

Identity, Interest and Information Search in a Dynamic Rural Economy *

Paulo Santos[†] Christopher B. Barrett[‡]

October 2008

Abstract

This paper examines how farmers in a developing country search for information. Using data on farmers' decisions to connect with other farmers in order to ask for information about different agricultural problems, we explore the role played by identity in accessing information. We show that farmers target different subsets of acquaintances when searching for different types of information, supporting the view that identity cannot be the main determinant of such decisions.

*This is a much revised version of a paper that circulated under the title "Interest and Identity in Network Formation and Activation". We thank J. S. Butler, Andrew Foster, Rachel Kranton, Chris Udry, and conference participants at the 2004 American Agricultural Economics Association annual meeting, the Pew Workshop on "Social Dynamics and the Microeconomics of Poverty" held in Bellagio in April 2005 and seminar participants at Cornell University and the University of Oxford for comments that greatly improved previous versions of this paper. Santos thanks the Fundação Calouste Gulbenkian and the Fundação para a Ciência e Tecnologia for financial support and the Centre for the Study of African Economies at the University of Oxford for the hospitality during the last stages of writing this paper. Barrett thanks The Pew Charitable Trusts, through the Christian Scholars Program of the University of Notre Dame for financial support. The usual disclaimers apply.

[†]Lecturer, Agricultural and Resource Economics, University of Sydney. Correspondence Address: 107 Watt Building, Science Road, Camperdown, NSW 2006, Australia. Email: p.santos@usyd.edu.au

[‡]Professor, Applied Economics and Management, Cornell University.

“In 88 BC, King Mithriadates VI of Pontus invaded Roman territory in Asia Minor. He encouraged Asian debtors to kill their Roman creditors. Happy to reduce their credit card bills, the Asians massacred 80,000 Romans.”¹

Identity – the vector of an individual’s identifiable characteristics that are irrelevant to his or her productivity – is an empirically elusive concept: In the citation above, did the “Asians” kill the “Romans” or did the “debtors” kill the “creditors”? What motive ruled, identity or the material interests of the debtors?¹ The difficulty in answering questions like this has long been recognized in the economic literature on discrimination: what one interprets as effects associated with identity may merely reflect some unobserved, payoff-relevant characteristic that is correlated with observable indicators of identity but that is not itself easily measured by the outsider (Heckman, 1998). Identity almost surely matters; but its effects on individual behavior can be too easily exaggerated.

These frictions acquire special salience in the economics literatures on social learning, social insurance, the provision of public goods, conflict and other phenomena where social context is not a second-order consideration. The recognition that interactions between people matter fundamentally to such processes prompted the interest in understanding how people choose the networks to which they belong and, in particular, whether identity can be an impediment to exchange and growth. For example, Munshi and Myaux

¹For the remainder of this article, we’ll omit the reference to the “material” nature of the interest motive under analysis, as only this narrower definition allows us to separate interest from identity.

(2006) use shared religious affiliation to understand the diffusion of contraceptive practices, while Romani (2003) argues that ethnic differences act as barriers to information flows in Côte d’Ivoire. Miguel (2004) suggests that the higher provision of local public goods in one district in Tanzania, when compared with a similar one in Kenya, can be explained by Tanzanian nation-building politics that helped overcome ethnic differences that, by contrast, were exacerbated in Kenya (see also Miguel and Gugerty, 2005). Finally, Easterly and Levine (1997) find that ethnic heterogeneity is one of the causes of Africa’s growth tragedy, leading to subsequent standard use of a measure of ethnic diversity as a regressor in cross-country growth studies. Such behaviors are usually interpreted as typical of pre-modern societies. Indeed, one of the classical economists’ arguments in favor of capitalism was its ability to replace “hot passions” with the “cold feelings” of trade (Hirschman, 1997).² If differences in identity pose real obstacles to the conduct of interpersonal exchange – whether of information or real resources – then these sociocultural issues are of first-order importance in questions of economic development.

This article exploits an unusual data set to ask whether the behavioral importance of individuals’ identity has been somewhat exaggerated in the literature. We follow a strategy somewhat akin to that of Bertrand and Mullinathan (2004), who confronted employers with resumé of job applicants

²Although a more recent literature argues that the political relevance of ethnic divisions is a result of growth, reversing the causality argument just identified (Bates, 2000, Miguel and Posner, 2006)

that were identical except for the individuals' names, which clearly signalled different ethnic origins (along racial lines). They found significant differences in the resulting calls to interview associated with these identity differences, interpreted as discriminatory behavior given the absence of any other systematic difference among the candidates. In this study, we employ a complementary strategy, holding identity constant while varying the prospective payoffs of an interpersonal link. More precisely, we explore how farmers in a developing country construct the networks through which valuable information may flow on any of several different issues. By asking our respondents whether they would contact specific others to gain access to different sorts of information, we can test not only whether differences in identity significantly affect their choices but, more robustly, whether these effects are stable across different types of information. As such this article not only adds to a small empirical literature that addresses how people choose with whom to connect,³ it also explores the relative importance of identity to individual behavior in a manner no other study has been able to do.

