where $N$ denotes the number of members of a neighborhood. Under rational expectations it is immediate that one joint probability measure for agents' choices is

$$
\mu(\omega) \sim \exp(\beta h \sum_i \omega_i + \beta NJ \sum_i \omega_i E(\omega)) \quad (33)
$$

where

$$
E(\omega) = \tanh(\beta h + \beta NJ E(\omega)) = E(\omega_i) \forall i. \quad (34)
$$

which implies the following theorem.

**Theorem 3. Relationship between global and local interactions models**

Any equilibrium expected individual and average choice level $m$ for the global interactions model is also an equilibrium expected individual and average choice in a homogeneous local interactions model.

This result might initially appear odd, given the explicit local interaction structure of preferences. In fact, the equivalence is not surprising. When all expectations are identical, and the sample mean is required to equal the population mean, then agents are all implicitly connected to one another through the expectations formation process. To be clear, the local interactions model can exhibit equilibria which are different from that of the global case.

Focus on the case where each individual is required to possess identical $E(\omega_i)$'s is not required by the logic of the local interactions model. There has been little work on the existence and characterization of asymmetric equilibrium $E(\omega_i)$'s, i.e. equilibria where the expected values differ across agents. Examples of asymmetric equilibria of this type may be found in Blume and Durlauf (1998b).
A trivial example can be produced by taking two environments which exhibit global interactions and multiple equilibria and defining them as a common population.

Finally, it is worth noting that when interactions between decisions are all intertemporal, then the assumption of extreme-valued random utility increments can be dropped. The equilibrium properties of the dynamic models in this section can be recomputed under alternative probability densities such as probit which are popular in the discrete choice work. In fact, under the mean field analysis of global interactions, alternative specifications incorporate probit or other densities as well. In both cases, the large scale properties of models under alternative error distributions are largely unknown.

iv. relationship to statistical mechanics

The models we have thus far outlined bear a close relationship to models in statistical mechanics. A standard question in statistical mechanics concerns how a magnet can exist in nature. A magnet is defined as a piece of iron in which a majority of the atoms are either spinning up or down. Since there is no physical reason why atoms should be more likely to spin up or down when considered in isolation, the existence of a natural magnet, which requires literally billions of atoms to be polarized towards one type of spin, would seem extraordinary unlikely by the law of large numbers. As a result, statistical mechanics models are based on the primitive idea that the probability that one atom has a given spin is an increasing function of the number of atoms with the same spin within the atom's neighborhood. For the Ising model of ferromagnetism, the assumption is that atoms are arrayed on a 2 – (or higher) dimensional integer lattice, so that

\[ \mu(\omega_i \mid \text{spins of all other atoms in material}) = \]
\[ \mu(\omega_i \mid \omega_j \text{ such that } |i - j| = 1) \sim \exp(\beta J \omega_i \sum_{|i - j| = 1} \omega_j) \] (35)

For the Curie—Weiss model, the physical interaction structure is assumed to be such that each atom’s spin is probabilistically dependent on the average spin in the system, so that

\[ \mu(\omega_i \mid \text{spins of all other atoms in material}) \sim \exp(\beta J \omega_i \bar{\omega}) \] (36)

Hence our models of binary choice with social interactions are mathematically quite similar to physical models of magnetism.

An important difference, however, does exist. While our socioeconomic model embeds pairwise interactions via the products of individual choices \( \omega_i \) with the expected choices of others, the physical models are based upon conditional probabilities which depend on the products of the realized individuals choices for all pairs of individuals. Interestingly, the physics literature has also dealt with expectations-based interactions. It turns out that models with interactions across realizations are extremely difficult to analyze, so physicists have developed what is referred to a “mean-field approximation” to various ferromagnetism models. A mean field approximation amounts to replacing certain terms in an original model with their mathematical expectation. Hence, the mean field approximation for the conditional probability of the spin of a given atom for the Curie-Weiss model is

\[ \mu(\omega_i) \sim \exp(\beta J \omega_i E(\bar{\omega})) \] (37)

which is of the same form as eq. (12) when agents possess identical \( \xi_i \)'s and \( J_{i,j} = \frac{J}{I} \). Of course, what is an approximate model in a physical context is an exact model in the socioeconomic context we have been analyzing, at least given our behavior primitives. This difference occurs because our behavior assumption is that individuals interact through their expectation's of one another's behavior,
rather than through realizations.

