Discrete Choice with Generalized Social Interactions

Oscar Volpe*

June 2, 2025

[Frequently updated. **Access the latest version**.]

Abstract

This paper examines how individual identity influences group behavior through social interactions. I study a discrete choice model in which people are affected differently by different members of their network, conforming to the actions of some peers while deviating from the actions of others. Under this generalized framework, I explore what aggregate outcomes arise from noncooperative decisionmaking. I analyze uniqueness and stability of equilibria, and I characterize how negative spillovers impact social welfare. I then show how to take the model to data, introducing a novel identification strategy that leverages within-network variation in individual characteristics to account for unobserved network effects. I also show how to construct internal instruments to overcome the issue of measurement error, which is a primary source of endogeneity in models with incomplete information. Lastly, I apply my method to data from the large-scale education experiment Project STAR, where I find strong evidence that classroom peer effects differ by gender.

Key Words: social interactions, identity, polarization, strategic complementarity, non-negative matrix, contextual effects, measurement error, nonparametric identification, Project STAR.

I. Introduction

The role of social interactions in individual decisionmaking has received widespread attention in economics. This work has largely focused on settings with uniform strategic complementarities, where all agents seek to conform to the average action in the rest of their network. Meanwhile, less attention has been given to cases with nonuniform interactions, where agents conform to some people, while deviating from others. These types of

I am grateful to Steven Durlauf, whose guidance and encouragement has made this project possible. I also thank Isaiah Andrews, Stéphane Bonhomme, Ozan Candogan, Arun Chandrasekhar, Michael Dinerstein, Ali Hortaçsu, Robert Moffitt, Magne Mogstad, Philip Reny, and Alex Torgovitsky for their insightful comments, as well as participants in the North American Urban Economics Association Meeting, the Applied Micro working group, IO working group, and third year research seminar at the University of Chicago. All errors are my own.

*Department of Economics, University of Chicago, 1126 East 59th Street, Chicago, IL 60637, USA. Email: ovolpe@uchicago.edu.

interactions are more challenging to study, particularly when agents make discrete choices, since the nonlinear dependence between outcomes and potential for multiple equilibria can complicate the theoretical analysis of economic outcomes and also pose challenges to identification and estimation. Nevertheless, nonuniform interactions appear in many real-world contexts. Consider, for example, the impact of gender on peer effects in schools. Male and female students often face pressure to conform to members of their own gender, while there may be less pressure to conform across genders. In some cases, students might even wish to distinguish themselves from the opposite gender—a preference that can be modeled by a negative interaction effect. As Hoxby (2000) and others discuss, these social influences may contribute to gender differences in behavior and academic achievement. More generally, as Akerlof & Kranton (2000) argue, social identities—such as gender, race, class, and political affiliation—are relevant for most everyday decisions that people make and are therefore likely to inform a wide range of economic outcomes. Given these considerations, it is worth understanding how nonuniform, and even negative, interactions shape economic behavior.

In this paper, I develop a framework for studying generalized social interactions, where agents are affected differently by different people in their network. I focus on a model with binary choices where agents experience positive and negative spillover effects. I characterize what aggregate outcomes can occur in equilibrium, and I show how the model is empirically tractable. This work may be especially useful for understanding the impact of identity on social and economic behavior, as well as the emergence of polarized or segregated networks.

I begin my analysis by modeling a network of agents who make binary choices subject to social influences and private, idiosyncratic preferences. Following Brock & Durlauf (2001), I assume that agents act with incomplete information, such that they are influenced by the expected average outcomes in the network, rather than by every person's realized outcome. However, unlike previous work, I assume that the social interaction effects differ, and might even be negative, across agents. To achieve this objective, I categorize the agents into different social groups (or "identities"). These classifications play a role in determining private payoffs, and they also form the basis for nonuniform social interactions. In particular, the interaction effects can be expressed in terms of a matrix $\mathbf{J}$, where each entry $J_{k\ell}$ indicates how an agent in group k is influenced by the expected average action of people in group ℓ.

I examine equilibrium properties of the model under noncooperative decisionmaking. These properties are well-studied in settings with uniform complementarities, where the spillover effects are all positive; e.g., consider Glaeser, Sacerdote & Scheinkman (1996, 2003), Brock & Durlauf (2001), Calvo-Armengol, Patacchini & Zenou (2009), and Cabrales, Calvo-Armengol & Zenou (2011), among others. In these settings, the interactions produce social multipliers, which can lead to large differences in aggregate outcomes across similar populations. Also, models with uniformly positive interactions may exhibit multiple, locally stable equilibria, which are generally well-ordered, in the sense that they form a complete lattice.[1]

[1]The microeconomic foundations for models of complementarity are largely developed by Donald Topkis (see Topkis (1998) for a summary), Vives (1990), and Milgrom & Roberts (1990). This work leverages key results

Overall, there is less known about the equilibrium properties of models with both positive and negative interaction effects. One issue is that the tools economists use to analyze settings with complementarities do not readily transfer to settings with strategic substitutabilities, i.e., negative spillovers. For example, models involving negative interaction effects may fail to possess a pure strategy Nash equilibrium. Moreover, even when equilibria do exist, it may be that none of them are dynamically stable, which means that they are unlikely to be observed by researchers.[2] In much of the networks literature, e.g., Ballester et al. (2006), Galeotti et al. (2010), Bramoullé et al. (2014), contraction mapping arguments are used to prove that there exists a unique, stable equilibrium if the spillover effects are sufficiently weak. I derive a similar, albeit slightly more general, result for my model by using an index theorem. However, the bulk of my theoretical analysis focuses on cases with strong interaction effects where unstable equilibria arise and where aggregate outcomes become harder to interpret.

I identify two conditions—one strictly weaker than the other—that each guarantees the existence of a locally stable equilibrium in contexts with strong interaction effects. Both conditions have meaningful economic interpretations. In either a strict or weak sense, they require that agents are not repelled by their own social group. Using these two conditions, I show how key results in the complementarities literature extend to a broader class of models that involve substitutabilities. I start by establishing necessary and sufficient conditions for the existence of multiple equilibria. I show that multiplicity depends on a single parameter: the spectral radius of the Jacobian matrix of the equilibrium system. This number measures the intensity of the cycles in a network, thereby quantifying the cumulative strength of the interaction effects. I find that a unique, stable equilibrium exists if the spectral radius is below a certain threshold, while multiple, locally stable equilibria exist if the spectral radius is above this threshold. Next, I consider the social welfare of agents at different equilibria. I find that negative interaction effects can introduce welfare trade-offs, such that it impossible to jointly maximize the expected utility of agents in different social groups. Specifically, if two agents who otherwise prefer the same action are negatively influenced by one another, then the equilibrium that maximizes welfare for one may not maximize welfare for the other. This result is instructive for policy analysis as it carries implications about social inefficiency.

In the second half of the paper, I turn to the question of identification. That is, given data on individual choices across multiple networks, what can be said about the role of social interactions? The identification of network-based discrete choice models with incomplete information is carefully studied by Brock & Durlauf (2001, 2007), and more recently by Bhattacharya et al. (2023). Paula (2017) and Kline & Tamer (2020) also give recent reviews of

from lattice theory, e.g., Tarski's fixed point theorem, to study equilibrium behavior in settings with uniform complementarities. Cooper & John (1988) also provide an early discussion about the role of complementarity in economics. Additional contributions have been made by Milgrom & Shannon (1994) and Athey (2001, 2002).

[2] Jackson & Zenou (2015) describe several unresolved questions regarding how and why network-based models with substitutabilities behave differently from models with uniform complementarities. Of course, strategic substitutabilities appear frequently throughout economics, albeit outside of the social interactions literature; e.g., see Bramoullé & Kranton (2007), Bramoullé et al. (2014), and Elliott & Golub (2019) for recent contributions.

the literature.[3] My analysis builds on this work by tackling two key identification problems.

First, I address the issue of unobservable network effects. These contextual factors may prevent researchers from uncovering the role of social interactions. For example, when comparing student achievement across different classrooms, it is hard to distinguish the impact of peer effects from unobserved teacher quality or unknown class characteristics. Much of the applied literature tends to rule out these unknown network factors.[4] Alternatively, some papers have used a panel to difference-out the fixed effects between time periods; e.g., see Hoxby (2000) and Brock & Durlauf (2007). However, this approach relies on having access to panel data, as well as the assumption that the model parameters do not change over time.

To handle the issue of unobservable network effects, I introduce a novel approach that allows for the partial identification of social interactions. Importantly, this method places no restrictions on the network-level determinants of people's actions. Rather, it leverages a panel structure that is inherent in the model, whereby members of different social groups interact in the same network. Since these agents face the same contextual factors, I can control for network effects by comparing the outcomes of different social groups in a given network. Using this approach, I recover the differences between any two social interaction effects, which I can then use to measure the amount of polarization in a network.[5] This technique of exploiting within-network variation in individual-level characteristics is new to the social interactions literature. However, similar approaches have been proposed in other areas of economics. For example, Berry & Haile (2022) show how to use data about heterogeneity within markets to reduce the number of instruments needed to estimate systems of demand.

Next, I address an issue that is common to most network-based models with incomplete information, which is that the expected average outcomes in a network are never directly observed. Instead, a researcher only sees the average actions among finitely-many agents. Since these observed averages converge to the true expectations as the networks grow large, previous research has treated this issue as an estimation problem rather than as a barrier to identification; e.g., see Bhattacharya et al. (2023). These estimation strategies rely on double asymptotic arguments, such that the number of networks and the size of each network must both tend to infinity. I argue that these approaches are not advisable given that replacing expectations with observed averages always leads to measurement error. Indeed, I show that, in any setting with finite networks, the observed average outcomes are endogenous in the model. Moreover, by failing to account for the endogeneity, most common estimation methods would produce biased estimates—even as the number of networks tends to infinity.

I approach the issue as a classical measurement error problem (Wooldridge, 2013, sec.

[3] Blume et al. (2011) also review issues related to identification of linear and nonlinear network-based models.

[4] For example, Brock & Durlauf (2001, 2007) assume that the network effect is a constant linear function of observed variables, whereas Bhattacharya et al. (2023) assume that it exhibits a specific linear factor structure.

