineapple have an average nonland wealth of only 0.4 million cedis versus the average of 2.1 million cedis for pineapple farmers.[16] Pineapple farmers' wealth reduces the potential importance of credit constraints for fertilizer decisions (mean fertilizer use on an average-sized farm is 0.08 million cedis, a small fraction of average pineapple farmer nonland wealth). The clan indicator variables denote membership in a particular *abusua*, or matrilineal clan. The church indicator denotes membership in the most popular church. Members of that church and of one of the matrilineages are overrepresented among experienced pineapple farmers. We collected information on the soil chemistry (pH and organic matter content) of approximately 80 percent of the plots. Approximately one-third of farmers report that they have received advice from an agricultural extension agent (from the Ghanaian Ministry of Food and Agriculture) in the past; we do not know when such advice was received.

[16] Cedis are small units. The exchange rate during the sample period ranged from 1,700–2,300 cedis/US dollar.

Table 1—Descriptive Statistics, Estimation Sample

	Estimation sample	Novice	Veteran	t (H_0: equal means by experience)
Fertilizer use (cedis per sucker)	2.05	3.92	1.48	−1.88
	(5.76)	(9.82)	(3.66)	–
Change in fertilizer use ($\Delta x_{i,t}$)	0.52	0.52	0.52	0.04
	(0.50)	(0.50)	(0.50)	–
Indicator of change in fertilizer $\neq 0$	−0.33	−1.45	0.01	0.60
	(10.62)	(20.43)	(4.85)	–
Wealth (million cedis)	2.13	0.90	2.50	2.78
	(2.59)	(0.72)	(2.84)	–
Clan 1 indicator	0.35	0.16	0.40	2.26
Clan 2 indicator	0.44	0.52	0.41	−0.92
Church 1 indicator	0.49	0.20	0.57	3.41
pH	5.81	5.97	5.75	−1.42
	(0.62)	(0.45)	(0.67)	–
Soil organic Matter (percent)	2.78	2.53	2.87	1.82
	(0.78)	(0.59)	(0.82)	–
Contact with extension Agent indicator	0.30	0.28	0.31	0.24
Avg. dev. of lagged use from geographic Neighbors' use ($\gamma_{i,t}$)	0.52	−3.00	1.58	2.56
	(8.01)	(14.45)	(4.13)	–
Indices of good news				
Input levels				
M ask advice	−0.23	−1.48	0.15	1.80
	(3.99)	(8.14)	(0.79)	–
M roster of contacts, farm info only	−0.18	−1.65	0.27	2.48
	(3.47)	(6.46)	(1.58)	–
M roster of contacts, full list	−0.13	−1.65	0.34	2.57
	(3.47)	(6.46)	(1.57)	–
M predicted ask for advice	−0.02	−0.35	0.08	2.05
	(0.93)	(1.82)	(0.35)	–
Share of good news at lagged fertilizer use $s(good, x = x_{i,previous})$	0.15	0.24	0.11	−1.72
	(0.33)	(0.43)	(0.29)	–
Share of bad news at lagged fertilizer use $s(bad, x = x_{i,previous})$	0.05	0.11	0.04	−1.45
	(0.21)	(0.30)	(0.17)	–
Share of good news at alternative fertilizer use $s(good, x \neq x_{i,previous})$	0.07	0.08	0.06	−0.39
	(0.24)	(0.28)	(0.23)	–
Share of bad news at alternative fertilizer use $s(bad, x \neq x_{i,previous})$	0.03	0.05	0.02	−1.04
	(0.12)	(0.20)	(0.08)	–
Novice farmer indicator	0.23	1.00	0	–
Number of observations	107	25	82	–
Number of farmers	47	15	32	–

Notes: Unless otherwise indicated, cells contain means of the indicated variable for the indicated sample. Standard deviations in parentheses.

A. *Communication and Knowledge*

One of our main innovations is that we are able to use the survey data to define information neighborhoods. We base our measure of information availability on direct data about conversations between individuals.

Table 2—Information Connections by Cohort of Pineapple Adoption
(*Proportion of pairs of individuals in each other's information neighborhood*)

	Not farming pineapple	Novice pineapple farmer	Veteran pineapple farmer	Neighborhood metric
Not farming pineapple	0.02	0.02	0.07	Response to "Have you ever gone to ____ for advice about your farm?"
Novice pineapple farmer	0.02	0.09	0.13	
Veteran pineapple farmer	0.03	0.13	0.21	

Each respondent was questioned about a random sample (without replacement) of seven other individuals from our own sample in the same village. The samples of individuals produced responses to the question: "Have you ever gone to ___ for advice about your farm?" In this case, we say an information link exists between farmers i and j either if i responded "yes" to this question about j or if j responded "yes" to this question about i. We use responses to this question as our benchmark definition of information neighbors, because during the field research it appeared reliably answered and it is transparently related to the learning process under study. Farmers are of course included in their own information neighborhoods.[17] Not counting farmers themselves, the median number of information neighbors is two.

There are several systematic patterns in information links. Here we briefly summarize estimates of a model predicting our benchmark neighborhood connections given exogenous farmer characteristics that is fully reported in Appendix A. Spatial proximity is correlated with the presence of information links but it is not their sole determinant. Information links occur over long as well as short distances. These longer-distance information links are essential to our ability to distinguish the impact of information from that of spatially correlated shocks in growing conditions. Cross-gender links are rare and links are positively correlated with common clan membership and similarity in age. Individuals with different levels of wealth are more likely to be linked, reflecting the strong vertical patron-client ties that exist in these villages. There is no evidence that religion influences information links. Pineapple farmers—especially veteran pineapple farmers—are more likely to be in each other's information neighborhood than would be expected by chance. Table 2 provides a summary of our baseline information link distribution by experience. Over 20 percent of veteran pineapple farmers (within each village) have approached each other for advice about farming, while only 2 percent of nonpineapple farmers are in each others' information neighborhood. A similar pattern is observed using our other information metrics. It may be the case that these information connections were important determinants of the adoption process; however, we have too few instances of new adoption during our sample period to address this question formally. In Section IV, we discuss the possibility that farmers vary in their ability to learn from others, and in particular that the pineapple farmers who comprise our sample are selected along that dimension.

In Section IV, we check the robustness of our main results to varying definitions of the information neighborhood by using three alternative measures of information flow. Two of these measures are based on lists of interactions between respondents during the course of the survey (discussing farming, buying or selling goods, hiring labor, exchanging gifts, etc.). Our third definition of the information neighborhood is based on predicted links between individuals using the

[17] There are relatively few plantings with timing that allows individuals to learn from their past plantings, so we do not distinguish between such "learning by doing" events and observations in our benchmark specification. We do however, examine learning by doing events separately in Section IV.

model discussed in Appendix A. Information neighborhoods based on predicted neighbors are less subject to concerns that observed links are endogenously formed in anticipation of obtaining input advice. All of these measures are fully described in Appendix A.

Finally, we note the surprising fact that pineapple exporters, though they might have an incentive to provide input information, do not appear to be an information source. There is no evidence that farmers receive advice on fertilizer application from the exporters to whom they sell their harvest (nor is there any contract farming). Farmers were asked a series of open-ended questions about sources of information regarding farming, including fertilizer application. In no instance did any farmer mention pineapple exporters as a source of information or advice.[18]

B. *Inputs and Outputs*

We focus on farmers' decisions about the intensity of input use in pineapple production. The two key inputs are fertilizer and labor, which are used in essentially fixed proportions but with varying intensities per pineapple plant. There is agronomic evidence that pineapple yields are very responsive to fertilizer (John W. Purseglove 1972; W. S. Abutiate 1991). In informal interviews, individuals in the sampled villages expressed substantial uncertainty and conflicting views regarding the optimal intensity of fertilizer and associated labor use.[19]

Our empirical measure of input intensity is based on fertilizer usage since it is better measured than labor. During the period from six weeks to six months after planting, pineapples are extremely sensitive to nutrient availability (Duane P. Bartholomew and Saleh Kadzimin 1977; Soler 1992). Our input measure is the amount of fertilizer per plant applied during this period. This corresponds to one to four survey rounds after planting. We adopt the convention of indexing plantings by the round when input application is complete, four rounds after the pineapple was planted. Thus our input measure for plot i, $x_{i,t}$, is the per-plant amount of fertilizer applied over the reference period: $t - 3$ through t. We observe many plantings of pineapple during each of our survey rounds because pineapple production in southern Ghana is not strongly seasonal.

Per-plant fertilizer application is uniform within plots. The plots in our sample are close to the minimal viable scale for export farmers. The median plot size in our data is approximately 0.5 hectares. The novice pineapple farmers who exhibit the most evidence of learning have a median plot size of 0.25 hectares; exporters are reluctant to harvest and export crops from plots any smaller (only 5 plantings in our data were as small as 0.125 ha.). Plots have to be harvested on a single day for efficient export of the fresh fruit by air to Europe. It is essential that the fruits mature simultaneously, which requires common treatment across plants within the plot. As a consequence, there is no scope for experimentation with different levels of fertilizer within plots.

Pineapple takes approximately five of our survey rounds to mature after the application of fertilizer is completed. Profit for plot i with fertilizer application completed at time t, denoted $\pi_{i,t+5}(x_{i,t})$, is obtained by deducting from harvest revenue the value of all inputs, including family labor valued at the relevant gender-specific wage. In addition to plantings with harvest revenue data, we have some plantings that were unharvested at the end of the survey period but far enough along to enable an accurate forecast of their value at harvest. Our measure of $\pi_{i,t+5}(x_{i,t})$ for these plantings (about a third of the total observed) is constructed using the farmer's forecast

[18] This pattern has changed since the survey. Suri (2008) describes the recent emergence of contract farming among pineapple farmers in this area.

[19] While there are official recommendations on fertilizer use from the extension service of the Ghanaian Ministry of Agriculture (which we interpret as the yield-maximizing level of fertilizer use on test plots), these far exceed the levels of application in the sampled villages. The recommendation is 400 kilograms of fertilizer per hectare which, is more than 10 times the mean fertilizer use observed in our sample. Only 4 of 208 plantings we observed exceeded the recommended level of fertilizer application.

of revenue at harvest. Respondents had no trouble providing such forecasts, perhaps because preharvest crop sales are routine for traditional crops like cassava.

