Empirics of Social Interactions
June 7, 2006


Abstract

Empirical studies of social interactions address a multitude of definitional, econometric and measurement issues associated with role of interpersonal and social group influences in economic decisions. Applications range from studies of crime patterns, neighborhood influences on upbringing and conformist behavior, mutual influences among classmates and keeping up with roommates in colleges regarding academic and social activities, to herding and to learning about social services. The entry reviews several instances of successful identification of effects emanating from others’ behavior as distinct from characteristics of others. Data sets with increasingly rich contextual information will allow estimation of complex models of economic decisions.

*Related entries:* educational finance, geographic information systems, herd behavior, natural and quasi-natural experiments, neighbors and neighborhoods, policy experiments, psychology of social networks, social multipliers, social network formation, sociology of social networks, theory of social interactions, Tiebout hypothesis.

*Keywords:* Social interactions, peer effects, contextual effects, neighborhood choice, neighbors, neighborhoods, neighborhood effects, laboratory experiments, field experiments, self selection, social networks.

*JEL Codes:* C25, I30, R00

1 Introduction

The empirical economics literature on social interactions addresses the significance of the social context in economic decisions. Decisions of individuals who share a social milieu are likely to be interdependent. Recognizing the nature of such interdependence
in a variety of conventional and unconventional settings and measuring empirically the role of social interactions poses complex econometric questions. Their resolution may be critical for a multitude of phenomena in economic and social life and of matters of public policy.

The social context enters in a variety of ways. One is that individuals care not only about their own purely private outcomes, e.g. the kinds of cars they drive or the education they acquire, but also about outcomes of others, such as the kinds of cars or the education of their friends. This type of interpersonal effect is known as endogenous social effect (or interaction), because it depends on decisions of others in the same social milieu. Individuals may also care about personal characteristics of others, that is whether they are young or old, black or white, rich or poor, trendy or conventional, and so on, and about other attributes of the social milieu that may not be properly characterized as deliberate decisions of others. The latter is known as exogenous social or contextual effect. In addition, individuals in the same or similar social settings tend to act similarly because they share common unobservable factors. Such an interaction pattern is known as correlated effects. This terminology is due to Manski (1993).

Emergence of social interdependencies is natural if individuals share a common resource or space in a way that is not paid but still generates constraints on individual action. This is also known as pecuniary externalities. Individuals who try to form expectations about future outcomes of current decisions, like occupational choice, may rely on lessons from the actions of others and therefore end up mimicking their behavior. Endogenous social interactions are a case of real externalities, a pervasive feature of economic behavior.

Theorizing in this area must lie in the interface of economics, sociology and psychology and often is imprecise. Terms like social interactions, neighborhood effects, social capital and peer effects are often used as synonyms although they may have different connotations. Empirical distinctions between endogenous, contextual and correlated effects are critical for policy analysis because of the “social multiplier,” as we see further below.

Joint dependence among individuals’ decisions and characteristics within a social milieu is complicated further by the fact that in many interesting circumstances indi-
viduals in effect choose the social context. E.g., individuals choose their friends and their neighborhoods and thus their neighborhood effects as well. Such choices involve information that is in part unobservable to the analyst, and therefore require making inferences among the possible factors which contribute to decisions [Brock and Durlauf (2001) and Moffitt (2001)]. The present entry focuses on highlighting the significance of key empirical findings and owes a lot to Durlauf (2004), the most comprehensive review to date that examines the methodological basis, statistical reliability and conceptual and empirical breadth of the neighborhood effects literature.