[5] My definition of polarization contributes to a burgeoning literature documenting how people's choices are divided along cultural or political lines; e.g., see Boxell et al. (2022) and Bertrand & Kamenica (2023). In this literature, there is still no clear consensus on how to measure polarization as it is a relatively abstract concept. One benefit of my approach is that I measure polarization in a way that is motivated by an economic model.

15.4). Specifically, I note that the observed average outcomes are noisy approximations of the true expectations and are therefore endogenous. To correct for this endogeneity, I use internal instruments. This procedure involves randomly partitioning the network into two parts and then computing the sample average action within each part. Using these sample averages, I treat one as the endogenous variable and the other as an instrument. Since each network is partitioned randomly, this IV strategy ensures that the model parameters are identified even for small networks. I then define an IV estimator and prove it is consistent. I also demonstrate the efficacy of the estimation strategy by running Monte Carlo simulations.

Finally, I provide an empirical application of the model and identification strategy using data from the class size reduction experiment Project STAR. Prior research has used this data to measure peer effects in classrooms, e.g., see Boozer & Cacciola (2001) and Graham (2008). However, I leverage the generalized interactions framework to study a new question: How do peer effects differ by gender? For this application, I find evidence of significant gender differences in peer effects, where male and female students are both more likely to conform to members of their own gender than to members of the opposite gender. This finding is notable as it contributes to a longstanding literature about the impacts of gender and social pressure on academic achievement, e.g., see Hoxby (2000), Lavy & Schlosser (2011), and Bostwick & Weinberg (2022). While previous work has primarily focused on the effects of gender composition, I explore how students are directly affected by the expected achievement of their male and female peers. Moreover, I perform this analysis while accounting for unobserved classroom-level determinants of student outcomes. Overall, my empirical findings illustrate the advantages of using a generalized interactions framework to learn about systematic heterogeneity in social interactions. They also contribute to a growing literature about identity in economics; see Charness & Chen (2020) and Shayo (2020) for discussions.

This paper proceeds as follows. Section II describes the binary choice model. Section III explores the equilibrium properties of the model, such as existence, uniqueness, dynamic stability, and social welfare. Section IV considers alternative specifications and also explores how the equilibrium properties generalize to a broader class of models. Section V explains the identification strategy, along with the estimation procedure and the simulation results. Section VI presents the empirical application and key findings. Lastly, Section VII concludes.

II. A Model with Generalized Social Interaction Effects

I study a binary choice model with interaction effects that vary on the basis of group identity. This model extends Brock & Durlauf's (2001) binary choice framework by assuming that individuals are influenced differently—perhaps even negatively—by different people. For example, a person might seek to conform to the average behavior in certain groups, while distinguishing herself from others. Unlike previous work, I allow the distribution of idiosyncratic preferences to be nonparametric. This feature ensures that the equilibrium properties of the model will be robust under relatively loose functional form assumptions.

II.A. Individual Preferences

Consider a network of I agents, where each agent i belongs to one of K social groups. Every agent chooses a binary action $\omega_i \in \{0, 1\}$ at a common time. Let $\bar{\omega}^k$ be the average action in group k, and let $\bar{\omega}^k_{-i}$ be the average action among members of group k excluding i.

When forming decisions, people are influenced by the expected behavior of everyone else in their network. For any person i in group k, the utility from choosing an action ω_i is:

$$U_i(\omega_i|k) = v_k(\omega_i) + J_{kk}\omega_i \,\mathrm{E}_i(\bar{\omega}^k_{-i}) + \sum_{\ell \neq k} J_{k\ell}\omega_i \,\mathrm{E}_i(\bar{\omega}^\ell) + \epsilon_i(\omega_i). \tag{1}$$

Here, $\mathrm{E}_i(\bar{\omega}^k_{-i})$ and $\mathrm{E}_i(\bar{\omega}^\ell)$ represent i's subjective expectations about $\bar{\omega}^k_{-i}$ and $\bar{\omega}^\ell$, respectively. The term $v_k(\omega_i)$ is the private utility associated with a choice. This utility may vary by group membership. Finally, $\epsilon_i(\omega_i)$ is an idiosyncratic preference that is independent across agents.

Under this framework, utility exhibits proportional spillovers, so there is a multiplicative interaction between an agent's choice and the expected average choice in every group.[6] Each term $J_{k\ell}$ captures how much members of group k seek to conform to the mean behavior in ℓ.

Since the action is binary, I can write $v_k(\cdot)$ as an affine function: $v_k(\omega_i) = h_k\omega_i + \eta_k$. Also, I can write $\epsilon_i(\omega_i) = \varepsilon_i\omega_i + \xi_i$ without loss of generality, where ε_i and ξ_i are random coefficients in the model. Note that h_k parameterizes the deterministic private utility bias toward $\omega_i = 1$ for an agent in group k, while ε_i captures the agent's idiosyncratic payoff from this action.

I assume that the idiosyncratic payoffs ε_i may be distributed differently in every group:

$$\mathrm{P}(\varepsilon_i \leq z|k) = F_{\varepsilon|k}(z), \quad \text{for } k = 1, \ldots, K, \tag{2}$$

where $F_{\varepsilon|k}(\cdot)$ is continuously differentiable, symmetric about zero, and has positive density everywhere. I make no further parametric assumptions about these distributions. So, this framework applies for a variety of empirical specifications, e.g., logistic or Gaussian errors.

Three quantities are especially important for characterizing the model. First, there is a vector $h = (h_1, \ldots, h_K)'$ of private utility terms, which specifies each group's intrinsic preference over the two actions. Second, there is a vector of distribution functions $\{F_{\varepsilon|k}\}_{k=1}^K$, which determine how likely it is that any idiosyncratic payoff is realized in each group. Third, there is a matrix $\mathbf{J} \in \mathbb{R}^{K\times K}$ containing all the social interaction effects:

$$\mathbf{J} = \begin{bmatrix} J_{11} & J_{12} & \cdots & J_{1K} \\ J_{21} & J_{22} & \cdots & J_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ J_{K1} & J_{K2} & \cdots & J_{KK} \end{bmatrix}. \tag{3}$$

[6] The spillover term $J_{k\ell}\omega_i \,\mathrm{E}_i(\bar{\omega}^\ell)$ may also be generated from a quadratic conformity effect $-\frac{1}{2}J_{k\ell}[\omega_i - \mathrm{E}_i(\bar{\omega}^\ell)]^2$ as studied by Bernheim (1994). See Brock & Durlauf (2001) and Blume et al. (2015) for additional discussion.

Throughout this paper, I will refer to $\mathbf{J}$ as the *interaction matrix*. It may also be interpreted as the adjacency matrix of a directed graph $(\mathcal{K}, \mathbf{J})$, where the nodes $\mathcal{K} = \{1, \ldots, K\}$ represent different groups of individuals.[7] The entries of $\mathbf{J}$ specify the nature and intensity of the relationships between groups. Note that the interaction effects may not be symmetric, so $J_{k\ell}$ need not equal $J_{\ell k}$ for $k, \ell \in \mathcal{K}$. In addition, any of the interaction effects could be negative.

II.B. Equilibrium under Noncooperative Decisionmaking

When analyzing this model, I focus on (pure strategy) Bayesian Nash equilibria where agents act noncooperatively. In other words, agents do not coordinate with one another when making decisions. Each agent i in group k chooses the action $\omega_i = 1$ with probability:

$$\mathrm{P}(\omega_i = 1|k) = F_{\varepsilon|k}\Big(h_k + J_{kk}\,\mathrm{E}_i(\bar{\omega}^k_{-i}) + \sum_{\ell \neq k} J_{k\ell}\,\mathrm{E}_i(\bar{\omega}^\ell)\Big). \tag{4}$$

Since ω_i takes values in the set $\{0, 1\}$, the expected action $\mathrm{E}(\omega_i|k)$ also equals $\mathrm{P}(\omega_i = 1|k)$.

I assume that everyone has rational expectations about other people's choices. So, while agents cannot directly observe the actions of others, they do correctly infer these actions in expectation, i.e., $\mathrm{E}_i(\omega_j|k) = \mathrm{E}(\omega_j|k)$ for all agents i and j, and all groups k. By symmetry of the conditional expected choice equations, it follows that $\mathrm{E}(\omega_i|k) = \mathrm{E}(\omega_j|k)$ for all i, j, and k.

An equilibrium is defined by the expected average choices $\{\mathrm{E}(\bar{\omega}^k)\}_{k=1}^K$ that are consistent with individually optimal decisions. Letting m^{k*} denote $\mathrm{E}(\bar{\omega}^k)$, I can write this condition as:

$$m^{k*} = F_{\varepsilon|k}\bigg(h_k + \sum_{\ell=1}^{K} J_{k\ell} m^{\ell*}\bigg), \quad \text{for } k = 1, \ldots, K. \tag{5}$$

Any fixed point solution $m^* = (m^{1*}, \ldots, m^{K*})$ to this system of equations is an equilibrium.

III. Equilibrium Properties of the Model

I now examine various properties of an equilibrium, such as existence, dynamic stability, and uniqueness. I also explore the implications of equilibrium outcomes for social welfare.

III.A. Existence

Since the distributions $\{F_{\varepsilon|k}\}_{k=1}^K$ are continuous and the support of m^* is $[0, 1]^K$, Brouwer's fixed point theorem guarantees that there is at least one solution to the equilibrium system.

Property 1. There exists at least one equilibrium m^* in the binary choice model.

III.B. Dynamic Stability

Another key equilibrium property is dynamic stability. This property asserts that any iteration on best response dynamics would converge to an equilibrium. Formally, consider a

[7] For linear-in-means models, this matrix is sometimes called a "sociomatrix" (e.g., see Blume et al., 2015).

dynamic analogue of the equilibrium system: $m_t^k = F_{\varepsilon|k}\big(h_k + \sum_{\ell=1}^K J_{k\ell} m_{t-1}^\ell\big)$ for $k = 1, \ldots, K$. An equilibrium m^* is defined to be *locally stable* if it is a limiting solution to this dynamical system, where the initial iterate m_0 lies within some sufficiently small neighborhood of m^*. Alternatively, an equilibrium is *unstable* if there is a neighborhood of m^* such that, for any m_0 arbitrarily close to m^*, there is some eventual iterate m_t that lies outside of the neighborhood.