C. *Growing Conditions*

We expect shocks to growing conditions to be positively correlated across both space and time. Our concern about spatial correlation is motivated by the observation in these villages that growing conditions vary spatially on the scale of hundreds of meters. Soil types and topographical features are highly correlated across neighboring plots, but vary over the village as a whole. Therefore, common village-level weather shocks can have varying impacts across the village. Moreover, rainfall realizations can be different on opposite sides of a single village. Finally, weeds spread in a broadly continuous manner across space, and soil moisture and pest and disease environments are often much more similar on nearby plots than on more distant plots within villages. Positive serial correlation in growing conditions is also to be expected across the six-week periods. At this time scale, there is substantial correlation in soil moisture, weeds, and pest conditions on a given plot over multiple periods. Thus, we anticipate substantial correlation in growing conditions for physically close plots planted at different but near points in time, due to the overlap in much of the environmental conditions they experience. As detailed in Section I, it is essential for us to control for changes in growing conditions in our empirical work.

Our control for the change in growing conditions faced by farmer i between his current and previous plantings is based upon the difference between the observed input choices of other farmers with plots that are close to i's current planting and the earlier input choice of farmer i. The assumption underpinning this is that strong positive spatial and temporal correlations in growing conditions will induce farmers to make highly correlated input choices. Therefore, input choices for plots proximate to plot i at time t will contain a strong signal regarding the growing conditions at this place and time. We define $x_{i,t}^{close}$ as the (plant-weighted) average of fertilizer input on plots sufficiently close to plot i and time t.[20] Our regressor for the change in growing conditions for plot i at time t since i's previous planting at time t_p is then defined as

$$\Gamma_{i,t} = x_{i,t}^{close} - x_{i,t_p} \tag{6}$$

The regressor $\Gamma_{i,t}$ should be a good predictor of changes in input choice by farmer i, $\Delta x_{i,t} \equiv x_{i,t} - x_{i,t_p}$, that are due to growing conditions. This regressor can be interpreted as a gap between a target input use of those close to (i,t) and the farmer's previous input choice. We use this construction rather than a measure of changes in geographic neighbors' input choices because of the unbalanced nature of our panel. Many farmers have single plantings, and those with multiple plantings are irregularly staggered in time. Single-planting farmers remain useful in constructing our target term. We also employ a regressor $\tilde{\Gamma}_{i,t}$ to reflect the level of absolute discrepancies between x_{i,t_p} and inputs used on plots close to (i,t). $\tilde{\Gamma}_{i,t}$ is a plant-weighted average (across plots close to i,t) of the absolute difference between inputs used on the plot and x_{i,t_p}.

In addition, we construct a regressor analogous to $\Gamma_{i,t}$ with the neighborhood definition based on financial rather than geographic neighborhoods. Two farmers belong to each other's financial neighborhood if they lend to, borrow from, or exchange gifts or hold assets in common with each

[20] "Sufficiently close" means within one kilometer and at time t to $t-3$. The median number of physically proximate plots is 12, the maximum is 25. See Figure 4 for a scale map of plot centers within one of the villages. Our results are not very sensitive to the radii chosen here; qualitatively identical results obtain with a range of 500 to 1,500 meters and one to four time periods. See Appendix B.

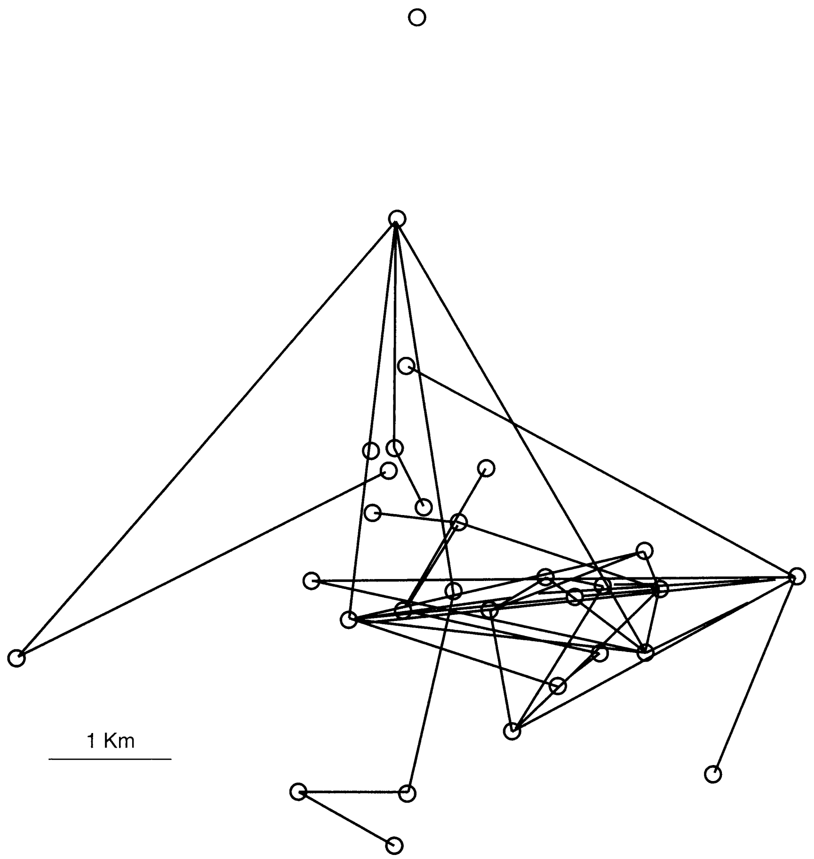

Figure 4. Roster Connections and Average Pineapple Plot Coordinates, Village 3

other at any point during the two-year survey period. This regressor is motivated by alternate explanations that would suggest that significant β_1 estimates might be caused by omitted variable bias because information neighbors share common access to credit arrangements.

D. *Innovations to Information*

In this section we discuss our measures of farmers' subjective expectations and our operational definition of innovations in information regarding input productivity. A key object in this measure is $E_{i,t_p}(\pi_{j,t+5}(x_{j,t}))$, which is farmer i's subjective expectation, at the time of his previous planting opportunity t_p, of profitability for his information neighbor farmer j, who used input level $x_{j,t}$ at t. Farmer i's information includes knowledge of farmer j's inputs $x_{j,t}$ and growing conditions $w_{j,t}$. We use observed inputs and profits to approximate $E_{i,t_p}(\pi_{j,t+5}(x_{j,t}))$, relying on physical and temporal proximity to effectively condition upon $w_{j,t}$ without observing it.[21] Our

[21] We note that while this method allows us to over come the absence of data on w, it has the disadvantage of forcing our estimate of expected profitability to be based on information relatively near the time t to $t + 5$ interval. Farmers in our model and, we presume, in real life have relatively more past information. However, with our data it appears impossible to incorporate more past information in these estimated expectations.

approximation, $\hat{E}_{i,t_p}(\pi_{j,t+5}(x_{j,t}))$, is the median profits of all plots whose observed input choices are close to $x_{j,t}$ and whose time and location are close to plot j and dated period t, or shortly before, so that growing conditions are approximately the same as $w_{j,t}$.[22] Armed with these approximations to expected profits, we construct indicator variables for whether profits $\pi_{j,t+5}$ exceeded or were below i's expectations, which we refer to as good or bad news events, respectively.

Our characterization of the information that i receives between his planting opportunities depends on the number of plantings that he observes and their precise timing relative to his planting times. Let farmer i have plantings at time t_p and later at time t_c ("c" for "current"). In the modal case (half of all observations), i gets information from only a single planting, say by farmer j, and the profits of j's planting are revealed between t_c and t_p, as depicted in Figure 5. Recall from equation (3) that the change in farmer i's input use will be

$$(3') \qquad x_{i,t_c} - x_{i,t_p} = [1\{\pi_{j,t+5}(x_{j,t}) - E_{i,t_p}(\pi_{j,t+5}(x_{j,t})) > c_{i,t_c}(x_{j,t})\}](x_{j,t} - x_{i,t_p}).$$

That is, the input change will be nonzero if j's profit exceeds expectations (it is good news) and does so by an amount that also exceeds the threshold $c_{i,t_c}(x_{j,t})$. This threshold is implicitly a function of the characteristics of farmers i and j, including of course farmer i's beliefs about all input levels and his experience, and of the growing conditions w_{i,t_c}, which are not observed by us.

It is necessary for profits to exceed expectations for the bracketed term in (3′) to equal one, but of course this is not sufficient. Good news about an input level may not exceed the threshold $c_{i,t_c}(x_{j,t})$. In particular, this threshold will increase with farmer i's experience, as it attenuates updates in expectations. We construct a rough empirical analogue to the right-hand side of (3′) using our good news indicator and a term in farmer i's experience, $Experience_{i,t_c}$, as follows:

$$(7) \qquad M_{i,t_c} \equiv \frac{1}{Experience_{i,t_c}}[1\{\pi_{j,t+5}(x_{j,t}) - \hat{E}_{i,t_p}(\pi_{j,t+5}(x_{j,t})) > 0\}](x_{j,t} - x_{i,t_p}).$$

If there is learning of the type we have proposed, M_{i,t_c} should be positively correlated with changes in inputs conditional on changes in growing conditions. If planting j,t results in good news about an input level $x_{j,t}$ which is much higher than farmer i's previously used level (x_{i,t_p}), M_{i,t_c} is large and positive; if, instead, it contains good news about a level $x_{j,t}$ that is much lower than x_{i,t_p}, it is large and negative; and it will be near zero if good news concerns input levels near x_{i,t_p} (or zero in the absence of good news). M_{i,t_c} will of course be nonzero for some observations when in fact the innovation in information is not sufficiently large to induce a change in fertilizer use by farmer i. Nevertheless, M_{i,t_c} should be a good predictor of both the direction and magnitude of changes in inputs. We refer to M_{it} as our "index of good news input levels." Equation (7) defines M_{i,t_c} for the modal case. Slight variants of this definition are used when i observes multiple harvests and/or there is different timing. When the farmer observes more than one harvest between t_c and t_p, our baseline definition of M_{i,t_c} uses a plant-weighted average of (7) across all plots j,t farmer i observed between times t_p and t_c.[23] When a planting that i observes is harvested shortly after t_c, as is the case for farmer k in Figure 5, profits $\pi_{j,t+5}$ are not fully observed when i's input decisions are finalized at t_c, but i has an excellent signal of the value of impending harvest.

