INFORMATION NETWORKS IN DYNAMIC AGRARIAN ECONOMIES

KAIVAN MUNSHI
Department of Economics, Box B, Brown University, 64 Waterman Street, Providence, RI 02912, USA
and
NBER, USA

Contents

Abstract 3086
Keywords 3086
1. Introduction 3087
2. The adoption of new agricultural technology 3088
   2.1. A simple theoretical framework 3089
   2.2. Empirical applications 3091
   2.3. Identifying social learning 3097
3. The fertility transition 3100
   3.1. Social norms and social learning in the fertility transition 3102
   3.2. Identifying social learning 3106
4. Health and education 3108
5. Conclusion 3111
Acknowledgements 3112
References 3112
Abstract

Over the past 50 years, people living in developing countries have gained access to technologies, such as high yielding agricultural seed varieties and modern medicine, that have the potential to dramatically alter the quality of their lives. Although the adoption of these technologies has increased wealth and lowered mortality in many parts of the world, their uptake has been uneven. The traditional explanation for the observed differences in the response to new opportunities, across and within countries, is based on heterogeneity in the population. An alternative explanation, which has grown in popularity in recent years is based on the idea that individuals are often uncertain about the returns from a new technology. For example, farmers might not know the (expected) yield that will be obtained from a new and uncertain technology and young mothers might be concerned about the side effects from a new contraceptive. In these circumstances, a neighbor’s decision to use a new technology indicates that she must have received a favorable signal about its quality and her subsequent experience with it serves as an additional source of information. Because information must flow sequentially from one neighbor to the next, social learning provides a natural explanation for the gradual diffusion of new technology even in a homogeneous population. Social learning can also explain the wide variation in the response to external interventions across otherwise identical communities, simply as a consequence of the randomness in the information signals that they received. Recent research described in this chapter indicates that social learning can play an important role in the adoption of new agricultural technology, the fertility transition, and investments in health and education in developing countries.

Keywords

social learning, technology adoption

JEL classification: D80, O30
1. Introduction

The post-colonial era has witnessed many dramatic technological changes in the developing world. The introduction of high yielding varieties of wheat and rice in the 1960s dramatically increased farm incomes. Mass immunization programs and the availability of modern medicines led to a sharp decline in mortality. These changes in mortality and incomes, together with aggressive family planning initiatives, saw many countries enter the fertility transition.

While this change is very encouraging, it masks substantial variation in the response to new opportunities across and within countries. Countries at similar levels of economic development have been observed to display very different patterns of fertility behavior. Entire communities sometimes stubbornly oppose the use of modern medicine or contraceptives. And while new agricultural technology might have spread widely, it took as long as two decades in some cases for the new technology to be adopted.

One explanation for these commonly observed patterns of behavior is that heterogeneity in the population, in terms of wealth, education, or individual ability, sometimes leads different groups to respond to new opportunities at different rates. Persistent patterns of behavior might arise due to liquidity constraints or because traditional institutions, such as community-based networks, hold their members back. Starting from the early 1990s, however, an alternative information-based explanation has been proposed for the patterns of behavior described above. The seminal contributions to this “social learning” literature are Banerjee (1992) and Bikhchandani, Hirshleifer, and Welch (1992), and the basic idea behind this line of research is that individuals are often uncertain about the returns from the new opportunities that are available to them. For example, farmers might not know the (expected) yield that will be obtained from a new and uncertain technology and young mothers might be concerned about the side-effects from a new contraceptive. In these circumstances, neighbors’ decisions and experiences can be extremely valuable. A neighbor’s decision to use the new technology indicates that she must have received a favorable signal about its quality and her subsequent experience with it serves as an additional source of information.

Since information must flow sequentially from one neighbor to the next and since there is typically a lag between the adoption decision and the subsequent outcome, social learning provides a natural explanation for the gradual diffusion of new technology even in a homogeneous population. Furthermore, it can be shown that the entire population could end up choosing the wrong investment, when multiple options are available, if the signals received by neighbors about the new technology cannot be completely backed out from their decisions. Social learning can thus explain the wide variation in the response to external interventions across otherwise identical communities, simply as a consequence of the randomness in the information signals that they received.

The early work on social learning gave rise to an enormous theoretical literature (see Bramoulle and Kranton, 2004 for an excellent summary). This chapter, however, is concerned with a smaller empirical literature on social learning in dynamic agrarian economies that has emerged in recent years. The initial contributions to this literature
focussed on the adoption of new agricultural technology. Agriculture provides a natural setting in which to test for social learning since agricultural production occurs at fixed frequency and is a relatively simple and well-understood process. The domain of the information network is also conveniently defined as the village, at least for the case of Indian agriculture which is the setting for these early studies.

Subsequent research has extended this work to the fertility transition, investment in education, and the adoption of new health technology. It is reasonable to assume that farmers operate in competitive input and output markets and that there are no social restrictions on the adoption of new crops. In contrast, many traditional societies had norms in place that prohibited fertility regulation. The availability of modern contraceptives would have made fertility regulation more attractive and encouraged some women to deviate from the norm, opening up the possibility for multiple reproductive equilibria. The object of interest in this case might not be the performance of the new contraceptive, but the nature of the social equilibrium that the community ultimately ends up in. Women gradually learn about this equilibrium over time as they interact with each other and so social change, and the gradual weakening of traditional restrictions in some communities, can also be characterized by a process of social learning.

Learning about health technology differs from agricultural adoption due to the externality that accompanies the individual’s decision to participate in a public health program, for example. More information about the new technology becomes available as more members of the information network vaccinate themselves, but the immunity to the rest of the community that this provides implies that the benefit to the individual from vaccination will decline as well. If the group-immunity is sufficiently large, an increase in adoption within the network could actually lower the individual’s propensity to adopt the technology.

Despite these important differences across the applications that we consider in this chapter, the characterization of social learning as a signal extraction process remains the same and we will see that the same basic framework can be used, with some modification, to test for the presence of social learning in each case. The chapter is organized in five sections. The second section lays out a theoretical framework for social learning, due to Banerjee (1992), and then proceeds to apply this framework to the adoption of new agricultural technology. The third and fourth sections extend this framework and the empirical tests that are proposed to the fertility transition, investment in education, and the adoption of new health technology. Section five concludes.

2. The adoption of new agricultural technology

This section describes the role played by information networks in the adoption of new agricultural technology. The past five decades have witnessed tremendous technological progress in agricultural production throughout the developing world, starting with the Green Revolution in the 1960s. The Green Revolution is associated with the introduction of high-yielding varieties (HYVs) of wheat and rice. These dwarf varieties have a
high grain-to-straw ratio and thus provide much higher average yields than the traditional varieties. However, their performance is relatively sensitive to expensive inputs such as fertilizer and irrigation and is generally associated with greater uncertainty than the traditional varieties.

The HYVs did ultimately spread widely, but this diffusion process was not entirely smooth. There were long lags in adoption for some crops and the rate of diffusion varied widely with geography and by crop. Such delays have been observed historically in US agriculture as well; for example, Ryan and Gross (1943), in an influential study that spawned an enormous literature in rural sociology estimate that it took 14 years before hybrid seed corn was completely adopted in two Iowa communities. The traditional explanation for this lag focussed on heterogeneity in the population (Griliches, 1957; Mansfield, 1968). The idea here is that some people may be more receptive to new ideas and innovations (Rogers, 1962) or, alternatively, that individuals face different economic opportunities which lead them to adopt at different speeds (Griliches, 1957). An alternative explanation, which has grown in popularity in recent years, is based on the idea that farmers will learn from their neighbors’ experiences (their previous decisions and outcomes) about a new and uncertain technology. Since the individual’s information network—the set of neighbors that he learns from—is limited, this social learning process generates natural lags in adoption even in a homogeneous population. Moreover, we will show below that otherwise identical communities can end up at very different long-run adoption levels, depending on the (random) pattern of information signals that their neighbors receive.

This section begins with a theoretical framework, due to Banerjee (1992), that characterizes the social learning process. Subsequently, we describe social learning and the adoption of new agricultural technology in a number of different settings. This section concludes with a discussion on the identification of social learning in agriculture.

2.1. A simple theoretical framework

All individuals are identical and risk neutral in Banerjee’s model. They must choose an asset from a set of assets indexed by numbers in [0, 1]. Only one asset $i^*$ yields a positive return, while all other choices in the unit interval yield zero return. Thus, everyone wants to invest in $i^*$, but no one knows which asset it corresponds to. Individuals have uniform priors on [0, 1] and so there is no likely candidate for $i^*$. However, a positive fraction of the population receive an information signal informing them about the value of $i^*$. Signal quality does not vary across those individuals who receive a signal, with a positive fraction receiving the correct signal. When a signal is false, it is uniformly distributed on [0, 1] and so is completely uninformative.