The question of local stability is fundamental to comparative statics. If a locally stable equilibrium is slightly perturbed, then the average outcomes in a network would return to that equilibrium. Meanwhile, if an equilibrium is unstable, then nearby outcomes would diverge from it. In practice, locally stable equilibria are the ones that a researcher observes, while unstable equilibria represent tipping points between different equilibrium outcomes.

To assess when an equilibrium is locally stable, I need to introduce some terminology. First, I define the *Jacobian matrix* of the right-hand-side of the equilibrium system (5) to be:

$$\mathbf{D}(m^*) = \boldsymbol{\beta}(m^*)\mathbf{J}. \tag{6}$$

Here, $\boldsymbol{\beta}(m^*) = \operatorname{diag}\big[f_{\varepsilon|1}\big(h_1 + \sum_{\ell=1}^K J_{1\ell} m^{\ell *}\big), \ldots, f_{\varepsilon|K}\big(h_K + \sum_{\ell=1}^K J_{K\ell} m^{\ell *}\big)\big]$ is a diagonal matrix of densities $f_{\varepsilon|k}\big(h_k + \sum_{\ell=1}^K J_{k\ell} m^{\ell *}\big)$, each representing the relative likelihood that an agent in group k is close to indifferent between the two actions at an equilibrium m^*. So, $\mathbf{D}(m^*)$ equals the interaction matrix $\mathbf{J}$ where each row k is weighted by the expected fraction of group k that is near indifferent at m^*. Since all the density functions $f_{\varepsilon|k}$ are strictly positive, the entries $D_{k\ell}(m^*)$ of the Jacobian matrix have the same signs as the interaction effects $J_{k\ell}$.

Next, I define the *spectral radius* of the Jacobian matrix $\rho(\mathbf{D}(m^*))$ as the largest eigenvalue of this matrix in absolute value. Formally, for any square matrix $\mathbf{A}$, this quantity is equal to:

$$\rho(\mathbf{A}) = \max\big\{|\lambda| : \lambda \text{ is an eigenvalue of } \mathbf{A}\big\}. \tag{7}$$

A matrix $\mathbf{A}$ is convergent, in the sense that $\lim_{t\to\infty} \mathbf{A}^t = \mathbf{0}$, if and only if $\rho(\mathbf{A}) < 1$. So, a larger spectral radius corresponds to a more expansive matrix. In this model, $\rho(\mathbf{D}(m^*))$ allows me to measure the collective strength of the social interaction effects within and across groups. As seen through the next property, this quantity also governs the local stability of equilibria.

Property 2. If $\rho(\mathbf{D}(m^*)) < 1$, then m^* is locally stable. If $\rho(\mathbf{D}(m^*)) > 1$, then m^* is unstable.

This property almost gives a necessary and sufficient condition for local stability. It falls short only when $\rho(\mathbf{D}(m^*))$ equals one. However, this case is not especially relevant in this model, since it is not economically meaningful and it occurs with probability measure zero.[8]

III.C. Uniqueness

This model has the potential to exhibit multiple equilibria. In other words, for some fixed parameter values, there may be multiple solutions to the equilibrium equations. In

[8]When $\rho(\mathbf{D}(m^*)) = 1$, the stability of m^* is governed by the Hessian matrices of the equilibrium equations. However, for almost all $\{F_{\varepsilon|k}\}_{k=1}^K$, the Jacobian $\mathbf{D}(m^*)$ evaluated at any equilibrium m^* has no unit eigenvalues.

this subsection, I examine what social environments lead to uniqueness versus multiplicity.

III.C.1. Social Environments with a Unique Equilibrium

When is there only one equilibrium? This situation arises whenever the social interaction effects are not strong enough to generate any unstable equilibria. To see why, consider the next property, which follows from Sard's Theorem and the Poincaré-Hopf Index Theorem.

Property 3. For almost all distributions $\{F_{\varepsilon|k}\}_{k=1}^{K}$, the number of equilibria is finite and odd. Also, if there are d_s locally stable equilibria, then there are at least $d_s - 1$ unstable equilibria.

Using Property 3, I recover a sufficient condition for uniqueness. Namely, if all equilibria are locally stable, then there can only be one. This reasoning leads to the following corollary.

Corollary. If $\rho(\mathbf{D}(m^*)) < 1$ at all equilibria m^*, then there is a unique, locally stable equilibrium.

Recall that $\rho(\mathbf{D}(m^*))$ measures the intensity of social interactions, weighted by the likelihood that agents are indifferent between the two choices. So, a unique equilibrium exists if the social interactions are relatively weak and/or if agents have strong private preferences.[9]

III.C.2. Social Environments with Multiple Equilibria

When is there more than one equilibrium? By Property 3, multiplicity can only occur in settings with strong interaction effects where unstable equilibria exist. However, strong interactions do not always imply multiplicity. In some cases, there is a unique equilibrium that is unstable.[10] To interpret equilibrium behavior when there are strong interaction effects, I must first examine the role of strategic complementarity and substitutability in networks.

Negative Spillovers and Global Instability of Equilibria

If all the entries of the interaction matrix $\mathbf{J}$ are non-negative, then the model exhibits strategic complementarity between all agents. In this case, each person's utility is a supermodular function of individual choices, which means that the marginal payoff from one's action (weakly) increases when anyone else chooses that same action. This type of model has important properties. Most notably, it almost always has a locally stable equilibrium.[11]

If the interaction matrix has negative entries, then these same equilibrium properties do not generally apply. In particular, some interaction effects are incompatible with locally stable equilibria. This instability has an economic interpretation. It arises when agents are

[9]By using the Poincaré-Hopf Index Theorem, I obtain a sufficient condition for uniqueness that is weaker than the conditions that are adopted in much of the literature, i.e., those that are implied by the Banach contraction mapping theorem. Note that Property 3 is consistent with a longstanding literature about the oddness of Nash equilibria in finite games; for example, see Wilson (1971) Harsanyi (1973), and Kohlberg & Mertens (1986).

[10]One example of this phenomenon occurs when the interaction matrix $\mathbf{J}$ is symmetric and only has eigenvalues with non-positive real parts. For such interactions, the model always has a unique equilibrium m^*, which is locally stable when $\rho(\mathbf{D}(m^*)) < 1$ and becomes unstable when $\rho(\mathbf{D}(m^*)) > 1$. See the Appendix for justification.

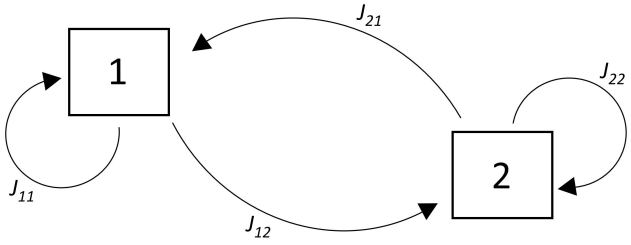
[11]See Milgrom & Roberts (1990, 1994) and Milgrom & Shannon (1994) for properties of supermodular games.

strongly repelled by the choices in their own group. In these cases, the group can never settle on one average action because its members will always be discontented with the outcome.

To better understand this point, assume there is one group in the network, i.e. $K = 1$, and suppose, for simplicity, that there is no private utility bias, i.e. $h = 0$. In this case, an equilibrium m^* is defined as the fixed point solution to $m^* = F_\varepsilon(Jm^*)$ (group subscripts are removed for notational convenience). When $J < 0$, agents are repelled by the expected average action in the population. Therefore, if $\mathrm{E}(\bar{\omega})$ is high, then people tend to prefer the low action. Conversely, if $\mathrm{E}(\bar{\omega})$ is low, then most people prefer the high action. When $J < 0$ is sufficiently large in magnitude, the equilibrium behavior in the model becomes unstable.[12]

When there are multiple groups, the role of negative interactions is more complicated. To illustrate why, suppose that there are now two groups, i.e. $K = 2$. The interactions in this case are depicted in Figure 1, where arrows indicate the direction of each effect. By the same reasoning as before, a stable equilibrium cannot exist if the within-group effects J_{11} and J_{22} are very negative. Additionally, consider what happens when $J_{12} > 0$ and $J_{21} < 0$. Members of group 1 seek to conform to the mean behavior in group 2, while members of group 2 want to distinguish themselves from group 1. These social influences, if they are strong enough, lead to self-contradictory behavior: if $\mathrm{E}(\bar{\omega}^2)$ is high, then $\mathrm{E}(\bar{\omega}^1)$ is also high, which means that $\mathrm{E}(\bar{\omega}^2)$ is low, and so forth. In this setting, equilibrium outcomes again become unstable.

Figure 1: Social Interactions with Two Groups



Social Interactions that Maintain Stable Equilibria

Going forward, it will be useful to classify what types of social interactions are conducive to stable behavioral outcomes. The discussion above suggests that global instability arises when preferences are self-contradictory in the sense that a population responds negatively to its own average choice. So, it makes sense to rule out these cases. The following condition captures the idea that all groups are overall positively affected by their own average choices.

Assumption A. There is an invertible matrix $\mathbf{B}$ such that $\mathbf{BJB}^{-1}$ has all non-negative entries.

This assumption states that the interaction matrix $\mathbf{J}$ is similar to a non-negative matrix. When two matrices are similar, they represent the same linear operator under different bases. In this way, Assumption A ensures that the model behaves similarly to one where all the interactions are non-negative, i.e. where there is strategic complementarity between all agents.

[12]Specifically, global dynamic instability will occur when $J < -f_\varepsilon^{-1}(0)$. For an explanation, see the Appendix.

On its face, Assumption A is hard to interpret. Nevertheless, it covers a range of contexts where agents do not wish to deviate from their own group. I highlight two special cases.

Example 1. The way that agents are affected by their own group depends on the cycles in a network. These cycles dictate how an agent's choice is reflected back onto itself via social interactions. Suppose that the product of interaction effects along any cycle is non-negative.

Assumption A.1. For any k and positive integer M, $J_{kj_1}J_{j_1j_2}\cdots J_{j_Mk} \geq 0$ for all $j_1, \ldots, j_M \in \mathcal{K}$.