This last remark relates to a more general consideration in the use of statistical mechanics methods by social scientists. A basic conceptual difference exists between social and physical environments which contain interactions. Physical (and many) mathematical models of interactions typically take as primitives the conditional probabilities linking elements of a system, i.e. \( \mu(\omega_1 | \omega_{-1}), \ldots, \mu(\omega_1 | \omega_{-1}) \). Analysis of the model considers the existence and (if so) properties of whatever joint probability measures are consistent with the conditional ones. In socioeconomic contexts, it is more natural to take preferences, beliefs, and technologies as primitives and from them determine what conditional probability relationships will hold. Hence, statistical mechanics and related models cannot be employed in socioeconomic contexts without determining what socioeconomic primitives will lead to a particular conditional probability representation. Further, the purposefulness of the objects of analysis in social science contexts also means that issues of the endogeneity of neighborhoods and the potential for the existence of institutions which coordinate collective action will naturally arise. These issues have no analog in physical contexts and are suggestive of the limitations in importing methods from physics into socioeconomic studies.

v. social planning problem

Our analysis thus far has assumed that individual decisions are not coordinated. An alternative approach is to examine how decisions would be made when coordinated by a social planner. Beyond its use in developing welfare comparisons and other contrast with the noncooperative case, the social planner's solution may have empirical content in some contexts. As described by Coleman (1988,1990 chapter 12), the evolution of social capital, defined to include aspects of social structure which facilitate coordination across individuals and which may
be embedded either in personal mores or organizations such as churches or schools, implies that in many types of social situations, coordinated behavior can emerge.

In order to do this, it is necessary to be more precise in the formulation of the underlying game played by members of the population. As before, we consider a population of $I$ individuals each with payoff function $V(\omega_i; Z_i; \mu_i^\tau(\omega_{-i}), \epsilon_i(\omega_i))$. The random functions $\epsilon_i(\cdot)$ are assumed to be observed by the members of the population, so that each agent $i$ knows the realizations of $\epsilon_j(\cdot) \quad \forall \ j \neq i$. We further assume that the distribution of these random components is described by eq. (2). Hence in terms of timing, nature draws the random functions $\epsilon_i(\cdot)$ and reveals them to the entire population. Second, players play the game $G$ defined by

$$G = \{V(\omega_i; Z_i; \mu_i^\tau(\omega_{-i}), \epsilon_i(\omega_i)), i = 1...I\}$$

(38)

where $\mu_i^\tau(\omega_{-i})$ denotes their beliefs about the behavior of other agents and is conditioned on nature’s draw of the random functions.

With respect to this environment, an obvious benchmark is a perfect foresight Nash equilibrium. By this, we mean that each player knows the $\epsilon_i(\cdot)$ functions for every agent and forms beliefs about the resultant choices in the population $\mu_i^\tau(\omega_{-i})$ which are confirmed in equilibrium. If each player is playing a pure strategy, this means that $\mu_i^\tau(\omega_{-i}) = \omega_{-i}$ so that a perfect foresight pure strategy equilibrium is a set of choices $\omega$ such that for all $i$

$$\omega_i = \arg\max_{\gamma \in \{-1, 1\}} V(\gamma; Z_i; \omega_{-i}; \epsilon_i(\gamma)).$$

(39)

For the analogous mixed strategy equilibrium, let $\pi_i = (\pi_{i, -1}, \pi_{i, 1})$ denote the row vector of probability weights assigned by agent $i$ to the two choices. Then $\Pi_i = (\pi_1 ... \pi_I)$ denotes a perfect foresight Nash equilibrium if each $\pi_i$ is consistent with
\[ \pi_i = \arg \max_{\gamma_i} \] 
\[ \gamma_{i,1} V(1, Z_i \cup -i, \varepsilon_i(1)) + \gamma_{i,1} V(-1, Z_i \cup -i, \varepsilon_i(-1)) \] 

such that \( \gamma_{i,1} \geq 0 \) and \( \gamma_{i,1} + \gamma_{i,-1} = 1 \), 

where \( E_{\tilde{\gamma}} \) means that agent \( i \) plays the mixture \( \gamma_i \) against the mixtures played by the other agents, \( \cup_{-i} = (\mathcal{Z}_1, \ldots, \mathcal{Z}_{i-1}, \mathcal{Z}_{i+1}, \ldots, \mathcal{Z}_f) \). Mixture \( \gamma_i \) means that \( i \) chooses 1 with probability \( \gamma_{i,1} \) and chooses \( -1 \) with probability \( \gamma_{i,-1} \). It is a standard result that a mixed Nash equilibrium of this type will always exist, although a pure strategy Nash equilibrium may not.