2 Empirical Framework

Let individual $i$’s outcome $\omega_i$, a scalar, be a linear function of a vector of observable individual characteristics, $X_i$, of a vector of contextual effects, $Y_{n(i)}$, which describe $i$’s neighborhood $n(i)$, and of the expected value of the $\omega_j$’s of the members of neighborhood $n(i)$, $j \in n(i)$. It is straightforward to incorporate social interactions into economic models in a manner that is fully compatible with economic reasoning, that is by positing that individuals maximize a utility function subject to constraints and obtain a behavioral equation such as:

$$\omega_i = k + cX_i + dY_{n(i)} + Jm_{n(i)} + \epsilon_i,$$  \hspace{1cm} (1)

where $\epsilon_i$ is a random error and $k$ a constant. Abstracting at the moment from the issue that individual $i$ may have deliberately chosen neighborhood, $n(i)$, and stating that conditional on individual characteristics, contextual effects and the event that $i$ is a member of neighborhood $n(i)$ the expectation of $\epsilon_i$, is zero, allows to focus on the estimation of such models. Critical next steps for translating theoretical models to empirical applications is to assume social equilibrium and that individuals hold rational expectations over $m_{n(i)}$. That is, individuals’ expectations are confirmed in that they are exactly equal to what the model predicts. So, taking the expectation of $\omega_i$ and setting it equal to $m_{n(i)}$ allows us to solve for $m_{n(i)}$. Substituting back into (1) yields a
reduced form, an expression for individual \( i \)'s outcome in terms of all observables:

\[
\omega_i = \frac{k}{1 - J} + cX_i + \frac{J}{1 - J}cX_{n(i)} + \frac{d}{1 - J}Y_{n(i)} + \epsilon_i.
\] (2)

This simple linear model obscures the richness that nonlinear social interactions models make possible, like multiplicity of equilibria [Brock and Durlauf (2001)]. Yet, it does facilitate studying other aspects. For example, it does confirm that endogenous social effects generate feedbacks which magnify the effects of neighborhood characteristics. That the effect of \( Y_{n(i)} \) is \( \frac{d}{1 - J} \), and not just \( d \). It also confirms why it is tempting for empirical researchers to study individual outcomes as functions of all observables. Following the pioneering work of Datcher (1982), a great variety of individual outcomes have been studied in the context of different neighborhoods and typically significant effects have been found. Deriving causal results requires suitable data.

Manski (1993) emphasized that the practice of including neighborhood averages of individual effects as contextual effects, \( Y_{n(i)} = X_{n(i)} \), may fail to identify endogenous separately from exogenous interactions, that is to estimate \( J \) separately from \( d \). However, partial identification is possible. That is, if the neighborhood attributes are restricted to the neighborhood averages of its inhabitants’ characteristics, or \( Y_{n(i)} = X_{n(i)} \), then regressing individual outcomes on neighborhood averages of individual characteristics as contextual effects allows us to estimate \( \frac{Jc + d}{1 - J} \). A statistically significant estimate of the coefficient of \( X_{n(i)} \) implies that at least one type of social interaction is present, either \( J \) or \( d \) or both are nonzero.

If it is plausible to exclude some of neighborhood averages of individual covariates, then identification may be possible. Also, if nonlinearities are inherent in the basic model specification, identification again may be possible. A noteworthy case in point here is Drewianka (2003) who studies two-sided matching in the marriage market and finds that it allows identification of endogenous and exogenous social interactions. The logic of the model requires that the two sides of the market contain an additional source of variation: the greater the number of potential marriage partners, the higher the probability that a match will occur. There is an inherent multiplier effect at work here. This likelihood depends on the rate at which other people match up, an endogenous social effect. Drewianka’s results show that a 10 percent increase in the
fraction of the population that is unmarried causes the marriage rate of never-married
men to fall by 10 percent and that of never-married women by 7 percent.