Banerjee makes the following assumptions to rule out ties:

A1: Whenever an individual has no signal and everyone else has chosen $i = 0$, he always chooses $i = 0$.

A2: When individuals are indifferent between following their own signal and following someone else’s choice, they always follow their own signal.
A3: When an individual is indifferent between following more than one of the previous decision makers, he chooses the one with the highest \( i \).

Notice that these assumptions, particularly A2, are chosen deliberately to minimize the potential for herding, which we will see below could nevertheless easily occur in this set up.

The equilibrium decision rule is straightforward to derive in this framework. Assume that individuals make choices sequentially. The first individual will choose \( i = 0 \) if he has no signal. If he has a signal, he will certainly follow it. If the second individual has no signal, he will certainly follow the first individual. If he has a signal and the first individual made a choice that was inconsistent with that signal (including \( i = 0 \)), he will follow his own signal from assumption A2.

The decision rule for the third individual is more complicated since in general one of four cases could occur:

**CASE 1. The first and second individual chose \( i = 0 \).**

In this case, the third individual will follow his own signal if he received one, else he will choose \( i = 0 \).

**CASE 2. One of the two chose \( i = 0 \).**

The third individual will follow the individual that chose \( i \neq 0 \) from assumption A3 if he did not receive a signal. If he did receive a signal and it does not match the previous decision maker with \( i \neq 0 \), he will follow his own signal from assumption A2. If the two signals match, then there is no inconsistency and the third individual will of course follow his own signal.

**CASE 3. The first and second individuals both disagreed and chose some \( i \neq 0 \).**

If the third individual did not receive a signal, he should follow the higher \( i \) from assumption A3. If he did receive a signal, he will follow it regardless of whether or not it coincides with either of the choices made previously.

**CASE 4. The first and second individuals both agreed and chose some \( i \neq 0 \).**

If the third individual did not receive a signal, he will of course follow the first two individuals. If he did receive a signal and it coincides with the previous choices, then there is no ambiguity and he can simply follow his own (correct) signal. The more interesting situation, which is key to herding, arises when the third individual’s signal does not coincide with the previous choices. If the third individual knew with certainty that the second individual did not receive a signal, then he would be indifferent between the first individual’s signal and his own signal; from assumption A3 he would then follow his own signal. However, as long as there is even a small probability that the
second individual received a signal, the balance will shift in favor of the first two choices since two signals could only have coincided if they were correct.\footnote{Recall that false signals are uniformly distributed on the unit interval.}

Since there is indeed a positive probability that the second individual received a signal, the third individual will abandon his own signal and follow the first two choices. Using the same reasoning, all the individuals that follow will do likewise, regardless of whether or not they receive a signal. Such herding can lead the entire community to the incorrect choice if the first individual received a false signal and the second individual received no signal. Herding can also occur with the more general case in which the first \( n \) individuals choose different options. If the next decision maker does not have a signal, he will choose the highest \( i \) among those that have already been chosen, and all the individuals that follow will do the same unless they receive a signal and it happens to match one of the first \( n \) options that were chosen (two signals can only match if they are correct and so the individual would certainly follow his own signal in that case).

The basic reason for the herding described above is that the unobserved signals that individuals receive cannot be backed out from the decision that they subsequently make. In contrast, the empirical literature on social learning and the adoption of new agricultural technology has assumed that the decision function is invertible; the population always ends up at the correct choice in the long run and the only issue is how long it takes before convergence occurs. While this assumption might be appropriate in the specific contexts that have been studied, it rules out the pathological outcomes that make the theoretical framework so interesting and I will return to this point in Section 5.

2.2. Empirical applications

Following Munshi (2004), consider a simple model of agricultural investment in which there are two technologies: a new risky HYV technology and a safe traditional technology. The traditional technology provides a certain yield \( y_{TV} \), where the yield is defined as the profit per unit of land. The yield from the new HYV technology for grower \( i \) in period \( t \) is specified as

\[
y_{it} = y(Z_i) + \eta_{it}
\]

where \( y(Z_i) \) is the yield under normal growing conditions and \( Z_i \) is a vector of soil characteristics and prices. \( \eta_{it} \) is a mean-zero serially independent disturbance term with variance \( \lambda_i^2 \). \( \eta_{it} \) is determined by a combination of rainfall, temperature and other growing conditions. While the farmer may be aware of some of the individual factors that contribute to the growing conditions in a given year, their net effect on the yield and, hence, \( \eta_{it} \) cannot be observed.

The grower arrives at his (optimal) expected utility maximizing acreage \( A^*_i \) immediately when the expected yield \( y(Z_i) \) is known with certainty. Under reasonable conditions, \( A^*_i \) will be increasing in \( y(Z_i) - y_{TV} \) and decreasing in \( \lambda_i \). A role for social
learning does arise, however, when the grower has incomplete knowledge of the new technology. Specifically, assume now that the expected yield $y(Z_i)$ is no longer known to the grower with certainty. The acreage function can then be expressed as

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

(2)

where $\hat{y}_{it}$ is the grower’s best estimate of the expected yield in period $t$ and $\sigma_{it}^2$ is the variance of the grower’s expected yield estimate. For the risk averse grower, the chosen acreage $A_{it}$ is increasing in $\hat{y}_{it} - y_{TV}$ and decreasing in $\lambda_i, \sigma_{it}$. We will see below that $\hat{y}_{it}$ slowly converges to $y(Z_i)$ as the grower receives more information over time, accompanied by a corresponding convergence in $A_{it}$ to $A^*_i$.

The risk-averse grower is always interested in minimizing the uncertainty in $y(Z_i)$ since his expected utility is declining in $\sigma_{it}$. He will thus utilize all the information about $y(Z_i)$ that is available to him to arrive at his best estimate $\hat{y}_{it}$ in each period. Three sources of information are available to the grower. First, he receives an exogenous information-signal, perhaps from the local extension agent, about the value of his expected yield in each period. We assume that this signal provides an unbiased estimate of the yield that the grower should expect on his own land. We also assume that all growers in the village receive signals of equal precision. Second, he can use his neighbors’ decisions to infer the signals that they received. Third, he can learn directly from his own and his neighbors’ yield realizations.

The timing of receipt of the alternative sources of information is as follows. At the beginning of a period the grower receives his private information-signal. In a Bayesian setting, the grower combines that signal with his prior at the beginning of the period to compute the best-estimate of his expected yield. This in turn determines the acreage that he allocates to HYV in that period. Subsequently, he observes his neighbors’ acreage decisions, which reveal the signals that they received, as well as all the yields that are realized in the village. With social learning the new information from neighbors’ decisions as well as their yield realizations is used to update the grower’s prior about the value of his expected yield for the next period.

To begin with, assume that expected yields are constant across growers in the village so neighbors’ information-signals and yield realizations provide an unbiased estimate of the grower’s own yield, $y$. The information-signals that arrive in the village in a given period and the yields that are subsequently realized are pooled together by each grower and applied to update his prior in the next period. All growers in the village begin with a common prior $\hat{y}_0$ in period 0. Since they utilize the same information to update their priors in each period, all growers have a common prior in subsequent periods as well. Over time these beliefs converge to the true yield $y$.

Following the timing of various information-sources outlined above, each grower combines the common prior at the beginning of period $t$, $\hat{y}_t$, with the private

---

$^2$ $\sigma_{it}$ declines over the course of this convergence process, which explains the increase in acreage allocated to the new technology $A_{it}$ over time that is commonly observed.
information-signal that he receives at the beginning of that period, $u_{it}$, to arrive at his best-estimate of the expected yield on his land. Beliefs at the time of planting thus vary across growers in the village in each period. When $\eta_{it}$, $u_{it}$ are normally distributed, $\hat{y}_{it}$ will be a weighted average of $\hat{y}_t$ and $u_{it}$ in a Bayesian setting:

$$\hat{y}_{it} = \alpha \hat{y}_t + (1 - \alpha) u_{it}. \quad (3)$$

The next step in describing the learning process is to study how $\hat{y}_t$ is determined. For this we need to go back one period in time. After all the growers in the village have made their decision in period $t - 1$, the mean information-signal that they received in that period $\bar{u}_{t-1}$ can be extracted from their decisions when the acreage function is invertible. Since no one is systematically misinformed, the mean-signal provides more information about the expected yield than any individual signal. Thus $\bar{u}_{t-1}$ supersedes $u_{it-1}$ when it becomes available at the end of period $t - 1$ and so is used to compute each grower’s prior for the next period $\hat{y}_t$. Of course, growers subsequent yield realizations also appear as an additional source of information by the end of period $t - 1$. Applying Bayes’ Rule once more, the expression for $\hat{y}_t$ is consequently obtained as

$$\hat{y}_t = (1 - \beta - \gamma) \hat{y}_{t-1} + \beta \bar{u}_{t-1} + \gamma \bar{y}_{t-1} \quad (4)$$

where $\hat{y}_{t-1}$ captures all of the information about the expected yield that was received in the village up to the beginning of period $t - 1$ in Eq. (4). The village-means, $\bar{u}_{t-1}$, $\bar{y}_{t-1}$, represent the new information that became available in that period.$^3$