Alternatively, a limited information Nash equilibrium can be characterized when agents make choices without knowledge of the \( \varepsilon_i(\cdot) \) functions for agent's other than themselves. In terms of timing, one can think of agents forming beliefs \( \mu^e_i(\mathcal{Z}_{-i}) \) before any \( \varepsilon_i(\cdot) \)'s are realized, nature then drawing the \( \varepsilon_i(\cdot) \)'s, revealing \( \varepsilon_i(\cdot) \) to agent \( i \), and each \( i \) then choosing \( \omega_i \). For this case,

\[ \omega_i = \arg \max_{\gamma \in \{-1,1\}} V(\gamma, Z_i \cup \mu^e_i(\mathcal{Z}_{-i}), \varepsilon_i(\gamma)), \]  

when \( \mu^e_i(\mathcal{Z}_{-i}) = \mu(\mathcal{Z}_{-i} | Z_j \forall j) \forall i \), so that each agents belief's are consistent with the model. This is the equilibrium concept we have employed above.

In contrast to these noncooperative environments, we may characterize a social planner's perfect foresight problem as choosing \( \mathcal{Z} \) in order to maximize total utility in the population, i.e.

\[ \max_{\mathcal{Z}} \sum_{i=1}^{L} V(\omega_i, Z_i, \mu^e_i(\mathcal{Z}_{-i}), \varepsilon_i(\omega_i)) \]  

From (42), one can in principle compute quantities such as the expected average payoff under a social planner and contrast it with their counterparts under the...
two noncooperative environments.

In order to perform such a comparison, however, analytical tractability becomes a problem. To see this, notice that for our global interactions model, the social planner's problem becomes

$$\max_{\omega} \sum_{i=1}^{I} \left( h_i \omega_i - \frac{I}{2}(\omega_i - \bar{\omega})^2 + \epsilon_i(\omega_i) \right)$$

Unfortunately, \(\sum_{i=1}^{I} \epsilon_i(\omega_i)\) is not independent and extreme value distributed over the \(2^I\) possible configurations of \(\omega\) even though the individual \(\epsilon_i(\omega_i)\)'s are distributed that way. One way around this problem is, following Brock and Durlauf (1995), to replace this original social planner's problem with an approximate problem

$$\max_{\omega} \sum_{i=1}^{I} \left( h_i \omega_i - \frac{I}{2}(\omega_i - \bar{\omega})^2 + \epsilon_i^*(\omega_i) \right)$$

where \(\sum_{i=1}^{I} \epsilon_i^*(\omega_i)\) is itself extreme value distributed. One can require that the variance of the errors in the approximate social planner's problem to equal those in the original problem in order to achieve some calibration between the two problems.

This approximate social planner's problem can be analyzed by replacing \(E_i(\bar{\omega}_I)\) with \(\bar{\omega}_I\) in eq. (19). The probability measure characterizing the joint choice of \(\omega\) equals

$$\mu(\omega) = \frac{\exp(\beta(\sum_{i=1}^{I} h_i \omega_i + \frac{I}{2}(\sum_{i=1}^{I} \omega_i)^2))}{\sum_{\nu_1 \in \{-1, 1\}} \sum_{\nu_I \in \{-1, 1\}} \exp(\beta(\sum_{i=1}^{I} h_i \nu_i + \frac{I}{2}(\sum_{i=1}^{I} \nu_i)^2))}.$$

In order to analyze this probability measure, which is known in the statistical mechanics literature as the Curie-Weiss model, it is necessary to
eliminate the \((\sum_{i} \omega_i)^2\) terms in eq. (45). This calculation is complicated and may be found in the Mathematical Appendix; further analysis appears in Brock (1993). A result currently exists only for the case \(h_i = h\) and only for the large economy limit; however, Amaro de Matos and Perez (1991) suggests that for the large economy limit generalization to heterogeneous \(h_i\)’s is possible. The appendix verifies Theorem 4.