[22] We use a variable band with to define "sufficiently close." For the large majority it means within 1 kilometer of location j and from time t to $t-3$. About ¼ of $x_{j,t}$ are on relatively isolated plots and for these we expand the geographic neighborhood to 3 kilometers. In both cases our baseline definition of proximate inputs are those within coarse categories of $x = 0$ and $x > 0$. The robustness of results to these assumptions is examined in Appendix B.

[23] It turns out in our data there are no multiple plantings with good news observations disagreeing about the direction of movement. Thus, we never have to average Ms of opposite sign.

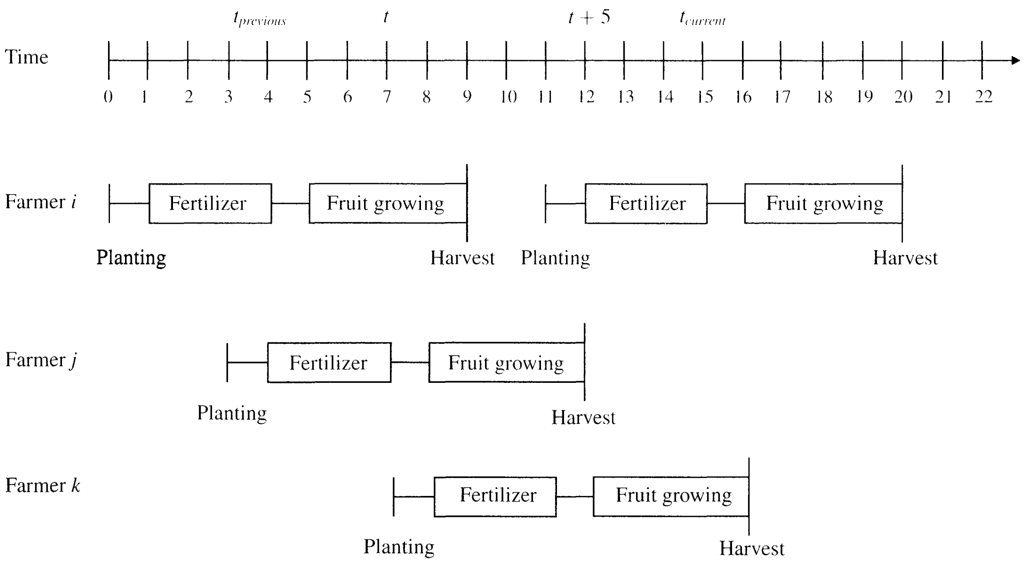

Figure 5. Timing of Cultivation and Information Revelation

Therefore, we are motivated to make use of these observations. Our baseline measure M_{i,t_c} is formed using (7) and discounted to reflect the reduced information compared to fully revealed harvests. The tedious details of these weighting schemes are described in Appendix A, and the robustness of our results to changes in weighting is examined in Appendix B.

III. Empirical Results

We begin by presenting some simple cross tabulations to illustrate our basic empirical strategy. Table 3 provides a cross tabulation of changes in fertilizer use and the two measures of information flow with the most robust estimated effects. The first row in each cell provides a simple count of the number of transitions in each category. The most populous category is one maintaining zero fertilizer use, while transitions from positive use to zero use are more common than transitions in the opposite direction. The second row in each cell is the within-cell average of M_{it}, our index of good news input levels. A farmer whose previous input level was above that of his "good news" neighbors input level will have a negative value for M, whereas those who observe good news at a higher input level than they had previously used will have M positive. Cell-average M is strongly predictive of changes in fertilizer inputs: it is strongly negative for those farmers who reduced fertilizer use to zero, positive for those who increased fertilizer use from zero, and near zero for those who did not change categories. The third row in the table provides a measure of the amount of bad news information about previously chosen input levels, the average (within-cell) share of new information that falls in this category. This bad news regarding past choices is clearly correlated with switching input categories.

The correlations evident in Table 3 are consistent with the implications of our model of social learning. However, as we have noted above (Implication 5), they could also be generated by spatially and serially correlated growing conditions. Therefore, we move beyond cross tabulations and estimate regressions predicting the occurrence of a change in input use and the change in inputs—specifications that include regressors controlling for this confounding source of variation in input changes.

Table 3—Transitions in Fertilizer Use and Receipt of Information
(*Proportion of pairs of individuals in each other's information neighborhood*)

		Current fertilizer use	
		Zero	Positive
Previous fertilizer use	Zero count	49	12
	Avg $M_{i,t}$	0.443	0.86
	Avg $s(bad, x = x_{i,previous})$	0.029	0.264
	Positive count	17	29
	Avg $M_{i,t}$	−3.301	−0.025
	Avg $s(bad, x = x_{i,previous})$	0.059	0.008

Notes: Count is the number of transitions in each category Avg $M_{i,t}$: within-cell average of $M_{i,t}$, our index of good news input levels Avg $s(bad, x = x_{i,previous})$: within-cell average of the share of plots observed by i that had bad news about the input level he used in his previous planting.

We now recall the base regression specifications that we use to examine the relationships between changes in fertilizer use and information shocks that were introduced in Section I. Alternative specifications and a discussion of robustness are considered in Section IV and Appendix B. We let the characteristics of i and his plot that we use for conditioning be contained in a vector $z_{i,t}$. These characteristics include the farmer's wealth, soil characteristics, and indicators for religion, clan, village, round of the planting, and the novice farmer indicator.

We first estimate a logistic model of the probability of a change in x:

$$(4') \qquad \Pr\{\Delta x_{i,t} \neq 0\} = \Lambda \begin{bmatrix} \alpha_1 s(good, x = x_{i,t_p}) + \alpha_2 s(good, x \neq x_{i,t_p}) \\ + \alpha_3 s(bad, x = x_{i,t_p}) + \alpha_4 s(bad, x \neq x_{i,t_p}) \\ + \alpha_5 \tilde{\Gamma}_{i,t} + z'_{i,t}\alpha_6 \end{bmatrix}.$$

The first four terms reflect the share of new information to the farmer in good/bad news and both input categories, as described in Section IID. Recall that $s(good, x = x_{i,p})$ is the share of plants on plots associated with good news and with inputs equal to the farmer's previous choice, with the other three s terms defined likewise. Our model implies α_1 and α_4 are negative and α_2 and α_3 are positive. Our control for growing conditions defined in Section IIC is represented by $\tilde{\Gamma}_{i,t}$. When the absolute discrepancy between farmer i's growing conditions w_{i,t_p} is very different from the growing conditions nearby (i,t), $\tilde{\Gamma}_{i,t}$ will tend to be large. Therefore, it should be correlated with changes induced by changes in growing conditions, and α_5 is anticipated to be positive.

Our baseline regression predicting changes in fertilizer use is

$$(5') \qquad \Delta x_{i,t} = \beta_1 M_{i,t} + \beta_2 \Gamma_{i,t} + z'_{i,t}\beta_3 + v_{i,t},$$

where $\Gamma_{i,t}$ defined in (6) is our crucial control for movements in $x_{i,t}$ induced by correlated growing conditions. As in (4′), $z_{i,t}$ includes wealth, soil characteristics, indicators for religion, clan, village, and round of the planting, and novice indicator. In addition, it includes the financial analog to $\Gamma_{i,t}$ with the neighborhood definition based on financial rather than geographic neighborhoods. Finally, the error term $v_{i,t}$ is permitted to be conditionally heteroskedastic and spatially correlated across plots as a general function of their physical distance using the spatial GMM approach of Conley (1999).

Recall from Section I that our key identification assumption is that conditional on our measure of changes in growing conditions (Γ_{it}), our measures of information shocks are uncorrelated with unobserved determinants of changes in input use. We cannot test this assumption directly. However, it is true that *observed* determinants of input changes are uncorrelated with this component of our information measures. Conditional on changes in growing conditions, our measures of information shocks are uncorrelated with *any* of the characteristics of farmers (experience, wealth, matrilineage, religion, land characteristics) that might influence input choice. Thus, we argue it is plausible that, conditional on our measure of changes in growing conditions (and our other covariates), surprisingly high profits achieved by a given farmer's information neighbors influence his decision to change his input level only through that information link: a significant, positive coefficient β_1 is evidence of social learning.

A. *Logit Results*

Table 4 presents the coefficient and spatial standard error estimates from equation (4′).[24] We show results for three descriptors of a change in inputs that vary with respect to their sensitivity to small changes in input use.

