Imperfect Social Learning Among Kenyan Smallholders

Heidi Hogset
Molde University College (Norway)
and
Christopher B. Barrett§
Cornell University (USA)

March 2007 draft: comments greatly appreciated

§Corresponding author: Department of Applied Economics and Management, Cornell University, 315 Warren Hall, Ithaca, NY 14853-7801 USA Tel. +1-607-255-4489, Fax. +1-607-255-9984, E-mail: cbb2@cornell.edu

This research was made possible by funding from the Social Science Research Council, with funding from John D. and Catherine T. MacArthur Foundation, The Pew Charitable Trusts, the Mario Einaudi Center for International Studies, the Cornell International Institute for Food, Agriculture, and Development, the United States Agency for International Development through grant LAG-A-00-96-90016-00 to the Broadening Access and Strengthening Input Market Systems/ Collaborative Research Support Program (BASIS CRSP), and the Strategies and Analysis for Growth and Access (SAGA) cooperative agreement, number HFM-A-00-01-00132-00, as well as the Coupled Natural and Human Systems Program of the Biocomplexity Initiative of the National Science Foundation, grant BCS - 0215890. The International Center for Research on Agroforestry (ICRAF) in Nairobi, Kenya Agricultural Research Institute (KARI) and Kenya Forestry Research Institute (KEFRI) generously provided practical support during the fieldwork in Kenya. Annemie Maertens and Paulo Santos provided helpful comments. Any remaining errors are the authors' alone.

Imperfect Social Learning Among Kenyan Smallholders

Abstract

This paper explores passive and active forms of social learning by Kenyan farmers. We document high rates of error and bias in respondents' proxy reporting of their network members' behavior. Proxy reporting of peer behavior proves no more accurate than random guesses based on known adoption rates in the broader population, suggesting limited individual- or network-specific knowledge of behaviors. Moreover, perceptions of others' behaviors are systematically related to respondents' own behaviors, confirming Manski's (1993) concern about reflection problems that might overstate social networks effects on learning in standard regression models. The bias and inaccuracy of reporting on network members' behaviors calls into question the importance of passive social learning. We then show, however, that these same respondents actively pursue and accurately report obscure information when given a direct incentive to do so. They mobilize social networks, especially to access information available from socially distant people, the weak ties that Granovetter (1973) suggests are the most valuable information sources within a social network. These results suggest a need to revisit how development economists study social learning, as active social learning may be more important and passive social learning less important than implied by prevailing models and empirical tests of social learning.

Introduction

Economic growth and poverty reduction depend heavily on technological change. The pace of technological progress depends, in turn, on the speed and accuracy with which economic agents learn about and adopt new innovations. In economic models, agents generally either learn from their own experience ("learning by doing"), from passive observation of others ("learning from others"), or both (Armantier 2004). Development economists' interest in whether learning from others about technologies is widespread and important has grown considerably in recent years (e.g., Behrman, Kohler and Watkins 2002; Miguel and Kremer 2004; Munshi 2004; Mwakubo *et al*. 2004; Conley and Udry 2001, 2005; Bandiera and Rasul 2006). In this paper, we explore the apparently imperfect process by which small farmers in Kenya learn from one another and the instrumental role(s) social networks play in active and passive social learning. In particular, we seek not to establish *whether* or not people learn from others, but rather *how* they learn from others.

Agents embedded in a social network may learn from and imitate the behavior of others within their network, as suggested by the well-known contagion model for adoption of innovations (Rogers 1995). These network-mediated effects are often decomposed into social learning effects and social influence effects (Montgomery and Casterline 1996). Social learning reflects an active process between persons who identify and evaluate each other's contributions to the information generation process. Active social learning often involves seeking out socially distant people with whom one otherwise has infrequent contact, so-called "weak ties" one mobilizes selectively (Granovetter 1973). By contrast, social influence arises through passive acquisition of (sometimes inadvertent) casual observations of others, typically the most socially proximate members of one's social network ,with whom one interacts most frequently, i.e., those with whom one shares “strong ties”. Because of its passive nature – individuals are often not purposefully seeking out the information they obtain— social influence sometimes yields merely awareness of the prevailing pattern of behavior within a relevant reference group, rather than specific, accurate knowledge of each individual’s behavior.

This distinction between passive social influence relying primarily on strong ties and active social learning that depends disproportionately on weak ties has gone unexplored in the recent development economics literature on social learning and technology adoption. Yet the distinction may make a difference. First, there are implied differences in the cost of information diffusion. If passive social learning accurately diffuses information through networks, that suggests a more modest role for extension programs to disseminate new information accurately – at least within well-connected subpopulations – than if passive learning works less well, in which creating incentives for people to seek out accurate information via active learning may be more necessary. At present, we have little empirical evidence on active or passive social learning processes among small farmers in the developing world.

This leads to the second, more methodological reason why this distinction matters. In designing field data collection and analyzing the resulting data, researchers inevitably make assumptions about learning processes. Those assumptions, if erroneous, may significantly influence inference and thus the design of policies and projects based on empirical evidence. Our concern is that contemporary economic analysis of technological change and innovation seems to largely assume passive learning based on the social influence of strong ties within respondents’ social networks, without having tested carefully to ensure that assumption is correct. If instead people rely heavily on active social learning through selectively activated weak ties, then learning models based on standard measures of one's social network characteristics (e.g., network density) may mislead.

Moreover, researchers studying learning from others need information about peers' behaviors. In many cases, the research subjects themselves are the source of this information via proxy reporting of peer behavior, despite longstanding warnings about problems associated with such data (Montgomery and Casterline 1996). Recent examples of social network studies that use proxy reporting of peer behavior include several eminent scholars and widely cited papers (e.g., Behrman *et al.* 2002, Bandiera and Rasul 2006). Proxy reporting implicitly assumes, however, that people accurately observe behaviors within their social network. But errors generated by imperfect peer observation may not be random but, instead, systematically related to respondent attributes in a way that might distort statistical inference. For example, if a respondent who uses a particular technology is more likely to believe a member of his social network likewise uses that technology, regardless of the match's actual behavior, then what appears as learning effects may be mere reflection effects (Manski 1993).

This paper presents field data which demonstrate that commonplace assumptions about active and passive social learning and about the accuracy of proxy reporting may not always be empirically supportable. Our objective in this paper is therefore modest and cautionary. We aim merely to challenge economists and other social scientists to pause to consider carefully the processes by which individuals learn, the social ties of instrumental value in those processes, and the implications for research design and policy recommendations from empirical research.

Our contribution is twofold. First, in the next section, we offer a simple heuristic model that allows for both costless, passive learning and costly active learning. Passive learning is the sort conventionally considered in the social capital literature, wherein the mere existence of social contacts leads to information flow. In active learning, an agent mobilizes his or her social network to acquire information. In active learning, having (direct or indirect) access to persons with the right skills or knowledge is highly important, while the density of such persons may be less important. If the accuracy and cost of active and passive observation differ based on individual characteristics, different people may predictably choose one form over the other.

Second, the remainder of the paper presents empirical evidence from rural Kenya that suggests that the emerging quantitative literature on social learning may be making overly strong implicit assumptions that merit more direct investigation. Using data from standard network mapping exercises, we find that passive learning, manifest in proxy reporting of peer behavior within respondents' social networks, is highly imprecise – consistent with an understanding of broad patterns in the population but no specific knowledge of individual network members' behaviors – and significantly biased in a way that raises questions about prospective reflection problems in standard regression analyses of social learning. We then present evidence from an experiment designed to measure rural villagers' ability to locate previously unknown information, prospectively by activating social networks. These data suggest that active learning is significant and varies predictably according to individual respondent characteristics, and that social networks play an instrumental role, not only as direct sources of information, but perhaps especially as intermediaries through which one can learn from socially distant persons, recalling Granovetter's classic (1973) study of the importance of "weak" ties relative to "strong" ones. If our sample is at all representative of broader patterns of agent observation and learning, it would seem that active social learning processes may be more important than often acknowledged and that social network effects through passive learning may be somewhat exaggerated.