Under this condition, agents would never be repelled by their own group. For example, with two groups, A.1 implies: $J_{11} \geq 0$, $J_{22} \geq 0$, and $J_{12}J_{21} \geq 0$. This restriction does not rule out negative interactions, but it requires that J_{12} and J_{21} have the same sign. In other words, between-group relations are mutual: agents either "agree to agree" or "agree to disagree".

More broadly, Assumption A.1 holds if the indirect spillover effect $J_{km}J_{m\ell}$ has the same sign as the direct spillover effect $J_{k\ell}$, i.e. if $\operatorname{sgn}(J_{k\ell}) = \operatorname{sgn}(J_{km}J_{m\ell})$ for all $k, \ell, m \in \mathcal{K}$.[13] In this way, A.1 invokes the phrases "the friend of my friend is my friend" and "the enemy of my friend is my enemy". This condition is rooted in balance theory (Cartwright & Harary, 1956).

To see how Assumptions A and A.1 are related, consider the following lemma. It reveals that A.1 is a special case of Assumption A where the change-of-basis matrix $\mathbf{B}$ is diagonal.

Lemma 1. Assumption A.1 holds if and only if $\mathbf{BJB}^{-1}$ is non-negative for a diagonal matrix $\mathbf{B}$.

Example 2. Suppose that the interaction matrix is symmetric, positive semi-definite. In this case, it satisfies Assumption A since it can be diagonalized so that $\mathbf{J} = \mathbf{B}^{-1}\mathbf{\Lambda B}$, where $\mathbf{B}$ is orthogonal and $\mathbf{\Lambda}$ is a diagonal matrix of the eigenvalues of $\mathbf{J}$, which are all non-negative.

One example in this class of matrices is the *diagonally dominant* matrix, which satisfies:

$$J_{kk} \geq \sum_{\ell \neq k} |J_{k\ell}|, \quad \text{for } k = 1, \ldots, K. \tag{8}$$

This condition may be interpreted as saying that the level of cohesion in a group is strong relative to the between-group influences. Hence, agents tend to conform to their own group.

Deriving a Sufficient Condition for Multiple Equilibria

In settings with strong interaction effects, Assumption A is instrumental in characterizing equilibrium behavior. This assumption ensures that there is almost always one locally stable equilibrium. Also, the locally stable equilibria outnumber the unstable equilibria.

[13] To motivate this restriction, consider a random effects version of the model where the interaction effects are unweighted. In particular, let $J_{ij} \in \{-1, 1\}$, where J_{ij} varies across all agents i and j. Interpret $\mathbf{J}$ as a matrix of average interactions $\mathbf{J}_{k\ell} = \mathrm{E}(J_{ij}|i \in k, j \in \ell)$. In this case, the restriction must only hold in an average sense, such that $\mathrm{P}(J_{i_0i_1}J_{i_1i_2} = J_{i_0i_2}|i_0 \in k, i_1 \in m, i_2 \in \ell) \geq 0.5$ for all $k, \ell, m \in \mathcal{K}$. See the Appendix for justification.

Property 4. Suppose that Assumption A is satisfied. Then, for almost all distributions $\{F_{\varepsilon|k}\}_{k=1}^{K}$, there is exactly one more locally stable equilibrium than there are unstable equilibria.

To justify Property 4, I first give a proof in the case where $\mathbf{J}$ is non-negative. This proof relies on the Perron-Frobenius theorem, as well as other mathematical results related to non-negative matrices. I then show how the proof extends to settings where Assumption A holds (where $\mathbf{J}$ is similar to a non-negative matrix). These arguments are laid out in the Appendix.

Under Assumption A, I obtain a sufficient condition for multiplicity. Specifically, multiple equilibria arise when agents have a strong desire to conform to the average action in their own group. In these social environments, aggregate behaviors become self-reinforcing, which leads to multiple locally stable equilibrium outcomes from the same fundamentals.

Corollary. Suppose that Assumption A is satisfied. If $\rho(\mathbf{D}(m^*)) > 1$ at some equilibrium m^*, then the model has multiple equilibria, and at least two of these equilibria are locally stable.

III.D. Comparisons of Aggregate Welfare across Social Groups

In this model, there is no strict Pareto ranking across equilibria since extreme realizations of the random payoff ε_i can dominate an agent's utility. So, to learn about social welfare, I consider the expected utility of agents in a group. In doing so, I can assess which equilibrium makes agents better off on average. In group k, the expected utility at an equilibrium m^* is:

$$\mathrm{E}\left(\max_{\omega_i} U_i(\omega_i|k)|m^*\right) = \mathrm{E}\left(\max_{\omega_i}\left\{h_k\omega_i + \eta_k + \sum_{\ell=1}^{K} J_{k\ell}\omega_i m^{\ell*} + \varepsilon_i\omega_i + \xi_i\right\}\right). \tag{9}$$

To evaluate this expected utility, it helps to rescale the choices so that $\omega_i \in \{-1, 1\}$. This modification has no impact on equilibrium behavior; however, it makes the welfare calculations easier to interpret. For example, it implies that—in absence of private preferences, i.e., when $h_k = \varepsilon_i = 0$ for all i, k—an agent's realized payoff when everyone selects the high action is the same as it would be when everyone selects the low action. In this way, there is no negative externality inherent to an equilibrium in settings where all agents are ambivalent between their choices. In general, welfare analysis in models of social interactions is highly sensitive to the way that utility is specified, even when different specifications yield the same expected choice functions. The reason is that welfare depends on the exact mechanisms that give rise to spillover effects (e.g., social learning, pressure to conform, or free-riding), not just the spillover effects themselves; see Bhattacharya et al. (2023) for a detailed explanation.

When the action takes values in the set $\{-1, 1\}$, the expected utility in group k equals:

$$\mathrm{E}\left(\max_{\omega_i} U_i(\omega_i|k)|m^*\right) = \mathrm{E}\left(\left|h_k + \sum_{\ell=1}^{K} J_{k\ell}m^{\ell*} + \varepsilon_i\right|\right) + \eta_k + \mathrm{E}(\xi_i|k). \tag{10}$$

As seen through the next result, the equilibrium that generates the highest expected utility

is the one where most group members choose the same action, i.e., where $\mathrm{E}(\bar{\omega}^k)$ is largest in magnitude. In addition, if agents are privately biased toward the high (low) action, then they tend to maximize their expected utility at the equilibrium where $\mathrm{E}(\bar{\omega}^k)$ is highest (lowest).

Property 5. Let $\mathcal{M}^*$ be the set of equilibria. For any group k, $\operatorname{argmax}_{m^* \in \mathcal{M}^*} \mathrm{E}(\max_{\omega_i} U_i(\omega_i|k)|m^*)$ equals $\operatorname{argmax}_{m^* \in \mathcal{M}^*} |\mathrm{E}(\bar{\omega}^k)|$. Also, there always exists some threshold T_k for which:

(i) If $h_k > T_k$, then $\operatorname{argmax}_{m^* \in \mathcal{M}^*} \mathrm{E}(\max_{\omega_i} U_i(\omega_i|k)|m^*) = \operatorname{argmax}_{m^* \in \mathcal{M}^*} \mathrm{E}(\bar{\omega}^k)$.[14]

(ii) If $h_k < T_k$, then $\operatorname{argmax}_{m^* \in \mathcal{M}^*} \mathrm{E}(\max_{\omega_i} U_i(\omega_i|k)|m^*) = \operatorname{argmin}_{m^* \in \mathcal{M}^*} \mathrm{E}(\bar{\omega}^k)$.

When do different social groups favor the same equilibrium? To answer this question, it will be useful to characterize when two groups are positively or negatively influenced by one another. These notions are generally hard to define because social influences reflect the sum of direct spillover effects (e.g., $J_{k\ell}$) and indirect spillover effects (e.g., $J_{km}J_{m\ell}$), which arise through interactions with other groups. Fortunately, using Assumption A.1, I can more easily interpret the nature of relations between groups. Consider the following definitions.

Definition 1. Group k is *connected* to group ℓ if $J_{kj_1}J_{j_1j_2}\dots J_{j_M\ell}$ is nonzero for some $j_1, j_2, \dots, j_M \in \mathcal{K}$.

Definition 2. Group k is *positively* (*negatively*) *influenced* by group ℓ if, for any positive integer M, $J_{kj_1}J_{j_1j_2}\dots J_{j_M\ell} \geq 0$ (≤ 0) for every $j_1, j_2, \dots, j_M \in \mathcal{K}$, with at least one inequality strict.

Whenever group k is positively (negatively) influenced by group ℓ, its members seek to conform to (deviate from) the average behavior in ℓ. If all groups are connected, i.e., if $\mathbf{J}$ is an irreducible matrix, then Assumption A.1 ensures: (1) every group is positively influenced by itself and (2) any two groups are either positively or negatively influenced by one another.

When Assumption A.1 holds, the equilibria in the model are ordered in a distinctive way. This ordering determines how the relative welfare of equilibria varies across social groups.

Property 6. Suppose that Assumption A.1 is satisfied, and consider any two social groups k and ℓ.

(i) Suppose that k and ℓ are positively influenced by one another. Then the equilibrium where $\mathrm{E}(\bar{\omega}^k)$ is highest (lowest) is the same equilibrium where $\mathrm{E}(\bar{\omega}^\ell)$ is highest (lowest).

(ii) Suppose that k and ℓ are negatively influenced by one another. Then the equilibrium where $\mathrm{E}(\bar{\omega}^k)$ is highest is the same equilibrium where $\mathrm{E}(\bar{\omega}^\ell)$ is lowest, and vice versa.

To understand the implications of Property 6, consider any social environment where A.1 applies and there are multiple equilibria. In this setting, there are two *extremal equilibria* (call them $\underline{m}^*$ and $\overline{m}^*$), where $\mathrm{E}(\bar{\omega}^k)$ is either maximized or minimized for all groups k. As I prove in the Appendix, both $\underline{m}^*$ and $\overline{m}^*$ are always locally stable. So, under appropriate initial conditions, any fixed-point iteration on (5) will converge to an extremal equilibrium.