The dependent variable in column A is an indicator equal to one if the farmer changed his fertilizer use from zero in his previous planting to a positive value at t, or vice versa . We see that the direction of the influence of our information and experience variables upon the predicted probability of changing is as implied by our model. Observations of bad news at the farmer's previous input level strongly increase the predicted probability that he will change input levels. The estimated coefficient is positive, significantly different from zero, and large. A one-standard-deviation increase (0.12) in a farmer's observation of bad news at that farmer's previous level of fertilizer use is associated with an increase in the probability of changing of 15 percentage points, calculated at the median probability of changing fertilizer use (which is 13 percent). Similarly, observations of bad news at alternative levels of fertilizer use strongly decrease the predicted probability of changing. A one-standard-deviation increase (0.12) in the share of bad news at alternative fertilizer levels is associated with a reduction in the probability of changing fertilizer use of 9 percentage points (off of the same base probability of 13 percent). The point estimate of the effect of good news at alternative levels of fertilizer use on the probability of changing is positive, and that of the effect of good news at the previous level of use is negative, as anticipated, but these coefficients are relatively small (compared to those for bad news) and these estimates' precision is too low to statistically distinguish them from zero. As would be expected in virtually any model of learning, novice farmers are much more likely to change input levels. In column B, the dependent variable is an indicator that the absolute value of the change in fertilizer use is larger than one cedi per plant, Mean fertilizer use is two cedis per plant, and the twenty-fifth percentile of changes in fertilizer use for those farmers whose fertilizer use changed is one cedi per plant. Hence, this column focuses on relatively

[24] The standard errors in all our specifications use limiting results for cross-section estimation with spatial dependence characterized by physical distance between the centroids of each farmer's set of plots. Serial dependence is allowed only by use of time (round) dummies. Specifically, spatial standard errors are calculated using the estimator in Conley (1999) with a weighting function that is the product of one kernel in each dimension (North-South, East-West). In each dimension, the kernel starts at one and decreases linearly until it is zero at a distance of 1.5 kilometers and remains at zero for larger distances. This estimator is analogous to a Maurice S. Bartlett (1950) or Whitney K. Newey and Kenneth D. West (1987) time series covariance estimator and allows general correlation patterns for distances shorter than the cutoff. Note that plantings by the same farmer are allowed to be arbitrarily correlated, as they are all distance zero from each other. The inferences reported below are robust to cutoff distances between 1and 2 kilometers. This is largely due to the fact that there is little spatial correlation in our regression errors because we are intentionally conditioning on indices involving geographic neighbors' actions that provide a good signal of spatially correlated growing conditions.

TABLE 4—DETERMINANTS OF CHANGING INPUT USE

	A	B	C
	Dependent variable: Indicator for change between zero and positive	Dependent variable: Indicator for $\|change\|$ $> 1 Cedi/Plant$	Dependent variable: Indicator for nonzero change in fertilizer
Good news at previous input use $s(good, x = x_{i,previous})$	−0.94 (1.24) [−0.04]	−0.08 (0.95) [−0.01]	−0.34 (0.84) [−0.03]
Good news at alternative fertilizer use $s(good, x \neq x_{i,previous})$	1.15 (0.81) [0.03]	1.64 (0.78) [0.09]	2.35 (1.80) [0.14]
Bad news at lagged fertilizer use $s(bad, x = x_{i,previous})$	6.38 (2.86) [0.15]	4.32 (1.93) [0.20]	4.16 (1.80) [0.22]
Bad news at alternative fertilizer use $s(bad, x \neq x_{i,previous})$	−6.72 (3.04) [−0.09]	−5.90 (2.57) [−0.15]	−3.05 (1.85) [−0.09]
Ave. abs. dev. from geog. neighbors' fertilizer use $[\Gamma_{i,t}]$	0.09 (0.10) [0.07]	0.15 (0.07) [0.24]	0.08 (0.04) [0.15]
Novice farmer	2.32 (0.75) [0.26]	1.97 (0.89) [0.43]	1.22 (0.92) [0.30]
Talks with extension agent	−0.48 (0.61) [−0.05]	−1.35 (0.67) [−0.29]	−1.38 (0.76) [−0.34]
Wealth (million cedis)	0.20 (0.10) [0.06]	0.18 (0.13) [0.10]	0.10 (0.12) [0.06]
Clan 1	1.62 (1.14) [0.18]	1.59 (1.10) [0.35]	2.15 (1.03) [0.54]
Clan 2	4.54 (1.45) [0.51]	2.15 (1.23) [0.47]	2.51 (0.99) [0.63]
Church 1	1.84 (0.93) [0.21]	−0.29 (0.73) [−0.06]	−0.24 (0.77) [−0.06]

Notes: Logit MLE point estimates, spatial GMM standard errors in parentheses, and marginal effects in brackets below. Standard errors allow for heteroskedasticity and correlation as a function of physical distance, (see footnote 24 for details). Sample size = 107. Pseudo R^2's 0.40, 0.26, and 0.31 for columns A, B, and C, respectively. A full set of village and round dummies were included but not reported. Information neighborhoods defined using responses to: "Have you ever gone to farmer ____ for advice about your farm?" Marginal effects: calculated at the median probability of a change for each column (0.13, 0.32, and 0.53, respectively). For continuous variables, this is the change in probability associated with a one-standard-deviation increase in the variable; for dummy variables a one-unit increase in the variable.

large changes in fertilizer use. The results are qualitatively similar to those in column A, with the main exceptions that the estimated probability of changing fertilizer levels is significantly increasing in good news at alternative fertilizer levels and in the average absolute deviation of farmer i's lagged inputs from his geographic neighbors. The latter provides evidence of the importance of positively serially and spatially correlated unobserved shocks to the productivity of fertilizer. In column C the dependent variable is a indicator variable equal to one whenever the change in input use is nonzero. Obviously, the median probability of a nonzero change is higher in this case: it is now 53 percent. The pattern of statistical significance of the estimates remains similar to those of the previous specifications. As in the other columns, bad news at the farmer's lagged fertilizer use, bad news at alternative levels of fertilizer use, the absolute deviation of the farmer's fertilizer use from that of his geographic neighbors, and his experience are all quantitatively important determinants of the likelihood of changing fertilizer use, although the effect of experience is no longer statistically significant.

Finally, all three specifications include an indicator for whether the respondent has ever received advice from the local extension agent. We do not know when any such conversation occurred. In columns B and C, the estimate indicates that those who have received advice from an extension agent are significantly less likely to adjust their fertilizer use.

B. *Regression Results*

Table 5, column A, presents the results of estimating equation (5′). The coefficient on the index of good news input levels in the farmer's information neighborhood is positive and statistically significant, as implied by our model. A one-standard-deviation increase in M (about four) is associated with an increase in fertilizer use of approximately four cedis per plant, which is greater than the median level of fertilizer use per plant of those farmers who use fertilizer.

Round indicator variables are included, but not reported. There is no evidence that changes in input use are significantly related to inputs used by financial neighbors. In this, as in the following columns, changes in fertilizer use are strongly in the direction of the use by one's geographic neighbors.

In column B, we examine the relationship between experience and a farmer's responsiveness to information on the profitability of fertilizer. The coefficients on M for veteran and novice farmers are statistically different from each other at the 3 percent level. There is no evidence that veteran pineapple farmers respond at all to good news about alternative levels of fertilizer use. For novice farmers, in contrast, a one-standard-deviation increase in M is associated with an increase in fertilizer use of approximately four cedis per plant. We raised the possibility in Section IIB that farmers might be heterogeneous in their ability to learn from others, and in particular that lower ability farmers adopt pineapple more slowly (or not at all) and react less to information from their neighbors. If this is the case, then our use of a sample of current pineapple farmers overstates the responsiveness of farmers in general to information from neighbors. The results in column B provide some evidence on the importance of this kind of selection. If there is selection such that later adopters of pineapple are less responsive to news, then this selection is sufficiently weak that recent adopters are still very responsive to new information.

Columns C–F present the results of an investigation of the influence of the source of information on farmer i's reactions. In alternate specifications, we use variants of M defined on partitions of farmer i's information neighborhoods based on i's information neighbors' experience, farm size, relative wealth, and relative soil type. Our novice/veteran indicators are as described above. We define large farms as those with plantings of at least 60,000 total pineapple plants over our sample period (27 percent of farmers have large farms).[25] Finally, we define a classification of wealth as rich/poor, with rich as those whose nonland wealth at the start of the survey is greater than the mean nonland wealth (30 percent of farmers are rich by this definition).

Column C defines M separately for novice and veteran farmers in i's information neighborhood. The coefficient on M using veteran farmers' results is large and significant, and that corresponding to M comprising novice farmers' information is not. Column D presents a partition depending on whether i's information neighbor is in i's wealth category (both rich or both poor). Wealth-partitioned M is an important and significant predictor for same category

[25] Median and mean numbers of plants planted by farmers in our sample are 22,000 and 41,000, respectively.

TABLE 5—PREDICTING INNOVATIONS IN INPUT USE, DIFFERENTIAL EFFECTS BY SOURCE OF INFORMATION
(*Dependent variable: Innovation in per plant fertilizer use*)

	A	B	C	D	E	F
Index of good news input levels ($M_{i,t}$)	1.05					
	(0.20)					
$M_{i,t}$ × novice farmer		1.07				
		(0.22)				
$M_{i,t}$ × veteran farmer		−0.46				
		(0.34)				
Index of good news input levels by novice farmers			−0.05			
			(0.39)			
Index of good news input levels by veteran farmers			1.05			
			(0.20)			
Index of good news input levels by farmers with same wealth				1.06		
				(0.22)		
Index of good news input levels by farmers with different wealth				−0.32		
				(0.32)		
Index of good news input levels on big farms					1.17	
					(0.19)	
Index of good news input levels on small farms					0.92	
					(0.20)	
Index of good news input levels, farmers with same soil						1.08
						(0.23)
Index of good news input levels, farmers with different soil						0.93
						(0.22)
Novice farmer		3.97	4.03	4.02	3.96	3.94
		(2.67)	(2.68)	(2.67)	(2.69)	(2.77)
Avg. dev. of geog. neighbors from previous use [$\Gamma_{i,t}$]	0.52	0.56	0.56	0.57	0.56	0.57
	(0.07)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
Avg. dev. of financial neighbors from prev. use	0.52	0.55	0.38	0.41	0.23	0.24
	(0.59)	(0.57)	(0.58)	(0.54)	(0.62)	(0.61)
Village 1	−7.50	−8.09	−7.97	−8.10	−7.68	−7.79
	(1.22)	(1.50)	(1.42)	(1.48)	(1.39)	(1.36)
Village 2	−0.47	−1.91	−1.94	−1.98	−1.60	−1.59
	(1.53)	(2.07)	(1.99)	(2.07)	(1.99)	(2.03)
Wealth (million cedis)	0.10	0.41	0.36	0.40	0.24	0.26
	(0.25)	(0.17)	(0.18)	(0.18)	(0.21)	(0.21)
Clan 1	−2.36	−2.44	−2.43	−2.32	−2.24	−2.33
	(1.41)	(1.25)	(1.27)	(1.23)	(1.28)	(1.30)
Clan 2	−0.35	0.00	−0.10	−0.13	−0.26	−0.24
	(1.44)	(1.35)	(1.34)	(1.35)	(1.31)	(1.31)
Church 1	0.13	0.63	0.48	0.41	0.69	0.74
	(1.31)	(1.13)	(1.09)	(1.13)	(1.15)	(1.15)