Two important variables that characterize the learning process, $\hat{y}_{t-1}$, $\bar{u}_{t-1}$ are not directly observed by the econometrician. Assuming that $A_{it}$ is an additively separable function of $\hat{y}_{it}$ in Eq. (2), Munshi proceeds to derive $A_{it}$ as a function of observed characteristics using Eqs. (3) and (4):

$$A_{it} = \pi_0 + \pi_1 A_{it-1} + \pi_2 \bar{A}_{t-1} + \pi_3 \bar{y}_{t-1} + \epsilon_{it}, \quad (5)$$

with $\pi_1$, $\pi_2$, $\pi_3$ derived in terms of $\alpha$, $\beta$, $\gamma$. The specification for the acreage function described above is very intuitive. When information is pooled efficiently within the village, $A_{it-1}$ contains all the information about the expected yield that was available at the beginning of period $t - 1$; specifically the entire history of information-signals and yield realizations up to that time. Conditional on $A_{it-1}$, $\bar{A}_{t-1}$ represents the new information that was received by the village in period $t - 1$ through the exogenous signals. Similarly, $\bar{y}_{t-1}$ represents the information that was obtained from the yield realizations in that period. Note that the grower’s own lagged yield $\hat{y}_{it-1}$ does not enter independently in Eq. (5) since it is superseded by the mean-yield in the village $\bar{y}_{t-1}$.

The previous scenario that we discussed was the most suitable for social learning. Neighbors’ signals and yields provide an unbiased estimate of the grower’s own expected yield and so can be utilized without modification. In fact, information from a

---

$^3$ The grower will actually place more weight on his prior as he grows more confident about the yield level on his land. Thus, $\alpha$ will be increasing, while $\beta$, $\gamma$ are decreasing, over time. We ignore the time subscripts on $\alpha$, $\beta$, $\gamma$ to simplify the exposition.
neighbor is as useful as the information that the grower receives himself. However, the expected yield will more generally depend on the farmer’s characteristics; the fact that the technology worked well for a neighbor does not necessarily imply that it will work well for you. Ellison and Fudenberg (1993) use this argument to justify simple rules of thumb where individuals learn from similar neighbors only, slowing down the rate of diffusion, but the grower could in principle do better than that by conditioning for differences between his own and his neighbors’ observed characteristics when learning from them. However, the prospects for social learning decline immediately once we allow for the possibility that some of these characteristics may be unobserved, or imperfectly observed.

The grower may be prepared to accept transitory errors in his yield estimates; associated with the \( \eta \) disturbance term in Eq. (1) or noise in the exogenous information-signals. However, mistakes that arise because he is unable to control for differences between his own and his neighbors’ characteristics when learning from their yields are persistent, and therefore more serious. Take the case where all the neighbors’ characteristics are unobserved by the grower. He now has two choices. He could rely on his own information-signals and yield realizations, ignoring information from his neighbors. Consistent but inefficient estimates of the expected yield would be obtained with such \textit{individual} learning. Alternatively, he could continue to utilize information from his neighbors, measured by the mean-acreage and the mean-yield, as before. The efficiency of his estimates increases with \textit{social} learning since more information is being utilized, but some bias will inevitably be introduced since the grower cannot control for variation in the underlying determinants of the yield when learning from his neighbors. The grower will ultimately choose between individual learning and social learning on the basis of the trade-off between bias and efficiency. The testable prediction in this case is that the grower will choose individual learning if the population is heterogeneous and the yield with the new agricultural technology (crop) is sufficiently sensitive to unobserved characteristics, otherwise he will prefer to learn from his neighbors.

While Munshi (2004) and Besley and Case (1994) assume that the yield (or profit) \( y(Z_i) \) is exogenous and uncertain, Foster and Rosenzweig (1995) and Conley and Udry (2005) assume that the grower’s objective is to learn his optimal (profit-maximizing) input use \( Z_i \). The point of departure for their work is the target-input model of Jovanovic and Nyarko (1996), but we will see that the signal extraction aspect and, hence, the basic structure of the learning process remains the same across these different models of learning.

With a slight change of notation, Foster and Rosenzweig assume that the grower attempts to learn the optimal or target input use on his land \( \theta^* \),

\[
\hat{\theta}_{ijt} = \theta^* + u_{ijt}
\]

where \( \hat{\theta}_{ijt} \) is the optimal input use on plot \( i \) for farmer \( j \) in period \( t \), \( u_{ijt} \) is an i.i.d. random variable with a (known) variance \( \sigma_u^2 \), and the farmer’s prior on \( \theta^* \) is \( N(\hat{\theta}_{j0}, \hat{\sigma}_{\theta_{j0}}^2) \). Notice the similarity with Eq. (1), where the grower’s objective was to learn the value of the (expected) yield \( y(Z_i) \). Previously he was seen to collect information on \( y(Z_i) \) from
various sources to finally arrive at the optimal acreage. In the current set up, the grower
collects information on $\theta^*$ from various sources to arrive at his profit-maximizing input
level.

The yield (or profit) per plot from HYV on the $i$th most suitable plot for HYV for a
farmer with $A_j$ plots is

$$\eta_a + \eta_h - \eta_h a \frac{i}{A_j} - (\theta_{ijt} - \tilde{\theta}_{ijt})^2$$

where $\eta_a$ is the yield from traditional varieties, $\eta_h$ is the maximum yield from HYV on
the plot most suitable for the new technology, and $\eta_h a$ represents the loss from using
land less suitable for HYV. $(\theta_{ijt} - \tilde{\theta}_{ijt})^2$ is the loss due to sub-optimal input use, which
declines over time with social learning.

Based on the expression above, the profit from all the farmer’s plots can be specified
as,

$$H_{jt} \sum_{i=1}^{H_{jt}} \left( \eta_a + \eta_h - \eta_h a \frac{i}{A_j} - (\theta_{ijt} - \tilde{\theta}_{ijt})^2 \right) + (A_j - H_{jt}) \eta_a$$

where $H_{jt}$ is the number of plots allocated to HYV. The profit expression can be sim-
plified further as

$$\eta_h H_{jt} - \frac{\eta_h a H_{jt}^2}{2} + A_j \eta_a - \sum_{i=1}^{H_{jt}} (\theta_{ijt} - \tilde{\theta}_{ijt})^2.$$  

Taking expectations, the grower’s expected profit can finally be expressed as,

$$\pi_{jt} = \left[ \eta_a - \frac{\eta_h a H_{jt}}{2} - \sigma_{\theta_{jt}}^2 - \sigma_{\tilde{\theta}_{jt}}^2 \right] H_{jt} + \eta_a A_j.$$  

(7)

The term in square brackets above represents the yield from HYV, which is declining
in $\sigma_{\theta_{jt}}^2$ and $\sigma_{\tilde{\theta}_{jt}}^2$, which corresponds to $\lambda^2$ in Munshi’s model, is a source of (natural)
uncertainty that the grower cannot avoid. However, $\sigma_{\tilde{\theta}_{jt}}^2$, which corresponds to $\sigma_c^2$ in
Munshi’s framework, will go to zero as the grower learns the value of $\theta^*$ from his own
and his neighbors’ experiences, and so the HYV yield will increase over time with social
learning. The next step in the analysis is consequently to characterize this change in $\sigma_{\theta_{jt}}^2$
over time.

Foster and Rosenzweig assume that the optimal input on each plot $\tilde{\theta}_{ijt}$ is revealed
to the farmer at the end of each year. The farmer also learns from his neighbors’ input
decisions and profit realizations, although the signals received from neighbors’ plots are
assumed to be less precise than the signals received from the grower’s own plots. Under
the normality assumption, the variance of $\theta^*$ can be expressed in a Bayesian setting as,

$$\sigma_{\theta_{jt}}^2 = \frac{1}{\rho + \rho_0 S_{jt} + \rho_1 \bar{S}_{jt}}$$  

(8)
where $S_{jt}$ is the cumulative experience on the grower’s own land and $\bar{S}_{-jt}$ is the average of the cumulative experience of the neighbors. As experience with the new technology grows, mistakes in input use and hence the variance in $\theta^*$ get smaller, which increases profits from Eq. (7) and Eq. (8).

While Munshi shows that the effect of neighbors’ past decisions and experiences on the grower’s current decision will vary across crops, depending on growing conditions and the technology, Foster and Rosenzweig derive predictions for changes in the pattern of learning over time.

$$\frac{\partial \pi_{jt}}{\partial S_{jt}} = \frac{\rho_0}{\rho_v}$$

and so own and neighbors’ acreage effects change at the same pace over time.