[14] Why would $T_k \neq 0$? In any group k, the payoff from choosing ω_i depends on h_k, as well as on h_ℓ for each group ℓ that influences k. So, even if $h_k > 0$, most agents in group k may still prefer the low action if (1) $h_\ell < 0$ for some ℓ that attracts k or if (2) $h_\ell > 0$ for some ℓ that repels k. Only if h_k is strong enough to overcome these external influences, i.e., if h_k lies above some T_k, is $\mathrm{E}(\max_{\omega_i} U_i(\omega_i|k)|m^*)$ maximized where $\mathrm{E}(\bar{\omega}^k)$ is highest.

Taken together, Properties 5 and 6 reveal that social interactions can introduce welfare trade-offs, such that it is impossible to maximize aggregate welfare jointly in every group. In particular, if two groups are biased toward the same action, i.e., if $h_k > T_k$ and $h_\ell > T_\ell$, and if they are negatively influenced by one another, then they will maximize their expected utility at different equilibria. A similar trade-off arises when two groups are biased toward different actions, i.e., if $h_k > T_k$ and $h_\ell < T_\ell$, and are positively influenced by one another.

IV. Extensions and Alternative Network-Based Models

IV.A. Games on Networks

The properties in Section 3 are not exclusive to the binary choice framework. They also have implications for a much broader class of models where agents interact in a network. To see how, consider a game with K players. Each player k chooses a_k from a compact action space $A_k \in \mathbb{R}$. Given a profile of actions $a \in A_1 \times A_2 \times \cdots \times A_K$, the player's best response is $a_k^* = q_k\big(\sum_{\ell=1}^K J_{k\ell} a_\ell\big)$, where $q_k(\cdot)$ is some non-decreasing function that maps from $\mathbb{R}$ to A_k.

This game encompasses an abundance of economic models. For example, an action a_k could represent a person's investment into a public good, with everyone benefiting from how much their neighbors contribute.[15] Alternatively, a_k could be the output of a firm that competes in an oligopoly, where each firm's action influences the market price. These types of models are well-studied, and they both involve strategic substitutabilities between agents.

Another interpretation of this game is that each player represents a community of individuals. In the binary choice model, the players are social groups, where the members of each group make one of two choices subject to social influences and idiosyncratic biases. Agents act noncooperatively, and a_k refers to the average choice within group k. This framework would also apply to a different type of model, in which the residents of a country or local institution make a collective decision. For example, consider modeling spillover effects in US state policy, where voters support more liberal or conservative agendas based on the laws enacted in other states. Here, a_k would represent the collective action taken in state k.

I focus on pure strategy Nash equilibria, which are defined by the action profiles a^* where no player k wishes to deviate from a_k^*. It is well known that an equilibrium exists if there are continuous best responses and compact, convex action spaces. However, even without these restrictions, an equilibrium would still exist if the interaction matrix $\mathbf{J}$ satisfies Assumption A.1. To prove this result, I use Tarski's fixed point theorem, which ensures existence if the best responses are increasing, i.e., if $\mathbf{J}$ is a non-negative matrix. I then show that this property extends to settings where A.1 holds.[16] Crucially, this approach does not rely

[15] Bramoullé et al. (2014) study a type of public goods game that is nested by my framework. In their paper, players choose from an interval $[0, 1]$ and the best responses are $a_k^* = \max\{0, 1 - \delta \sum_{\ell \neq k} g_{k\ell} a_\ell\}$, where $\delta \in [0, 1]$ and $g_{k\ell} \in \{0, 1\}$ indicates whether two players k and ℓ are linked. Notice that their paper focuses exclusively on contexts with pure strategic substitutes, whereas my analysis allows for both positive and negative spillovers.

[16] Specifically, under A.1, the best response functions map to an alternate system of equations with a non-negative interaction matrix. Any equilibrium in the original model corresponds to an equilibrium in a different model that has supermodular payoffs. Given this mapping, I use Tarski's theorem to prove that equilibria exist.

on continuous best responses. So, in the binary choice model, Assumption A.1 is enough to ensure existence even when the random utility component ε_i is not continuously distributed.

If the best responses are differentiable at each equilibrium a^*, then the spectral radius of the Jacobian matrix is all that is needed to determine uniqueness. In particular, let $\tilde{a}^*$ be the equilibrium at which $\rho(\mathbf{D}(a^*))$ is greatest. If $\rho(\mathbf{D}(\tilde{a}^*)) < 1$, then $\tilde{a}^*$ is the unique equilibrium. Additionally, whenever Assumption A holds, there are multiple equilibria if $\rho(\mathbf{D}(\tilde{a}^*)) > 1$.

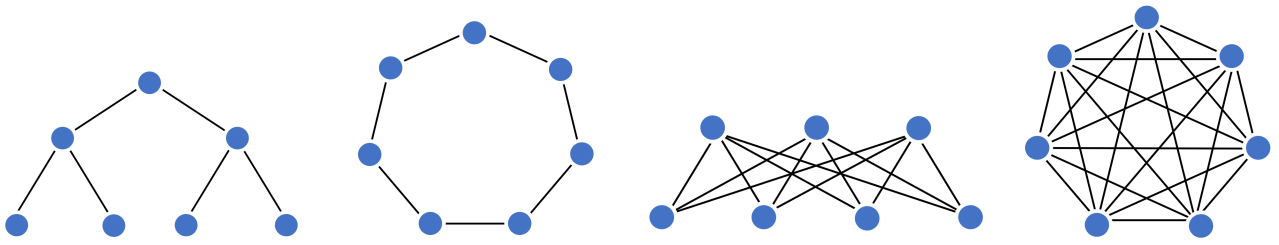
The equilibria of this game are generally not well-ordered. For example, they often do not form a lattice structure, which is useful for drawing conclusions about social welfare. As seen through Property 6, ensuring well-ordered equilibria can be achieved by imposing Assumption A.1. This condition guarantees that there always exist extremal equilibria. It also implies that there are trade-offs between different players in the network: if k and ℓ are negative influenced by one another, then a_k^* is highest wherever a_ℓ^* is lowest, and vice versa.

One convenient feature of this game is that each player's best response either (weakly) increases or decreases monotonically in another player's action. So, if the best responses are continuously differentiable, then the entries of the Jacobian matrix $\mathbf{D}$ retain the same sign for every value of a. However, this monotonicity assumption may be weakened. For example, suppose that a player's current action determines whether they conform to or deviate from someone else. In this case, characterizing equilibrium properties would require evaluating whether Assumptions A and A.1 are satisfied locally in certain regions of the support of a.

IV.B. Stable Network Structures

Given the implications of Assumptions A and A.1 for equilibrium behavior, it is worth understanding what network structures would satisfy these conditions. Namely, what types of exclusion restrictions (entries of 0 in the interaction matrix) are associated with stable equilibrium outcomes? I explore this question by examining four canonical types of networks.

Figure 2: Graphs with Seven Nodes



(a) Tree (b) Circle (c) Bipartite Graph (d) Complete Graph

Example 1 (Tree). If the links between agents are sparse, then Assumptions A and A.1 are likely to hold even in very large networks. To see why, consider a tree with K nodes. This network is used to study peer effects in social hierarchies or firm interactions in vertical production networks. Trees also encompass two common types of network structures: lines and stars. Since a tree has no cycles, any walk to and from the same node requires retracing

the same edges. So, if the interactions are symmetric or even *weakly mutual*, in the sense that either $J_{k\ell}, J_{\ell k} \geq 0$ or $J_{k\ell}, J_{\ell,k} \leq 0$ for all $k, \ell \in \mathcal{K}$, then Assumption A.1 is always satisfied.[17]

Example 2 (Circle). Consider a circle with K nodes. This network is used to study domino effects that arise when agents only interact with close contacts, rather than with the entire population; e.g., see Ellison (1993). Suppose that the interactions are weakly mutual, and let d_{edge} be the number of edges involving negative interactions. If d_{edge} is even, then A.1 always holds. If d_{edge} is odd, then each agent is negatively influenced by herself. In this case, Assumptions A and A.1 both fail. Therefore, the stability of equilibria for this type of network will depend on the number of agents, which determines the parity of negative interactions.

Example 3 (Bipartite Graphs). Consider an environment with pure strategic substitutes, i.e., let $J_{k\ell} \leq 0$ for any $k, \ell \in \mathcal{K}$. In this setting, A.1 holds if and only if the corresponding graph is bipartite. More generally, A.1 applies if and only if the agents can be partitioned into two teams, such that negative interactions only exist between members of different teams. Note that, even if A.1 fails, Assumption A may still be used to ensure equilibrium stability.[18]

Example 4 (Complete Graphs). Consider a network where all agents are linked. As there are no exclusion restrictions, Assumption A.1 is unlikely to hold if there are many agents in the network. For example, if the interactions are weakly mutual, then the fraction of complete graphs for which A.1 holds is $1/2^{\gamma_K}$, where $\gamma_K = (K-1)(K-2)/2$ for $K \geq 2$. With three agents, this fraction is $1/2$. With seven agents (depicted in Figure 2d), it is $1/32,768$. So, Assumption A.1 is more likely to apply in settings with very few agents and/or where the interaction effects are fairly uniform. In the binary choice model, I take the latter approach by partitioning the network into a few subgroups, in which the spillover effects are constant.

IV.C. Preferences over Network Composition

Until now, I have assumed that agents are influenced by the expected average action in each group, regardless of which group comprises a larger share of the total population. I now consider an alternative setup, where utility depends on the expected composition of people who choose an action. Namely, suppose agents care about $\mathrm{E}(k|\omega_i)$ instead of $\mathrm{E}(\omega_i|k)$.

This reformulation may be used to study how social influences affect network selection. For example, suppose that agents are choosing whether to enter a new environment, such as a school, and they care about what types of people they are likely to encounter there. This scenario invariably leads to negative interaction effects, since a preference that one group

[17] If the interactions are not weakly mutual, then $J_{k\ell}J_{\ell k} < 0$ for some $k, \ell \in \mathcal{K}$, which means that A.1 fails. This scenario is strategically similar to a matching pennies game, which has no pure strategy Nash equilibrium.