An interesting consequence of endogenous social interactions is in amplifying differences in the average neighborhood behavior across neighborhoods. In fact, Glaeser et al. (2003) use directly such patterns in the data to estimate a social multiplier. This is defined for a change in a particular fundamental determinant of an outcome as the ratio of a total effect, which includes a direct effect to an individual outcome plus the sum total of the indirect effects through the feedback from the effects on others in the social group, to the direct effect. It is easy to see that as the ratio of the “group level” coefficient, the coefficient of $Y_{n(i)}$ in Equ. (2), to the “individual level” coefficient, the coefficient of $Y_{n(i)}$ in Equ. (1): $\frac{d_1 - J_1 d}{1 - J_1} = \frac{1}{1 - J}$. It follows that a social multiplier greater than one implies endogenous social interactions, $0 < J < 1$. This approach must deal, in practice, with dependence across decisions of individuals belonging to the same group, which is implied by non-random sorting in terms of unobservables. It is particularly useful in delivering ranges of estimates for the endogenous social effect and when individual data are hard to obtain.

This is the case with crime data. Glaeser, Sacerdote and Scheinkman (1996) motivate their study of crime and social interactions by the extraordinary variation of incidence of crime across US metropolitan areas over and above differences in fundamentals. If social interactions are present, variations in observed outcomes are larger than what would be expected from variations in underlying fundamentals. Glaeser, Sacerdote and Scheinkman (2003) regress actual crime rates against predicted crime rates, which are formed by multiplying percentages of US individuals in each of eight age categories by the crime rate of persons in that category. They perform such regressions at the level of county and state cross-sectionally and for the entire US over time. Their results imply large social multipliers, which increase with the level of aggregation exactly as their basic theory would predict are consistent with large endogenous social interaction coefficients.

It is possible to modify this basic model in order to study several other areas involving economic decisions akin to social interactions. For example, diffusion of innovations, herding and adoption of norms or other institutions by a population involve
ideas that are conceptually related to social interactions. Also, $J$ may be negative, as in the case of land development, which is conceivably due to congestion.

3 Identification of Social Interactions Using Observational Data on “Natural Experiments”

Several researchers have sought to identify social interactions by exploiting uniquely suitable features of observational data, that are often referred to as “natural experiments.” For example, consider outcomes for children from families with several children who share the common influence of unobservable family factors, such as parental values and competence, taste for education and time spent with children, and other unobservables that affect upbringing of household members living in close proximity. They also share the variation in neighborhood effects that is produced by families’ residential moves. By using observations on several children from the same family who are separated in age by at least three years, Aaronson (1998) controls for family-specific characteristics. This obviates the need to control for the impact of self-selection in terms of unobservable neighborhood characteristics. He uses data from the Panel Study of Income Dynamics and finds large and statistically significant contextual neighborhood effects, but his models exclude endogenous social effects. His results are robust to changes in estimation techniques and in sample and variable definitions, but are sensitive to the formulation of neighborhood characteristic proxy. Incomplete specification of family characteristics is an important concern, and its consequences for the robustness of estimated relationships are aptly demonstrated by Ginther et al. (2000).

Grinblatt et al. (2004) use data for all residents of two large Finnish provinces, that is millions of observations, and establish that automobile purchase decisions by close residential neighbors influence one another. The measured endogenous neighborhood effects are strongest among individuals belonging to the same “social class” (especially if they belong to lower income classes), or when the cars they purchase are of the same make or even the same model. These findings mitigate in favor of information sharing instead of “keeping up with the Joneses.” We note that excluding neighborhood means
of demographics as contextual effects are reasonably plausible in this case: there is no reason why the average age of my neighbors should affect directly my taste in cars.

Luttmer (2005) uses data from the U.S. National Survey of Families and Households, augmented with Census data from the Public Use Microdata Areas, and examines how self-reported well being varies with own and neighbors’ incomes and of other characteristics. He interprets his findings as direct evidence that people have taste over their neighbors’ incomes. That is, after controlling for an individuals own income, higher earnings of neighbors are associated with lower levels of self-reported happiness in terms of a variety of measures.