**Theorem 4.** Expected average choice under social planner for binary choice model with interactions

Let \(m^*\) denote the root of \(m^* = \tanh(\beta h + \beta Jm^*)\) with the same sign as \(h\). If eq. (39) characterizes the joint distribution of individual choices as determined by a social planner, then

\[
\lim_{I \to \infty} E(\bar{\omega}_I) = m^*.
\]

(46)

One aspect of this theorem is intuitive, in that a planner would choose that average choice level in which the interaction effects and the private deterministic utility comparisons work together. What is perhaps surprising is that the social planner’s equilibrium is sustainable as an equilibrium in the limited information noncooperative environment. However, this result is somewhat special to the functional form originally assumed for individual deterministic social utility. If the original social utility term had been \(J \omega_i E(\bar{\omega}_I)\), then the noncooperative equilibrium average choice level would be the same as for the case we have studied, but the analogous social planner’s problem would choose that root of \(m^* = \tanh(\beta h + 2\beta Jm^*)\) with the same sign as \(h\) (Brock and Durlauf (1995)), which would mean it is not supportable in the limited information noncooperative environment.
vi. linear-in-means model

Much of the empirical work on interaction effects has assumed that the behavior variable \( \omega_i \) has continuous support and depends linearly on various individual and neighborhood effects. These assumptions permit a researcher to use ordinary least squares methods, which will be discussed below. While these empirical papers generally do not consider what decision problems generate their econometric specifications, it is straightforward to do so. For a trivial example, suppose that an individual solves

\[
\max_{\omega_i \in (-\infty, \infty)} -\frac{1}{2} (\omega_i - \omega_i^*)^2
\]

(47)

where \( \omega_i^* \) is a reference behavior level to which individual \( i \) prefers to conform. When this reference behavior level equals \( h_i + JE_i(\bar{\omega}_I) + \epsilon_i \), it is immediate that

\[
\omega_i = h_i + JE_i(\bar{\omega}_I) + \epsilon_i.
\]

(48)

This is the type of equation studied by Manski (1993), Moffitt (1998), Duncan and Raudenbusch (1998), among others.

2. Identification: Basic issues

In this section, we describe the identification of interactions-based models in cross-sections. Identification is a concern in these cases because of the likelihood that group versus individual determinants of individual behavior are likely to be correlated. Hauser (1970) provides an early and clever analysis of how these correlations can, if not properly accounted for, lead to spurious inferences. We recommend this paper as an example of how powerful intuitive reasoning (as well as good common sense) can complement and foreshadow formal analysis.
Manski (1993a,b,1997) has pioneered the study of the identification of interaction effects, and we will follow his treatment closely. In his work, Manski, distinguishes between three explanations for correlated behavior within groups:

"endogenous effects, wherein the propensity of an individual to behave in some way varies with the behaviour of the group...exogenous (contextual) effects, wherein the propensity of an individual to behave in some way varies with the exogenous characteristics of a group...correlated effects, wherein individuals in the same group tend to behave similarly because they have similar individual characteristics or face similar institutional environments" (Manski (1993a), pg. 532)

The treatment of identification problems in terms of the ability to distinguish these different effects in data seems to us very useful and so we employ it throughout.

For purposes of discussing identification and other econometric aspects of interaction-based models, we begin with a baseline set of data assumptions which will apply both to the binary choice model and to the linear-in-means model. We assume that the econometrician has available a set of observations on \( I \) individuals. We assume that each individual is drawn randomly from a set of neighborhoods. Within each neighborhood, all interactions are global. For notational purposes, we denote individuals as \( i \) and the neighborhood (which means the set of other individuals who influence \( i \) through interactions) as \( n(i) \). We assume that our original \( z_i \) vector can be partitioned into an \( r \)-length vector of individual-specific observables \( X_i \) and an \( s \)-length vector of exogenously determined neighborhood observables \( Y_{n(i)} \) associated with each individual in the sample. This will allow us to replace the private utility component \( h_i \) in our theoretical discussion with a linear specification

\[
h_i = k + c'X_i + d'Y_{n(i)}.
\]

Notice that this specification means that none of the individual-specific observables \( X_i \) or neighborhood observables \( Y_{n(i)} \) contains a constant term. We
will maintain this assumption throughout. Within a neighborhood, all interactions are assumed to be global and symmetric, so that there is a single parameter \( J \) which indexes interactions.