Let $y_{jt+1} = \pi_{jt+1}/H_{jt+1}$ and $y_{jt} = \pi_{jt}/H_{jt}$ be the profit per unit of HYV acreage in period $t + 1$ and $t$, respectively. Then it is easy to verify that

$$\frac{\partial y_{jt+1}}{\partial S_{jt+1}} = \frac{(\rho + \rho_0 S_{jt} + \rho_v \bar{S}_{-jt})^2}{(\rho + \rho_0 S_{jt+1} + \rho_v \bar{S}_{-jt+1})^2} < 1$$

and so returns per hectare from own experience are decreasing over time. By a similar calculation, it can be shown that returns per hectare from neighbors’ experience are also decreasing in the same fashion.

The grower’s objective, in Munshi’s framework, is to learn the value of the expected yield $y$, which we take to be constant across all farmers in the simplest case. Each neighbor’s expected yield estimate in period $t - 1$ can be backed out from his acreage decision in that period, assuming that the acreage function is invertible, while his accompanying yield realization provides an additional estimate of $y$. Foster and Rosenzweig’s learning problem is more challenging since the grower is presumed to infer the optimal input on each plot in each period for himself and his neighbors. This seems to be a difficult task with a single observation per plot-period and at best we might imagine that the grower obtains an unbiased estimate of $\tilde{\theta}_{ijt}$; I suspect that the implications of Foster and Rosenzweig’s model would in fact go through in this case as well. However, obtaining even an unbiased estimate of the optimal input with a single observation might sometimes be a difficult task. A paper by Conley and Udry (2005) consequently proceeds to relax this assumption.

To understand Conley and Udry’s model of learning it is convenient to ignore idiosyncratic variation in the optimal input use $u_{ijt}$ and assume instead a common and constant optimal level $\theta^*$ on all plots and in all periods. The grower observes the input level and the associated profit on his own and his neighbors’ plots in each period. With “parametric” learning, the grower could specify a particular functional relationship between profits and inputs and then estimate these parameters with the data obtained from his own and his neighbors’ experiences. Assuming that the profit function is correctly specified, the estimated parameters would converge over time to the true parameters and the grower would arrive at the optimal level of input use. Conley and Udry assume instead that the grower learns “nonparametrically”; profits observed at a particular input level
provide no information about the profit-input relationship at any other level. Nonparametric learning avoids potential misspecification in the profit function, but convergence will now occur relatively slowly. The grower will trade off efficiency and potential bias when choosing between parametric and nonparametric learning, and this choice will in practice be motivated by growing conditions and the nature of the crop technology (as in Munshi, 2004). While they do not provide such a motivation, Conley and Udrey's nonparametric learning model nevertheless yields a number of testable implications for the relationship between the grower’s input (fertilizer) use and his neighbors’ past use.

First, suppose that a neighbor uses the same level of input as the grower’s current level, which is presumably his best estimate of the optimal level. If the neighbor’s profit exceeds the grower’s expected profit at that input level, then this will only reinforce the grower’s belief that he is at the optimal level and there will be no change in inputs in the future. If the neighbor’s profit is below the grower’s expected level, then this increases the probability that the grower will shift to a new level in the future. Second, suppose that a neighbor uses a different level of input than the grower. If his profit exceeds the grower’s prior belief about the profit at that level, then this increases the probability that the grower will switch to his neighbor’s input level in the future. If the neighbor’s profit is below what the grower expected, then this only reinforces his prior that it is a sub-optimal input level and there will be no response in the future.

Munshi generates predictions for differences in the pattern of learning across crops. Foster and Rosenzweig place restrictions on the pattern of learning over time. Conley and Udrey’s learning model distinguishes between the response to input use at the grower’s own level and at other levels. We will see below that these testable predictions play an important role in the identification of social learning.

### 2.3. Identifying social learning

The simplest test of social learning in agriculture, following Eq. (5), is to see whether the grower’s current acreage is determined by his neighbors’ past acreage decisions and yield realizations. We thus estimate regressions of the form

\[
A_{it} = \Pi_0 + \Pi_1 A_{it-1} + \Pi_2 \bar{A}_{t-1} + \Pi_3 \bar{y}_{t-1} + \epsilon_{it}
\]

where \(A_{it}\) is the acreage allocated to the new technology by grower \(i\) in period \(t\), \(A_{it-1}\) is the acreage allocated in the previous period, \(\bar{A}_{t-1}\) is the average HYV acreage in the village in that period, and \(\bar{y}_{t-1}\) is the corresponding average yield. As discussed earlier, \(A_{it-1}\) collects all the information about the new technology that was available in the village up to the beginning of period \(t - 1\), \(\bar{A}_{t-1}\) represents the new information that became available through the external signals \(\bar{u}_{t-1}\) in period \(t - 1\), and \(\bar{y}_{t-1}\) represents the new information obtained from yield realizations in the village in that period.

The acreage decision in period \(t\) depends on the grower’s best estimate of the HYV yield \(\hat{y}_{it}\), which is determined in turn by his prior \(\hat{y}_{t}\) and the new information signal \(u_{it}\). The prior \(\hat{y}_{t}\) is represented by the \(A_{it-1}, \bar{A}_{t-1}, \bar{y}_{t-1}\) terms and \(u_{it}\) is included in \(\epsilon_{it}\). As long as information signals are correlated over time and across growers in the village,
\( \hat{A}_{t-1} \) will be correlated with \( \epsilon_{it} \) and learning from neighbors’ signals cannot be distinguished from learning from own signals.\(^4\) Unlike Banerjee (1992), Munshi assumes that growers are never systematically misinformed, \( E(u_{it}) = y \), and so \( \bar{y}_{t-1} \) and \( \epsilon_{it} \) will be correlated as well.

One solution to this problem would be to difference \( \bar{y}_{t-2} \) from \( \bar{y}_{t-1} \), leaving us with \( \bar{y}_{t-1} - \bar{y}_{t-2} \) from Eq. (1). Under the assumptions of the model, \( \bar{\eta}_{t-1}, \bar{\eta}_{t-2} \) measure unobserved (to the grower) deviations from normal growing conditions, which will be uncorrelated with the information signals. However, this differencing procedure leaves an additional \( \pi_3 \bar{y}_{t-2} \) term in the residual of the acreage regression, which is negatively correlated with \( \bar{\eta}_{t-1} - \bar{\eta}_{t-2} \) by construction and so will generate conservative estimates of the yield effect. A potentially more serious problem is that changes in village yield over time could be due to factors other than serially independent deviations from normal growing conditions. By specifying yield to be the sum of a constant term \( y(Z_i) \) and an idiosyncratic shock \( \eta_{it} \), we are implicitly assuming that input markets function smoothly and that input and output prices do not change over time. In practice, changes in the yield from period \( t-2 \) to \( t-1 \) could reflect changes in prices or access to scarce resources that are unobserved by the econometrician but directly determine the grower’s period-\( t \) acreage decision. A spurious yield effect could in that case be obtained.

We could control to some extent for unobserved determinants of current acreage by including prices and access to seeds, fertilizer, and irrigation, as well as the grower’s wealth.\(^5\) An alternative solution takes advantage of the distinction between parametric and nonparametric learning described above. Social learning will be weaker in a heterogeneous population, particularly when the performance of the new technology is sensitive to neighbors’ unobserved characteristics. The rice growing areas of Peninsular India are characterized by wide variation in soil characteristics. In contrast, conditions are fairly uniform in the Northern Plains, where wheat is grown traditionally. The early rice varieties were also quite sensitive to soil characteristics such as salinity, as well as to managerial inputs, which are difficult to observe. The rice grower would thus have found it difficult to control for differences between his own and his neighbors’ characteristics when learning from their experiences. Consistent with this view, Munshi (2004) using farm-level data over a three-year period at the onset of the Green Revolution finds that HYV acreage responds to lagged yield shocks in the village with wheat but not rice. While the acreage effects are more difficult to interpret, the coefficient on own lagged acreage is larger for rice than for wheat, whereas the pattern across crops is reversed for the coefficient on lagged mean HYV acreage in the village, consistent with the view once again that social learning was stronger for wheat than for rice.

\(^4\) Following Manski (1993), neighbors’ past acreage allocations will be correlated with the grower’s current acreage decision if any unobserved determinant of the acreage decision is correlated across neighbors and over time.

\(^5\) Along the same lines, we could include the grower’s own lagged yield-shock in the acreage regression to distinguish social learning from individual learning.
If access to credit and other scarce inputs varied systematically across the wheat- and rice-growing areas of the country, then the observed differences in the yield effect across crops could still be explained without appealing to differences in underlying social learning. Munshi responds to this potential concern by testing for social learning at the district level, which reflects underlying learning at the farm level, using districts that grow both wheat and rice. The set of neighbors is now defined by the set of geographically contiguous districts and the acreage regressions are estimated over the 1969–1985 period. The estimated pattern of acreage and yield coefficients across crops, with district fixed effects, matches the patterns described above at the farm level, providing independent support for the presence of social learning in the adoption of new agricultural technology.