The remainder of the article proceeds as follows. After a brief conceptual discussion of identity and social distance (section 1), we present the data in section 2, show that the choice of a network link for a particular purpose is embedded in the set of individuals' pre-existing social acquaintances, and

³For example, DeWeerd (2004) uses data on the set of links within a Tanzanian village to explain who insures with whom. Fafchamps and Gubert (2007) study the formation of insurance networks in the Philippines, while Udry and Conley (2005) explore the formation of the networks that shape access to information, credit, land and labor in rural Ghana.

discuss the econometric specification. Section 3 discusses our econometric results and shows that while identity certainly appears to play some role in people’s decisions to interact with others when searching for valuable information, material self-interest guides the choice of social acquaintances for instrumental purposes. In particular, we show that farmers use different sets of social contacts in pursuit of different types of information, clearly signalling that information search is not primarily determined by identity. Section 4 concludes.

1 Network formation and identity: A conceptual discussion

The theoretical literature on network formation (see Jackson (2007) for a recent review) approaches the question of how people choose which links to form with each other by assuming that agents perform a cost-benefit analysis of each link

$$L_{ijk} = \begin{cases} 1 & \text{if } B(X_{ij}) - C(Y_{ij}) \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where $B(\bullet)$ represent the benefits of a link between i and j with respect to activity k (L_{ijk}), $C(\bullet)$ represent the costs of forming such a link, X_{ij} and Y_{ij} are prospectively distinct (or identical) matrices of characteristics of individuals i and j , and $L_{ijk} = 1$ if a link is formed and 0 if it is not.⁴ Based on this frame-

⁴This literature then quickly turns to explore how network architectures are chosen and the welfare effects of such choices. Such explorations extend well beyond our purposes. In

work, the social networks literature has analyzed extensively the importance of homophily (see McPherson, Smith-Lovin, and Cook, 2001, for a recent review), following early experimental results in psychology and sociology that showed that attraction and interaction are affected by perceived similarity, a result popularized in the proverb “birds of a feather flock together”. This concept is powerful and, at least implicitly, widely used as it justifies why, in the absence of more detailed data, one may think of groups defined by shared individuals characteristics (e.g., age, ethnicity, gender, religion) as social units. It also links to the contemporary economics literature on identity, most notably with the work of Akerlof and Kranton (2000). Their concept of identity effectively extends the notion of *homo economicus* by augmenting his utility function with an exogenous vector of identity-based “roles” (for example, gender). Society (exogenously) establishes “prescriptions” (for example, men do not wear skirts) that bias preferences in favor of certain actions.⁵

Implicit in this approach is an understanding of such roles as *intrinsically* irrelevant to one’s productivity, similar to Arrow (1972) analysis of the economics of discrimination. This requirement is essential in order to escape tautological interpretations of the importance of identity and clearly defines

our empirical application we focus only on dyad choice and are unable – and therefore do not try – to quantify the welfare effects of differences in access to information.

⁵By accepting exogenous identities, we rule out the idea of individual agents as relationally constructed, in which case networks can be “constitutive” in the sense that they shape agents’ identities and their capacities to act. As argued by Rauch and Hamilton (2001), this is a perspective that is harder to integrate within economics.

how this concept can augment – instead of simply rephrasing – the analysis of a variety of individual behaviors.⁶ Yet it is terribly difficult to maintain in empirical analysis. Take for example the discussion of the importance of identity for the economics of education by Akerlof and Kranton (2002). That, in this model, students have to choose their social categories (taking into account their own characteristics and the ideal characteristics of the category to which they would like to belong, students can choose between leading crowd, nerds or burnouts) *and* their effort level (by considering the ideal behavior of the chosen category) offer a good example of how productivity-irrelevant characteristics can end up inexorably mixed with productive ones. To unbundle the consequences of identity and of productivity-relevant characteristics in the functioning of the labor market, in variables such as getting a specific wage, would require very detailed data, covering all productivity relevant characteristics, in order to argue that identity did play a role in such decisions (Heckman, 1998). *Mutatis mutandis*, the same is true in an analysis of who contacts whom when searching for information. In the absence of such implausibly detailed observational data, alternative strategies have to be sought. We offer one such strategy in this article.

The final conceptual issue we must address concerns how to measure the effect of identity (or any other variable, for that matter) when explaining

⁶The classical definition of *homo economicus* specifies “Every member of the society is to act as an individual only, in entire independence of all other persons. To complete his independence he must be free from social wants, prejudices, preferences, or repulsions [...]” (Knight (1921, p. 78) *cit in* Swedberg (2004)).

the formation of dyadic relations. As made clear by others (for example, Jenkins, 1997, Kirman and Teschl, 2004) the concept of identity adopted by economists is that of social identity and is inherently relational: it tells us where, in the social space (conceived as a set of exogenous marks or “roles”, to use the terminology of Akerlof and Kranton (2000)) do agents place themselves, with one’s position necessarily measured with respect to the position of others.