Recall that \( m^e_{n(i)} \) is agent \( i \)'s subjective expectation of the average choice in neighborhood \( n(i) \). In the subsequent discussion, it will be useful to distinguish between \( m^e_{n(i)} \) and \( m_{n(i)} \), the mathematical expectation of the average choice in a neighborhood under self-consistency. (We will specify the information sets under which self-consistency is calculated below.) The reason for this is that we will have need to distinguish between the data in a statistical exercise and the mathematical solution to a model. Of course, \( m^e_{n(i)} = m_{n(i)} \) is part of our maintained assumption in the analysis, so there is no loss of generality in doing this. For purposes of discussion of identification, we are therefore either implicitly assuming that the neighborhoods are arbitrarily large, so that the neighborhood sample average can be used in place of the expected value or that accurate survey data are available. We finally assume that the errors are independent across individuals and that \( \mu(\epsilon_i(\omega_i) - \epsilon_i(-\omega_i) \mid X_i; X_{n(i)}; m^e_{n(i)}) = \mu(\epsilon_i(\omega_i) - \epsilon_i(-\omega_i)) \) for the binary choice model, and \( E(\epsilon_i \mid X_i; X_{n(i)}; m^e_{n(i)}) = 0 \) for the linear-in-means model.

Our strategy of using individual level data has an important advantage: to the extent that the parameters of the individual model are identified, one can infer whether or not multiple equilibria exist with respect to population aggregates. This can be done without consideration of an equilibrium selection rule because population aggregates are always treated as independent variables in the analysis. Hence we can circumvent some of the problems described in Jovanovic (1989).

\section*{i. binary choice}

For the binary choice model, we consider the identification based on a
naive estimator of the parameters of the model. By naive, we refer to the case where a logistic regression is computed which does not impose the relationships between neighborhood means. In this case, the conditional likelihood function for the set of individual choices will have a standard logistic form. Using our theoretical model of global interactions (and exploiting symmetry of the logistic density function), the likelihood is

\[
L(\omega | X_i, Y_{n(i)}, m_{n(i)}^e, \forall i) = \\
\prod_i \mu(\omega_i = 1 | X_i, Y_{n(i)}, m_{n(i)}^e) \cdot \mu(\omega_i = -1 | X_i, Y_{n(i)}, m_{n(i)}^e) \\
= \prod_i \exp(\beta k + \beta c X_i + \beta d Y_{n(i)} + \beta J m_{n(i)}^e) \cdot \exp(-\beta k - \beta c X_i - \beta d Y_{n(i)} - \beta J m_{n(i)}^e)
\]

(50)

As is standard for logistic models, the complete set of model parameters is not identified as \( k, c', d' \) and \( J \) are each multiplied by \( \beta \). We therefore proceed under the normalization \( \beta = 1 \).

The reason that identification is a concern in a model like this is the presence of the term \( m_{n(i)}^e \) in the likelihood function. Since this term embodies a rationality condition, it is a function of other variables in the likelihood function. Specifically, we assume that

\[
m_{n(i)}^e = m_{n(i)} = \int \tanh(k + c' X_i + d' Y_{n(i)} + J m_{n(i)}) dF_X | Y_{n(i)}.
\]

(51)

Here \( F_X | Y_{n(i)} \) denotes the conditional distribution of \( X \) in neighborhood \( n(i) \) given the neighborhood characteristics \( Y_{n(i)} \). What this means is that each agent is assumed to form the conditional probabilities of the individual characteristics in a neighborhood given the aggregates which determine his or her payoffs. Since one
can always add elements of $\mathcal{X}_{n(i)}$ with zero coefficients to the payoff equation for agents, this is without loss of generality.

Rather than prove identification for the particular case where the theoretical model is logistic (see McFadden (1974) and Amemiya (1993, chapter 9) for proofs for this case) we prove identification for an arbitrary known distribution function for the random payoff terms. Specifically, we assume that the conditional probability of individual $i$'s choice can be written as

$$
\mu(\epsilon(\omega_i) - \epsilon(-\omega_i) \leq z \mid X_i, Y_{n(i)}, m_{n(i)}^e) = F(z \mid k + c'X_i + d'Y_{n(i)} + Jm_{n(i)}^e)
$$

where $F$ is a known probability distribution function that is continuous and strictly increasing in $z$.

We consider identification based on a naive estimator of the parameters of the model. By naive, we refer to the situation where parameter estimates for the model are computed which do not impose the rational expectations condition between neighborhood means and neighborhood characteristics, but rather uses these variables as regressors. Hence, we assume that $m_{n(i)}^e$ is known to the researcher; see discussion below for the case when $m_{n(i)}^e$ is not observable.