Munshi’s test of social learning is based on the relationship between the grower’s current HYV acreage and his neighbors’ lagged HYV yields. In contrast, Foster and Rosenzweig derive implications for the relationship between the grower’s profit (yield) with HYV and cumulative experience with the new technology:

$$
\pi_{jt} = (\eta_h + \beta_{ot} S_{jt} + \beta_{vt} \bar{S}_{-jt}) H_{jt} + \eta_a A_{jt} + \xi_{jt}
$$

where the term in parentheses represents the profit (yield) from HYV, which is increasing in the cumulative experience with the new technology on own land $S_{jt}$ and neighbors’ land $\bar{S}_{-jt}$. $H_{jt}$, $A_{jt}$ measure acreage allocation in period $t$ to the new technology and the traditional technology, and $\pi_{jt}$ measures crop profits. The potential sources of spurious correlation that arise in Munshi’s analysis evidently apply here as well. To begin with, fixed grower characteristics could jointly determine HYV yields and acreage allocations. Since farm level data are available over a three-year period, grower fixed effects can be included in the profit regression. However, unobserved time-varying changes in growing conditions or access to scarce resources, which would affect both the current HYV yield as well as $S_{jt}$, $\bar{S}_{-jt}$, would not be accounted for by the fixed effects. Foster and Rosenzweig use inherited wealth and lagged values of $S_{jt}$, $\bar{S}_{-jt}$ as instruments for $S_{jt}$, $\bar{S}_{-jt}$, but these instruments will only be valid if the unobserved determinants of the yield are serially uncorrelated, which may not be the case in practice.

Once again it is possible to appeal to the restrictions from the theory to provide additional support for the presence of social learning. Foster and Rosenzweig’s learning model generates the predictions that (i) $\beta_{ot}$, $\beta_{vt}$ will be declining over time, and (ii) $\frac{\beta_{ot}}{\beta_{vt}} = \frac{\beta_{ot+1}}{\beta_{vt+1}}$. These predictions are successfully tested, consistent with the presence of social learning. Conley and Udry take a similar approach, testing the restrictions placed on the relationship between the grower’s input use and his neighbors’ lagged input use. They find that farmers are more likely to change their fertilizer use when members of their information network using similar levels of fertilizer achieve unexpectedly low profits. Farmers also increase (decrease) fertilizer use when their neighbors using more (less) fertilizer than them do unexpectedly well, as predicted.

Finally, differences in patterns of experimentation across farmers or crops generate additional testable restrictions that can be taken to the data. All the studies discussed in
this section allow for experimentation on the grower’s own plot, as well as for strategic adoption, when deriving testable predictions for social learning. Munshi, in particular, generates predictions for patterns of experimentation across crops to support his argument that social learning was stronger for wheat than for rice. If the view that rice growers were informationally disadvantaged is correct, then we would expect such growers to have compensated for their lack of information by experimenting on their own land. Munshi shows that rice growers who do adopt HYV allocate a greater amount of land to the new technology than comparable wheat growers. This is despite the fact that farms are smaller in the rice growing areas of the country and the likelihood of HYV adoption is significantly higher for wheat growers.

Along the same lines, Bandiera and Rasul (2006) use strategic delays to explain the inverted U-shaped relationship between the number of neighbors that adopt and the individual’s own decision to adopt a new agricultural technology in Mozambique. On the one hand, an increase in the number of adopters in the individual’s information network increases the amount of (social) information that is made available, increasing his propensity to adopt. On the other hand, having many adopters in the network increases the individual’s incentive to delay adoption and free-ride on the information that is made available by his neighbors. Bandiera and Rasul show that the second effect dominates once the adoption level within the network grows beyond a cut off point.6

While the early studies on social learning in agriculture such as Besley and Case (1994), Foster and Rosenzweig (1995), and Munshi (2004), treat the village as the exogenous domain of the information network, recent studies such as Conley and Udry (2005) and Bandiera and Rasul (2006) use a self-reported list of social contacts to construct the network. All plots are clustered together in India and so all decisions and experiences in the village are readily observable. In contrast, agriculture is more spatially dispersed in sub-Saharan Africa and so information will flow less smoothly within the village. The advantage of using the actual (self-reported) network links as opposed to the potential (village-wide) links is that tests of social learning will have more power. The potential cost of using a self-reported network is that the omitted variable problems discussed above could be worsened. This point will be discussed in greater detail in Section 3 in the context of the identification of social learning in the fertility transition.

3. The fertility transition

Declining mortality, together with rising incomes and access to modern contraceptives throughout the developing world have led to a substantial decline in fertility over the

6 Bandiera and Rasul depart from the set up in previous studies by analyzing the relationship between the grower’s decision to adopt the new technology and the proportion of his self-reported network that adopts contemporaneously. The interesting predictions that they derive are based on the grower’s response to exogenous variation in his neighbors’ adoption. With cross-sectional data, what they identify instead is the equilibrium correlation between the individual’s adoption and his neighbors’ adoption, which is difficult to interpret.
past decades. However, long delays and wide differentials in the response to family planning programs have been frequently observed, both across countries and within countries (Bulatao, 1998; Cleland et al., 1994; NRC, 1993). One explanation for such slow adoption is based on the idea that individuals may be uncertain about the efficacy of the new contraceptive technology or, perhaps more importantly, they may be concerned about its potential side-effects. Such concerns are very relevant given the long-term nature and potential medical risks of contraceptives such as IUD and injectables that are typically distributed in developing countries. In these circumstances, a neighbor’s decision to adopt the new contraceptive technology indicates that she must have received favorable information about its performance. Her subsequent experience with the new technology serves as an additional source of information.

The social learning just described is conceptually no different from the learning about new agricultural technology that was discussed in some detail in the previous section. The performance of the new contraceptive technology is likely to be relatively insensitive to the individual’s socio-economic characteristics and so learning from a neighbor’s experience with the new technology, at the very least, should proceed without any need to condition for differences in characteristics across individuals. What slows down the fertility transition is not the nature of the technology, as in the agricultural case, but the restrictions on fertility regulation put in place in most traditional societies. While such social regulation may have had advantages of its own historically, the drawback is that it may prevent individuals from responding immediately to the new opportunities associated with the availability of modern contraceptives. Social learning will be seen in a moment to play an important role in the gradual weakening of these restrictions as well.

Social norms are seen to emerge in environments characterized by multiple equilibria to keep the community in a preferred equilibrium (Kandori, 1992). Changes in the economic environment, such as the unexpected availability of modern contraceptives, could reopen the possibility for such multiple equilibria. The point of departure for Munshi and Myaux’s (2006) model of the fertility transition is a social uncertainty following the introduction of a family planning program: the individual does not know the reproductive equilibrium that her community will ultimately converge to. This uncertainty is gradually resolved as individuals interact sequentially with each other over time. There are only two types of individuals in their simple model, which is constructed so that only two possible equilibria can emerge in the long-run. No one regulates fertility prior to the intervention. While this remains a potential equilibrium, a new equilibrium in which a sufficient fraction of the community regulates fertility is also shown to emerge. The discussion that follows describes the social learning process that leads some communities from the traditional equilibrium to the modern equilibrium, while others remain where they were. Testable implications will subsequently be derived and empirical results discussed from a particular setting, as we did in the previous section with the adoption of agricultural technology.
3.1. Social norms and social learning in the fertility transition

Each community consists of a continuum of individuals. Each individual chooses from two actions at the beginning of every period: the traditional \((t)\) action corresponding to unchecked fertility and the modern \((m)\) action, which refers to fertility control. Subsequently she is randomly matched with a member of the community. When reproductive behavior is socially regulated, the individual’s payoff from a particular action depends not only on the intrinsic utility that she derives from that action, but also on the social pressure or sanctions that go with it. Following Kandori’s (1992) characterization of social norms, the individual’s payoff depends on her own action, as well as her partner’s action, which determines the social sanction that she will face in that period.

Since there are two possible actions, and the individual matches with a single partner in each period, payoffs corresponding to four combinations of actions must be considered:

\[
\begin{align*}
V_i(m, m) &= U_i, \\
V_i(m, t) &= U_i - l, \\
V_i(t, t) &= 0, \\
V_i(t, m) &= g.
\end{align*}
\]

\(V_i\) is individual \(i\)’s payoff at the end of the period, where the first term in parentheses refers to the individual’s own action, and the second term refers to her partner’s action. \(U_i\) is the intrinsic utility that the individual derives from the modern action. There are two types of individuals in this simple model, conformists and reformists, with reformists comprising a fraction \(P\) of the community. Conformists are assumed to have internalized the opposition to reproductive control that is built into the social norm. Thus they derive lower intrinsic utility from the \(m\) action than the reformists; \(U_i = v\) for the conformists and \(U_i = w > v\) for the reformists.