A Simple Learning Model

Agents can learn from others actively, passively or both. For the purposes of simplification, we assume an economic agent seeks to learn about some phenomenon of interest, u, perhaps a target input rate (Foster and Rosenzweig 1995, Conley and Udry 2006) or optimal cultivation and marketing methods for a new crop (Santos and Barrett 2005, Bandiera and Rasul 2006). Agent j has prior beliefs over u that are distributed $N(\hat{u}_{0j},\sigma^2_{0j})$. We assume her well-being is decreasing in the deviation of $\hat{u}_j$ from the true population parameter u* and in the variance of her posterior beliefs, σ^2_{1j}, as would be true of the examples above. For simplicity's sake, we abstract from learning from one's own experience and focus just on the agent's Bayesian updating based on potentially imperfect observation of others' behavior. The information j absorbs can arrive either via costless, passive observation of her network or through active and potentially costly searching, both of which may depend on her network density, n, which determines the sources of information available to her, directly or indirectly, and her educational attainment, e, which affects her ability to process and absorb the information she receives.

The passive learning process works as follows. The agent costlessly but imperfectly observes the vector of relevant behaviors of the n_j members of her social network, u_{-j}. We assume errors of observation are independent across network members, although they might be correlated with attributes of the agent. That is, agent j passively learns $\hat{u}_{-j} = u_{-j} + \varepsilon^P_{-j}$, where errors of passive observation, ε^P_{-j}, are distributed $N(b(x_j), \sigma^2_{Pj}(x_j))$, with bias b and positive, finite variance σ^2_{Pj}, both of which may be related to characteristics of agent j, such as educational attainment, proximity to the social network member, or other factors affecting the accuracy of one's observations. If we assume all agents choose u following the same expected welfare-maximizing rule, and any bias in observation is subconscious – i.e., people do not notice and account for their biases – then $\hat{u}_{-j}$ provides the best available estimate of u* since no one in j's social network should be intentionally erring in their practice of u.

Agent j may also or instead pursue costly active learning, incurring cost $c(d_j,n_j,w_j)$ to collect information through a number of direct approaches, d_j, with the cost equaling zero if she chooses not to engage in active learning (i.e., $c(0,n_j,w_j)=0$). We assume costs are decreasing in social network density, under the assumption that more dense networks provide (at least weakly) superior access to key information, and increasing in the opportunity cost of j's time, as reflected by her wage rate, w_j, which may itself be a function of educational attainment and other individual and site-specific characteristics. Active learning generates information from selected others who provide unbiased but noisy signals. Agent j thus actively learns $\hat{u}^*=u^* + \varepsilon^A_j$, where errors of active observation, ε^A_j, are distributed $N(0, \sigma^2_{Aj}(x_j))$, with no bias and positive, finite variance σ^2_{Aj}, which may again be related to characteristics of agent j.

Following Bayes' rule for updating the moments of a normally distributed random variable, agent j thus learns u_{-j} (passively) from each member of her social network, u* actively through direct approach to knowledgeable individuals, or both, generating posterior beliefs about the mean and variance of u:

$$\hat{u}_{1j} = \frac{\lambda_0 \hat{u}_{oj} + \lambda_P \hat{u}_{-j} + \lambda_A \hat{u}^*}{\lambda_0 + \lambda_P + \lambda_A} \tag{1}$$

$$\sigma^2_{1j} = \frac{1}{\lambda_0 + \lambda_P + \lambda_A} \tag{2}$$

where $\lambda_0=1/ \sigma^2_{0j}$, $\lambda_P=n_j/ \sigma^2_{Pj}$, and $\lambda_A=d_j/ \sigma^2_{Aj}$. In words, j's posterior belief about u* is the weighted sum of her prior belief, passive observations of members of her social network, and active learning associated with costly search. The weights directly reflect the density of the social network (which puts more emphasis on passive learning), the extent of direct contacts undertaken (which emphasizes active learning), and inversely reflect the relative precision of each type of observation. In this formulation, those who have no social network to observe passively and who choose not to undertake active information collection do not update their posterior, i.e., $\hat{u}_{1j} = \hat{u}_{0j}$ and $\sigma^2_{1j} = \sigma^2_{0j}$. Others learn. Possession of a social network yields dividends from passive learning if observations are unbiased, so that $\hat{u}_{1j}$ is nearer u* than $\hat{u}_{0j}$ if $b(x_j)=0$ (for $n_j>0$).

Although direct approaches to unbiased experts inevitably lead to more accurate beliefs, there may not be an economic incentive to learn actively. The value of the learning is simply the difference between the expected welfare generated by optimal behavior conditional on belief $\hat{u}_{1j}$ and that generated by optimal behavior conditional on belief $\hat{u}_{0j}$, accounting for any cost of active learning $c(d_j,n_j,w_j)$. That value is decreasing monotonically in the bias of observations, $b(x_j)$, in the variance of those observations, $\sigma^2_{Pj}(x_j)$ and $\sigma^2_{Aj}(x_j)$, for passive and active learning, respectively, and in the costs of active learning. Indeed, if observations are biased but precise, posterior beliefs may lead to divergence from, rather than convergence towards, the true optimum, u*. Thus biased learning can even be welfare decreasing. And if passive learning is highly imprecise – e.g., if direct observation of an individual is no more accurate than a random guess based on the prevailing population characteristic – then it may convey no additional information beyond one's prior beliefs. Under our assumption that active learning is more precise and less biased, its gross benefits weakly dominate those of passive learning. Still, rational agents should only undertake costly active learning when the added welfare generated by active learning exceeds the cost. Because costs surely vary across individuals according to attributes such as the opportunity cost of one's time and educational attainment, propensity to invest in active learning may well vary predictably within a population.

This simple heuristic model raises several questions amenable to empirical inquiry. First, how accurate and unbiased is passive observation of the behaviors of members of one's social network (i.e., what do we know about $b(x_j)$ and σ^2_{Pj})? If observation is accurate and unbiased, the informational benefits of a social network could be considerable as passive social learning via social influence can be a (nearly) costless means of disseminating valuable information. But if passive learning is neither accurate nor unbiased, then learning and technology diffusion may require more costly actions. This leads to our second set of questions: will people undertake active learning in response to incentives and how do patterns of active pursuit of information vary with agent attributes, such as the opportunity cost of their time, and what role(s) do social networks play in active social learning? If there are patterns in who pursues active learning and who does not, this may help account for cross-sectional variation in patterns of technology adoption, market participation and other behaviors commonly associated with learning about new opportunities. It may also offer important clues for targeting of extension efforts.

Passive Social Learning

Data from Kenya's Highlands

The data with which we explore passive and active learning patterns were collected during 2003-2004 in two sites in Kenya: Manyatta Division in Embu District (Eastern Province), and the former Madzuu Division (now divided between several new political entities) in Vihiga District (Western Province), in the country's central and western highlands, respectively. The sample of research subjects consists of 120 households in each site who had been previously surveyed under a separate research project. All households were active farming units, but some households had substantial nonfarm income sources as well.

We ran two different data collection exercises among the same households. The first survey aimed to enumerate respondents'[1] social networks of different sorts and to collect information on respondents' perceptions of their network contacts' use of any of four different natural resource management (NRM) practices that may improve soil fertility and agricultural productivity: terracing, fallowing, use of organic fertilizers, and planting crops in deep, manure-filled pits, a practice known by the Swahili word *tumbukiza*, which means "submerge". We describe these four NRM practices in a bit more detail below.

The survey respondents (whom we refer to as "egos") were asked to identify the people with whom they engage in borrowing and lending (their "transfers" network), the people they (currently) "like" to discuss issues of farming with (their "information" network), and their neighbors, i.e., the household head of each farm bordering the respondent's. Respondents were asked a set of questions about each of these network contacts (labeled "alters"). For each alter, irrespective of network, respondents were asked whether they had discussed the four NRM practices under study with alter, and whether alter had adopted the practice.

A subset of the sample (55 respondents) – one village from each site – was selected for a follow-up round of "snowball sampling" in which alters identified in the first round were tracked down and interviewed, using the same questionnaire to which egos responded in the first round. This method permits us to check the accuracy of respondents' perceptions of the technology adoption behaviors of members of their social networks.

The second study, fielded among the same 55 respondents, elicited their answers to questions asking for often-obscure data available in the area – prices, extension

[1] Within each household, we interviewed the person who had the main responsibility for day-to-day farm management decisions. In many households both spouses were farm managers, either working together, or having separate enterprises. Therefore, about one tenth of the sampled households were selected for interviews with both spouses, selected among those where both spouses were farm managers.

recommendations, identity of particular officials, etc. Respondents were given a modest and time-varying financial incentive to locate the correct information in a timely fashion. They were visited at weekly intervals to see if they could now correctly answer any questions they had previously missed. At the end of each visit, we asked respondents a variety of questions about the means they used, if any, to track down the correct answer(s). These data thus provide us with direct evidence of learning and the methods used to seek out and acquire accurate information. We provide further detail on this experiment below, when we introduce those empirical results.