Notes: OLS point estimates, spatial GMM (Conley 1999) standard errors in brackets allow for heteroskedasticity and correlation as a function of physical distance; see footnote 24 for details. Sample size = 107. A full set of round dummies included but not reported. Information neighborhoods defined using responses to: Have you ever gone to farmer ____ for advice about your farm?

neighbors but not for different category neighbors. For each of the pairs of M partition coefficients in columns C and D, their difference is statistically significant with a p-value under 2 percent. Column E presents analogous estimates with a partition of M depending on the size of the farms in i's information neighborhood. Both coefficient estimates are large, positive, and statistically significant. Point estimates suggest that the responsiveness of input use to news from large farmers may be stronger than it is to similar news from small farmers. However, these estimates are not statistically different from each other. Finally, Column F presents estimates with M defined using a partition based on whether i's neighbor has the same soil type as i (sandy or clay). These estimates provide no significant evidence that news from others with the same soil type matters more to a farmer. In summary, novice farmers appear to be the ones reacting to good news and they tend to react to information revealed by neighbors who are veterans and who have similar wealth.[26]

IV. Robustness Checks and Extensions

A. *Learning-by-Doing, Alternative Information Neighborhoods, Endogenous Sorting*

There are 19 cultivators in our data who have multiple plantings sufficiently far apart in time for the fruit on the earlier planting to be growing before fertilizer is applied on the later planting. These farmers present the opportunity to identify learning-by-doing alongside the social learning that is the key focus of the paper. Column A of Table 6 presents regression results with a partition of M using only an individual's past history and using only other farmers' information. In column A, we see that there is no important or statistically significant difference between the impact of good news on one's own plot and that of good news on a neighbor's plot.

In columns B–D, we examine whether our finding that M predicts innovations in fertilizer is robust to changes in the definition of an information link. Full definitions of each of these alternatives are provided in Appendix A. In column B, j is considered to be in i's information neighborhood if j is named by i when asked a series of open-ended questions about who taught them to farm and from whom they have received farming advice (or vice versa). In column C, we use the broader definition of an information link if either i or j is listed anywhere in the other's roster of interactions with other sample members. In column D, we define information neighborhoods based on the predicted probabilities for going to another farmer for advice (corresponding to the estimates in Table A1).

Regardless of the precise definition of the information neighborhood, the coefficient on M is statistically significant and large for novice farmers (the standard deviation of M is approximately 3.5 for the two roster-of-contacts neighborhoods, and about 1 for the "predicted advice" neighborhood). In each case we find that when i is a novice, good news experiments in i's information neighborhood tend to be followed by i changing his fertilizer use in the direction of those experiments, conditional on our growing conditions control, village, and round effects, and i's wealth, clan, and church membership. In contrast, there is evidence of responsiveness to information by veteran pineapple farmers only for one metric: predicted ask-for-advice.

The robustness of our main results to the use of predicted neighborhoods provides some assurance that they are not driven by sorting/selection effects. It is also reassuring that, across alternate definitions, our strong partial correlations of M with changes in fertilizer are driven as much by farmers moving down in response to good news at lower levels of fertilizer use as by upward movement in response to good news at higher fertilizer levels. Endogeneity due to positive, associative sorting along a characteristic correlated with fertilizer use (e.g., unobserved wealth) could produce a tendency for good news to be associated with either high or low fertilizer use, but we were unable to think of a scenario where it would induce both. For example, assume high fertilizer levels are more productive than low levels, and assume a positive correlation between unobserved wealth and the amount of fertilizer used. In such a scenario, positive sorting would produce high-fertilizer farmers with neighbors prone to good news events at high levels of fertilizer, but not farmers lowering their fertilizer use being prone to receipt of good news from their low-fertilizer neighbors.

[26] Of course, an important caveat to this summary is that our small sample size constrains us to examine these partitions one variable at a time rather than jointly.

TABLE 6—LEARNING-BY-DOING AND ALTERNATE DEFINITIONS OF THE INFORMATION NETWORK
(*Dependent variable: Innovation in per plant fertilizer use*)

Information neighborhood metric	Learning-by-doing farm learning by others A	Roster of contacts: Information only B	Roster of contacts: Full set of contacts C	Predicted advice D
M from own plots only	1.46 (0.56)			
M from information neighbors	1.22 (0.31)			
$M \times$ novice farmer		1.50 (0.28)	1.50 (0.28)	6.40 (1.13)
$M \times$ veteran farmer		0.19 (0.21)	0.15 (0.22)	4.75 (1.84)
Novice farmer		4.66 (2.83)	4.64 (2.84)	4.00 (2.76)
Ave. dev. of lagged use from geographic neighbors' use ($\Gamma_{i,t}$)	0.49 (0.10)	0.49 (0.09)	0.49 (0.09)	0.32 (0.12)
R^2	0.73	0.72	0.72	0.73

Notes: OLS point estimates, spatial GMM (Conley 1999) standard errors in brackets allow for heteroskedasticity and correlation as a function of physical distance; see footnote 24 for details. Sample size = 107. All of the variables in Table 5 were included, but coefficients are not reported. Alternative information neighborhoods are as defined in Section IIIA and Appendix A.

We are also confident our results are not driven by reverse-causality sorting. For example, suppose farmers first decided whether to increase or decrease their fertilizer use and then sought out the friendship/advice of surprisingly successful farmers who tended to dogmatically use their chosen amount. Such a scenario is unlikely in our setting, since our data on information connections were collected at the beginning of a two-year survey. It is implausible that these farmers premeditated to the extent that they planned fertilizer choices one to two years in advance and chose their contacts accordingly. Moreover, the results using the predicted information neighborhoods could not be generated by this type of scenario.

B. *Learning about Optimal Labor Use in Pineapple and in Established Crops*

There should be no learning about optimal inputs in the maize-cassava cultivation that occurs in our study villages. A standard maize-cassava intercrop pattern has been the foundation of the economy here since the local decline of cocoa cultivation in the 1930s. The characteristics of the maize-cassava production function are well known to farmers in these villages. This provides a "placebo" environment in which to test our methodology. The only nonseed variable input into maize-cassava production is labor; no chemical inputs are used on any maize-cassava plot in these villages (Markus Goldstein and Udry 2008).

In this section we estimate a model of learning about optimal labor use in maize-cassava cultivation. To verify that our results regarding learning about labor use is not an artifact of a peculiar aspect of our data on labor inputs, we also estimate a model of learning about optimal labor use on pineapple farms. Since labor and fertilizer are used in approximately fixed proportions in pineapple cultivation, we will see the same patterns associated with learning regarding labor that we found with fertilizer.

We estimate a regression of changes in labor inputs for pineapple plots and for maize-cassava plots with a specification analogous to (5′):

$$\text{(8)} \qquad \Delta x_{i,t}^{labor} = \delta_1 M_{i,t}^{labor} + \delta_2 \Gamma_{i,t}^{labor} + z_{i,t}' \delta_3 + u_{i,t},$$

where x^{labor} is the labor input per plant for pineapples and per hectare for maize-cassava, and $M_{i,t}^{labor}$ and $\Gamma_{i,t}^{labor}$ are constructed exactly as above for this labor input.[27] We expect a positive δ_1 for pineapple plots if pineapple farmers are learning from their neighbors about the productivity of labor. A nonzero δ_1 for maize-cassava cannot be attributed to learning, because this technology is well established.

We estimate (8) for the same sample of pineapple plots examined above. Labor inputs are measured over the crucial period early in the life cycle of the pineapple, during which fertilizer inputs also occur. All farmers change their labor inputs across plantings, so there is no need to estimate an analog of the logit (4′). M^{labor} is defined using the benchmark (asked for advice) information neighborhood. Column A of Table 7 presents the results of estimating (8). We condition on the average deviation of i's lagged labor use from the lagged labor used by his geographic neighbors, Γ^{labor}, and its analog for his financial neighbors.

For pineapple farmers, good news experiments in i's information neighborhood tend to be followed by i changing his labor use in the direction of those experiments' labor, conditional on geographic and financial neighbors' lagged labor use, plot characteristics, village and round effects, and i's wealth, clan, and church membership. The coefficient is also large: a one-standard-deviation increase in M^{labor} (which is 348) is associated with an increase in labor use of approximately 682 cedis per plant, which is 37 per cent of the median labor use per plant on pineapple plots (1,845). Pineapple farmers are learning about the productivity of inputs in the cultivation of pineapple from the experiences of their information neighbors. The data on labor show the same pattern we saw with fertilizer.[28]

In contrast, in column B, we see that there is no evidence that maize-cassava farmers adjust labor inputs to information from the cultivation of their information neighbors.[29] The coefficient of the learning index M^{labor} is virtually zero (at the point estimate, a one-standard-deviation increase in M^{labor} is associated with an increase in labor use of 5,000 cedis per hectare, while the mean labor use is 475,000 cedis, and its standard deviation is 571,000). M^{labor} has no significant predictive power for innovations in labor use in maize-cassava production, just as we expect given the familiarity of this farming system in the study region.

For both pineapple and maize-cassava, we find that there is a strong geographic correlation in innovations in labor use. There are important spatially and serially correlated shocks to the productivity of inputs. This underscores the value of direct data on communication for defining information neighborhoods. In the more typical case in which we had data only on geographic proximity, it would be tempting to rely on this to proxy for information links. The consequences of this are presented in column C. We construct a new variable, analogous to M^{labor}, but with

[27] Labor inputs include both the value of hired labor and that provided by the farmer's household. The labor input range was divided into two categories (above and below median) for determining whether profits were unusually high given inputs.