[18] Assumption A typically holds if the within-group spillovers J_{kk} are positive and large relative to the between-group spillovers $J_{k\ell}$, $\ell \neq k$. For example, let the interaction matrix $\mathbf{J}$ and the change-of-basis matrix $\mathbf{B}$ equal:

$$\mathbf{J} = \begin{bmatrix} \delta & 1 \\ -1 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

If $\delta \geq 3$, then $\mathbf{BJB}^{-1}$ is a non-negative matrix, which means that the model will have a locally stable equilibrium.

is in the majority is equivalent to a preference that other groups are in the minority. These behaviors can be analyzed using the same techniques that I outlined in the previous sections.

Consider a model with two social groups: a and b.[19] Let λ_a be the share of people in group a, and let $s_a(\omega_i)$ be the probability of being in group a given that someone chooses ω_i. The utility from a choice ω_i depends on the composition of agents who make that choice. Let:

$$U_i(\omega_i|k) = v_k(\omega_i) + J_k s_a(\omega_i) + \epsilon_i(\omega_i), \quad \text{for } k \in \{a, b\}. \tag{11}$$

Under this framework, the parameter J_k indicates how much the people in group k benefit from associating with members of group a. Both $v_k(\cdot)$ and $\epsilon_i(\cdot)$ are specified exactly as before.

In equilibrium, the expected composition of agents making a choice must be consistent with individually optimal decisionmaking. By Bayes' rule, any equilibrium should satisfy:

$$s_a(\omega_i) = \frac{\lambda_a \,\mathrm{P}(\omega_i|a)}{\lambda_a \,\mathrm{P}(\omega_i|a) + (1-\lambda_a)\,\mathrm{P}(\omega_i|b)}, \quad \text{for } \omega_i \in \{0, 1\}, \tag{12}$$

where $\mathrm{P}(\omega_i = 1|k) = F_{\varepsilon|k}\big(h_k + J_k(s_a(1) - s_a(0))\big)$. By Brouwer's FPT, an equilibrium exists.

Uniqueness and dynamic stability of equilibria depend on the Jacobian matrix $\mathbf{D} \in \mathbb{R}^{2\times 2}$ of the system (12). Letting $\beta_k = f_{\varepsilon|k}\big(h_k + J_k(s_a(1) - s_a(0))\big)$, I can define this matrix so that:

$$\mathbf{D} = \begin{bmatrix} D_{11} & -D_{11} \\ -D_{22} & D_{22} \end{bmatrix}, \quad \text{where: } \begin{cases} D_{11} &= s_a(0)[1 - s_a(0)] \left(\frac{J_a \beta_a}{\mathrm{P}(0|a)} - \frac{J_b \beta_b}{\mathrm{P}(0|b)} \right) \\ D_{22} &= s_a(1)[1 - s_a(1)] \left(\frac{J_a \beta_a}{\mathrm{P}(1|a)} - \frac{J_b \beta_b}{\mathrm{P}(1|b)} \right) \end{cases} \tag{13}$$

Suppose that agents prefer to associate with their own group, i.e., let $J_a \geq 0$ and $J_b \leq 0$. In this case, the matrix $\mathbf{D}$ satisfies Assumption A.1. Hence, there is almost always a locally stable equilibrium—even if the interaction effects are strong enough to generate multiplicity.

If agents prefer not to associate with their own group, i.e., if $J_a < 0$ or $J_b > 0$, then the model may not have any locally stable equilibria. This global instability occurs when agents' behavior is self-undermining, such that a group always seeks to deviate from its own action. As an example, consider Allende (2019), who studies how peer effects drive student sorting in Peru. In her model, all parents prefer to send their children to schools that are made up of wealthy, high-achieving peers. If the peer effects are sufficiently weak, then a unique, locally stable equilibrium would exist. However, if low-performing students overwhelmingly seek to associate with their high-achieving peers, then a locally stable equilibrium may not exist.

In this model, the amount of diversity in a network is tied to stability of an equilibrium. To see why, note that the rows of the Jacobian matrix $\mathbf{D}$ are scaled by $s_a(\omega_i)[1 - s_a(\omega_i)]$. This term grows larger as $s_a(\omega_i)$ approaches $1/2$, and it tends to zero as $s_a(\omega_i)$ approaches 0 or 1. So, if the agents who choose an action are more diverse, then the Jacobian is more expansive, and the realized equilibrium is more likely to be unstable. Conversely, if one group makes

[19] I study the two-group case for simplicity, but this framework may be generalized to cases with many groups.

up the majority of people choosing an action, then the equilibrium is more likely to be locally stable. Of course, this relationship only holds for a static model, where group membership is fixed. In the school choice example, some characteristics like student achievement are likely to change over time, while other traits, such as racial identity, may be constant. One area for future work is to explore how the dynamics of identity affect the formation of networks.

V. Identification and Estimation of Endogenous Interaction Effects

I now discuss how the model can be brought to data. Specifically, I show how to use data on individual choices to learn about the social interaction effects $\{J_{k\ell}\}_{k,\ell}$. These effects are of primary interest to applied researchers as they have major implications for public policy.

V.A. Empirical Setting

Suppose that a researcher has data pertaining to many networks. In practice, a network could represent a household, a school, a neighborhood, or a workplace. Agents only interact with other people in their network, and these interactions may differ on the basis of social identity, e.g., by gender or occupation. In the data, the researcher either observes the entire network or a random sample that is drawn from each network. In addition, the researcher knows which network n an agent belongs to, as well as an agent's social identity k. Let $I_{n,k}$ denote the total number of agents in the sample belonging to network n and social group k.

To ensure a broad scope for application of the model, I assume that an agent's private utility is heterogeneous across networks. So, for any agent i with social identity k who resides in a network n, the private payoff from an action ω_i depends on: (1) individual-level factors ε_i, (2) identity-specific factors h_k, and (3) contextual network effects α_n. Importantly, each of these features is unobserved by the researcher. Hence, an agent i's action ω_i equals:

$$\omega_i = \mathbb{1}\left\{ h_k + \alpha_n + J_{kk}\,\mathrm{E}_i(\bar{\omega}^k_{n,-i}) + \sum_{\ell \neq k} J_{k\ell}\,\mathrm{E}_i(\bar{\omega}^\ell_n) + \varepsilon_i \geq 0 \right\}, \tag{14}$$

where h_k, α_n, ε_i, and $\{J_{k\ell}\}_{k,\ell}$ are unknown coefficients. The model can also be adapted to incorporate observed individual-level covariates. However, since including covariates does not meaningfully change my analysis, I omit them for now and give details in the Appendix.

Throughout this section, I make two assumptions about the error structure in the model. First, I assume that the idiosyncratic payoffs ε_i and ε_j are independent for any two agents i and j within and across networks. Hence, there is no covariation between the error terms. Second, I assume that ε_i, conditional on group membership k, is independent of the network effect α_n. This condition implies that there is no self-selection into networks based on agents' unobserved idiosyncratic payoffs.[20] Both of these assumptions are standard in the literature on social interactions, and they could also be relaxed to allow for selection on observables.[21]

[20] This restriction is stronger than needed to prove identification. As shown by Manski (1988) and elaborated on by Horowitz (2009), full conditional independence could be replaced by a quantile independence restriction.

[21] Suppose that $\mathrm{P}(\varepsilon_i \leq z | W_n, \alpha_n, k) = F_{\varepsilon|W_n,k}(z)$ for some observed network-level variable W_n. Even if $F_{\varepsilon|W_n,k}$

Assumption B.1. (i) the errors $\{\varepsilon_i\}_i$ are pairwise independent; (ii) $\mathrm{P}(\varepsilon_i \leq z|k, \alpha_n) = F_{\varepsilon|k}(z)$.

As before, agents act with incomplete information and consistent beliefs. While they do not see everyone's idiosyncratic payoff ε_i, they know about the generic parameters h_k, α_n, and $\{J_{k\ell}\}_{k,\ell}$, as well as the distribution functions $\{F_{\varepsilon|k}\}_{k=1}^K$. Therefore, they would accurately infer the expected average choices $\{\mathrm{E}(\bar{\omega}_n^k)\}_{k=1}^K$. In equilibrium, the expected outcomes equal:

$$m_n^{k*} = F_{\varepsilon|k}\bigg(h_k + \alpha_n + \sum_{\ell=1}^{K} J_{k\ell} m_n^{\ell*}\bigg), \quad \text{for } k = 1, \ldots, K \text{ and } n = 1, \ldots, N. \tag{15}$$

As shown in Section 3, multiple equilibria can arise if the social interaction effects $\{J_{k\ell}\}_{k,\ell}$ are sufficiently large in magnitude. When multiple equilibria exist, I assume that agents know which one is realized, so there is no coordination involved in selecting an equilibrium.

V.B. Overview of the Identification Strategy

The model is point identified if there is only one set of parameter values $\{h_k\}_k$, $\{\alpha_n\}_n$, $\{J_{k\ell}\}_{k,\ell}$, and $\{F_{\varepsilon|k}\}_k$ that is consistent with the data under the equilibrium equations. Even if there is no self-selection into networks based on ε_i, two obstacles to identification remain. In the rest of this section, I describe both of these obstacles and explain how to overcome them.

Network-Level Unobservables

The first barrier to identification is the presence of unobservable network effects α_n, which impede my ability to learn about social interaction effects. To account for this issue, I propose a new technique that allows for the partial identification of social interactions while imposing no added assumptions on the network-level determinants of agent's choices. Specifically, I am able to difference-out the network fixed effects for members of two social groups residing in the same network. I outline this procedure in detail in the next subsection.

Using my approach, I show that I can recover the differences between two interaction effects, i.e., $\{J_{k_1\ell} - J_{k_2\ell}\}_{\ell=1}^K$ for any groups k_1 and k_2. These parameters are economically meaningful since they specify how the members of any two groups k_1 and k_2 differ in their desire to conform to another group ℓ. Furthermore, I can use these parameters to construct an economic measure of polarization. Specifically, for any groups k_1 and k_2, I can compute $\delta_{k_1k_2} = J_{k_1k_1} + J_{k_2k_2} - J_{k_1k_2} - J_{k_2k_1}$. This term quantifies how much agents want to resemble their own group plus how much they want to distinguish themselves from the other group.