The 55 primary respondents and 267 secondary respondents generated a total of more than 1200 relations amongst each other, i.e., where both ego and alter were among the respondents (the "snowball sample", based on alter self-reporting). This study also uses data collected for an earlier study. These data are only available for primary respondents. When secondary respondents are excluded, the data comprise 654 relations. The relational data we use for studying passive learning thus do not suffer from small sample problems. By contrast, the analysis of active learning uses data on only the 55 primary respondents. We therefore interpret these latter results as suggestive only, inviting corroboration or refutation using other, larger data sets.

The Natural Resources Management Practices Under Study

Before presenting the first set of empirical results, let us briefly describe the four technologies studied. By design, these differ in the ease with which outsiders might observe the practice and with respect to the prevalence of each practice within the population. In a matrix of high/low ease of observation and high/low prevalence, each of the technologies can be assigned to a separate cell (Figure 1).

Terracing was introduced in Kenya nearly a century ago by the British colonial government in response to severe land degradation in parts of the highlands. Despite the heavy-handed interventions of the colonial government and some policy reversals in the immediate post-independence period, adoption of terracing increased steadily as farmers discovered the benefits of terracing (Kamar, Mburu and Thomas 1999). Adoption of terracing has therefore likely reached its long-term equilibrium level, at around 90% adoption in both sites. Terracing is easy to observe as it involves clear, prominent alterations of the physical landscape, so one would expect passive observation of terracing to be relatively accurate and unbiased.

Fallowing was an important component of the traditional slash-and-burn agriculture practiced in the past. Increasing population pressure on the land has led to increased continuous cultivation of land and a decline in fallowing (Drechsel *et al.* 2001). Fallows are more common in Embu (31%) than in Vihiga (23%), reflecting a higher population density in the latter. Like terracing, fallowing is easy to observe by any passerby, especially in the high population density areas we study, where unexploited private land is relatively uncommon.

For generations, farmers in the Kenyan highlands have incorporated organic matter into their cultivated plots as a way of boosting soil fertility. Thus organic fertilizer use is a mature technology, like terracing. Almost all (96% of) respondents apply at least one type of organic fertilizers.[2] However, many have very limited access to organic material for production of organic fertilizers, so quantities applied are often quite small, making observation of organic fertilizer application difficult.

Tumbukiza represents deep incorporation of organic fertilizer into the soil on a small piece of land, thereby increasing the long-term effects of the treatment but requiring significantly

[2] We used a relatively broad definition of organic fertilizers: farm-yard manure, compost, and mulching using organic matter collected on or off-farm.

more labor effort than top dressing with organic or inorganic fertilizer. This relatively new practice is being promoted by the government extension service in the study areas. But most survey respondents were unfamiliar with *tumbukiza*. In Embu, about 45% of our respondents practiced *tumbukiza*, while in Vihiga, only 12.7% did. *Tumbukiza*, like organic fertilizer application, can be difficult to observe passively, thus one might reasonably expect relatively greater inaccuracy in proxy reporting of this practice.

The Accuracy of Passive Learning

The first question we study is the accuracy of egos' reporting of alters' behaviors, a key determinant of passive learning through social influence. The more people routinely and seriously misperceive others' behaviors, the more limited their capacity to learn passively through social influence. Because our snowball sample permits us to match egos' proxy reporting of alters' behavior with alters' self-reported behaviors (which we assume to be correct), we can establish the accuracy of passive social learning. The accuracy of proxy reporting reflects the accuracy of passive social learning because egos did not have occasion or incentive to seek out this information directly in response to our survey, thus by design the information elicited was not the product of active learning associated with search.

What is the appropriate benchmark against which to judge the accuracy of proxy reporting? The naïve approach is simply to look at the percentage of correct answers. The associated null would be that proxy reporting is accurate, i.e., the probability of correct responses is one versus the alternate that it is significantly less than one. But this can easily conflate knowledge of broader patterns in a population – in the present case, knowledge of whether certain practices are widespread or not – with knowledge of the specific behavior of an individual member of one's self-reported social network. If adoption of terracing or organic fertilizers is widespread, then uninformed guessing that everyone practices these techniques may yield few errors. Conversely, for an infrequently adopted practice like *tumbukiza*, one may erroneously report some network members' behaviors in spite of knowledge that makes proxy reporting far more accurate than a random guess. Therefore, a better measure is whether the probability of accurately reporting the behavior of a network member is higher than a mere random guess based on knowledge of the broader population patterns.

Toward that end, we use the following approach. An ego, j, answers "yes" to the question "does alter employ {one of the four NRM practices}?" with frequency f_j. If the true adoption rate among j's network contacts is α_j, j answers "yes" correctly with probability $f_j*\alpha_j$, and "no" correctly with probability $(1 – f_j)*(1 - \alpha_j)$, thus the probability of reporting proxy behavior correctly is the sum of these:

$$p_j = f_j*\alpha_j + (1 – f_j)*(1 - \alpha_j) \qquad (3)$$

If j knows the behavioral pattern in the broader population, α, but not the specific behaviors of individual alters, the natural f_j would be α. Assume f_j is distributed $N(f, \sigma_f^2)$ in the population. An appropriate benchmark for the accuracy of proxy reporting – do respondents know anything specific about their network members, beyond their knowledge of the population as a whole? – is the null that $f = \alpha$. Assume proxy reporting follows a stochastic process that generates correct answers that are distributed Binomial(n, p), where n is the number of observations, and p is the probability of guessing correctly. Then under the null hypothesis that people have no specific knowledge of the behavior of their networks' members,

$$p = \alpha^2 + (1 - \alpha)^2 \qquad (4)$$

Given the above reasoning, we test two different hypotheses. The first is that respondents perfectly report the behaviors of their own social network members, i.e., a null hypothesis that proxy reporting errors do not exceed a standard (albeit arbitrary) tolerance threshold of 5%. The second hypothesis is that, in aggregate, proxy reporting reflects no specific knowledge about the behaviors of those in one's network, that it does no better than mere guessing based on knowledge of the population adoption rate. The null is thus that reporting is consistent with guessing based on a correct belief about overall adoption rates, i.e., $f=\alpha$. Rejection of the first hypothesis implies imperfect reporting of network members' behavior, and thus some prospect for bias and associated reflection problems in inference, which we investigate in the next sub-section. Failure to reject the second null implies respondents lack individual-specific knowledge about NRM adoption behavior, but may know overall adoption rates relatively well (i.e., $f = \alpha$). This has important implications for reliance on social learning to accurately disseminate valuable information, for example, about new NRM practices or production technologies.

The alternate hypothesis to the second null is that $f \neq \alpha$, which means that respondents are guessing *without* knowing overall adoption rates. However, when $f \neq \alpha$, the rate of correct reporting can deviate from what is expected under the null either way. The hypothesis (both the null and the alternate) is rejected if the whole pattern of correct and incorrect answers is inconsistent with the hypothesis. A higher rate of correct reporting than expected under the null can be interpreted as evidence that respondents do have some network-specific knowledge about peer behavior if incorrect answers are not biased. The discussion below will illustrate this point. Only a lower than expected rate of correct answers can be interpreted straight forward as evidence that respondents know neither network-specific behavior nor overall adoption rates.

Table 1 reports rates of correct and incorrect proxy reporting by site and NRM practice.[3] We see that in both sites and for all four practices, errors in reporting (failure rates) exceed 5%, so the first null is rejected. Respondents do not accurately report peer behavior. Indeed, the two highest failure rates in the table are close to 50%.

We use the sum of correct positive and false negative answers as an unbiased estimate of α. The final row of Table 1 reports the t-statistic for the test of the second null hypothesis that $f=\alpha$. A positive test statistic means the rate of correct reporting is higher than expected under the null, and vice versa if the test statistic is negative. In only two of eight cases (fallowing in Embu and *tumbukiza* in Vihiga) respondents have reported correctly more frequently than expected under the null. In three of the cases, proxy reporting of network members' behavior is statistically significantly less accurate than a random guess based on knowledge of the prevailing population adoption rate.