[28] Note, again, that because labor and fertilizer move together in approximately fixed proportions (subject to the additional measurement error in labor), this is not independent evidence of learning about inputs on pineapple farms.

[29] There are two differences in specification between the pineapple and maize-cassava regressions: first, in contrast to pineapple, the maize-cassava intercrop system is seasonal. Hence, we compare inputs across successive seasons and replace the round indicators in A with season indictors in B. Second, the maize-cassava mixture is grown in all four of our survey villages, while pineapple is grown in only three villages; hence, there is an additional village indicator for maize-cassava.

TABLE 7—PREDICTING INNOVATIONS IN LABOR FOR PINEAPPLE AND MAIZE-CASSAVA PLOTS
(Dependent variable: First difference in labor inputs for pineapple and maize-cassava)

Crop	Pineapple (labor and cost in cedis per plant) A	Maize-cassava (labor cost in 1000 cedis per hectare) B	Maize-cassava (labor cost in 1000 cedis per hectare) C
Index of good news input levels: M^{labor}	1.96	0.02	
	(0.86)	(0.16)	
Index of good news input levels in the the geographic neighborhood			0.32
			(0.12)
Average deviation of lagged use from geographic neighbors' use $[\Gamma_{i,t}^{labor}]$	0.49	0.74	
	(0.20)	(0.09)	
Average deviation of lagged use from financial neighbors' use	0.52	0.01	
	(0.29)	(0.07)	
Village 1	162.12	−111.66	−79.29
	(301.85)	(77.24)	(87.40)
Village 2	−432.70	−160.23	−246.19
	(289.83)	(112.79)	(109.97)
Village 3		−196.84	−198.11
		(71.40)	(82.96)
Wealth (million cedis)	166.42	41.66	2.44
	(91.55)	(30.49)	(35.13)
Clan 1	−159.86	1.10	−378.15
	(247.73)	(267.59)	(299.52)
Clan 2	463.35	−48.84	−67.78
	(239.96)	(71.20)	(65.29)
Church 1	−552.22	−86.44	−48.52
	(253.94)	(83.54)	(91.96)
Sample size	405	405	
R^2	0.55	0.42	0.24

Notes: OLS point estimates, spatial GMM (Conley 1999) standard errors in brackets allow for heteroskedasticity and correlation as a function of physical distance; see footnote 24 for details. Round/season dummies included but not reported. Information neighborhood from: Have you ever gone to farmer ——- for advice about your farm?

information links replaced with an indicator of geographic proximity. We see in column C that maize-cassava farmers adjust labor inputs in the direction of successful "experiments" in their geographic neighborhood. The coefficient is large (a one-standard-deviation increase in the index of experiments in the geographic neighborhood is associated with an increased labor input of 153,000 cedis/hectare, compared to mean labor input of 475,000 cedis/hectare) and statistically significant. This result has nothing to do with learning; it is induced entirely by the strong correlations in growing conditions. However, without our direct data on communication, we might incorrectly infer the existence of social learning about labor productivity in this well-established farming system.

V. Conclusion

This paper presents evidence that social learning is important in the diffusion of knowledge regarding pineapple cultivation in Ghana. We take advantage of data that combine agronomic and conventional economic information with details regarding relationships between farmers to address the challenge of identifying learning effects in an economy undergoing rapid technological change. We find that farmers are more likely to change input levels upon the receipt of bad news about the profitability of their previous level of input use, and less likely to change when they observe bad news about the profitability of alternative levels of inputs. Farmers tend to increase (decrease) input use when an information neighbor achieves higher than expected profits when using more (less) inputs than they previously used. This holds when controlling for correlations in growing conditions, for common credit shocks using a notion of financial neighborhoods, and across several information metrics. Support for the interpretation of our results as indicating learning is provided by the fact that it is novice farmers who are most responsive to news in their information neighborhoods. Additional support is provided by our finding no evidence of learning when our methodology is applied to a known maize-cassava technology.

Further evidence of learning is provided by changes in profits that correspond to input changes that appear to be mistakes and those that appear to be correct, subject to a conjecture regarding the optimal level of input use. Learning implies that farmers respond to both signal and noise. So, particularly in the early stages of learning, we expect to see novice farmers make mistakes by switching to what is truly a suboptimal input level after seeing it perform surprisingly well in a small number of experiments. We have a strong belief that optimal fertilizer use for pineapple in Ghana is much higher than the levels we observe in our sample; certainly it is greater than zero.[30] Thus, we believe that many of the movements from positive to zero fertilizer use are mistakes. About a quarter of the farmers moved toward our conjectured optimal input levels in response to good news about high levels of fertilizer use. On average, these farmers have the highest growth of profit in our sample, with average profit growth of 122 cedis per plant. We also observe approximately the same number of farmers who make (we conjecture) mistakes by reducing their level of fertilizer input in response to good news about low levels of fertilizer. They achieve substantially lower average profit growth at 62 cedis per plant. Though such mistakes provide evidence in favor of learning, they inherently undermine the ability of our data to provide us with good estimates of the value of learning via estimating profits associated with optimal inputs. The span of our data is simply too short for most farmers to have learned optimal input choices.

We have presented evidence that social learning plays a role in the cultivation decisions of these farmers. Information, therefore, has value in these villages, as do the network connections through which that information flows. This raises the possibility that farmers consider the availability of information when forming the connections that underlie their information neighborhoods. If so, measurement of the extent of social learning is not sufficient for adequate evaluation of policy regarding the diffusion of technology. It is necessary, in addition, to understand the endogenous process of information network formation, making this a very important topic for future research. For example, consider the impact of a subsidy offered to one farmer in a village that induces him to use an optimally large amount of fertilizer and (with high probability) get high profits. The speed with which this information spreads, and hence the value of the subsidy, depends upon the choices of the subsidized farmer and others in the village to make and maintain information linkages. These choices may depend on the value of the information available to each farmer and upon the costs of information links, which may depend on a rich array of characteristics of the farmers and the social structure of the village. In some contexts, differing religions may be an effective barrier to communication. In others, gender, wealth, or family ties may be the most salient determinants of the shape of the information network.

[30] This belief is based mainly on follow-up visits to these villages several years after the survey period, and the observation that virtually all pineapple farmers now use positive amounts of fertilizer.

Appendix A: Construction of Alternate Information Neighborhoods and the Good/Bad News Indices

In this Appendix we provide details concerning the information neighborhoods mentioned in Section IIA and the construction of our index of good news in those cases in which farmer i might be deciding on his fertilizer input before the outcome on farmer j's plot is fully revealed by his harvest.

Two of our alternate indicators for the existence of an information link between two farmers are based on a listing of all the individuals named by each respondent in a number of different contexts. These data include people named in response to questions designed to record all "significant" conversations about farming between individuals, and people who were hired by, borrowed from, lent or sold output to, or exchanged gifts, transacted land, or jointly held assets with the respondent.[31] We construct two metrics from this information, first defining an information link to exist between two farmers if either reports learning about farming from the other. Because important information might be transmitted during a quite casual conversation, we also define a broader information neighborhood that defines a link as existing if either farmer lists the other anywhere in his or her roster of contacts.

Both our baseline "ask for advice" metric and these "roster of contacts"–based measures have potential drawbacks. The ask for advice measure is based on a random sample of other farmers, and so yields estimates of the information neighborhood of a farmer that are smaller than his actual information neighborhood. The roster of contacts measures include some pairs who probably do not discuss farming activities, and depends on the respondents' subjective understanding of "significant conversations about farming." In addition, there is some concern that observed information links might be endogenously formed in anticipation of receiving input advice (see Venkatesh Bala and Sanjeev Goyal 2000; Armin Falk and Michael Kosfeld 2003). Therefore, we also construct a predicted information neighborhood based on estimates reported in Table A1 of the probability of a link, based on the question "Have you ever gone to ___ for advice about your farm?" given pair characteristics. There is evidence of spatial correlation in link patterns, as the marginal effect of proximity is to increase link probability, but distance is not the sole determinant of links. Individuals are more likely to have information links if they are of the same gender, same clan, and similar ages. Individuals with different levels of wealth are more likely to be linked, reflecting the strong vertical patron-client ties that exist in these villages. There is no evidence that religion influences information links.[32] We construct the predicted probability that farmer j is in i's information neighborhood from the parameter estimates in Table A1, and $M_{i,t_c}^{predicted}$ is constructed as a weighted average of (7) across all plots j, t planted in farmer i's village between times t_p and t_c, with weights equal to this predicted probability (times the number of plants on j's plot).

Turn now to the timing of information revelation. In about half of our observations, farmer i observes farmer k, whose profits $\pi_{k,t+5}$ are not fully observed when farmer i's input decisions are finalized at t_c, as illustrated in Figure 5. In this case, farmer i may have an excellent signal about

[31] Significant conversations include, for example, discussions of techniques for using agricultural chemicals or seeds, dealing with agricultural problems, or crop choice.

[32] For the sake of discussion of the quantitative importance of the determinants of link probability, take as a base pair one with the mean values of wealth difference, age difference, and distance (2.9, 10.9, and 1.25 respectively) with the same gender and soil but different clans, religions, and where neither party holds an office. The point estimate of the link probability for this base pair is 22 percent. This point estimate would drop to 14 percent if one of the parties held some office and increase to 31 percent if instead they were from the same clan. A reduction in estimated probabilities to around 15 percent would accompany an approximate doubling of the base pair's distance or age difference, individually. Likewise, an approximate doubling of the wealth difference would result in an increase to 31 percent. If the pair is not of the same gender, the predicted probability of one asking the other for advice drops dramatically to 5 percent.