This suggests that, by analogy with physical distance, one can express the effect of differences between two individuals in terms of social distance (Akerlof, 1997). The parallelism may have been taken a bit too far, however, in the empirical literature that addresses the question of who connects with whom. By measuring social distance as a Euclidean norm to capture the effect of differences in observable characteristics between network partners, one implicitly imposes the restriction that the effect of differences between the two parties to a link is symmetric. For example, this assumes that the prescriptions in place make it equally likely for a man to contact a woman for information about farming practices as vice versa, an assumption that seems unrealistically restrictive.

We therefore measure social distance through a simple modification of the Euclidean norm, using indicator variables to reflect ordinal relations between the pair.⁷ Let x be any of the continuous variables on which information is

⁷After developing this refinement and presenting an earlier version of this article, we discovered that Fafchamps and Gubert (2007) employ a similar refinement. An earlier use of a similar approach, but limited to one variable - gender - can be found in Burger,

available. We then measure the social distance between two agents i and j by the following two variables:

$$I_{(x_i - x_j < 0)} \times |x_i - x_j| + I_{(x_i - x_j \geq 0)} \times |x_i - x_j| \quad (2)$$

where $I_{(\bullet)}$ is an indicator function taking value one if true, zero otherwise: the effect of age, for example, would be expressed by the interaction between the absolute value of the difference in age with two indicator variables, older and younger, of which, for each dyad, at least one has to be zero. For categorical variables, social distance is defined by a set of dummy variables that consider the several possible characterizations of the match: hence the analysis of the effect of gender requires the definition of a dummy variable for each of the four possible combinations (male–male, male–female, female–male, female–female).

2 Data and econometric questions

We use data on economic activities and social interactions between people living in four villages in southeastern Ghana. This region has a long tradition of commercial agriculture and, in the early 1990s, initiated a process of conversion from cassava and maize production, directed towards domestic markets, into pineapple production, directed to European export markets. The transition in crops and markets brought with it new inputs (primarily,

Collier, and Gunning (1993).

inorganic fertilizers) and a process of differentiation between those farmers who adopted the new crop (mainly, male and wealthier farmers) and those who did not. Udry and Conley (2005, p.246) summarize this dynamic by saying that “economic development in this region is being shaped by the networks of information, capital and influence that permeate these communities”.

The data, discussed in Goldstein and Udry (1999), include detailed information on different characteristics of the respondents and their households (socio-demographic variables, assets, agricultural and non-agricultural production, etc.) and, more unusually, information on respondents’ patterns of personal interaction, in particular when searching for information.⁸ Recent work on the role of learning and information diffusion in rural Africa – namely about the productivity of fertilizer application (Conley and Udry, 2005) – build on such data.

The data on information networks that we use was experimentally collected in the following way. First, respondents were matched with seven individuals randomly chosen from the sample.⁹ Then they were asked about their acquaintance with that person (through the question “Do you know . . . ?”), establishing a random sample of the respondents’ set of social

⁸The data and the survey instruments are available at <http://www.econ.yale.edu/cru2/ghanadata.html>.

⁹The respondents were also non-randomly matched with three individuals considered to be focal in their village (a farmer in the survey and two other persons not previously surveyed). We do not use data on those three prospective matches, both due to its non-random nature and the fact that information is generally not available on most of the focal individuals, making the sort of analysis we perform impossible.

acquaintances (hereafter, social networks). Finally, they were asked about whether they would contact the match when faced with some specific problem, through the following questions (our emphasis):

“Could you go to ... for advice about when to *apply a new kind of fertilizer?*”

“Could you go to ... if you wanted to *discuss changing your method of planting?*”

“Could you go to ... if you wanted to *find a buyer for any of your crops?*”

These questions identify information links (hereafter, information networks), that is, those individuals chosen as a source of information needed to solve a particular problem.¹⁰

This data collection strategy sheds unusual light on the question of where in one’s social space people search for information. By asking about social networks and then about information networks, and not only about the latter, this design speaks directly to the well-know comment on social capital by

¹⁰Goldstein and Udry (1999) report two other potential sources of identification of information networks. One is the question on with whom the respondent had significant discussions on agricultural matters. We think this measure provides no better insights into our question for three reasons. First, not all of the listed individuals were part of the sample. Hence no information is available on the match’s identity. Second, the period when each of the listed individuals was contacted is not known, making it impossible to estimate the effect of differences in characteristics that change over time, such as experience or wealth. Third, it is not easy to identify why someone was not chosen – the assumption that everyone knows everyone else is clearly needed here. The second possibility would be to use the answers to the question “Have you ever gone to ... for advice about your farm?” This poses similar problems to the one just described, to which one may add that a negative answer may just indicate that there were some other source of information that the respondent valued more and on which we may not have any information (including its existence).

Arrow (2000, p.4, emphasis added) that “it may be more fruitful to think of the *existing* social relations as a *preexisting network* into which new parts of the economy (...) have to be fitted”. That economic activity involving interpersonal relations is commonly nested within social relations that are built without specific, instrumental, purposes is the core of the concept of embeddedness, advanced by sociologists (namely Granovetter, 1985) and now accepted by economists.¹¹

These data offer important advantages but also potential shortcomings. The major advantage is that we know the characteristics of *all* respondents and matches; we are not forced to disregard information on matches’ identity simply because they do not belong to the original sample. The major shortcoming is that our information on the dependent variable reflects hypothetical dyads. One might reasonably wonder how well potential information networks elicited in this manner reflect the decision process underlying the formation of real information networks. We cannot answer this question directly in this setting, but Santos and Barrett (2007) use data on links elicited through this approach and real links to show that the inference with respect to the determinants of social interactions decisions is not significantly affected by the experimental nature of the question.¹² The appeal of using randomly

¹¹See Durlauf and Fafchamps (2005) or Jackson (2007) for recent reviews.