The case with the lowest error rate (organic fertilizer use in Vihiga) is *not* among those with a statistically significantly better-than-expected result because the population adoption rate there is very high thus accurate reporting need not reflect network-specific knowledge. Similarly, the case with the second highest error rate (*tumbukiza* in Embu) is not among those with a significantly worse-than-expected result, and the NRM practice with the highest proxy reporting error rate (fallowing in Vihiga) is less significantly worse than expected than two other cases with lower error rates. This simply reflects that as α approaches its limits (0 or 1), the variance of stochastic proxy reporting falls.

Under our general hypothesis that respondents are guessing and answering "yes" with frequency f_j, the expected rate of incorrect "yes" responses is $f_j*(1-\alpha_j)$, and the expected rate of

[3] We use site-specific adoption rates under the maintained hypothesis that respondents are more likely to know site-specific adoption rates rather than rates spanning multiple sites hundreds of kilometers apart. The qualitative results are reinforced by using the full sample rates.

incorrect “no” answers is $(1 – f_j)*\alpha_j$. Under the null that f=α, these rates will be equal, so we expect the rate of incorrect positive and negative answers to be equal when we fail to reject the null. But if the null is rejected in favor of the alternate that f≠α, a certain pattern of divergence in the frequencies of incorrect positive and negative answers is consistent with the theory that respondents are guessing. Let us define reporting bias, b, as the difference between incorrect positive and negative answers, i.e. (dropping indexing to indicate aggregate measures),

$$b = f*(1 - \alpha) - (1 - f)*\alpha = f - \alpha \qquad (5)$$

Closer inspection of Table 1 reveals that the test statistic is always negative when the reporting bias is positive and $\alpha < 50\%$ or when the reporting bias is negative and $\alpha > 50\%$. This follows quite naturally. Per equation (3), $\partial p/\partial f = 2\alpha - 1$, implying $\partial p/\partial f > 0$ if and only if $\alpha > 0.5$. Therefore, guessing “yes” (“no”) for all network members when one knows that a majority of the population adopts (does not adopt) the practice maximizes the probability of correct answers, leading to both positive (negative) bias and rejection of the f=α null. This merely underscores how important it is to consider the underlying population patterns when evaluating the accuracy of proxy reporting and how inaccurate proxy reporting appears to be in this sample, even for easily observable NRM practices.

But a rate of correct reporting that is higher than expected than under the null, associated with unbiased errors, is not consistent with the guessing hypothesis. This may be the case for fallowing in Embu, where reporting is signifcantly better than expected under the null, while bias is relatively small (14% incorrect positives and 15% incorrect negatives), suggesting that respondents’ reporting is actually better than random guessing. But in general, our results suggest that respondents have some idea about true adoption rates, but this knowledge is imprecise and does not seem to incorporate much network- or individual-specific information on network members' behaviors.

Bias and Reflection Problems in Proxy Reporting

Our next question is, does f_j vary in any systematic way with characteristics of the respondent or alter? If proxy reporting errors vary systematically with variables commonly used in social network studies on technology adoption, then inferences made in those studies may be misleading. In order to get a sense of whether that is the case, for all four NRM practices under study we regress the error in proxy reporting (proxy's answer minus alter's own answer) on four dummy variables, one measuring ease of observation (alter is a family member, defined as a next of kin – parent, sibling, or own child), one reflecting the respondent's own adoption behavior with respect to the NRM practice in question, one indicating whether ego and alter have spoken about the NRM practice in question, and the last one indicating the research site. Table 2 reports descriptive statistics for the relevant variables from the snowball sample, which is used in these regressions.

There are indeed statistically significant patterns to proxy reporting errors (Table 3). The dependent variable can take three values – 0 if the is no error, -1 if negative error (underreporting adoption), and 1 if positive error (overreporting adoption). Thus, we have chosen a multinomial probit model for these regressions. The base outcome is no error, so the table reports coefficient estimates on making negative and positive errors, respectively. The sign on a coefficient estimate indicates whether proxies are more or less likely to make a negative or positive error when reporting alters' adoption when the variable in question is positive. We see that the constant term is consistently negative, sizable and in all cases but one highly significant, indicating that non-adopters who are not close family members and have not discussed the NRM practice with alter are significantly less likely to report alters' adoption of all four NRM practices incorrectly. On the other hand, the coefficient estimates on having discussed the technology indicate that egos are more likely to overreport alters' adoption of uncommon technologies the two have discussed (i.e., fallowing and *tumbukiza*), and less likely to underreport alter's adoption of common technologies the two have discussed (terracing and organic fertilizers). In the case of *tumbukiza*, having discussed makes ego both more likely to overreport, and less likely to underreport, systematically mistaking discussion of the technology for adoption of it. The coefficient estimate for the effect of one's own adoption behavior on reporting errors indicate that egos who practice the most uncommon technologies are more likely to overreport alter's adoption while egos who have terraced their farms are less likely to underreport alter's terracing., This suggests reflection bias, i.e., ego projects his own behavior on alter.

The absense of significant coefficient estimates on the family member variable indicates that although their reporting is not significantly biased, it is also no better than the reporting of non-family members. In contrast, the site variable has a strong effect on reporting errors. Respondents in Embu are more likely to overreport alter's adoption of terracing than those in Vihiga. The actual adoption rate of terracing is lower in Embu (87%) than in Vihiga (92%), so there is more scope for overreporting adoption there. Their reporting of alter adoption of organic fertilizers is significantly more biased, with a higher probability of underreporting and a lower probability of overreporting. Both of these practices are very common in both sites, but they differ with respect to ease of observation. It looks like the practice that is easy to observe has been overreported, while the one that is hard to observe is underreported.

The coefficient estimates on the site variable for fallowing is consistent with the observation made above, that this reporting is better than expected under the guessing hypothesis for fallowing in Embu. Respondents in Embu have both overreported and underreported adoption of fallowing significantly less than respondents in Vihiga. This may at least partly be the result of a combination of actual adoption rates and ease of observation. The adoption rate for fallowing is less than 50% in both sites, but lower in Embu (22%) than in Vihiga (34%), and thus closer to the extreme. In addition, this practice is easy to observe. In contrast, respondents in Embu have been significantly less precise in their reporting of *tumbukiza* adoption than those in Vihiga, reflecting the opposite situation. *Tumbukiza* is difficult to observe, and adoption rates, while less than 50% in both sites, is much higher in Embu (38%) than in Vihiga (9%).

Next, we want to know the consequence of using proxy reported data on network behavior in regressions to make inferences about relationships between peer behavior and respondents' adoption choices. The standard hypothesis for such regressions is that respondent's own behavior is influenced by peer behavior. Our concern is that such effects may appear to be evident when regressions are based on proxy reported data, while they may be absent when regressions use self reported data on peer behavior. To demonstrate this difference, we have ran a series of regressions where alter adoption is either represented by a self reported or a proxy reported variable. Our null hypothesis is that proxy reported data and self reported data are drawn from the same distribution, and generate the same regression results, so it does not matter whether data were collected one way or the other. Table 4 reports summary statistics of data for primary respondents, which are the data used in these regressions.

Table 5 a) and b) report the results of two series of regressions of respondent's own adoption behavior on explanatory variables, including variables representing peer behavior and direct communication about the technology in question. One of these series of regressions uses self reported peer behavior (a), the other uses proxy reported peer behavior (b). We have dropped the regressions for organic fertilizers here, because of singularities due to the near universal adoption of this practice. In particular, we could not use the self reported alter adoption variable in regressions. At the bottom of Table 5 b), we report a test statistic for whether the adoption variables are significantly different. These test statistics were obtained by including both adoption variables in the same regressions, keeping all other variables unchanged, and test the difference between them. We see that the difference is significant for fallowing and *tumbukiza*, but not for terracing. The reason for this may be the nearly universal adoption of terracing, which means even guessing can produce relatively accurate data.