Sacerdote (2001) exploits the fact that freshman year roommates and dormmates are randomly assigned at Dartmouth College, thus producing a natural quasi-experimental setting for studying peer effects. Sacerdote posits that an individual’s grade point average is a function of an individual’s own academic ability prior to college entrance, of social habits, and of the academic ability and grade point average of his roommates. Sacerdote finds that peers have an impact on each others’ grade point average and on decisions to join social groups such as fraternities. He does not, however, find residential peer effects in other major college decisions, such as choice of college major. He finds peer effects in grade point average at the individual room level — you keep up with your roommates! — whereas peer effects in fraternity membership occur both at the room level and the entire dorm level — dorms are conformist! These data provide strong evidence for the existence of peer effects in student outcomes, even among highly selected college students who may be otherwise quite homogeneous albeit in close proximity to one another. Peer effects are smaller the more directly a decision is related to labor market activities.

4 Peer Effects in Classrooms and Schools

Social interactions in classrooms, peer effects, are particularly interesting in understanding schooling as an economic activity and its consequences for inequality of social outcomes. Whether students benefit from classmates with different characteristics and academic performance and whether the effect is different depending upon whether one’s
classroom peers are more or less able is important for education policy and the actual functioning of schools. In other words, deciding whether or not students should be “tracked,” that is administratively segregated in terms of different characteristics, are the sort of policy questions which rest on understanding peer effects quantitatively.

Hoxby (2000) posits a relationship between individual academic achievement by a male student in a particular school and grade as the sum of what the mean achievement among males would have been in the absence of peer effects, of a term that is proportional to the percentage of females in the classroom, plus an error. She extends such a relationship to the case of several racial groups, which is particularly appropriate for the Texas Schools Project data that she uses. Her identification strategy involves exploring the panel structure of the data under the plausible assumption that there is natural idiosyncratic variation across successive cohorts in terms of gender, race and other individual attributes. Hoxby finds that students are affected by the achievement levels of their peers: an exogenous one point increase in peers’ reading scores raises a student’s own score between 0.14 and 0.4 points. Peer effects are stronger intra-race, and there is evidence of contextual effects: both male and female students perform better in classrooms that are more female despite the fact that females’ math performance is about the same as that of males.

The role of gender is corroborated by research by Arcidiacono and Nicholson (2005), who use data on the universe of students admitted to US medical schools for a particular year. One positive peer effect in US medical schools that they find pertains to female students, who benefit from attending medical schools that have other female students with relatively high scores on the verbal reasoning section of the Medical College Admission Test.

Of particular interest recently have been studies of the impact of school racial integration in the US on student performance. Let us consider Boston’s Metropolitan Council for Educational Opportunities (METCO) program, a voluntary desegregation program. The program allows mainly black inner-city kids from Boston public schools to commute to mainly white suburban communities in the Boston area that accommodate them in their public schools. Angrist and Lang (2004) show that although the receiving districts, which tend to higher mean academic performance, experience a
mean decrease due to the program, the effects are merely “compositional”, and there is little evidence of statistically significant effects of METCO students on their non-METCO classmates. Analysis with micro data from a particular receiving district (Brookline, Massachusetts) generally confirms this finding, but also produces some evidence of negative effects on minority students in the receiving district. METCO is a noteworthy social experiment, which was initiated by civil rights activists seeking to bring about defacto desegregation of schools. Lack of evidence of negative peer effects is particularly useful for informing desegregation policy. Still, there is self-selection in the participants on both sides.

5 Estimation of Social Interactions in Experimental Settings

Experimental data used by social interactions studies come from two types of deliberate experiments, field and laboratory experiments. A well known field experiment is Project STAR, an experimental program in the U.S. State of Tennessee that randomly assigned entering kindergarten students into three different class sizes and then randomly assigned teachers to them. A recent study that utilizes Project STAR data is Graham (2005). He seeks to estimate a relationship like (1) by measuring “excess” variance patterns across groups of exogenously given, but varying, sizes of classrooms that are associated with randomly assigned students and teachers. Graham compares contrasts in excess variance across small and large classrooms and finds social multipliers between 1.07 and 2.31, and 1.05 to 3.07, for math and reading achievement, respectively. Studies of this type need to discriminate between excess between-classroom variance, which is due to social interactions, from that due to group-level heterogeneity.