¹²This is not an entirely surprising result. An extensive literature on stated choice methods suggests that when properly contextualized, elicitation of hypothetical behaviors can provide an accurate view of actual behaviors (Arrow et al., 1993). The benefits of using experimental data in the study of social capital (a concept closely related to that of social networks) are also emphasized by Durlauf and Fafchamps (2005).

matched respondents thus seems to outweigh the prospective pitfalls that the answers might reflect merely hypothetical links.

Table 1 presents the answers to the last three questions disaggregated by whether or not the respondent knows his/her match.¹³ Three key facts emerge from this table. First, contrary to the widespread practice of using common village membership as a proxy for a social network, not everyone knows everyone else, even in a small, rural village setting.

Second, survey respondents do not anticipate addressing those with whom they are not already acquainted in order to gain access to information on these specific questions. Given that the set of acquaintances is relatively large (only in 9.8% of the cases is the match unknown by the respondent), these respondents concentrate information search within the known social space. This behavior has important consequences for econometric estimation of the determinants of information networks in that it configures a sequential decision process. Since there is effectively no variation in the decision to contact those individuals not known by the respondent – in less than one percent of cases would a respondent ask a question of someone not known to them – we follow Maddala (1983, p.124) and restrict the estimation of the determinants of information networks to the subsample of those who know their match.

Third, the links through which rural Ghanaian farmers actively acquire

¹³A third possibility in answering the last three questions, “Yes, but he wouldn’t know”, was never reported.

information comprise only a small subset of their social networks. Only 30% of the known matches are selected as sources of information for solutions to specific problems. Understanding what limits the formation of information networks is therefore central to understanding the determinants of differential access to information and is the object of the next section's econometric investigation.

In order to estimate equation 1 we make the following assumptions

$$\text{Prob}(L_{ijk} = 1) = \Lambda(I_{(X_i - X_j < 0)} \times |X_i - X_j|, I_{(X_i - X_j \geq 0)} \times |X_i - X_j|) \quad (3)$$

where X is the vector of characteristics of the social space over which we will measure the distance between the respondent (indexed by i) and the random match (indexed by j), $\Lambda(\bullet)$ is the logistic cumulative distribution function, and $L_{ijk} = 1$ when i would ask j question k . Regarding the unobserved error term in this regression, ε_{ij} , we further assume that:

$$E(\varepsilon_{ij}, \varepsilon_{ih}) \neq 0 \text{ if } j \neq h \quad (4)$$

$$E(\varepsilon_{ih}, \varepsilon_{jh}) \neq 0 \text{ if } i \neq j \quad (5)$$

These assumptions imply a logit model that incorporates the possibility of correlation across matches' unobservables, that is, that the error term is also dyadic.¹⁴

¹⁴An alternative way of modeling the error term is to assume that the personal network is a complex attribute of the individual and that relations are nested within individuals

Most of the studies that account for the effect of correlation across unobservables do so by correcting the covariance matrix using the estimator suggested by Conley (1999). We follow a different strategy, using a nonparametric permutation test known as Quadratic Assignment Procedure (QAP) (Hubert and Schultz, 1976, Krackhardt, 1988, 1987) to obtain correct p-values.¹⁵ The basic intuition behind this procedure is that the permutation of the data on the dependent variable must maintain its clustered nature. In practice, this means that the same permutation must be applied to respondents and matches. We can then estimate the above model when all correlation between dependent and independent variables is broken through resampling – that is, when the null hypothesis that all slopes equal zero is known to be true – and compare our first estimates with their empirical distribution obtained through the repetition of this exercise (in our case, 500 times), to generate a sampling distribution for the parameter estimates.

(Valente, 2005). This assumption imply a logit model estimated by clustering the observations on the identity of the respondent, that is, that $E(\varepsilon_{ih}, \varepsilon_{jh}) \neq 0$ if $i \neq j$. The record of whether such simplification matters is mixed. Fafchamps and Gubert (2007), studying insurance networks in the Phillipines, and Udry and Conley (2005), using the same data we use, find no significant differences from estimates that do not account for correlation across matches' unobservables. Other studies, for example Arcand and Fafchamps (2007), find that allowing for correlation across matches' unobservables does matter to inference.

¹⁵Each of our respondents is matched with seven other individuals. With such a small number of matches, it does not seem credible that the assumptions for the asymptotic properties of an estimator such as the one introduced by Conley (1999) would hold. In an earlier version of their analysis, Fafchamps and Gubert (2007) used QAP to derive correct p-values. As they mention, inference was similar to using the correction that they ultimately report.