\(l\) and \(g\) refer to the punishment and rewards that have been put in place to regulate reproductive behavior. When a woman who chooses the modern action meets another woman who continues to follow the traditional action, she faces some sort of social censure. The reward \(g\) may be associated with enhanced social standing, possibly within a very restricted peer group, for having punished a deviator. Notice that there are no social sanctions when two deviators meet each other.

Munshi and Myaux impose the following conditions on the payoffs prior to the external intervention: \(v > 0, w - l < 0, w < g\). Under these conditions it is easy to verify that a unique equilibrium is obtained in each period, in which both conformists and reformists choose the \(t\) action. Subsequently they allow for the availability of modern contraceptives, which reduces the inconvenience associated with fertility control, increasing the individual’s intrinsic utility from the \(m\) action by an amount \(S\). The conditions on payoffs in the new regime are the same as what we described above, with one important exception: \(v + S > 0, w + S - l < 0, w + S > g > v + S\).
It is easy to verify that the traditional equilibrium, without fertility control, continues to be sustainable in the new regime. A new modern equilibrium can also be sustained if the proportion of reformists in the community \( P \) is sufficiently large. In this equilibrium all the reformists choose \( m \) and all the conformists choose \( t \). A reformist will not deviate from this equilibrium if the expected payoff from choosing \( m \) exceeds the expected payoff from choosing \( t \)

\[
P(w + S) + (1 - P)(w + S - l) \geq Pg. \tag{9}
\]

Simplifying the expression above, a necessary condition to sustain the modern equilibrium is obtained as \( P \geq P^* = \frac{l - (w + S)}{l - g} \). Communities with \( P > P^* \) must choose between two equilibria, while only the traditional equilibrium can be supported in communities with \( P < P^* \).

The basic source of uncertainty in Munshi and Myaux’s model is that the proportion of reformists \( P \) is not known to begin with, since each individual’s type is private information and both conformists and reformists chose the same traditional action prior to the intervention. To simplify the equilibrium dynamics they assume that there are two types of communities: stable communities with \( P < P^* \) reformists and unstable communities with \( \bar{P} > P^* \) reformists. We will see in a moment that information about \( P \) is gradually revealed over time as individuals interact with each other, with unstable communities moving to the modern equilibrium while stable communities remain where they were.

All communities continue to remain in the traditional equilibrium following the introduction of the family planning program. Since the program has increased the payoff from the \( m \) action by an amount \( S \), it is evidently interested in making sure that the reformists in the unstable communities take advantage of the new opportunities that are available. Most family planning programs employ health workers to persuade women to adopt the new contraceptive technology. In our context, these workers play a critical role in initiating the transition from the traditional equilibrium to the modern equilibrium in the unstable communities, just as extension workers must provide the exogenous information signals to initiate the adoption of new agricultural technology. The health worker visits a fraction \( \theta \) of the community, drawn at random, in each period and persuades any reformist that she meets to switch to the \( m \) action, but for a single period only. Since there is a continuum of individuals in each community, this implies that a constant fraction \( \theta P \) of the community, where \( P = \bar{P} \) in unstable communities and \( P = P \) in stable communities, deviates exogenously in each period. We will see that this exogenous deviation provides the seed for subsequent endogenous deviation in the unstable communities, which ultimately moves them to the new social equilibrium.

Let \( \alpha_t \in [0, 1] \) be the individual’s belief about the state of the world, the probability that \( P = \bar{P} \), in period \( t \). Begin with a degenerate distribution of beliefs \( \alpha_0 \) in period 0, in both stable and unstable communities, such that no reformist deviates endogenously. This leaves only exogenous deviation in the first few periods: a proportion \( \theta \bar{P} \) of the individuals in the unstable communities and a corresponding proportion \( \theta P \).
of the individuals in the stable communities choose the \( m \) action in each period.\(^7\) While contraceptive prevalence might be constant in these early periods, the distribution of beliefs within each community will spread out over time as different individuals are faced with a different sequence of matches. To derive the evolution of individual beliefs during these early periods without endogenous deviation apply Bayes’ Rule to an individual with belief \( \alpha_t \) in period \( t \) who matches with an \( m \) in that period. Her belief \( \alpha_{t+1} \) in the subsequent period is then expressed as:

\[
\alpha_{t+1} = \Pr(P = \bar{P} \mid m) = \frac{\alpha_t(\theta \bar{P})}{\alpha_t(\theta \bar{P}) + (1 - \alpha_t)(\theta P)}.
\]  

(10)

\( \theta \) is common knowledge, and while the individual might not know the type of community that she belongs to, she does know the values of \( \bar{P}, P \). Since the term in the denominator of Eq. (10) is a weighted average of \( \theta \bar{P} \) and \( \theta P \), it is easy to verify that \( \alpha_{t+1}/\alpha_t > 1 \). As the individual matches with \( m \)'s in the community, her belief that \( P = \bar{P} \) grows. The right-hand support of the distribution in any period \( t \) is thus defined by the beliefs of the individuals in the community who have matched with a continuous sequence of \( m \)'s up to that period, and so will shift steadily over time.

A reformist will choose the \( m \) action in any period, without persuasion from the health worker, if the expected probability of matching with an \( m \) exceeds \( P^* \) (from Eq. (9)). The expected probability of matching with an \( m \) in these early periods is simply \( \alpha(\theta \bar{P}) + (1 - \alpha)(\theta P) \), where \( \alpha \) is the individual’s belief that \( P = \bar{P} \). This expected probability is evidently increasing in \( \alpha \). As long as \( \theta \bar{P} > P^* \), there exists a threshold belief \( \alpha^* \), for which the individual is indifferent between the \( t \) and the \( m \) action, satisfying the following condition:

\[
\alpha^*(\theta \bar{P}) + (1 - \alpha^*)(\theta P) = P^*.
\]  

(11)

If the individual’s belief that \( P = \bar{P} \) exceeds \( \alpha^* \), then the expected probability of matching with an \( m \) will exceed \( P^* \), and she will deviate endogenously. If not, she will only choose \( m \) when she meets the health worker. Following the discussion above, the support of the belief distribution will shift steadily over time in the early periods until ultimately the right-hand support reaches \( \alpha^* \). The first wave of endogenous deviators will now appear, at the same time in both stable and unstable communities.

To describe actions in the community after the first wave of endogenous deviators appears, Munshi and Myaux begin by showing that the distribution of beliefs in the unstable communities shifts steadily to the right over time, whereas the distribution shifts in the opposite direction in the stable communities. In every period following the emergence of the first endogenous deviators there exists an \( \alpha^*_t \) such that all individuals with

---

7 If the health worker observes all the individual decisions, then \( P \) would be revealed to the external agency in the first period itself. Immediate withdrawal by the external agency would signal in turn whether a community was stable or unstable. Programatic constraints would typically rule out such early withdrawal by the external agency from a subset of communities.
beliefs to the right of this threshold belief will choose the $m$ action. Given the change in the distribution of beliefs described above, this implies that contraceptive prevalence in the unstable communities will increase steadily over time until all the reformists ultimately switch to the $m$ action. In contrast, contraceptive prevalence in the stable communities will decline after some initial endogenous deviation and ultimately these communities will end up where they were to begin with, in the traditional equilibrium.\footnote{This process of social learning is conceptually similar to Banerjee’s (1993) characterization of a rumor process. In his model, the delay before individual’s meet reveals the state of the world. In our case, individuals match every period; it is the sequence of partners’ decisions that ultimately reveals the type of community that the individual belongs to.}

The simple model just described can explain two stylized facts that are commonly observed in the fertility transition; relatively slow rates of change and wide variation in the response across communities to the same external intervention.\footnote{While we focus on uncertainty about social fundamentals (the underlying social structure of the community), a model based on strategic uncertainty could also generate some of the stylized facts observed in the data. Suppose, for example, that all the communities are unstable, with $\bar{P} > P^*$. We are still left with a coordination problem, since both the traditional and the modern equilibrium can be sustained in these communities. The standard approach to model this coordination problem would be to perturb the system by exogenously switching a fraction of the community to the $m$-action in period 0. If we assume that individuals mimic their partner’s action (in the next period) with a fixed probability, then we are essentially describing the beginning of a contagion. It is well known that if the initial perturbation is sufficiently large, then the community will “tip over” to the modern equilibrium, if not it will return to the traditional equilibrium after a temporary deviation. While the contagion model can explain the gradual change in reproductive behavior that is commonly observed, it cannot explain the wide variation in the response across communities to the same external intervention.}

The individual’s decision in period $t$ is determined by her belief, relative to the threshold belief $\alpha_t^*$. If her belief lies to the right (left) of $\alpha_t^*$ she will choose the $m(t)$ action. The individual’s belief in period $t$ is in turn determined by her belief in period $t-1$, augmented by the change in this belief through the social interaction in that period. It is easy to see that matching with an $m$ will shift her belief to the right by returning to Eq. (10) and replacing $\theta \bar{P}, \theta P$ with $\bar{x}_{t-1}, \Sigma_{t-1}$, the contraceptive prevalence in unstable and stable communities in period $t-1$. The transition dynamics described above indicate that $\bar{x}_{t-1} > \Sigma_{t-1}$, which implies from the equation that $\alpha_t/\alpha_{t-1} > 1$ when the individual matches with an $m$. Munshi and Myaux map these changes in beliefs into changes in actions to derive the individual’s decision rule: the probability that the individual chooses the $m$ action in period $t$ is specified to be the weighted average of her decision in period $t-1$ and the probability that she matched with an $m$ in that period. With random matching, this last probability is in turn measured by the proportion of $m$’s in the community in period $t-1$.