Unobserved Expected Average Choices

The second barrier to identification is that the expected average choices $\{m_n^{k*}\}_{k,n}$ are not actually observed. Instead, a researcher only sees the average choices among a finite number of agents in a network. This consideration is an inherent feature of network-based models with incomplete information. However, in the literature, identification proofs often assume

does not equal $F_{\varepsilon|k}$, identification still follows by comparing networks with the same observable characteristics.

that $\{m_n^{k*}\}_{k,n}$ are already known; e.g., see Brock & Durlauf (2001) and Blume et al. (2015). The basis for this assumption is that the expectation m_n^{k*} may be consistently estimated by the observed average choice $\bar{\omega}_n^k$ as the size of each network tends to infinity. So, the researcher's inability to see m_n^{k*} is treated as an estimation problem that is separate from identification.

I claim that it is not desirable to treat the expected average actions as known quantities. In many applications, the network sizes are small, which makes the observed averages be poor approximations for the true expectations. Even in large networks, $\bar{\omega}_n^k$ is a noisy measure of m_n^{k*}. So, by replacing m_n^{k*} with $\bar{\omega}_n^k$ when estimating the model, the estimates are biased due to measurement error. This bias exists even as the number of networks tends to infinity.

I explain how to correct for the bias with internal instruments. In particular, I randomly split each network into two subsamples, and I use the average choice in one subsample as an instrument for the (endogenous) average choice in the other subsample. By construction, these averages are both noisy measures of m_n^{k*}, and the measurement errors do not depend on one another. Hence, this IV procedure is valid. I demonstrate that this approach leads to consistent estimates of the social interaction effects, even as the network sizes remain small.

Remarks on Reflection and Multiple Equilibria

In linear simultaneous equation models, a primary threat to identification is the *reflection problem* (Manski, 1993). This issue arises whenever the expectation m_n^{k*} is linearly dependent on α_n such that—even if α_n were a deterministic function of observed variables—it would be impossible to disentangle the role of social interactions from contextual network effects. To overcome the reflection problem, researchers typically rely on exclusion restrictions, which are variables that only affect some agents in a network, while leaving others unaffected.

In my setting, the reflection problem does not arise. Indeed, Brock & Durlauf (2001, 2007) show that this issue is not a threat to identification in the binary choice framework because the data always uncovers a nonlinear relationship between m_n^{k*} and α_n. This nonlinearity is inherent to the model and it acts like an exclusion restriction, which allows me to distinguish between the social and contextual effects. So, after accounting for the other two issues raised above, I find that the parameters in the model are identified without additional restrictions.

Finally, the possibility of multiple equilibria would not interfere with identification. The model is identified as long as there is a many-to-one mapping from the data to the preference parameters. Moreover, when estimating the model, I do not need to solve the equilibrium equations (15) directly. So, my analysis is robust to any issues that stem from nonuniqueness; see Bhattacharya et al. (2023), sec. 5.5, for more discussion.[22] In most cases, it makes sense to assume that the realized equilibrium is locally stable; otherwise, it is unlikely to be observed by a researcher. However, no part of my identification strategy will rely on dynamic stability.

[22] In discrete choice models with full information, nonuniqueness can affect identification (see Tamer, 2003).

V.C. *Identification with Known Expected Average Choices*

As a first step, suppose that the expected average choices $\{m_n^{k*}\}_{k,n}$ are observed. In this case, the main obstacle to identification is the unobserved network effect α_n. To handle this issue, I propose to difference-out the fixed effects by contrasting the outcomes of two social groups in the same network. The intuition for this strategy is contained in the lemma below. I include the proof along with the result so that my approach can be more clearly interpreted.

Lemma 2. *(Sufficiency Property.)* For any network $n \in \{1, \ldots, N\}$ and for any social group $k \in \mathcal{K}$:

$$\mathrm{E}(\omega_i | k, \alpha_n, \{m_n^{\ell*}\}_{\ell=1}^K) = \mathrm{E}(\omega_i | k, \{m_n^{\ell*}\}_{\ell=1}^K). \tag{16}$$

Proof. Choose some social group $\tilde{k}$ such that $\tilde{k} \neq k$. The expected average choice $m_n^{\tilde{k}*}$ is defined according to equation (15). Also, since $F_{\varepsilon|\tilde{k}}$ is strictly increasing, it is invertible. So:

$$\alpha_n = F_{\varepsilon|\tilde{k}}^{-1}\big(m_n^{\tilde{k}*}\big) - h_{\tilde{k}} - \sum_{\ell=1}^{K} J_{\tilde{k}\ell} m_n^{\ell*}. \tag{17}$$

By plugging this expression for α_n into the definition of $\mathrm{E}(\omega_i | k, \alpha_n, \{m_n^{\ell*}\}_{\ell=1}^K)$, I conclude that:

$$\mathrm{E}(\omega_i | k, \alpha_n, \{m_n^{\ell*}\}_{\ell=1}^K) = F_{\varepsilon|k}\left(h_k - h_{\tilde{k}} + \sum_{\ell=1}^{K} (J_{k\ell} - J_{\tilde{k}\ell}) m_n^{\ell*} + F_{\varepsilon|\tilde{k}}^{-1}\big(m_n^{\tilde{k}*}\big) \right). \tag{18}$$

The network effect cancels out, and $\mathrm{E}(\omega_i | k, \alpha_n, \{m_n^{\ell*}\}_{\ell=1}^K)$ is now a constant function of $\{m_n^{\ell*}\}_{\ell=1}^K$. □

This lemma shows how the observed differences between individuals in a network can be used to control for unknown contextual effects. Consider any two agents i and j with different social identities (k and $\tilde{k}$, respectively) who both reside in the same network n. Since these agents share the same context, all the network-level determinants of i's choice are captured by j's decision. Any difference between ω_i and ω_j is driven by idiosyncratic preferences (i.e., ε_i versus ε_j), as well as factors that relate to social identity (i.e., k versus $\tilde{k}$). This framework offers a natural panel structure, which allows me to control for contextual effects by comparing the expected outcomes of different types of agents in the same network.

I give two versions of my identification result. First, I provide conditions for semiparametric identification, where the error distributions $\{F_{\varepsilon|k}\}_{k=1}^K$ are known by the researcher. Then, I give conditions for nonparametric identification, where $\{F_{\varepsilon|k}\}_{k=1}^K$ are unknown. While the nonparametric version allows for greater flexibility, it also requires that there is a lot of variation in the data. In both versions, identification is achieved without α_n being observed.

Conditions for Semiparametric Identification

Before stating the identification result, I first write down the following assumption:

Assumption B.2. The network effect α_n is a continuously distributed random variable on $\mathbb{R}$.

This assumption corresponds to condition A.4 in Brock & Durlauf (2007), which is based on Manski (1988). It requires that α_n varies across networks and takes infinitely-many values. By implication, the equilibrium outcomes $\{m_n^{k*}\}_{k=1}^K$ must be heterogeneous across networks. This heterogeneity is crucial for uncovering the nonlinear relationship between the expected average choices. It ensures there is ample variation to isolate the role of social interactions.

Theorem 1. Suppose that Assumptions B.1 & B.2 hold, and assume that m_n^{k*} is observed for all networks n and all social groups k. If the distribution functions $\{F_{\varepsilon|k}\}_{k=1}^K$ are known, then:

(i) Without further assumptions, $\{h_{k_1} - h_{k_2}\}_{k_1,k_2}$ and $\{J_{k_1\ell} - J_{k_2\ell}\}_{k_1,k_2,\ell}$ are identified.

(ii) If $\alpha_n = W_n'd$ for some observed vector W_n, then d, $\{h_k\}_k$, and $\{J_{k\ell}\}_{k,\ell}$ are identified.

This theorem has two parts. Part *(i)* leverages Lemma 2 by showing that the model is partially identified even if researchers have no prior knowledge about the network effects. In particular, I can recover the differences in identity fixed effects $h_{k_1} - h_{k_2}$, as well as the differences between the interaction effects $\{J_{k_1\ell} - J_{k_2\ell}\}_{\ell=1}^K$ for any social groups $k_1, k_2 \in \mathcal{K}$. This finding is new to the literature, and it is the main contribution of the theorem. Part *(ii)* considers a special case—studied by Brock & Durlauf (2001, 2007), among others—where α_n is a constant linear function of observed variables. In this case, the model is fully identified.

Remark 1. By subtracting $J_{k_1k_2} - J_{k_2k_2}$ from $J_{k_1k_1} - J_{k_2k_1}$, I obtain a measure of polarization:

$$\delta_{k_1k_2} = J_{k_1k_1} + J_{k_2k_2} - J_{k_1k_2} - J_{k_2k_1}. \tag{19}$$

This term specifies how much agents in social groups k_1 and k_2 prefer conforming to their own group over the other group. In the current literature, polarization is often viewed as an abstract concept, and there is still no clear consensus on how define it. One benefit of my approach is that I can measure polarization in a way that is motivated by an economic model.[23]

Remark 2. If the error distributions are known (e.g., if $\{F_{\varepsilon|k}\}_{k=1}^K$ are standard Gaussian or logistic), then—under appropriate normalizations—all the interaction effects $\{J_{k\ell}\}_{k,\ell}$ are identified up to scale. To see how, assume $J_{\ell\ell} = 1$ for some group ℓ. Since $\{J_{k\ell} - J_{\ell\ell}\}_{k=1}^K$ are point identified, every coefficient $J_{k\ell}$ is known relative to $J_{\ell\ell}$. Hence, by setting the diagonal entries of the interaction matrix $\mathbf{J}$ to one, each element $J_{k\ell}$ may be recovered from the data.[24]

Conditions for Nonparametric Identification

For nonparametric identification, I need an extra assumption. Specifically, there must be an exogenous, individual-level covariate that varies continuously over an unbounded support. Using this variation, I can recover each of the error distributions $\{F_{\varepsilon|k}\}_{k=1}^K$, which I then

[23] Under my framework, polarization is defined with respect to a particular choice ω_i. So, if social identity is more salient for some choices than others, then $\{\delta_{k_1k_2}\}_{k_1,k_2}$ will depend on the decision that agents are making.

[24] In many practical contexts, $\{F_{\varepsilon|k}\}_{k=1}^K$ are only known up to a scale parameter, e.g., the variance. In these cases, $\{J_{k_1\ell} - J_{k_2\ell}\}_{k_1,k_2,\ell}$ is already identified up to scale, and additional normalizations would not be advisable.

use to identify the rest of the model. This strategy closely follows Brock & Durlauf (2007).