To formally characterize identification, we employ the following notation. Define $supp(X, Y, m^e)$ as the joint support of the distribution of $(X_i, Y_{n(i)}, m_{n(i)}^e)$. Intuitively, the definition of identification we employ says that a model is identified if there do not exist two distinct sets of parameter values each of which produces (for all subsets of $X$ and $Y$ which occur with positive probability) identical probabilities for individual choices and which are also self-consistent.

**Definition.** Global identification in the binary choice model with interactions and self-consistent expectations

The binary choice model is globally identified if for all parameter pairs $(k, c, d, J)$
and \((\bar{k}, \bar{c}, \bar{d}, \bar{J})\)

\[
k + c'X_i + d'Y_{n(i)} + Jm^e_{n(i)} = \bar{k} + \bar{c}'X_i + \bar{d}'Y_{n(i)} + \bar{J}m^e_i
\]  

(53)

and

\[
m^e_{n(i)} = m_{n(i)} = 
\int \omega_i dF(\omega_i | k + c'X_i + d'Y_{n(i)} + Jm_{n(i)})dF_X | Y_{n(i)} = 
\int \omega_i dF(\omega_i | \bar{k} + \bar{c}'X_i + \bar{d}'Y_{n(i)} + \bar{J}m_{n(i)})dF_X | Y_{n(i)}
\]  

(54)

\[\forall (X_i; Y_{n(i)}, m^e_{n(i)}) \in \text{supp}(X, Y, m^e) \text{ imply that } (k,c,d,J) = (\bar{k}, \bar{c}, \bar{d}, \bar{J})\]

In order to establish conditions under which identification can hold we follow the argument in Manski (1988), Proposition 5, and state the following Proposition, whose proof appears in the Mathematical Appendix. The assumptions we make are clearly sufficient rather than necessary; weakening the assumptions is left to future work. In interpreting the assumptions, note that Assumption i is the one used by Manski to identify this model when there are no endogenous effects, i.e. if \(J\) is known a priori to be 0. The assumption, of course does nothing more than ensure that the individual and contextual regressors are not linearly dependent. The additional assumptions are employed to account for the fact that \(m_{n(i)}\) is a nonlinear function of the contextual effects.

**Theorem 5. Sufficient conditions for identification to hold in the binary choice model with interactions and self-consistent beliefs**

Assume
i. $\text{supp}(X_i, \mathcal{X}_{n(i)})$ is not contained in a proper linear subspace of $\mathbb{R}^r + \mathbb{R}^s$.

ii. $\text{supp}(\mathcal{X}_{n(i)})$ is not contained in a proper linear subspace of $\mathbb{R}^s$.

iii. No element of $X_i$ or $\mathcal{X}_{n(i)}$ is constant.

iv. There exists at least one neighborhood $n_0$ such that conditional on $\mathcal{Y}_{n_0}$, $X_i$ is not contained in a proper linear subspace of $\mathbb{R}^r$.

v. None of the regressors in $\mathcal{Y}_{n(i)}$ possesses bounded support.

vi. $m_{n(i)}$ is not constant across all neighborhoods $n$.

Then, $(k, c, d, J)$ is identified relative to any distinct alternative $(\bar{k}, \bar{c}, \bar{d}, \bar{J})$.

**ii. linear-in-means model**

Identification in the binary choice model with interactions can be contrasted with the case of the analogous linear-in-means model,

$$
\omega_i = k + c'X_i + d'\mathcal{Y}_{n(i)} + Jm_{n(i)}^e + \epsilon_i.
$$

(55)

The unique self-consistent solution $m_{n(i)}$ for the linear-in-means model is easily seen, by applying an expectations operator to both sides of the individual behavioral equation, to be

$$
m_{n(i)} = \frac{k + c'\mathbb{E}(X_i | \mathcal{Y}_{n(i)}) + d'\mathcal{Y}_{n(i)}}{1 - J}.
$$

(56)

where $\mathbb{E}(X_i | \mathcal{Y}_{n(i)})$ denotes the expected value of the individual controls given the
neighborhood characteristics. Hence, following the argument in Manski (1993),
one can construct a reduced form expression for individual choices,

$$\omega_i = \frac{k}{1 - J} + c' X_i + \frac{J}{1 - J} d' Y_{n(i)} + \frac{J}{1 - J} c' E(X_i | Y_{n(i)}) + \epsilon_i.$$  \hspace{1cm} (57)