The regressions include variables representing ego characteristics like gender, age, education level, household size and farm size (area of cultivated land in acres), which are typically thought to be important correlates to technology adoption. Coefficient estimates on these variables are not important to our argument in this paper, so we will only comment them briefly in this discussion. The coefficient estimate on gender (being female) is positive and significant for terracing. However, when the respondent is female, it means that the household head is either female or absent. In the latter case, the household head is usually employed off-farm, and thus may be among the wealthier with the resources needed to invest in such costly structures. Coefficient estimates on the age variable suggest that older respondents are more likely to be adopters of the older, more established technologies, but not the new innovation being promoted by the extension service (*tumbukiza*). The uneducated are less likely to be adopters of practices that are associated with larger land holdings (terracing and fallowing), possibly a wealth effect, but they are more likely to be adopters of *tumbukiza*, perhaps reflecting that they are targeted by the extension service. Large households are less likely to adopt *tumbukiza*, maybe because such households focus their farm production on maize cultivation, where this practice is seldom used. Finally, coefficient estimates on area of land under cultivation is positive for both fallowing (for obvious reasons) and *tumbukiza* (maybe because larger holdings have more diverse production, including productions where *tumbukiza* is usually applied).

The contrast between the network variables in the two sets of regressions is striking. Where self reported alter adoption variables were used, their coefficient estimates are consistently not significant, while the coefficient estimates on having discussed the technologies are. This may reflect that current adopters are more interested than non-adopters in discussing these practices, either to reinforce their own experiences or to learn from others. But alter adoption does not appear to matter. In contrast, where proxy reported alter adoption variables were used, their coefficient estimates are consistently positive and highly significant, suggesting that respondents have imitated alters' behaviors. For terracing and *tumbukiza* the coefficient estimates on the discuss variable are also positive and significant, indicating an additional or independent effect of discussing the practices in question.

To summarize, we have found that our respondents do not report peer behavior accurately, and reports about peer behavior appear to be the result of guessing, but respondents appear to have information about general adoption rates, albeit rather imprecise information. Errors in proxy reporting are systematically related to variables commonly used in studies on social network effects on technology adoption, and using proxy reported data generates regression results that may exaggerate those sought-after social network effects.

Active Social Learning

An Information Search Experiment

Having explored passive learning through the strong ties of one's established social networks, we now study active learning, perhaps by way of "weak" social ties. Toward that end, we designed a simple experiment to try to induce active learning based on information search by the 55 individuals who served as primary respondents in the preceding section. Respondents were given a set of questions we did not expect they could answer off the top of their heads. We gathered this information locally through mass media and selected local informants.

Three alternative sets of 24 questions were used in the experiment, in order to prevent collaboration between neighboring respondents. The question sets consisted of eight categories, and within each category we asked for information at each of three levels: local, district, and national. For example, one category was prices of farm outputs. The local level question asked for the price farmers receive for eggs delivered to a specific merchant in the local farmers' market. The district level question asked for the price farmers receive for green tea leaves delivered to the district's tea factory. The national level question asked for the prices the district's tea factory achieved at the most recent national auctions. Although the local questions were tailored for each site, the questions were standardized for maximal comparability between sites.[4]

In order to give participants an incentive to invest in retrieving the requested information in timely fashion, we offered them a cash reward for each correct answer. Respondents were visited weekly. Rewards decreased the longer respondents took to find the right answer. They

[4] Complete questionnaires and question sets are included in Hogset (2005b).

were paid KShs 30 per correct answer[5] in the initial visit or the first revisit, so as to reward quick retrieval of accurate information. Payments declined by KShs10 per correct answer for each of the two subsequent revisits. Thus the cash reward participants earned was a function of how many correct answers they found and how quickly they found them. At the first visit each respondent was given a copy of the questions being asked and informed of the payment structure. The most any participant earned was KShs 660 out of a maximum possible of KShs 720. The lowest was KShs190. All correct answers were then followed up with questions about how the respondent found this answer, i.e., the source of the information, why the respondent chose this source, and if the source was a person, what sort of relationship the respondent had with this person and how difficult the person was to reach.

The experiment payoffs thus directly measure respondents' performance in accurately and quickly searching for information. We use these data to explore the characterizes of those most likely to engage in active learning. We then use the information on search methods used in finding correct answers to explore the role social relations played in active learning.

Characteristics of Active Learners

Table 6 presents basic descriptive statistics on the 55 respondents. We emphasize again that this is a small sample and thus we recommend interpreting our findings as merely suggestive. But the simple ordinary least squares regression of the payoffs earned in the information search experiment on participant characteristics yields very interesting and intuitive results (Table 7).

---

[5] The prevailing exchange rate at this time was roughly 72 Kenya shillings (KShs) per U.S. dollar, thus maximal winnings were roughly US$10. In this context, prospective winnings were substantial. The official Kenyan rural poverty line was Kshs 1239/month per person. Thus a respondent who answered all questions correctly in the initial visit or revisit would win cash equivalent to more than 17 times the daily individual poverty line.

There is a statistically significant and large difference between average rewards in the two research sites, with Embu residents earning more than those in Vihiga. This may be due to systematic differences in the level of difficulty between the question sets used in these sites or to timing differences (they were fielded in sequence, not simultaneously[6]), but may also be attributable to other differences between them. For example, the population in the two sites belong predominantly to different ethnic groups. In Embu respondents belong primarily to a sub-tribe within the Kikuyu, while in Vihiga they were predominantly Maragoli Luyhas. The Kikuyus are typically described as more entrepreneurial than other Kenyans, while the Luyhas are widely considered relatively conservative and tradition-bound.

Female respondents earned significantly less than male respondents. This may reflect gender roles that place a relatively heavier workload on women than men, making their effort in the exercise more costly, or less geographic or social mobility than men, making women's search for information more difficult. It is also possible that particularly valuable sources of information, such as highly educated persons, are concentrated in the networks of men. An age variable was also available for all but one participant, but this variable was not significant in any regressions, and its presence did not influence the estimates of other parameters much, so we omit it in the interest of conserving scarce degrees of freedom.

The field team got the impression that the busiest businessmen, i.e., persons with a high opportunity cost of time, put less effort into the information search exercise. Such respondents clearly expressed that they thought the reward for correct answers was low. On the other hand, low-income participants were very impressed with the amounts they could make in the experiment, and were observed to put considerable effort into information search. The regression results corroborate the hypothesis of a statistically significant inverse relation between

[6] The information search experiment was fielded in Embu first, then in Vihiga.

income or wealth (the brick house dummy variable) and the amount rewarded. The wealth effect is more precisely estimated, most likely because it reflects a more long-term measure of well-being, while the income variable, which was calculated using seasonal production and earnings data, only represents a recent realization of a highly volatile income stream. Nonetheless, in such a small sample, the magnitudes and relative precision of these estimates suggest a strong relation between wealth or income, reflecting the opportunity cost of respondents' time, and information search effort.

Respondents' human capital endowments also have an economically and statistically significant effect. Those who had not completed primary school earned significantly less than those who had, with no economically or statistically significant difference between those who had and had not completed secondary school as well. The basic literacy, numeracy and critical thinking skills typically achieved in primary school in Kenya seem sufficient to generate information search skills. The density of social networks also positively (and significant at the 10% level) affected earnings from information search, suggesting that social networks increase a person's social reach and reduce search costs. Interestingly, the parameter estimates on membership in a Savings and Credit Co-Operative (SACCO) and in a rotating savings and credit association (ROSCA) are both statistically significant, but with opposite signs. A SACCO is the most formal financial institution found in Embu, and is not available in Vihiga. A ROSCA (or "merry go round", as they are known locally) is a very informal financial institution found in both study sites. Moreover, it is equally available to both sexes and to people even with relatively low incomes, although not the very poorest (Hogset 2005a). A key difference is that SACCOs are relatively anonymous institutions in which deposits are unobservable and inaccessible to network contacts, while ROSCAs are communal savings arrangements organized among social network members with transactions readily observable to all. Thus, self-selection into one or the other may reveal something about the nature of relations within a person's social networks, in particular the degree of openness and trust. The signs on the parameter estimates are consistent with this interpretation.

To summarize, success in the information search experiment depends heavily on the opportunity cost of the participants' time spent searching, as reflected by respondent wealth, income and social networks, as well as on individual search ability, related to both human and social capital. These findings corroborate the simple heuristic model laid out in the opening section of this paper. They are consistent with the Schultzian idea that education will preferentially enable poorer farmers to acquire valuable information rapidly (Schultz 1975).

Successful Search Strategies and Social Networks

Optimal information search strategies depend on how difficult the information is to locate. If one assumes the desired information is known by someone within one's established network of "strong ties" – those with whom one interacts frequently – one may simply pepper those closest with questions in the expectation that, eventually, someone will yield the answer. However, finding less accessible information may require a more strategic approach. Rogers (1995) emphasizes this in expounding on the greater innovativeness within heterophilous networks (i.e., networks with considerable diversity of members) as compared to homophilous networks. This is also the essence of Granovetter's (1973) seminal "strength of weak ties" hypothesis.