Notice that we have arrived at precisely the specification derived by Munshi (2004) in the context of agricultural technology adoption. This is not entirely surprising, since the fertility transition model would have yielded the same decision rule if the uncertainty...
revolved around the performance of the contraceptive technology $S$ rather than the composition of the community $P$. Munshi and Myaux’s characterization of changing social norms as a learning process, however, provides an additional testable implication that is absent with technology adoption. Social norms are organized at the level of the social group, defined by religion, caste, ethnicity, or some other group characteristic. Thus, contraceptive prevalence within the individual’s social group alone should determine her own contraception decision. In contrast, contraceptive prevalence outside the narrow social group should also provide useful information if the objective is to learn about a new contraceptive technology.

3.2. Identifying social learning

Demographers have recognized that the fertility transition is a relatively slow diffusion process for many decades. Early studies on this diffusion process concentrated on spatial patterns of fertility change, in some cases identifying remarkably strong ethnic and linguistic aspects to these spatial patterns, as in the case of the European demographic transition (see, for example, Lesthaeghe, 1977; Livi-Bacci, 1971, 1977). As discussed in the context of agricultural technology adoption, alternative explanations for such spatial patterns that do not rely on underlying social interactions are readily available, the simplest one being that the economic change or mortality decline that gave rise to the fertility transition occurred along ethnic or linguistic lines as well.

Recently there have been some attempts using micro-data to test directly for the role of social interactions in the fertility transition. Montgomery and Casterline (1993) is an early contribution to this literature, which has been followed by Entwistle et al. (1996) and Behrman, Kohler, and Watkins (2002), among other studies. A common approach in these studies is to ask individuals whom they talk to in general, or more specifically about health and contraception. An attempt is then made to establish a statistical link between the individual’s contraceptive use and the level of contraceptive use in the self-reported reference group. As discussed earlier, the basic problem with this approach is that contraceptive prevalence in the reference group could proxy for any unobserved determinant of individual contraceptive use, to the extent that it is correlated among the members of the group. Replacing current prevalence with lagged prevalence does not solve the problem when the unobserved determinants of contraceptive use are serially correlated. This is why lagged HYV acreage in the village was seen to be less useful than lagged yield (shocks) in identifying social learning in agriculture. Individuals will generally tend to interact with those that are similar to them and so the problem with correlated unobservables is exacerbated when a self-reported endogenous reference group is used to define the social unit, as in the demographic studies, instead of an exogenous group that the individual is born into such as the village, caste, or clan. Behrman, Kohler, and Watkins (2002) control to some extent for the omitted variable problem by using individual fixed effects, but this approach does not account for changes in the individual determinants of the contraception decision over time.
The model of the fertility transition as a process of changing social norms laid out in the previous section allows us to place additional restrictions on the relationship between the individual’s contraception decision and lagged contraceptive prevalence in the local area. In particular, we expect that social effects will be restricted to the narrow social group within which norms restricting fertility were traditionally enforced. Contraceptive prevalence outside that social group should have no effect on the individual’s contraception decision.

Munshi and Myaux (2006) test these predictions using data from rural Bangladesh. The International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR,B) launched a Maternal Child Health—Family Planning (MCH-FP) project in 1978, covering 70 villages in Matlab thana, Comilla district. Contraceptive use information for all married women of reproductive age (15–49) and capable of conceiving in the 70 villages is available at six-monthly intervals over the 1983–1993 period. The MCH-FP project is quite possibly the most intensive family planning program ever put in place: all households in the intervention area have been visited by a Community Health Worker (CHW) once every two weeks since the inception of the project in 1978, and contraceptives are provided to them free of cost. Not surprisingly, contraceptive prevalence increased substantially over the sample period, from 40% in 1983 to 63% in 1993, with an accompanying decline in the total fertility rate from 4.5 children per woman to 2.9 children over that period. Munshi and Myaux’s objective is to study the relationship between individual contraception decisions and lagged contraceptive prevalence within well defined exogenous social groups over the course of this transition.

Recall from the model of the fertility transition that the individual’s contraception decision is determined by her own lagged decision and lagged contraceptive prevalence in the social group when norms are breaking down. In rural Bangladesh, the traditional norm was characterized by early and universal marriage, followed by immediate and continuous child-bearing. Religious authority provided legitimacy and enforced the rules that sustained this equilibrium. Changes in social norms should then have occurred independently within religious groups within the village. The two major religious groups in rural Bangladesh are Hindus, who constitute 18% of the sample, and Muslims who account for the remainder of the sample. Munshi and Myaux thus estimate regressions of the form:

\[ C_{ijt} = \alpha C_{ijt-1} + \beta_I \bar{C}_{jt-1} + \beta_O \bar{C}_{jt-1} + X_{ijt} \gamma + \omega_{jt} \]  

(12)

where \( C_{ijt} \) = 1 if individual \( i \) belonging to religion \( j \) uses contraceptives in period \( t \), \( C_{ijt} = 0 \) if she does not. \( C_{ijt-1} \) is her decision in the previous period, \( \bar{C}_{jt-1} \) measures contraceptive prevalence in her own religious group within the village in that period, and \( \bar{C}_{jt-1} \) represents the corresponding statistic outside the religious group. \( X_{ijt} \) is a vector of individual characteristics that includes age, age-squared, and in some specifications, individual fixed effects and time-period dummies. \( \omega_{jt} \) collects all unobserved determinants of the contraception decision at the level of the religious group within the village in period \( t \). Taking expectations across individuals within the religious group in Eq. (12), and then lagging the expression that is obtained by one period, it is evident
that $\tilde{C}_{jt-1}$ will be correlated with $\omega_{jt}$ and, hence, with $\omega_{jt}$ when the unobserved term is serially uncorrelated.

The model of changing social norms, however, places additional restrictions on the coefficients in Eq. (12): $\beta_I > 0$, $\beta_O = 0$. While a positive within-religion effect ($\beta_I > 0$) could be easily obtained, as discussed above, Munshi and Myaux show that positive within-religion effects and zero cross-religion effects ($\beta_I > 0$, $\beta_O = 0$) could only be spuriously generated if the unobserved $\omega_{jt}$, $\omega_{jt}$ terms are uncorrelated. The intuition for this result is that $\tilde{C}_{jt-1}$ cannot completely proxy for $\omega_{jt}$ once additional controls, $C_{ijt-1}$, $X_{ijt}$, are included in the contraception regression. $\tilde{C}_{jt-1}$ then provides information about $\omega_{jt}$ unless it is uncorrelated with that unobserved term, which can only be the case if $\omega_{jt}$, $\omega_{jt}$ are uncorrelated.

Munshi and Myaux estimate Eq. (12) separately for Hindus and Muslims. They find that $\hat{\beta}_I > 0$, $\hat{\beta}_O = 0$ in each case. This result is obtained with and without fixed effects and time-period dummies, with six-monthly and annual data, as well as with restricted samples in which villages that are dominated by a single religion are excluded. In contrast, when they partition the village using age or education, positive and significant cross-group effects are consistently obtained. It is only when the village is partitioned by religion that cross-group effects are completely absent.

Munshi and Myaux use the absence of cross-religion effects to rule out alternative interpretations of these results, based on changing program effects, economic development, and learning about new contraceptive technology. They argue that while these unobserved determinants of the contraception decision might be correlated across individuals in the village, it is difficult to imagine that they would be uncorrelated across religious groups within the village, which is necessary to spuriously generate $\hat{\beta}_I > 0$, $\hat{\beta}_O = 0$. For example, program inputs are provided by the same health worker to all households in the village, and while she might have a differential impact on women from different groups, her influence will not be uncorrelated across these groups. Occupations are not segregated by religion within these villages and so once again economic change cannot explain the absence of cross-religion effects. Finally, while female mobility is severely restricted in rural Bangladesh and so social interactions will typically occur within the religious group, we would still expect information about the new contraceptive technology to ultimately cross religious boundaries within the village. It is only the norm-based motivation for the fertility transition that can easily explain the striking within-religion and cross-religion patterns that are obtained, since information flows across religious boundaries are irrelevant in this case.