In this equation, we have $2r + s + 1$ regressors and $r + s + 2$ parameters. The
possibility for identification in this model therefore will depend on which, if any,
of the regressors in the reduced form are linearly independent (i.e. their variance
covariance matrix is of full rank). For example if $E(X_i | Y_{n(i)})$ is linearly
dependent on $Y_{n(i)}$, then it is obvious that the model parameters are not
identified. More generally, it is necessary for identification that the dimension of
the linear space spanned by the regressors is at least equal to the number of
structural parameters, i.e. $r + s + 2$; otherwise, one cannot map the reduced form
coefficients back to the structural parameters. Hence, one can state the following
theorem.

**Theorem 6.** Necessary conditions for identification in the linear-in-means model
with interactions and self-consistent beliefs

In the linear-in-means model it is necessary for identification of the model’s
parameters that

i. The dimension of the linear space spanned by elements of $(1, X_i, Y_{n(i)})$ is
$r + s + 1$.

ii. The dimension of the linear space spanned by the elements of
$(1, X_i, Y_{n(i)}), E(X_i | Y_{n(i)})$ is at least $r + s + 2$.

Notice that the conditions of this theorem, while analogous to those in the
theorem for identification in the binary choice model, are now necessary and not
sufficient. This is because sufficient conditions will depend on the model parameters. For example, if \( c = 0 \), then the fact that \( E(X_i | Y_{n(i)}) \) is linearly independent of the regressors \( X_i \) and \( Y_{n(i)} \) will not eliminate collinearity of \( m_{n(i)} \) and \( Y_{n(i)} \) in the structural equation (55) and hence will leave only \( s + 1 \) regression coefficients in the reduced form available to identify \( k, J \) and \( d \), which is not enough.

This theorem is an extension of Manski's (1993) result on the nonidentifiability of contextual versus endogenous effects. Manski's analysis assumes that there is a one-to-one correspondence between the individual control variables \( X_i \) and the neighborhood control variables \( Y_{n(i)} \) so that for any individual-level variable that influences behavior, the neighborhood average of that variable also influences behavior. For example, if one controls for individual education, one also controls for average neighborhood education. In this case, \( E(X_i | Y_{n(i)}) = Y_{n(i)} \). Hence, \( m_{n(i)} \) is linearly dependent on \( Y_{n(i)} \) and so the model is not identified. Notice as well that the Theorem requires that \( E(X_i | Y_{n(i)}) \) is a nonlinear function of \( Y_{n(i)} \); this is analogous to the condition for identification of some interaction effect in Manski (1993), Proposition 1 and Corollary. (Manski's results have to do with the identification of either an endogenous or contextual effect in the presence of individual effects, but does not allow for identification between these two effects, whereas our result gives conditions under which the two group effects can be distinguished.)

Why is there this difference between the binary choice and the linear-in-means frameworks? The answer is that the binary choice framework imposes a nonlinear relationship between the group characteristics and the group behaviors whereas the linear-in-means model (of course) does the opposite. Intuitively, suppose that one moves an individual from one neighborhood to another and observes the differences in his behavior. If the characteristics and behaviors of the neighborhoods always move in proportion as one moves across neighborhoods, then clearly one could not determine the respective roles of the characteristics as opposed to the behavior of the group in determining individual outcomes. This
can never happen in the logistic binary choice case given that the expected average choice must be bounded between $-1$ and $1$. So, for example, as one moves across a sequence of arbitrarily richer communities, the percentage of high school graduates cannot always increase proportionately with income.

One can develop analogous identification conditions for alternative information assumptions in the linear-in-means model. For example, suppose that $\bar{X}_{n(i)}$, the sample average of the individual characteristics in neighborhood $n(i)$, is known to all members of the neighborhood. In this case,

$$m_{n(i)} = \frac{k + c' \bar{X}_{n(i)} + d' \bar{Y}_{n(i)}}{1 - J}.$$  \hspace{1cm} (58)

This equation makes clear that if the elements of $\bar{X}_{n(i)}$ lie in the linear space spanned by $\bar{Y}_{n(i)}$, then the linear-in-means model will not be identified. Hence, we have the following corollary.