So how did respondents successfully locate obscure factoids in rural Kenya? Table 8 presents the answers to questions asked of successful participants in the information search experiment. Three sample national-level questions (i.e., administered to respondents in both

Embu and Vihiga) appear at the top of the table. The figures beneath the questions report the percentage of (ultimately) correct answers received. For those who correctly answered the question, Table 8 then reflects their reported information source(s).

The first question (A) was the one with the lowest rate of success for all questions: the price of tea achieved by the local tea factory at recent auctions in Nairobi. About 45% of respondents never found these prices, even with three weeks and the promise of a cash reward as an incentive. This is surprising, given that both study sites are in tea growing districts, many of the respondents are themselves tea farmers, and national newspapers publish the tea auction prices weekly. Revealingly, those who found the correct answer relied on social information channels rather than impersonal, published reports. Instead, they contacted the management of the local factory, directly or indirectly through others in their social network.

In contrast, 76% managed to find currency exchange rates (question B), and 19% of those who did, found the information in newspapers, while 24% got it from TV or radio. People are plainly aware of mass media as sources of information, and most have access to them. Nonetheless, a majority still relied on social contacts to get the answer to a question widely known to be publicly available in the media. The third question (C) pertains to health recommendations promoted by the Government in campaigns that make heavy use of electronic media. About 91% of the participants got these answers right. More than one-third of respondents already knew the answer to this question before they were asked, suggesting these health messages are disseminated quite effectively, even to relatively poor rural households. Of those who needed to search out the correct answer, however, social contacts once again played a central role, accounting for two-thirds of the sources of correct answers for those did not already know it.

Two different search strategies emerge from the data. Among the 30 participants who found the answer to question A, 47% had asked a family member or another person with whom they have a self-described “very close” relationship and 60% said they chose the person they asked due to the nature of their relationship, as opposed to the qualifications of that person. But these closely related persons were generally not themselves sources of the information. Rather, they commonly served as intermediaries, asking competent persons on behalf of the participant. The same pattern is found for question B, but to a lesser degree for question C, the health campaign question for which detailed information has been quite widely broadcast. For the latter, most of the family members were themselves sources of the information, so there was no need to contact “experts” or for the family members to serve as intermediaries.

Alternatively, some participants contacted “experts” themselves, without assistance from intermediaries. Among those who got question A right, 23% had contacted the tea factory themselves, 33% had talked to someone they chose due to where that person was employed, 20% had talked to a total stranger, and 20% had to travel to meet their informants. These figures are probably overlapping, describing the same search events, although a few of the persons who had been approached due to where they worked may have been among the intermediaries. Many of those who asked family members or close others to serve as intermediaries probably did so for purely practical reasons, e.g., that person was going to the store or to town anyway and could more easily contact a socially distant expert for the correct information.

These results suggest that having a network of close friends and family members may increase a person’s social reach, not just by increasing the probability that someone within one’s network can provide sought-after information, but perhaps especially by increasing the probability that one can successfully identify and contact outside persons, through a trusted intermediary if necessary. The network contacts serve an important role as intermediaries or messengers, thereby reducing information search costs associated. But the key to getting accurate, timely information appears to be access to well-informed individuals, many of whom one will have at best infrequent contact with, i.e., Granovetter's (1973) "weak ties".

This pattern of mobilizing social networks to reach out to reliable information sources outside one's network of regular "strong ties" has implications for inference with respect to social learning. For example, models that regress some measure of technology adoption by a respondent on the density of that individual's social network or the frequency of adoption within one's social network may be capturing not (or not only) direct learning from the socially proximate other – as posited by standard "learning from others" models (e.g., Foster and Rosenzweig 1995), but rather (or in addition) the effect of access to information from distant experts (an omitted variable), as facilitated by both one's social network density and one's network's familiarity with the relevant experts. The apparent instrumental use of social networks to access information from unknown or socially distant sources and the possibility of omitted relevant variables bias it suggests, just like the reflection problem reported in the previous section, suggests that the magnitude of social learning from one's network members may be overstated in the existing development economics literature. This conjecture merits further detailed investigation as it carries important implications for the design of information dissemination campaigns, such as those associated with diffusion of new agricultural or health technologies to poor rural households.

Discussion and Conclusions

A basic premise in much contemporary research on farmers' choice between learning by doing and learning from others (e.g., Foster and Rosenzweig 1995) is that the former involves costly experimentation while the latter is costless, but relatively less accurate. In this paper, we refine this distinction, by unbundling learning from others into two distinct processes, passive and active learning, where the former is costless, reflecting information picked up through routine interactions with members of one's social network, and the later is costly, involving purposeful search for information, directly or often indirectly through intermediaries, and commonly from individuals with whom one has at best infrequent contact. We introduce a simple model to describe this multi-path process of social learning, in which the rate of learning may differ according to whether individuals learn passively or actively, and in which the optimal approach may depend on individual attributes such as the opportunity cost of time, educational attainment, and the density of one's social network. We then use original data collected among farmers in Kenya's highlands to test hypotheses associated with this distinction between active and passive learning and between the weak and strong ties that feature prominently under these different pathways for accessing information.

Passive learning may be costless in terms of effort, but it may also be highly inaccurate and subject to bias. The magnitude of both reporting errors and bias may depend on observer characteristics as well as properties of the information in question, such as ease of observation. Active learning from others requires effort, and is therefore costly, but may generate knowledge that is more accurate. The effort required to achieve these benefits may depend on learner characteristics, such as education and the opportunity cost of one's time, as well as the density of one's social network that can intermediate information retrieval through interpersonal channels.

Analysis of our data revealed that our respondents have only a rough idea about the NRM practice adoption behavior of those in their social networks. While they could not accurately identify who specifically in their networks adopted the NRM practices in question, they had a sense of the overall adoption rate within the broader population. We interpret this as evidence that people monitor broader technology adoption patterns in the population, but do not attach a high value to knowing precisely *who* the adopters are, even among those closest to them. Passive learning and proxy reporting seem associated with high rates of erroneous reporting, just as our model predicts.

In our data, respondents' beliefs about the behavior of their peers were systematically biased, depending especially on respondent's own adoption choices (i.e., a "reflection bias" problem) and on whether the respondent and network match had previously discussed the NRM practice in question. In regression analysis of respondents' NRM practice adoption patterns, the use of proxy reported data on network members' NRM adoption behaviors yields economically and statistically significantly different estimates of social learning than does use of network members' self-reported behaviors. These results illustrate that using proxy reported data about peer behavior in regression analysis may lead to seriously mistaken inference.

Our information search experiment among the same rural Kenyan population demonstrated quite clearly that active learning is indeed a costly activity that requires time, effort, and investment, often in terms of traveling or engaging intermediaries. Successful information acquisition is inversely associated with participants' opportunity cost of time, as reflected by the wealth proxy, income level and gender. Human capital helps considerably as well. Those who had not completed primary school (which generally signals illiteracy in our study site) earned significantly lower returns in the information search experiment. Social networks also prove helpful for active learning. The value of social networks appears less as a direct source of information, however, than for intermediation to reach sources of reliable information that the respondent did not know or had difficulty reaching. The larger the network, the greater the number of people available to provide such assistance, and the higher the payoff, on average. Overall, this paper raises questions concerning prevailing economic models of and empirical results related to social learning, which implicitly emphasize passive learning through strong social ties, commonly measured through proxy reporting of network members' behavior. Given the simple analytical model we develop and the small sample size we use, we emphasize that our findings are by no means definitive, but rather suggestive that researchers give more explicit thought to the processes of social learning that appear to guide information flows and technology adoption patterns in low-income communities. In particular, learning responds to incentives, so researchers need to identify the incentives for and the costs of learning. Learners actively and selectively choose from whom they seek information, thus researchers need to study endogenous network activation patterns more carefully (Santos and Barrett 2005). Finally, social networks are demonstrably helpful for learning, both through strong and weak ties, so research on social capital will benefit from looking at both types of ties.

While attempts to strengthen poor villagers' social networks may be helpful in stimulating information flow and technology uptake, a singular emphasis on strong social ties may be misplaced. Instead, individuals and groups need to learn how to acquire and utilize information that is often available mainly through weak social ties, in particular people who serve as "experts", such as extension officers and other key change agents. Proper identification of the value of expert or mass media information delivery will depend fundamentally on the development and estimation of properly specified models of social learning.