4. Health and education

Technological change increases farm profits and the returns to schooling, as established for example by Foster and Rosenzweig (1996) in the context of the Indian Green Revolution. Just as the grower was seen to be uncertain about the yield from the new HYVs, we would also expect him to be uncertain about the returns to schooling in this new
economic environment. And just as the grower’s neighbors were seen to provide information about the new crop technology above, we would expect them to provide information about the returns to schooling as well.

A recent paper (Yamauchi, 2007) studies social learning and investment in education with the same three-year farm panel at the onset of the Indian Green Revolution that was used by Foster and Rosenzweig (1995) and Munshi (2004). Schooling levels among the growers in the sample were determined long before the unexpected availability of the new HYV technology and so the returns to schooling can be estimated directly at the level of the village using realized incomes. A positive relationship between schooling enrollment among the children and the returns to schooling in the previous generation is then seen to be indicative of social learning.

One potential problem with this empirical strategy is that a spurious role for neighbors’ returns to schooling could be obtained. Suppose that the returns to schooling are correlated across households in the village and over time and that there is no uncertainty. Returns to education in the previous generation could still proxy for the corresponding returns in the current generation, which determine school enrollment. The implicit assumption in Yamauchi’s analysis is that the returns to schooling were essentially constant across villages prior to the Green Revolution, or that these returns were uncorrelated with the returns in the post-Green Revolution period. Under these conditions, the returns in the new economic environment can be consistently estimated using growers’ previously determined schooling levels. A second potential problem is that learning could occur at the level of the household rather than the village; if returns to schooling are spatially correlated then these alternative learning channels cannot be easily disentangled.

Yamauchi’s solution to the problems discussed above is to derive testable predictions from the theory that provide additional support for the presence of social learning at the level of the village. He shows formally that social learning will be faster when the income-variance is lower and when there is greater heterogeneity in educational attainment in the village. This last prediction is particularly interesting, but not inconsistent with Munshi’s observation that social learning will be slower in heterogeneous populations where neighbors’ characteristics are unobserved or imperfectly observed. Schooling is an easily observed characteristic and Yamauchi’s insight is that more variance in this characteristic leads to more precise estimates of the returns to schooling. Matching these predictions from the theory, schooling enrollment among the children is increasing in the returns to schooling in the village in the previous generation and, more importantly, is increasing in the interaction of the returns and the variance in educational attainment in the village.

In parallel with the introduction of new production technologies, vaccines and other medical treatments for infectious diseases have dramatically lowered morbidity and mortality throughout the developing world. However, the introduction of modern medicines has often met with community resistance and public health programs have sometimes been seen to collapse after an initial period of success. Recent research on a
deworming program in Western Kenya by Kremer and Miguel (2007) applies social learning to understand these dynamics.

The program that they evaluate covered 75 primary schools in Busia district, with over 30,000 enrolled students aged 6–18. The schools were randomly divided into three groups: Group 1 schools participated in the deworming program over the 1998–2001 period, Group 2 schools participated 1999–2001, and Group 3 schools began participating in 2001. A representative sample of parents of children in Group 2 and Group 3 schools were interviewed in 2001. The respondents were asked to list their closest social links: the five friends they speak with most frequently, the five relatives they speak with most frequently, additional social contacts whose children attend local primary schools, and individuals with whom they discuss child health issues. The listed individuals define the respondent’s set of social links.

Kremer and Miguel study the effect of access to “early” links, in Group 1 and Group 2 schools, on the individual’s participation in the program. The total number of reported links could reflect unobserved parental characteristics, such as how sociable they are, which could be correlated in turn with other characteristics that directly determine the choices that are made for the child. Conditional on the total number of links, however, the number of early links is determined by the random assignment of schools to different groups. Kremer and Miguel find that an exogenous increase in the number of early treatment links leads to a significant decline in the probability that the individual will participate in the program.

One explanation for this negative effect is based on the group-immunity externality that accompanies these early links. Previous research by Kremer and Miguel in the same setting indicates that deworming programs significantly reduce infection rates and increase attendance in both participating schools as well as neighboring schools. We might expect such externalities to be even stronger within the individual’s own social group, although the empirical results appear to indicate otherwise. An increase in program participation within the social group would then lower the benefit to the individual from deworming, since the child is less likely to get infected. If the group-immunity externality dominates the information externality that accompanies links’ participation in the program, as in Bandiera and Rasul, then the negative effect that is observed could be obtained.

Kremer and Miguel, in contrast, favor an alternative explanation based on the idea that members of the community were systematically misinformed about the private benefit from the deworming medicine. They argue that the agency providing this medicine might have described what is effectively the social benefit, inclusive of the group-immunity externality that is provided by participation, as opposed to the private benefit. This explanation is also consistent with the estimated negative social effect, with individuals gradually learning the true private value of the new technology over time and downgrading their priors. While Kremer and Miguel do not consider such a possibility, I suspect that there may exist conditions under which it is optimal for the external agency to inflate the benefit from the treatment, in response to the group-immunity externality which fails to be internalized by the individual. The inflated priors will ultimately come
in line with the true private value of the new technology, but in the short to medium term, at least, participation rates may be brought closer to the socially optimal level.

The deworming program also provides a convenient setting in which to compare non-experimental estimates with the experimental results described above. Kremer and Miguel study the relationship between the participation rate in the individual’s school and the individual’s own decision and find a positive association, in contrast with the negative experimental results that are obtained. While these differences between the experimental results and the non-experimental results are instructive, they do not undermine previous studies on social learning. Most of the non-experimental studies described in this chapter use panel data to estimate the effect of neighbors’ lagged decisions and outcomes on the individual’s current decision. Individual fixed effects, and even the individual’s lagged decision, are used in some applications to control for unobserved characteristics that may be correlated within the social group. Recognizing the inherent difficulties in identifying social learning from neighbors’ lagged decisions, many studies of agricultural adoption have concentrated on the grower’s response to neighbors’ yield shocks rather than decisions. It is not obvious that Kremer and Miguel’s spurious non-experimental results would hold up with all of these controls. Furthermore, previous non-experimental studies explicitly acknowledge the potential for omitted variable bias and so have used restrictions from the theory to rule out alternatives to social learning. For example, Munshi and Myaux use the context-specific result that cross-religion effects are completely absent in rural Bangladesh to rule out alternatives. The result that within-group effects are larger than cross-group effects in the non-experimental regressions, as reported by Kremer and Miguel, is not as useful in ruling out alternatives.

5. Conclusion

It has now been 10 years since the first empirical papers on social learning in developing countries were published. While the theoretical literature on this topic has grown rapidly over this period, the empirical literature has not generated a similar surge of research papers. As the discussion in this chapter suggests, the identification of social learning is a challenging problem, and while previous studies have made some progress along this dimension, their overall lack of success might have discouraged new entrants to the field. Here the use of randomized experiments, as in Kremer and Miguel’s work in Kenya, seems to be a promising direction for future research.

The empirical literature has also been unable to come up with hypotheses that go beyond straightforward tests of social learning, albeit in different contexts. What makes the theoretical literature so interesting is that rational individuals can end up making choices that lead the entire community to the wrong outcome. Whether such pathologies are widespread or not is an important empirical question, which the learning literature has not attempted to answer. More recent moves towards alternatives to the classical
model, as in Kremer and Miguel’s argument that the external agency provides exaggerated signals of the quality of the new technology, takes the literature in a new direction. But once again, it is not clear that such misinformation is empirically relevant.

Ultimately how important is social learning in the development process? Foster and Rosenzweig report results from a simulation exercise based on the estimated parameters from their learning model, which compares profits without learning, with learning from own experience, and with learning from own and neighbors’ experience. Profitability from the new HYV was lower than profitability from the traditional variety to begin with, but HYV profits exceed traditional profits after four years of experience without learning. With social learning, this point is reached one year earlier. Similarly, Yamauchi’s simulations indicate that an increase in schooling inequality within the village could increase enrollment levels by nearly 10 percent. Access to social information appears to be readily available in many practical applications, particularly since there is little cost to the individual from providing information to his neighbors. Thus, the value of interventions that provide the seed for the subsequent spread of such information could be quite high. Understanding how best to design such interventions would seem to be an important area for future research.

Acknowledgements

I thank Abhijit Banerjee, Esther Duflo, George Mailath, and Jack Porter for many discussions that have shaped my thinking on this topic. Rohini Pande and participants at the May 2005 Bellagio Conference provided helpful comments. I am responsible for any errors that may remain.

References

Bulatao, R. (1998). The Value of Family Planning Programs in Developing Countries. RAND, Santa Monica, CA.