Duflo and Saez (2005) study, using experimental data, how social interactions among employees of a large U.S. university may influence participation in a tax deferred account retirement plan. The experiment more than tripled the attendance rate of those who received a small monetary reward for participating, doubled that of those not thus “treated” but who belonged to the same departments as the treated, and
significantly increased participation in the target program by individuals from treated departments, and did so almost as much by those who did not receive direct encouragement. While clearly social interactions effect may coexist with differential treatment and motivational reward effects, social interactions are also relevant for the effect of treatment on attendance and of attendance on participation. The authors conclude that the role of social interactions in amplifying the effect of treatment is unambiguous, in spite of the fact that they cannot distinguish unambiguously between the three different effects.

Moving to Opportunity (MTO) is a set of large randomized field experiments that were conducted by the U.S. Department of Housing and Urban Development in several large U.S. cities. The experiments offered poor households, who were chosen by lottery from among residents of high-poverty public housing projects, housing vouchers and logistical assistance through NGOs for the purpose of relocating to precisely defined as “better” neighborhoods. Several studies based on data from these experiments show that outcomes after relocation have improved for children, primarily for females, on account of education, risky behavior and physical health, but the effects on male youth were adverse. Regarding outcomes for adults, such as economic self-sufficiency or physical health, the picture is more mixed. Kling, Liebman and Katz (2005) find that four to seven years after relocation families (primarily female-headed ones with children) lived in safer neighborhoods that had lower poverty rates than those of a control group that were not offered vouchers. Unfortunately, there is serious controversy over how to interpret these findings in the context of policy design for large scale policy interventions [Sobel (2006)].

Turning next to laboratory experiments, a notable study is Ichino and Falk (2006). The experiment involves workers in pairs stuffing envelopes, with control being provided by subjects working alone in a room. These authors find that standard deviations of output are significantly smaller within pairs than between pairs and that social interactions raise productivity: average output per person is greater when subjects work in pairs. They also show that social interactions are asymmetric: low productivity workers are more sensitive to the behavior of high productivity workers as peers. Their setting does reduce some of the noise associated with “natural” experiments but does
not allow for contextual effects.

6 Identification of Social Interactions with Self Selection to Groups and Sorting

Presence of non-random sorting on unobservables is a major challenge for the econometric identification of social interactions models. Brock and Durlauf (2001) turned adversity into advantage by recognizing that self-selection itself, the endogeneity of neighborhood $n(i)$ in Equ. (1), may be brought to bear additional evidence on identification. That is, if it is possible to estimate a neighborhood selection rule, then correction for selection bias via the mean estimated bias, the so-called Heckman term, introduces an additional regressor in the right hand side of (1) whose neighborhood average is not a causal effect. Ioannides and Zabel (2004) implement this method successfully using micro data for a sample of households and their ten closest residential neighbors from the American Housing Survey and contextual information for the census tracts in which these individuals reside. Endogeneity of the average of one’s neighbors’ housing demands, an endogenous social effect, is instrumented by treating housing demands by a group of close neighbors as a simultaneous system of equations. By choosing neighborhoods, census tracts in this application, individuals choose desirable social interactions. Ioannides and Zabel work with an otherwise standard housing demand model and find a very significant and large endogenous social effect along with very significant contextual effects in the form of unobservable group effects. Several other studies have sought to use instrumental variables to account for self-selection. The critical role of local public finance of education in the U.S. has been studied extensively as a link between sorting into residential communities and socioeconomic outcomes. See entries on “Educational Finance” and ”The Tiebout Hypothesis.”
6.1 Social Interactions and Social Networks