References

Armantier, Olivier. 2004. "Does Observation Influence Learning?" *Games and Economic Behavior* 46: 221-39.

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

Behrman, Jere, Hans-Peter Kohler and Susan Cotts Watkins. 2002. "Social Networks and Changes in Contraceptive Use over Time: Evidence from a Longitudinal Study in Rural Kenya." *Demography* 39 (November): 713-38.

Conley, Timothy G. and Udry, Christopher R. 2001. "Social Learning Through Networks: The Adoption of New Agricultural Technologies in Ghana." *American Journal of Agricultural Economics* 83 (3), 668–673.

Conley, Timothy G. and Christopher R. Udry. 2005. "Learning About A New Technology: Pineapple in Ghana." Yale University working paper.

Drechsel, Pay, Lucy Gyiele, Dagmar Kunze, and Olufunke Cofie. 2001. "Population Density, Soil Nutrient Depletion, and Economic Growth in Sub-Saharan Africa." *Ecological Economics* 38: 251-58.

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: 1176-209.

Granovetter, Mark. 1973. "The Strength of Weak Ties." *American Journal of Sociology* 78 (May): 1360-80.

Hogset, Heidi. 2005a. "Social Exclusion and Insurance Failure for the Poorest: On Informal Finance through Social Networks in Kenya's Smallholder Sector." Paper presented at the

ISSER (Institute of Statistical, Social and Economic Research)/ Cornell University/ World Bank Conference on "Shared Growth in Africa" in Accra, Ghana, July 21-22, 2005.

Hogset, Heidi. 2005b. "Social Networks and Rural Development: Theory and Applications in the Kenyan Highlands". Ph.D. dissertation. Ithaca, NY. Cornell University.

Kamar, Margaret J., Joseph K. Mburu, and Donald B. Thomas. 1999. "The Role of Incentives in Soil and Water Conservation in Kenya". In *Incentives in Soil Conservation; from Theory to Practice*, edited by Thomas Enters. Enfield, NH: Science Publishers, Inc.

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

Miguel, Edward and Kremer, Michael. 2003. "Networks, Social Learning, and Technology Adoption: The Case of Deworming Drugs in Kenya." University of California, Berkelye Center for Labor Economics Working Paper no. 61.

Montgomery, Mark R. and John B. Casterline. 1996. "Social Learning, Social Influence, and New Models of Fertility." *Population and Development Review* 22: 151-75.

Munshi, Kaivan. 2004. "Social learning in a heterogeneous population: Technology diffusion in the Indian Green Revolution." *Journal of Development Economics* 73(2): 185-213.

Mwakubo, Samuel M., Gideon A. Obare, John Omiti and Lutta Mohammed. 2004. "The Influence of Social Capital on Sustainable Agriculture in Marginal Areas, Kenya: A Case Study of Machacos and Taita-Taveta Districts." Research report presented to the International Food Policy Research Institute (IFPRI).

Rogers, Everett M. 1995. *Diffusion of Innovations*. New York, NY: The Free Press.

Santos, Paulo and Christopher B. Barrett. 2005. "Interest and Identity in Network Formation: Who Do Smallholders Seek out for Information in Rural Ghana?" Cornell University working paper.

Schultz, T.W., 1975. "Value of ability to deal with disequilibria". *Journal of Economic Literature* 13(3), 827-846.

Table 1: The accuracy of proxy reporting

(Percent)

		Terracing		**Fallowing**		**Organic fertilizers**		***Tumbukiza***	
Row	**Rate**	**Vihiga**	**Embu**	**Vihiga**	**Embu**	**Vihiga**	**Embu**	**Vihiga**	**Embu**
a	Correct positive	81.25	71.34	14.73	7.03	91.75	87.71	0.50	14.56
b	Correct negative	2.08	2.49	35.39	63.79	0.24	0.78	87.62	36.66
c	False positive	5.79	10.86	30.17	13.79	5.19	2.22	3.47	25.34
d	False negative	10.88	15.31	19.71	15.38	2.83	9.28	8.42	23.45
a+b	Accuracy	83.33	73.82	50.12	70.82	91.98	88.50	88.12	51.21
c+d	Error	16.67	26.18	49.88	29.18	8.02	11.50	11.88	48.79
c-d	Bias	-5.09	-4.45	10.45	-1.59	2.36	-7.06	-4.95	1.89
a+d	α (population adoption)	92.13	86.65	34.44	22.41	94.58	96.99	8.91	38.01
	t-statistic for test of H_0:f= α	-1.28	-1.99**	-1.95*	3.23***	1.52	-6.69***	2.37**	-0.91

*, **, *** statistically significant difference at 10%, 5% and 1% level, respectively.

H_0 under the maintained hypothesis that correct responses (x) are distributed Binomial(n, p).

Table 2. Descriptive statistics for snowball sample

Binary variables (0 = No, 1 = Yes)

Variable	**# obs**	**Mean**	**Std. Dev.**
Embu site	1222	0.64	0.48
Alter is a family member	1221	0.22	0.42
Ego (respondent) practices/uses:			
Terracing	1216	0.91	0.29
Fallowing	1220	0.27	0.44
Organic fertilizers	1178	0.95	0.21
Tumbukiza	1220	0.28	0.45
Ego and alter have discussed the technology:			
Terracing	1221	0.80	0.40
Fallowing	1215	0.37	0.48
Organic fertilizers	1202	0.83	0.37
Tumbukiza	1205	0.31	0.46

Table 3. Patterns in proxy reporting errors.

(Multinomial Probit estimation. No error is the base outcome)

	Terracing		**Fallowing**		**Organic fertilizers**		***Tumbukiza***	
Error	**Negative**	**Positive**	**Negative**	**Positive**	**Negative**	**Positive**	**Negative**	**Positive**
Alter is a family member	0.128	-0.179	-0.106	0.058	0.210	-0.309	-0.006	-0.240
	(0.160)	(0.184)	(0.164)	(0.147)	(0.214)	(0.283)	(0.166)	(0.176)
Ego has adopted technology	-0.401	-0.242	-0.032	0.846	-0.327	0.199	-0.171	0.583
	(0.217)*	(0.266)	(0.159)	(0.140)***	(0.462)	(0.568)	(0.171)	(0.167)***
Ego and alter have	-1.292	-0.013	-0.246	1.070	-1.732	-0.298	-0.622	1.468
discussed technology	(0.161)***	(0.202)	(0.151)	(0.136)***	(0.194)***	(0.279)	(0.172)***	(0.158)***
Embu site	0.204	0.511	-0.458	-0.698	0.874	-0.471	1.488	1.075
	(0.152)	(0.171)***	(0.139)***	(0.135)***	(0.235)***	(0.210)**	(0.165)***	(0.215)***
Constant	-0.341	-1.782	-0.670	-1.376	-1.283	-2.137	-1.965	-2.700
	(0.232)	(0.363)***	(0.132)***	(0.147)***	(0.458)***	(0.687)***	(0.137)***	(0.202)***
Number of observations		1189		1173		1164		1144
Wald chi-squared(8)		101.62***		169.93***		93.08***		305***

Robust standard errors in parentheses. *, **, *** statistically significant difference at 10%, 5% and 1% level, respectively

Table 4. Descriptive statistics for primary respondents

a) Binary variables (0 = No, 1 = Yes)

Variable	**# obs**	**Mean**	**Std. Dev.**	**Variable**	**# obs**	**Mean**	**Std. Dev.**
<u>Respondent (ego) practices/uses</u>				<u>Ego and alter have discussed</u>			
Terracing	654	0.94	0.23	Terracing	654	0.79	0.41
Fallowing	654	0.33	0.47	Fallowing	647	0.35	0.48
Org. fert.	632	0.93	0.25	Org. fert.	654	0.83	0.37
Tumbukiza	654	0.30	0.46	*Tumbukiza*	636	0.31	0.46
<u>Alter practices/uses (proxy reported)</u>				<u>Alter practices/uses (self reported)</u>			
Terracing	630	0.78	0.41	Terracing	523	0.90	0.30
Fallowing	613	0.30	0.46	Fallowing	523	0.29	0.46
Org. fert.	635	0.89	0.31	Org. fert.	516	0.97	0.16
Tumbukiza	587	0.21	0.40	*Tumbukiza*	523	0.27	0.44