3 Empirical results

The definition of the explanatory variables used and some descriptive statistics are presented on Table 2.¹⁶ Table 3 then summarizes the estimates of the decision to form a link with the match in order to obtain information in response to the three questions presented above, together with the QAP corrected p-values and the estimates of their marginal effect.¹⁷

The first conclusion is that, perhaps surprisingly, the variables expressing shared clan membership and similarities in migrant status are not statistically significant at the conventional levels of significance. The effect of shared clan affiliation is positive (as expected) but imprecisely estimated, with p-values that are never below 0.05 (although below 0.10 in two cases).

Gender differences matter a lot, both with respect to agricultural production and to marketing. Men are generally far less willing to ask questions of others than are women. More interestingly, these effects appear to be strongly asymmetric. Men are especially reluctant to ask questions of women, particularly if the questions are about farm production (fertilizer application or changes in planting techniques). A woman is more likely to ask agricultural

¹⁶In earlier versions of this article we experimented with the inclusion of individual attributes (gender, age, land, etc.) but these proved not to be statistically significant, either individually or jointly. We interpreted this result as evidence that it is relative position in the social space – and not the characteristics themselves – that matters in explaining the choice of who to contact. Other regressors were also included – for example, differences in experience with cocoa, differences in income from non-agricultural activities, whether the respondent and the matched cultivated modern varieties of pineapple. All of our core results are robust to inclusion of those regressors, so we only present the more parsimonious version of our estimates.

¹⁷Estimated at the average sample values of each variable

production questions of a man than is another man or of another woman. And when asking information about marketing, women are far more likely to talk with other women while men seem reluctant to ask such questions, regardless of their match’s gender. These differences would have been lost had we expressed the difference between match and respondent using the conventional Euclidean distance approach to social distance, in this dichotomous variable case through the variable “same gender”.

Differences in experience with pineapple production seem to significantly affect respondents’ search for information, as farmers appear significantly less likely to ask questions of matches with less experience than they have. The differences are much larger in both magnitude and in statistical significance with respect to experience in pineapple, a new crop in the region, than in maize, a long-established staple crop. Finally, differences in landholding matter as farmers with more land are less likely to ask for information from farmers with less land while differences in non-land wealth ¹⁸ play no statistically significant role.

In table 4 we test for the joint significance of subsets of these variables. In a first step, we define Identity as the union of clan membership, migrant status variables, gender, age and non-land wealth ¹⁹ and Interest as the joint

¹⁸Non-land wealth is the sum of the values of a variety of assets, including foreign currency, cash in bank accounts, bonds, money in informal savings and credit institutions, but excluding agricultural assets such as livestock and farm equipment.

¹⁹Our interpretation is that wealth, as defined in the previous footnote, is not *directly* productive, especially after we control for land ownership: our measure of wealth would then be a measure of “status”. The contrary argument, favoring the inclusion of wealth in the definition of Interest, is that monies are fungible and, as such, these assets also matter

effect of land ownership and experience with pineapple and maize. Although the reader may, at this point, quibble with our definition (and think about alternative arrangements of one variable or the other), we want to argue that, first, a consensual empirical definition of Identity is impossible to achieve and that, second, our testing strategy will, ultimately, make such a definition unnecessary. We do that after our discussion of the statistical significance of these two sets of variables.

We test the standard joint null hypothesis that

$$H_0 : \beta_X = 0 \tag{6}$$

where X is the subset of variables used to define Identity and Interest. The test statistic for this hypothesis, labeled D_1 , is defined as

$$D_1 = \sum_{x \in X} |\beta_x - 0| \tag{7}$$

where x is any variable belonging to the subset X . We can compute this statistic using the estimates obtained through the Quadratic Assignment Procedure and obtain p-values for tests of hypothesis such as equation 6 – see, for example, Good (2005) for a discussion of the use of permutation to test hypothesis.²⁰

for one’s productivity. As we’ll argue below, with regards to the interpretation of gender differences between respondent and match, this discussion only reflects how difficult it is to empirically define Identity. However, given our identification strategy, the specific set of variables that constitutes Identity (or Interest) will not matter.

²⁰Alternatively, and if we assumed that there is no correlation across matches’ unob-

Identity is statistically significant when we look at the decision to form a link through which advice on fertilizer application can flow and, marginally, in the case of finding a buyer. Interest, on the other hand, is important in explaining each of the link formation decisions. Recall that new fertilizer application is associated with the introduction of a new, commercial crop (pineapple) in the area. This makes the apparent relevance of identity to information search somewhat puzzling if one believes that modernization diminishes the importance of productivity-irrelevant characteristics in influencing individual behavior. Other results further underscore the difficulty, in practice, of effectively isolating purely payoff-irrelevant characteristics in empirical work.

Consider the following example. While women are more likely to ask men about fertilizer application, they are less likely to ask about marketing. How does one best interpret this result? An identity-based interpretation would explain these patterns through some “cultural norm”, but these patterns may also reflect gendered actions, where men take primary responsibility for on-farm production (for example, very few women cultivate pineapple) and women predominate in marketing activities. Differential patterns of willingness to establish a link may simply reflect differential assessments of the probability of receiving valuable information from different types of individuals. Although, in principle, one could always add more control variables

servables, as discussed in footnote 15, we could use the Wald statistic. It turns out that, in our case, the qualitative conclusions are identical.

for the productivity-relevant characteristics of respondents' matches, true isolation of the productivity-irrelevant characteristics associated with one's gender seems a futile pursuit. Rather like the Asians and Romans example with which we started this article, disentangling identity and interest is not an easy task.