Table A1—Determinants of Information Links

Either party holds traditional office	−0.502
	(0.25)
Same religion	0.169
	(0.32)
Same clan	0.438
	(0.24)
Same gender	2.138
	(0.75)
Same soil type	−0.176
	(0.26)
Absolute age difference (years)	−0.046
	(0.02)
Absolute wealth difference (million cedis)	0.153
	(0.03)
Distance between plot centers (meters)	−0.445
	(0.16)
Constant	−2.626
	(0.80)

Notes: Logit MLE estimates, sample size = 528, pseudo $R^2 = 0.12$. Dependent variable is one if either party answered yes to the question: Have you ever gone to ____ for advice about your farm?

a $t + 5$ harvest at time $t + 4$ and is likely to have some signal about a $t + 5$ harvest as early as $t + 1$. Therefore, we are motivated to make use of these observations. Our simplest treatment is to construct $M_{i,t_{current}}$ as in (7) separately for groups of observations with the same lag between $t + 5$ and $t_{current}$. Letting $k = t + 5 - t_{current}$ refer to this lag, we construct $M_{i,t_{current}}(k)$ for groups of $k \leq 1$ and $k = 2, 3$, or 4 and examine the robustness of our conclusions to using such a set of *Ms*. These results are presented in column A of Table A2 and are discussed in Appendix B. However, to conserve degrees of freedom in our baseline specification we define our $M_{i,t_{current}}$ regressor to be a linear combination of elements of this set with $M_{i,t_{current}}(1)$ getting a weight of one, $M_{i,t_{current}}(2)$ a weight of 0.75, and so on down to $M_{i,t_{current}}(4)$ with a weight of 0.25.[33]

Appendix B: Robustness to Assumptions on Construction of M_{it}

In Tables A2 and A3, we examine some of the assumptions we have made about the timing of learning, the categories of fertilizer used in constructing our proxy for subjective expectations, the size of the geographic neighborhood, and conditioning on soil characteristics. In each case we look at the impact of the specification change for our regression of Δx on M interacted with our experience indicator.

For Table A2 in column A, we examine the assumptions regarding the timing of information flows from neighbors' pineapple plots. We adopt a more flexible specification which permits the responsiveness of *i*'s fertilizer use to vary depending upon the lag between his planting and the planting of his information neighbor's plot. There is virtually no effect of the success of plots

[33] When we have both multiple harvests observed by farmer i and some of these harvests are at times after t_c, we use a weighted average for baseline $M_{i,t}$ that has this same pattern of weights based upon a lag between $t + 5$ and t_c and plant-weighting across plots.

TABLE A2—ROBUSTNESS TO CHANGES IN SPECIFICATION
(*Dependent variable: Innovation in per plant fertilizer use*)

	Flexible lags in learning (A)	Only largest good news plot enters M (B)	Drop observations with multiple good news plots (C)	Fertilizer categories zero, medium, and high (D)	Geographic neighborhood within 500m (E)
M × novice farmer		1.06	1.17	1.03	1.85
		(0.21)	(0.24)	(0.19)	(0.19)
M × veteran farmer		0.61	−0.59	−0.41	−0.13
		(1.26)	(0.33)	(0.35)	(0.40)
Novice farmer		3.98	4.02	4.01	2.84
		(2.68)	(2.89)	(2.71)	(2.73)
M (t − 1)	−0.19				
	(0.15)				
M (t − 2)	0.56				
	(0.14)				
M (t − 3)	0.66				
	(0.19)				
M (t − 4 and earlier)	0.90				
	(0.52)				
Lagged own fertilizer use					
Avg. dev. of lagged use from geog. neighbors' use ($\Gamma_{i,t}$)	0.54	0.56	0.52	0.58	0.10
	(0.11)	(0.09)	(0.10)	(0.08)	(0.06)
Sample size	107	107	94	107	107
R^2	0.69	0.73	0.75	0.68	0.68

Notes: OLS point estimates, spatial GMM (Conley 1999) standard errors in parentheses allow for heteroskedasticity and correlation as a function of physical distance (see footnote 24). All of the covariates listed in Table 5 are included in the regressions, but not reported. Alternative specifications are as defined in Appendix B.

planted in round $t - 1$ on i's round t planting. As the lag increases, so does the estimated effect of a successful experiment, until for plots planted at least four rounds previously the magnitude of the coefficient reaches 0.90.

In columns B and C, we examine the impact of averaging across multiple good news sources in $M_{i,t}$. For those 30 percent of our observations in which more than one plot enters into the calculation of M in our standard specification, we select only the one, largest plot (that is, the plot with the most pineapple plants) that provides information to the farmer. Thus, in column B, $M_{i,t}$ is defined as in (7), with $x_{j,t}$ defined as the level of fertilizer use on the largest successful plot in i's information neighborhood during the relevant time period. In column C, we take the more draconian step of dropping those observations for which more than one plot enters into the calculation of M in our standard specification. In neither case do the results change in any substantive way.

In column D, we modify our categorization of fertilizer use. Expectations about profitability of fertilizer use had been defined over the two coarse categories of $x = 0$ and $x > 0$. For 40 percent of the plantings in our sample, there is at least one planting in the farmer's information neighborhood that provides information about the profitability of $x = 0$; for 32 percent of the plantings there is at least one planting in the farmer's information neighborhood that provides information about $x > 0$. We now define expectations over three categories of input intensity: $x = 0, 0 < x \leq x_h, x_h < x$, where $x_h = 2.5$ (the eightieth percentile of fertilizer use is about 2.5). As can be seen in column D, this change in specification has no qualitative effect on the results. This conclusion is robust for any x_h less than the eighty-fifth percentile of fertilizer use. For larger x_h, the precision of the estimates falls enough that the coefficient on M is not significant at conventional levels. It does not appear to be feasible to define expectations over more than three meaningful categories given the size of our dataset.

In column E, we alter the definition of the geographic neighborhood so that only plots within 500 meters fall within a plot's geographic neighborhood. Again, we find that novice farmers change their fertilizer use in the direction of inputs associated with good news experiments by their information neighbors, but that experienced farmers do not. Very similar results are obtained when geographic neighborhoods are defined as within 1,500 meters.

For Table A3, in column A we include information on soil characteristics in the conditioning set. We lose some observations by doing so, because soil testing was not completed on all plots, but once again the core result is unchanged: the coefficient on M is positive, large, and statistically significant for novice farmers but not for experienced farmers.

In column B we examine the possibility that our results are an artifact of mean reversion in fertilizer use. Lagged own fertilizer use appears both in the dependent variable and in $M_{i,t}$, raising the possibility that mean reversion in fertilizer use, perhaps due to large measurement error, might lead us to find a spuriously significant coefficient on $M_{i,t}$. In the absence of any learning effects, the average $x_{k,t-1}$ across good news observations would be an estimate of the conditional mean of $x_{k,t}$, given $\pi_{k,t}$ was above its expectation. So $M_{i,t}$ could be interpreted as a noisy, biased estimate of whether $x_{i,t-1}$ is above or below its unconditional expectation, which might be positively correlated with $\Delta x_{i,t}$ solely due to mean reversion. A priori, we think this is an unlikely source of spurious results, as we include $\Gamma_{i,t}$ in (5′) in addition to $M_{i,t}$. The sample size within geographic neighborhoods is considerably larger than that in information neighborhoods. Despite a higher spatial correlation within geographic neighborhoods, averages within this larger neighborhood will have a smaller variance than averages within information neighborhoods. Therefore, if mean reversion were driving correlations, Γ_{it} should be a much less noisy measure of whether $x_{i,t-1}$ is above its long-run mean. Once $\Gamma_{i,t}$ is conditioned upon, $M_{i,t}$ should offer little or no predictive power for $\Delta x_{i,t}$ resulting from mean reversion.[34] However, to be confident that our results are not an artifact of mean reversion, in column B we add the lagged fertilizer level $x_{i,t-1}$ to the regression. The coefficients on $M_{i,t}$ change by a magnitude comparable to some of our other alternative specifications, and for novice farmers it remains a significant predictor. The coefficient on $\Gamma_{i,t}$ changes the most dramatically; this is unsurprising since the lagged input levels are spatially correlated and thus provide an alternate control for spatially correlated growing conditions.

These results should be robust to conditioning on lagged own profits. This is confirmed in column C, where we show that the key coefficients change little and that lagged profit realization has no influence on innovations in fertilizer use. This is consistent with our results on the credit neighborhood. We have no evidence that variations in the opportunity cost of capital are playing important roles in fertilizer choices.

In addition, we show in column D that the results are robust to constructing $\Gamma_{i,t}$ in a way that is strictly analogous to the way $M_{i,t}$ is constructed. In column D, the average deviation of previous use from geographic neighbor's use variable is constructed as $\ddot{\Gamma}_{it} = \ddot{x}_{it}^{close} - x_{i,t_{previous}}$,

[34] A special case of mean-reverting $x_{i,t}$ would result if our data on inputs were dominated by large amounts of classical measurement error. However, we think this case is unlikely to have occurred, as the field research was specifically designed to collect accurate data on farming inputs (including the number of plants planted) and output by sacrificing sample size in exchange for frequent and thorough visits to respondents.

We also examined the special case of measurement error by performing Monte Carlo experiments (available upon request), adding artificial measurement error to our fertilizer data. The mean of per-plant fertilizer use is four, and its standard deviation is seven; to our data we added a mean zero normal draws with standard deviation 1 to 7 (truncated so that measured fertilizer use is never negative). The estimated coefficient on M becomes insignificant at conventional levels when the standard deviation of the added noise is four, while Γ remains a significant predictor.