Variable	**# obs**	**Mean**	**Std. Dev.**
Female	654	0.50	0.50
Not completed primary school	654	0.23	0.42

b) Continuous variables

Variable	**# obs**	**Mean**	**Std. Dev.**	**Min**	**Max**
Age	612	55.29	13.12	25	80
Household size	654	5.27	2.17	1	12
Area of cultivated land (acres)	654	2.28	2.61	0.2	15

Table 5. NRM adoption and peer behavior, proxy or self-reported

(Probit estimation)

a) Self reported peer behavior	**Terracing**	**Fallowing**	***Tumbukiza***
Female	0.484	0.109	0.061
	(0.174)***	(0.150)	(0.153)
Age	0.016	0.019	-0.015
	(0.006)***	(0.006)***	(0.005)***
Not completed primary school	-1.116	-0.337	0.447
	(0.191)***	(0.170)**	(0.182)**
Household size	0.047	0.013	-0.145
	(0.039)	(0.030)	(0.033)***
Area of cultivated land (acres)	0.017	0.211	0.197
	(0.024)	(0.041)***	(0.030)***
Ego and alter have discussed	1.044	0.434	1.093
the technology	(0.217)***	(0.135)***	(0.140)***
Alter has adopted the technology	0.327	0.064	0.055
	(0.303)	(0.156)	(0.163)
Constant	-0.434	-2.055	-0.239
	(0.563)	(0.456)***	(0.403)
Number of observations	509	502	492
Wald chi-squared(7)	87.99***	43.79***	172.75***

b) Proxy reported peer behavior	**Terracing**	**Fallowing**	***Tumbukiza***
Female	0.478	0.165	0.171
	(0.176)***	(0.139)	(0.148)
Age	0.007	0.017	-0.014
	(0.006)	(0.005)***	(0.005)***
Not completed primary school	-1.238	-0.260	0.487
	(0.180)***	(0.155)*	(0.161)***
Household size	0.055	0.026	-0.102
	(0.039)	(0.028)	(0.030)***
Area of cultivated land (acres)	0.029	0.195	0.193
	(0.025)	(0.033)***	(0.023)***
Ego and alter have discussed	0.802	0.034	0.527
the technology	(0.242)***	(0.143)	(0.155)***
Alter has adopted the technology	1.066	0.616	1.123
	(0.243)***	(0.154)***	(0.172)***
Constant	-0.125	-2.078	-0.568
	(0.506)	(0.427)***	(0.391)
Number of observations	613	596	572
Wald chi-squared(7)	95.18***	57.19***	239.67***
<u>Testing self vs. proxy reported adoption variables</u>			
chi-squared(1)	2.18	6.73***	14.19***

Table 6. Descriptive statistics for participants of information search experiment

Variable	**# obs**	**Mean**	**Std. Dev.**	**Min**	**Max**
Continuous variables					
Amount rewarded in weak ties experiment	55	496.18	105.15	190	660
Household daily per-capita income (KShs)	55	192.18	163.14	6.48	704.94
Density of respondent's network	55	9.96	3.02	3	17
Binary variables (0 = No, 1 = Yes)					
Embu site	55	0.58	0.50		
Female	55	0.51	0.50		
Brick house	55	0.04	0.19		
Not completed primary school	55	0.24	0.43		
Has completed secondary school	55	0.15	0.36		
Member of SACCO	55	0.49	0.50		
Member of ROSCA	55	0.42	0.50		

Table 7. Information search experiment payoffs

(Estimated by ordinary least squares)

	Coefficient est.		Coefficient est.
Embu site	277.977	Has completed	19.75
	(52.664)***	secondary school	(29.483)
Female	-53.782	Density of respondent's	6.713
	(22.783)**	network	(3.801)*
Household daily per-	-0.146	Member of SACCO	-142.487
capita income (KShs)	(0.081)*		(45.778)***
Brick house	-186.606	Member of ROSCA	74.3
	(57.289)***		(22.195)***
Not completed primary	-143.98	Constant	399.87
school	(32.685)***		(39.332)***
Number of observations	55		
F(9, 45)	7.42***		

Standard errors in parentheses.

*, **, *** statistically significant difference at 10%, 5% and 1% level, respectively

Table 6. Descriptive statistics for participants of information search experiment

Variable	**# obs**	**Mean**	**Std. Dev.**	**Min**	**Max**
Continuous variables					
Amount rewarded in weak ties experiment	55	496.18	105.15	190	660
Household daily per-capita income (KShs)	55	192.18	163.14	6.48	704.94
Density of respondent's network	55	9.96	3.02	3	17
Binary variables (0 = No, 1 = Yes)					
Embu site	55	0.58	0.50		
Female	55	0.51	0.50		
Brick house	55	0.04	0.19		
Not completed primary school	55	0.24	0.43		
Has completed secondary school	55	0.15	0.36		
Member of SACCO	55	0.49	0.50		
Member of ROSCA	55	0.42	0.50		

Table 7. Information search experiment payoffs

(Estimated by ordinary least squares)

	Coefficient est.		**Coefficient est.**
Embu site	277.977	Has completed	19.75
	(52.664)***	secondary school	(29.483)
Female	-53.782	Density of respondent's	6.713
	(22.783)**	network	(3.801)*
Household daily per-	-0.146	Member of SACCO	-142.487
capita income (KShs)	(0.081)*		(45.778)***
Brick house	-186.606	Member of ROSCA	74.3
	(57.289)***		(22.195)***
Not completed primary	-143.98	Constant	399.87
school	(32.685)***		(39.332)***
Number of observations	55		
F(9, 45)	7.42***		

Standard errors in parentheses.

*, **, *** statistically significant difference at 10%, 5% and 1% level, respectively

Table 8: Search strategies of successful participants

(Percent)

Questions:

A. What is the highest price per 50-kg bag of tea class BP1 (Broken Pekoe 1) from [name of the local] Tea Factory fetched in the most recent auctions (US dollars)?

B. What is the CBK (Central Bank of Kenya) selling price for 1 US dollar/ 1 Euro/ 1 SA Rand?

C. What is the rule of thumb used as recommendation when people should be checked for tuberculosis? / How can you reduce the number of mosquitoes on your compound, and thereby reduce the probability of transmission of malaria? / What are the four rules recommended to go by to avoid transmission of HIV?

	A	**B**	**C**
Successful participants, incl. those who guessed or already knew, out of the 55 who participated	54.55	76.36	90.91
How did you get this information?			
Through TV or radio		23.81	14.00
Through newspapers, magazines, or other printed media		19.05	2.00
By contacting the shop/institution/etc. in question	23.33	2.38	6.00
By asking members of your family	46.67	30.95	24.00
By asking people outside of your family	26.67	23.81	18.00
You guessed or already knew it	3.33	2.38	36.00

(Note: columns do not have to sum to 100.0 due to rounding or multiple responses.)

If you asked someone, why did you choose this informant? Was it due to ..			
..where the person is employed	33.33	4.76	14.00
..the kind of education/skills this person has		11.90	
..who this person is known to have contact with	3.33	2.38	2.00
..the nature of your relationship (close friends, etc.)	60.00	38.10	30.00
Other			2.00
Not applicable or no answer	3.33	42.86	52.00
How will you characterize your relationship and level of contact with this informant?			
Very close - interact casually daily or frequently	46.67	30.95	24.00
Somewhat close - interact casually infrequently	10.00	7.14	8.00
Somewhat distant - interact only intentionally, not casually	3.33	9.52	8.00
Distant - interact only incidentally	16.67	4.76	6.00
Total strangers - never met before	20.00	4.76	2.00
Not applicable or no answer	3.33	42.86	52.00
To get in touch with this informant, did you need to do any of the following:			
Travel	20.00	4.76	8.00
Get an appointment			2.00
Use an intermediary to assist you	60.00	35.71	14.00
Not applicable or none of the above	20.00	59.52	76.00

Figure 1: Four natural resources management practices

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="2">Ease of Observation</td></tr>
<tr><td>Easy</td><td>Difficult</td></tr>
<tr><td rowspan="2">Prevalence</td><td>Widespread</td><td>Terracing</td><td>Organic fertilizer application</td></tr>
<tr><td>Limited</td><td>Fallowing</td><td>Tumbukiza</td></tr>
</table>