Besides exemplifying the difficulties in isolating identity from interest in a single regression, the preceding example also suggests one way to test empirically whether identity is the primordial force behind social learning decisions. Because identity is productivity irrelevant, its effect won't change with the question at hand. As a consequence, if identity were the main driver of interpersonal behavior, one would not expect to see important differences in the estimates of the determinants of the formation of different information networks. The respondent's interest in asking a question of a particular individual may change, however. For example, differences in area cropped may matter when searching for information on agricultural production questions but may matter less when searching for a buyer.

By asking whether people target questions differentially to distinct individuals, purposefully building different information networks, we are effectively asking whether identity is the main determinant of these networks. A simple test of this idea is a test of whether the coefficients associated with each variable are statistically identical across models, that is, whether

$$H_0 : \beta_{\text{change planting technique}} = \beta_{\text{fertilizer application}} = \beta_{\text{find a buyer}} \quad (8)$$

Table 5 presents the results of the test of this null hypothesis (and of the different possible combination of questions) using the test statistic D_2 , defined as

$$D_2 = \sum_{x \in X} |\beta_{x, \text{change planting technique}} - \beta_{x, \text{fertilizer application}} - \beta_{x, \text{find a buyer}}| \quad (9)$$

As above, we can use the results of the Quadratic Assignment Procedure to compute the p-value of this statistic.

We easily reject the null hypothesis that the variables explaining the formation of information networks have the same effect across all information sought. We can also easily reject the more restricted hypothesis that these variables have the same effect across any of the pairs of information networks under consideration. This signals that respondents selectively target a specific set of individuals within their social network who they consider most likely to provide beneficial information worth the cost of forming the network link. Since the identities of respondents and matches do not change with the nature of the information sought, the result that information networks appear to be constructed separately for distinct questions is strong evidence that interest plays a fundamental role in information search and one that is robust to the difficulties of interpretation of the estimates associated with specific identity-related variables.

4 Conclusions

This article analyzes the formation of information networks in a developing country, in a context of technological change (fertilizer application) associated with the introduction of a new crop (pineapple for export markets). The main purpose of this analysis is to discuss the role played by identity (e.g., clan affiliation, gender, migrant status) as potential barriers to information flow. We argue that, to escape tautological assertions of the importance of identity, the concept must be defined by contrast with interest, encompassing only characteristics that are inherently irrelevant to one's productivity. The challenge, of course, is that such a definition makes direct quantification of the practical importance of identity very difficult, as it is always possible to argue that what appears to the external observer as irrelevant to one's productivity may in fact subtly transmit relevant information regarding one's productive characteristics.

This implies that one needs to explore the impact of identity on economic behavior indirectly. In this article we use variation in survey respondents' decision to contact others for information on different agricultural problems, to identify the role played by identity in shaping such decision. Because identity is not problem-specific, the fact that respondents target different subsets of individuals as sources of information on different issues suggests that differences in identity cannot be the main motive behind information search in this setting.

There are no direct policy conclusions from this study and we don't want to claim that identity is irrelevant, certainly not in all contexts. We hope, however, that it may prompt some more skeptical interrogation of results that suggest that identity is a serious impediment to trade between individuals. If interest, and not identity, is the ruling motive underlying human interaction, that can only be good news for economic growth.

References

- Akerlof, G., and R. Kranton. 2000. "Economics and identity." *Quarterly Journal of Economics* 115:715–753.
- . 2002. "Identity and Schooling: some lessons for the economics of education." *Journal of Economic Literature* 40:1167–1201.
- Akerlof, G.A. 1997. "Social distance and social decisions." *Econometrica* 65:1005–1028.
- Arcand, J.L., and M. Fafchamps. 2007. "Matching in rural producer organizations." Unpublished, University of Oxford working paper.
- Arrow, K.J. 1972. "Models of job discrimination." In A. H. Pascal, ed. *Racial discrimination in economic life*. D. C. Heath.
- . 2000. "Observations on social capital." In P. Dasgupta and I. Serageldin, eds. *Social Capital. A multifaceted perspective*. World Bank, pp. 3–5.
- Arrow, K.J., R. Solow, P.R. Portney, E.E. Leamer, R. Radner, and H. Schuman. 1993. "Report of the NOAA Panel on contingent valuation." Washington, D.C., National Oceanic and Atmospheric Administration.
- Bates, R.H. 2000. "Ethnicity and Development in Africa: A Reappraisal." *American Economic Review* 90:131–34.