**Corollary 3.** Necessary conditions for identification in the linear in means model when $\bar{X}_{n(i)}$ and $\bar{Y}_{n(i)}$ are observable

If $\bar{X}_{n(i)}$ and $\bar{Y}_{n(i)}$ are observable, then a necessary condition for identification in the linear-in-means model is that the dimension of the linear space spanned by $(1, X_i, Y_{n(i)}, \bar{X}_{n(i)}, \bar{Y}_{n(i)})$ is at least $r + s + 2$.

Operationally, this corollary means that for the full information case, one needs one individual variable whose neighborhood level average is not an element of the individual behavioral equation. This average can then be used to instrument $m^e_{n(i)}$.

**iii. instruments for unobservable expectations**
The identification condition for the linear-in-means model suggests a set of instruments which may be used when \( m_{n(i)}^e \) is not observable, is measured with error, etc. Specifically, replacing \( m_{n(i)}^e \) with the projection of \( \omega_{n(i)} \), the sample average of behaviors in neighborhood \( n(i) \) onto \( H(\mathcal{X}_{n(i)}; E(\mathcal{X}_i | Y_{n(i)})) \), where \( H(a,b) \) denotes the Hilbert space generated by the elements of vectors \( a \) and \( b \), will not affect our identification results so long as \( \dim(H(\mathcal{X}_{n(i)}; E(\mathcal{X}_i | Y_{n(i)})) \ominus H(Y_{n(i)})) > 0 \), where for Hilbert spaces \( I \) and \( G \) such that \( G \subseteq I \), \( I \ominus G \) denotes the Hilbert space generated by those elements of \( I \) that are orthogonal to all elements of \( G \). An analogous procedure will apply when \( \mathcal{X}_{n(i)} \) is observable to individuals.

Of course, this assumes that the researcher has prior knowledge of what individual-level variables affect behavior when their neighborhood averages do not; otherwise, it would be the case that \( H(E(\mathcal{X}_i | Y_{n(i)})) \subseteq H(Y_{n(i)}) \) and so may be susceptible to Sim’s (1980) classic critique of “incredible” identifying restrictions; see Freedman (1991) for a similar critique of the sorts of regressions we describe here. The point remains, however, that identification in the linear-in-means model depends on the same classical conditions as does identification in general simultaneous equations models, as initially recognized by Moffitt (1998).

At the same time, we would argue that the issue of omitted variables is far from insuperable. Both the social psychology and sociology literatures have focused a great deal of attention as to which types of individual and group control variables are most appropriate for inclusion in individual level regressions through the determination of which variables seem to be proximate versus ultimate causes of individual behavior; indeed it is this distinction which is the basis of path analysis (Blau and Duncan (1967)); see Sampson and Laub (1995) for what we consider a persuasive example of such a study. In general, we find it likely that these literatures will be able to identify examples of individual variables whose group average analogs are not proximate causes of behavior, and hence are available as instruments. While these literatures are often not driven by formal
statistical modelling and further subjected to Sims/Freedman-type critiques (e.g. Freedman (1991)) when formal techniques are employed, this hardly means that these literatures are incapable of providing useful insights. In this respect, we find arguments to the effect that because an empirical relationship has been established without justification for auxiliary assumptions such as linearity, exogeneity of certain variables, etc., one can ignore it, to be far overstated. In our view, empirical work establishes greater or lesser degrees of plausibility for different claims about the world and therefore the value of any study should not be reduced to a dichotomy between full acceptance or total rejection of its conclusions. Hence the determination of the plausibility of any exclusion restriction is a matter of degree and dependent on its specific context, including the extent to which it has been studied.

iv. identification of individual versus neighborhood contextual effects

We now consider in more detail what is involved for identification of some type of neighborhood effects. What we mean is the following. Suppose that one wishes to determine whether any type of neighborhood effect exists, without distinguishing between endogenous and contextual effects, hence the only regressors in the model are a constant, \( X_i \), and \( X_{n(i)} \). Operationally, we define this as determining whether, for a statistical model which only includes contextual effects as controls, the parameters on these contextual effects are identified.\(^2\) A Corollary of the general identification Theorems 5 and 6 highlights the two conditions necessary to distinguish individual versus neighborhood contextual effects. A related result may be found in Manski (1993), corollary, pg. 535.

Corollary 4. Identification of individual versus neighborhood effects in the binary

\(^2\)When \( X_{n(i)} = E(X_i | i \in n(i)) \) the corollary can be interpreted as applying to identification of a group effect for the reduced form of the linear-in-means model.