The intuitive appeal of the notion that information transmitted through social discourse influences the behavior of individuals who interact socially has motivated recent research in labor markets, welfare program participation, and stock market participation. Ioannides and Loury (2004) review the labor market literature. Hong et al. (2004) use data from the U.S. Health and Retirement Survey and define social households as those who know at least some of their neighbors, interact with them at least occasionally and attend religious services. They show that controlling for wealth, race, education and risk tolerance, social households are more likely to invest in the stock market, with this effect being stronger in U.S. states where stock-market participation rates are higher. Aizer and Currie (2004) examine “network effects” in the utilization of publicly funded prenatal care. They find that pregnant women are most likely to be influenced in their use of public prenatal-care programs by new mothers from the same area and ethnic group. Such use is highly correlated within groups defined using race, ethnicity and residence in the same neighborhoods (defined as the areas of five-digit zip code), and persists even after accounting for unobserved characteristics by including zip code–year fixed effects. The richness of their data (from more than 3.5 million birth certificates from California) allows them to define fixed effects for the hospital of delivery interacted with the year of delivery. The estimates of network effects are then either reduced or eliminated. This casts doubt on the idea that the observed correlations can be interpreted as evidence of information sharing originating in ethnic and geographic proximity. They point instead to differences in the behavior of the low-income women involved, and of the institutions serving them, as the primary explanation for group-level differences in the take-up of publicly provided prenatal care. They examine the role of institutions by comparing the behavior of foreign-born with that of native-born Hispanic women. They find that such “network” effects are quite similar for both those groups of foreign-born and native-born Hispanic women. They conclude that it is differences in the behavior of institutions and not information sharing that explains the established correlations between neighborhood and ethnic group membership in prenatal care use.
7 Conclusions

Social interactions are ubiquitous and interest in estimating their effects is expanding rapidly in numerous areas of economics and is motivating important methodological advances. For econometricians, key challenges include social interactions effects on market outcomes coexisting with feedbacks from the characteristics of individual market participants via their impacts on prices, consequences of self-selection and the attendant role of presence of individual and group unobservables. Fundamentally and in the light of ever improving data availability, social interactions empirics will rely increasingly critically on careful theorizing that involves precise definitions of social interactions, possibly by calling on psychology and sociology to define appropriate boundaries, and their scope, and must facilitate use of data from different sources. The likely payoff is enormous: better understanding of social forces in the modern economy, with individuals sharing information while self-selecting into social groups and living and working in close proximity to one another as in firms and cities, the hallmark of modern economic life.

Yannis M. Ioannides

Tufts University

8 Bibliography


WORKING PAPER SERIES 2006

http://ase.tufts.edu/econ/papers/papers.html

2006-01  IOANNIDES, Yannis and Adriaan SOETEVENT; “Wages and Employment in a Random Social Network with Arbitrary Degree Distribution.”

2006-02  DAGEN, Richard and Daniel RICHARDS; “Merger Theory and Evidence: The Baby-Food Case Reconsidered.”

2006-03  PEPALL, Lynne, Daniel RICHARDS, John STRAUB and Michael DEBARTOLO; “Competition and Civic Engagement in the Religious Marketplace.”

2006-04  LOURY, Linda; “Job Search Among Informal Contacts.”

2006-05  BIR, Anupa and Karen EGGLESTON; “Measuring Selection Incentives in Managed Care: Evidence from the Massachusetts State Employee Insurance Program.”

2006-06  PEPALL, Lynne, Daniel RICHARDS and Liang TAN; “Advertising: The Good, the Bad and the Ugly.”

2006-07  METCALF, Gilbert E.; “Tax Incidence.”

2006-08  METCALF, Gilbert E.; “Value-Added Tax.”


2006-10  LOURY, Linda; “All in the Extended Family: Grandparents, Aunts, and Uncles and Educational Attainment.”

2006-11  IOANNIDES, Yannis; “Empirics of Social Interactions.”