TABLE A3—ROBUSTNESS TO CHANGES IN SPECIFICATION
(*Dependent variable: Innovation in per plant fertilizer use*)

	Soil characteristics (A)	Lagged fertilizer use (B)	Lagged profits (C)	Geographic neighbor defined as is M (D)	$i;t$ includes $t+1, t+2, t+3$ in addition to $t-3$ through t (E)
$M \times$ novice farmer	1.05	0.32	0.98	0.50	0.72
	(0.20)	(0.13)	(0.21)	(0.22)	(0.30)
$M \times$ veteran farmer	−0.77	−0.46	−0.28	−0.36	−0.45
	(0.58)	(0.32)	(0.52)	(0.34)	(0.34)
Novice farmer	5.21	4.02	3.27	4.03	4.32
	(2.46)	(2.66)	(2.13)	(2.71)	(2.75)
Avg. dev. of lagged use from geog.	0.52	0.10	0.65	1.74	0.77
neighbors' use ($\Gamma_{i,t}$)	(0.08)	(0.17)	(0.12)	(0.21)	(0.16)
Lagged own fertilizer use		−0.84			
		(0.22)			
Lagged own profits			0.01		
			(0.00)		
Soil organic matter	1.49				
	(0.57)				
Soil pH	4.92				
	(2.37)				
Soil type = loam	1.38				
	(1.12)				
Soil type = sandy	−5.91				
	(2.74)				
Sample size	89	107	107	107	107
R^2	0.80	0.75	0.75	0.72	0.72

Notes: OLS point estimates, spatial GMM (Conley 1999) standard errors in parenthesess allow for heteroskedasticity and correlation as a function of physical distance (see footnote 24). All of the covariates listed in Table 5 are included in the regressions, but not reported. Alternative specifications are as defined in Appendix B.

where $\ddot{x}_{it}^{close}$ is the (plant-weighted) average of fertilizer input on plots sufficiently close to plot i and time t that had surprisingly high profits, as defined in Section IID.

Column E demonstrates that results are robust to changing the definition of $\Gamma_{i,t}$ to include input choices of geographic neighbors at time $t+1$, $t+2$, and $t+3$, in addition to $t-3$ through t. We keep our base specification of $\Gamma_{i,t}$ as only $t-3$ through t comparisons because this corresponds to the period during which i is also applying fertilizer and thus provides the closest match to the unobserved $\Gamma_{i,t}$.

REFERENCES

Abutiate, W. S. 1991. *Guide to the Commercial Production of Pineapple, Papaya and Mango in Ghana.*

Acemoglu, Daron. 2009. *Introduction to Modern Economic Growth.* Princeton, NJ: Princeton University Press.

Aghion, Philippe, and Peter Howitt. 1998. *Endogenous Growth Theory.* Cambridge, MA: MIT Press.

Amanor, Kojo S. 1994. *The New Frontier: Farmer Responses to Land Degradation: A West African Study.* London: Zed Books.

Bala, Venkatesh, and Sanjeev Goyal. 2000. "A Noncooperative Model of Network Formation." *Econometrica*, 68(5): 1181–1229.

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

Barro, Robert J. 1991. "Economic Growth in a Cross Section of Countries." *Quarterly Journal of Economics*, 106(2): 407–43.

Bartlett, Maurice S. 1950. "Periodogram Analysis and Continuous Spectra." *Biometrika*, 37: 1–16.

Bartholomew, Duane P., and Saleh Kadzimin. 1977. "Pineapple." In *Ecophysiology of Tropical Crops*, ed. Paulo De T. Alvim and T. T. Kozlowski, 113–56. New York: Academic Press.

Bayer, Patrick J., Randi Pintoff, and David Pozen. 2009. "Building Criminal Capital Behind Bars: Peer Effects in Juvenile Corrections." *Quarterly Journal of Economics*, 124(1): 105–47.

Besley, Timothy J., and Anne Case. 1994. "Diffusion as a Learning Process: Evidence from HYV Cotton." Princeton Woodrow Wilson School Development Studies Working Paper 174.

Bindlish, Vishva, and Robert E. Evenson. 1997. "The Impact of T&V Extension in Africa: The Experience of Kenya and Burkina Faso." *World Bank Research Observer*, 12(2): 183–201.

Birkhaeuser, Dean, Robert E. Evenson, and Gershon Feder. 1991. "The Economic Impact of Agricultural Extension: A Review." *Economic Development and Cultural Change*, 39(3): 607–50.

Brock, William A., and Steven N. Durlauf. 2001. "Interactions-Based Models." In *Handbook of Econometrics Volume 5*, ed. James J. Heckman and Edward Leamer, 3297–3380. Amsterdam: Elsevier Science, North-Holland.

Conley, Timothy G. 1999. "GMM Estimation with Cross Sectional Dependence." *Journal of Econometrics*, 92(1): 1–45.

Conley, Timothy G., and Christopher R. Udry. 2005. "Learning about a New Technology: Pineapple in Ghana." Yale University Economic Growth Center Working Paper 817.

Coleman, James, Elihu Katz, and Herbert Menzel. 1957. "The Diffusion of an Innovation Among Physicians." *Sociometry*, 20(4): 253–70.

Coleman, James, Elihu Katz, and Herbert Menzel. 1966. *Medical Innovation: A Diffusion Study*. Indianapolis: Bobbs-Merrill Co.

Duflo, Esther, Michael Kremer, and Jonathan Robinson. 2006. "Understanding Technology Adoption: Fertilizer in Western Kenya, Evidence from Field Experiments." Unpublished.

Evenson, Robert E., and Larry E. Westphal. 1995. "Technological Change and Technology Strategy." In *Handbook of Development Economics Volume 3a*, ed. Jere Behrman and T. N. Srinivasan, 2209–99. Amsterdam: Elsevier Science, North-Holland.

Falk, Armin, and Michael Kosfeld. 2003. "It's all about Connections: Evidence on Network Formation." IZA Discussion Paper 777.

Feder, Gershon, Richard E. Just, and David Zilberman. 1985. "Adoption of Agricultural Innovations in Developing Countries: A Survey." *Economic Development and Cultural Change*, 33(2): 255–98.

Feltovich, Nick. 2000. "Reinforcement-Based vs. Belief-Based Learning Models in Experimental Asymmetric-Information Games." *Econometrica*, 68(3): 605–41.

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

Glaeser, Edward L., Hedi D. Kallal, Jose A. Scheinkman, and Andrei Shleifer. 1992. "Growth in Cities." *Journal of Political Economy*, 100(6): 1126–52.

Goldstein, Markus, and Christopher Udry. 2008. "The Profits of Power: Land Rights and Agricultural Investment in Ghana." *Journal of Political Economy*, 116(6): 981–1022.

Griliches, Zvi. 1957. "Hybrid Corn: An Exploration in the Economics of Technological Change." *Econometrica*, 25(4): 501–22.

Jacobs, Jane. 1969. *The Economy of Cities*. New York: Random House.

Jovanovic, Boyan, and Yaw Nyarko. 1996. "Learning by Doing and the Choice of Technology." *Econometrica*, 64(6): 1299–1310.

Kaelbling, Leslie P., Michael L. Littman, and Andrew W. Moore. 1996. "Reinforcement Learning: A Survey." *Journal of Artificial Intelligence Research*, 4: 237–85.

Kapteyn, Arie. 2000. "Saving and Reference Groups." Unpublished.

Kremer, Michael, and Edward Miguel. 2007. "The Illusion of Sustainability." *Quarterly Journal of Economics*, 122(3): 1007–65.

Lucas, Robert E., Jr. 1988. "On the Mechanics of Economic Development." *Journal of Monetary Economics*, 22(1): 3–42.

Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *Review of Economic Studies*, 60(3): 531–42.

Manski, Charles F. 1995. *Identification Problems in the Social Sciences*. Cambridge, MA: Harvard University Press.

Marshall, Alfred. 1890. *Principles of Economics*. London: Macmillan.

Moffitt, Robert A. 2001. "Policy Interventions, Low-Level Equilibria, and Social Interactions." In *Social Dynamics*, ed. Steven N. Durlauf and H. Peyton Young, 45–82. Cambridge, MA: MIT Press.
Munshi, Kaivan. 2004. "Social Learning in a Heterogeneous Population: Technology Diffusion in the Indian Green Revolution." *Journal of Development Economics*, 73(1): 185–213.
Munshi, Kaivan, and Jacques Myaux. 2006. "Social Norms and the Fertility Transition." *Journal of Development Economics*, 80(1): 1–38.
Newey, Whitney K., and Kenneth D. West. 1987. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica*, 55(3): 703–08.
Obeng, Irene S. 1994. "Effects of Domestic Policies on Production and Export of Non-Traditional Agricultural Commodities: A Case Study of Fresh Pineapples in Ghana." Unpublished.
Porter, Michael E. 1990. *The Competitive Advantage of Nations*. New York: Free Press.
Prescott, Edward C. 1972. "The Multi-Period Control Problem under Uncertainty." *Econometrica*, 40(6): 1043–58.
Purseglove, John W. 1972. *Tropical Crops: Monocotyledons*. London: Longman.
Rauch, James E., and Alessandra Casella, eds. 2001. *Networks and Markets*. New York: Russell Sage Foundation.
Richards, Paul. 1985. *Indigenous Agricultural Revolution: Ecology and Food Production in West Africa*. London: Hutchinson.
Rogers, Everett M. 1995. *Diffusion of Innovations*. 4th ed. New York: The Free Press.
Romani, Mattia. 2003. "Love Thy Neighbour? Evidence from Ethnic Discrimination in Information Sharing within Villages in Côte d'Ivoire." *Journal of African Economies*, 12(4): 533–63.
Romer, Paul M. 1986. "Increasing Returns and Long-Run Growth." *Journal of Political Economy*, 94(5): 1002–37.
Soler, Alain. 1992. *Pineapple: Quality Criteria*. CIRAD-COLEACP-OCAB, Paris.
Suri, Tavneet. 2006. "Selection and Comparative Advantage in Technology Adoption." Yale University Economic Growth Center Working Paper 944.
Suri, Tavneet. 2008. "Technological Change and the Evolution of Markets: The Case of Pineapple in Ghana." Unpublished.
Van den Bulte, Christophe, and Gary Lilien. 2001. "Medical Innovation Revisited: Social Contagion versus Marketing Effort." *American Journal of Sociology*, 106(5): 1409–35.
Woittiez, Isolde, and Arie Kapteyn. 1998. "Social Interactions and Habit Formation in a Model of Female Labour Supply." *Journal of Public Economics*, 70(2): 185–205.
Yamauchi, Futoshi. 2007. "Social Learning, Neighborhood Effects, and Investment in Human Capital: Evidence from Green-Revolution India." *Journal of Development Economics*, 83(1): 37–62.