- Bertrand, M., and S. Mullainathan. 2004. "Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination." *American Economic Review* 94:991–1013.
- Burger, K., P. Collier, and J.W. Gunning. 1993. "Social learning: an application to Kenyan agriculture." Unpublished, working paper, The Centre for the Study of African Economies.
- Conley, T., and C. Udry. 2005. "Learning about a technology: pineapple in Ghana." Unpublished, Yale University, working paper.
- Conley, T.G. 1999. "GMM estimation with cross-sectional dependence." *Journal of Econometrics* 92:1–45.
- DeWeerd, J. 2004. "Risk sharing and endogenous network formation." In S. Dercon, ed. *Insurance against poverty*. Oxford: Oxford University Press.
- Durlauf, S., and M. Fafchamps. 2005. "Social capital." In P. Aghion and S. Durlauf, eds. *Handbook Economic Growth*. Amsterdam: Elsevier.
- Easterly, W., and R. Levine. 1997. "Africa's growth tragedy: policies and ethnic divisions." *Quarterly Journal of Economics* 112:1203–1250.
- Fafchamps, M., and F. Gubert. 2007. "The formation of risk sharing networks." *Journal of Development Economics* 83:326–350.
- Goldstein, M., and C. Udry. 1999. "Agricultural innovation and risk management in Ghana." Unpublished, final report to IFPRI.

- Good, P. 2005. *Permutation, parametric and bootstrap tests of hypothesis*. New York: Springer-Verlag.
- Granovetter, M. 1985. "Economic action and social structure: the problem of embeddedness." *American Journal of Sociology* 91:481–510.
- Heckman, J.J. 1998. "Detecting discrimination." *Journal of Economic Perspectives* 12:101–116.
- Hirschman, A.O. 1997. *The passions and the interests: political arguments for capitalism before its triumph*. Princeton University Press.
- Hubert, L., and J. Schultz. 1976. "Predicting with networks: Nonparametric multiple regression analysis of dyadic data." *British Journal of Mathematical and Statistical Psychology* 29:190–241.
- Jackson, M. 2007. "The study of social networks in economics." Unpublished, forthcoming in *The Missing Link: Formation and Decay of Economics Networks*.
- Jenkins, R. 1997. *Social Identity*. Routledge.
- Kirman, A., and M. Teschl. 2004. "On the emergence of economic identity." *Revue de Philosophie Economique* 9:129–156.
- Knight, F.H. 1921. *Risk, uncertainty and profit*. Houghton Mifflin Company.
- Krackhardt, D. 1988. "Predicting with networks: Nonparametric multiple regression analysis of dyadic data." *Social Networks* 10:359–381.

- . 1987. “QAP partialling as a test of spuriousness.” *Social Networks* 9:171–186.
- Maddala, G.S. 1983. *Limited-dependent and qualitative variables in econometrics*. Econometric Society Monographs, Cambridge University Press.
- McPherson, J.M., L. Smith-Lovin, and J. Cook. 2001. “Birds of a Feather: Homophily in Social Networks.” *Annual Review of Sociology* 27:415–444.
- Miguel, E. 2004. “Tribe or Nation? Nation-building and public goods in Kenya versus Tanzania.” *World Politics* 56:327–362.
- Miguel, E., and M.K. Gugerty. 2005. “Ethnic diversity, social sanctions, and public goods in Kenya.” *Journal of Public Economics* 89:2325–2368.
- Miguel, E., and D.N. Posner. 2006. “Sources of Ethnic Identification in Africa.” Unpublished, working paper, UC Berkeley.
- Munshi, K., and J. Myaux. 2006. “Social norms and the fertility transition.” *Journal of Development Economics* 80:1–38.
- Rauch, J.E., and G.G. Hamilton. 2001. “Networks and markets: concepts for binding disciplines.” In J. E. Rauch and A. Casella, eds. *Networks and Markets*. Russell Sage Foundation Publications.
- Romani, M. 2003. “Love Thy Neighbour? Evidence from Ethnic Discrimination in Information Sharing within Villages in Cte d’Ivoire.” *Journal of African Economies* 12:533–563.

Santos, P., and C.B. Barrett. 2007. "Understanding network formation." Unpublished, Cornell University working paper.

Swedberg, R. 2004. *Interest*. Open University Press.

Udry, C., and T. Conley. 2005. "Social networks in Ghana." In C. B. Barrett, ed. *The social economics of poverty: identities, groups, communities and networks*. London: Routledge, chap. 10.

Valente, T.W. 2005. "Network models and methods for studying the diffusion of innovations." In P. J. Carrington, J. Scott, and S. Wasserman, eds. *Models and methods in social network analysis*. Cambridge University Press, pp. 98–116.

Table 1: Knowing and asking: a sequential process

Know	Planting Technique		Fertilizer Application		Finding a Buyer		Total
	No	Yes	No	Yes	No	Yes	
No	139	1	139	1	140	0	140
Yes	916	360	891	385	890	386	1276
Total	1055	361	1030	386	1030	386	1416

Table 2: Variable definitions and descriptive statistics

Variable	Definition	Mean (SD)
Same clan	Dummy variable, equal to 1 if both respondent and match belong to the same clan	0.301 (0.459)
Both migrant	Dummy variable, equal to 1 if both respondent and match are migrant	0.050 (0.218)
migrant, non-migrant	Dummy variable, equal to 1 if respondent is migrant and the match is not	0.171 (0.377)
non-migrant, migrant	Dummy variable, equal to 1 if respondent is not a migrant and the match is	0.173 (0.379)
Both male	Dummy variable, equal to 1 if both respondent and match are male	0.270 (0.444)
Male, female	Dummy variable, equal to 1 if respondent is male and the match is female	0.241 (0.428)
Female, male	Dummy variable, equal to 1 if the respondent is female and the match is male	0.230 (0.421)
Older	Absolute value of the age difference (in years) between respondent and match if the respondent is older than the match, 0 otherwise	7.086 (10.376)
Younger	Absolute value of the age difference (in years) between respondent and match if the respondent is younger than the match, 0 otherwise	7.062 (10.39)
Wealthier	Absolute value of the difference in wealth between respondent and match if the respondent is wealthier than the match, 0 otherwise (in million cedis)	0.174 (0.504)
Poorer	Absolute value of the difference in wealth between respondent and match if the respondent is poorer than the match, 0 otherwise (in million cedis)	0.204 (0.453)
More experience with maize	Absolute value of the difference in experience cultivating maize between respondent and match if the respondent has more experience than the match (in years), 0 otherwise	7.326 (10.960)
Less experience with maize	Absolute value of the difference in experience cultivating maize between respondent and match if the respondent has less experience than the match (in years), 0 otherwise	6.679 (10.048)
More experience with pineapple	Absolute value of the difference in experience cultivating pineapple between respondent and match if the respon-	1.856 (3.289)

Continued on next page...

... table 2 continued

Variable	Definition	Mean (SD)
	dent has more experience than the match (in years), 0 otherwise	
Less experience with pineapple	Absolute value of the difference in experience cultivating pineapple between respondent and match if the respon- dent has less experience than the match (in years), 0 otherwise	1.486 (3.099)
More land	Absolute value of the difference in land cropped be- tween the respondent and match if the respondent has more land than the match (in hectares), 0 otherwise	0.473 (1.139)
Less land	Absolute value of the difference in land cropped be- tween the respondent and match if the respondent has less land than the match (in hectares), 0 otherwise	0.946 (1.569)

Table 3: Estimates of the determinants of information network

Variable	Planting technique		Fertilizer application		Finding a buyer					
	Coef.	QAP p-value	Coef.	QAP p-value	Coef.	QAP p-value				
same clan	0.256	0.078	0.047	0.047	0.227	0.102	0.038	0.272	0.072	0.052
both migrant	-0.309	0.246	-0.057	-0.057	-0.664	0.070	-0.112	-0.066	0.464	-0.013
migrant, non-migrant	-0.054	0.420	-0.009	-0.009	-0.308	0.154	-0.051	-0.326	0.132	-0.063
non-migrant, migrant	-0.051	0.410	-0.009	-0.009	-0.210	0.178	-0.035	-0.037	0.458	-0.007
both male	-0.121	0.346	-0.022	-0.022	-0.262	0.192	-0.044	-0.965	0.000	-0.186
male, female	-1.214	0.000	-0.224	-0.224	-1.426	0.000	-0.241	-0.965	0.002	-0.138
female, male	0.298	0.080	0.055	0.055	0.364	0.046	0.061	-0.589	0.004	-0.114
older	0.003	0.376	0.001	0.001	0.002	0.424	0.000	0.019	0.094	0.003
younger	0.005	0.302	0.001	0.001	0.005	0.300	0.001	0.002	0.426	0.000
wealthier	0.022	0.448	0.004	0.004	0.070	0.384	0.012	-0.006	0.500	-0.001
poorer	0.052	0.316	0.010	0.010	0.006	0.472	0.001	0.127	0.156	0.025
more experience with maize	0.002	0.448	0.000	0.000	0.005	0.356	0.001	-0.017	0.106	-0.003
less experience with maize	-0.002	0.414	-0.001	-0.001	-0.000	0.478	-0.000	-0.010	0.236	-0.002
more experience with pineapple	-0.152	0.000	-0.028	-0.028	-0.161	0.000	-0.028	-0.076	0.014	-0.015
less experience with pineapple	-0.006	0.420	-0.001	-0.001	-0.012	0.356	-0.002	0.007	0.390	0.001
more land	-0.299	0.006	-0.055	-0.055	-0.298	0.010	-0.050	-0.179	0.064	-0.034
less land	0.050	0.242	0.009	0.009	0.063	0.196	0.011	0.095	0.068	0.018
Number of observations		1276		1276		1276			1276	
R^2		0.169		0.203		0.148			0.148	
Count R^2		0.749		0.771		0.748			0.748	

Table 4: Testing for the importance of Identity and Interest in the formation of information networks

Set of variables	Planting technique		Fertilizer application		Finding a buyer	
	D ₁	p-value	D ₁	p-value	D ₁	p-value
Identity	2.385	0.124	3.180	0.020	3.127	0.042
Interest	0.510	0.002	0.540	0.002	0.384	0.036

Table 5: Testing for interlinkage between different networks

	D ₂	QAP p-value
Planting × Fertilizer	1.134	0.002
Planting × Buyer	3.177	0.000
Fertilizer × Buyer	3.705	0.000
Planting × Fertilizer × Buyer	4.775	0.002