To set ideas, I first modify the choice equation to allow for exogenous covariates $X_i \in \mathbb{R}^r$. I define $\omega_i = \mathbb{1}\{X_i'c + h_k + \alpha_n + \sum_{\ell=1}^K J_{k\ell} m_n^{\ell *} + \varepsilon_i \geq 0\}$, such that $\mathrm{P}(\varepsilon_i \leq z | X_i, k, \alpha_n) = F_{\varepsilon|k}(z)$ and $\mathrm{P}(X_i \leq x | k, \alpha_n) = \mathrm{P}(X_i \leq x | k)$. Following Manski (1988), I impose the next assumption:

Assumption B.3. For any $k \in \mathcal{K}$, $\mathrm{supp}(X|k)$ is not contained in a proper linear subspace of $\mathbb{R}^r$; also, there is some component x_j of X—with a nonzero coefficient c_j—such that, for almost all values of $x_{-j|k}$, the distribution of $x_{j|k}$ given $x_{-j|k}$ has positive density everywhere on $\mathbb{R}$.

Assumption B.3 ensures that there is enough variation in the data to recover the model parameters even if the distributions $\{F_{\varepsilon|k}\}_{k=1}^K$ are unknown. Consider the following result.

Theorem 2. Suppose that Assumptions B.1, B.2, & B.3 hold, and assume that m_n^{k*} is observed for all networks n and social groups k. Then $\big(\{F_{\varepsilon|k}\}_{k=1}^K, c\big)$ is identified up to scale. Also:

(i) Without more assumptions, $\big(\{h_{k_1} - h_{k_2}\}_{k_1,k_2}, \{J_{k_1\ell} - J_{k_2\ell}\}_{k_1,k_2,\ell}\big)$ is identified up to scale.

(ii) If $\alpha_n = W_n'd$ for an observed vector W_n, then $\big(d, \{h_k\}_k, \{J_{k\ell}\}_{k,\ell}\big)$ is identified up to scale.

V.D. Identification with Unknown Expected Average Choices

In practice, researchers do not see the expected average choice m_n^{k*}. Instead, they would only see the average outcome $\bar{\omega}_n^k$ among finitely-many individuals. This quantity may be interpreted as a noisy measure of the true expectation, such that $\bar{\omega}_n^k = m_n^{k*} + u_{nk}$ where u_{nk} has mean zero and $\sqrt{I_{n,k}} \times u_{nk}$ converges in distribution to $N\big(0, m_n^{k*}(1 - m_n^{k*})\big)$ as $I_{n,k} \to \infty$.

If a researcher simply replaces m_n^{k*} with $\bar{\omega}_n^k$ without accounting for measurement error, then the previous identification arguments break down. To understand this point, it helps to re-write an agent's choice equation in terms of the quantities that are observed in the data.

$$\omega_i = \mathbb{1}\left\{h_k + \alpha_n + \sum_{\ell=1}^K J_{k\ell}\bar{\omega}_n^\ell + \tilde{\varepsilon}_i \geq 0\right\}, \quad \text{where } \tilde{\varepsilon}_i = \varepsilon_i - \sum_{\ell=1}^K J_{k\ell} u_{n\ell}. \tag{20}$$

In this equation, the sample averages $\{\bar{\omega}_n^\ell\}_{\ell=1}^K$ are correlated with the idiosyncratic term $\tilde{\varepsilon}_i$. Indeed, even if agent i's choice ω_i were excluded from the sample mean $\bar{\omega}_n^k$, it is still the case that $\mathrm{Cov}(\bar{\omega}_n^\ell, \tilde{\varepsilon}_i) = -J_{k\ell} \times \mathrm{Var}(u_{n\ell})$ for every $\ell \in \{1, \ldots, K\}$. Hence, $\{\bar{\omega}_n^\ell\}_{\ell=1}^K$ is endogenous in the model, which further implies that conditions *(i)* and *(ii)* of Assumption B.1 are violated.

To correct for this endogeneity, I propose an IV strategy that uses internal instruments. This procedure involves carrying out two steps. First, in each network n and social group k, I randomly split the sample into two subsets: a and b. In practice, there are many ways to form these subsets (e.g., by drawing Bernoulli random variables), and their sizes will not matter for identification. Second, I compute the average actions $\bar{\omega}_{n,a}^k$ and $\bar{\omega}_{n,b}^k$ in each subset.

By construction, the two sample averages $\bar{\omega}_{n,a}^k$ and $\bar{\omega}_{n,b}^k$ are both noisy measures of m_n^{k*}, where the measurement errors ($u_{nk,a}$ and $u_{nk,b}$, respectively) are independent of one another.

This independence is a consequence of the incomplete information setting, in which agents only respond to expectations rather than to the realized choices of others. Given this property, I can account for endogeneity in the model through IV estimation, where $\bar{\omega}_{n,b}^k$ is used as an instrument for $\bar{\omega}_{n,a}^k$. Before formalizing my result, I will first motivate it with an example.

Example (Brock & Durlauf, 2001). In their paper, Brock & Durlauf (2001) study the model:

$$m_n^* = \tanh\left(h + W_n' d + J m_n^*\right), \quad \text{for } n = 1, \ldots, N, \tag{21}$$

where W_n is a vector of observed contextual factors and (h, d', J) are unknown parameters.[25] In the case where m_n^* is known, the authors prove that the model is identified as long as there is sufficient variation in W_n for uncovering the nonlinear dependence between m_n^* and W_n.

However, the expectation m_n^* is generally unknown. So, any researcher who wishes to apply the model to data would instead be relying on the observed average choice $\bar{\omega}_n$. Unless one accounts for measurement error, many common estimation strategies (e.g., maximum likelihood estimation or OLS regression) will lead to biased estimates. To illustrate this point, suppose that someone wants to estimate the model via OLS using the observed means $\{\bar{\omega}_n\}_n$. By defining $u_n = \bar{\omega}_n - m_n^*$ and $v_n = \tanh^{-1}(\bar{\omega}_n) - \tanh^{-1}(m_n^*)$, I can re-write equation (21) as:

$$\tanh^{-1}(\bar{\omega}_n) = \tilde{h} + W_n' d + J\bar{\omega}_n + \tilde{\xi}_n, \quad \text{where: } \begin{cases} \tilde{h} &= h + \mathrm{E}(v_n) \\ \tilde{\xi}_n &= -J u_n + v_n - \mathrm{E}(v_n) \end{cases} \tag{22}$$

To ease notation, I define $\tilde{m}_n^*$, $\tilde{\bar{\omega}}_n$, and $\tilde{Y}$ to be the residuals from a least squares regression of m_n^*, $\bar{\omega}_n$, and $\tanh^{-1}(\bar{\omega}_n)$, respectively, on the vector $(1, W_n')$. The OLS estimand for J equals:

$$J^{\text{OLS}} = \frac{\mathrm{Cov}(\tilde{\bar{\omega}}_n, \tilde{Y})}{\mathrm{Var}(\tilde{\bar{\omega}}_n)} = J \times \frac{\mathrm{Var}(\tilde{m}_n^*)}{\mathrm{Var}(\tilde{m}_n^*) + \mathrm{Var}(u_n)} + \frac{\mathrm{Cov}(u_n, v_n)}{\mathrm{Var}(\tilde{m}_n^*) + \mathrm{Var}(u_n)}. \tag{23}$$

Observe that this estimand is "doubly-biased" because both the explanatory variable and the outcome variable are measured with error. As a result, $\mathrm{Var}(u_n) \neq 0$ and $\mathrm{Cov}(u_n, v_n) \neq 0$ in any setting with finite network sizes. So, the estimand J^{OLS} generally does not equal J.

To overcome this issue, I can use two-stage least squares. To do so, I start by partitioning each network into two parts (a and b), and then I compute the average choices $\bar{\omega}_{n,a}$ and $\bar{\omega}_{n,b}$. By construction, $\bar{\omega}_{n,b}$ is a valid instrument for $\bar{\omega}_{n,a}$ since $\bar{\omega}_{n,b} \perp u_{n,a}$. In addition, $\bar{\omega}_{n,b}$ will satisfy instrument relevance because $\mathrm{Cov}(\bar{\omega}_{n,a}, \bar{\omega}_{n,b}) = \mathrm{Var}(m_n^*) \neq 0$. Given these properties, I can recover J via TSLS. This strategy involves: (1) regressing $\bar{\omega}_{n,a}$ on $(1, W_n', \bar{\omega}_{n,b})$ in the first stage to estimate all the fitted values $\mathbf{L}(\bar{\omega}_{n,a}|1, W_n', \bar{\omega}_{n,b})$, and then (2) regressing $\tanh^{-1}(\bar{\omega}_{n,a})$ on $(1, W_n', \mathbf{L}(\bar{\omega}_{n,a}|1, \bar{\omega}_{n,b}))$ in the second stage. By the usual IV arguments, this procedure yields consistent estimates of the parameters d and J even in contexts with small networks.

An IV Estimator to Recover Endogenous Interaction Effects

[25] Brock & Durlauf (2001) allow for individual-level covariates X_i in their model, which I omit to ease notation.

To formalize this identification strategy, I begin by defining $\bar{\omega}^k_{n,a}$ and $\bar{\omega}^k_{n,b}$, which are the average outcomes in each randomly-generated subset of a network n and social group k.[26] I also define $u_{nk,a} = \bar{\omega}^k_{n,a} - m^{k*}_n$ and $v_{nk,a} = F^{-1}_{\varepsilon|k}(\bar{\omega}^k_{n,a}) - F^{-1}_{\varepsilon|k}(m^{k*}_n)$ to be the measurement error that enters the model when a researcher tries to approximate m^{k*}_n and $F^{-1}_{\varepsilon|k}(m^{k*}_n)$, respectively, using the average outcome $\bar{\omega}^k_{n,a}$. In the following theorem, I show that the social interaction effects are recovered from linear IV estimation, where $\bar{\omega}^k_{n,b}$ is used as an instrument for $\bar{\omega}^k_{n,a}$.

Theorem 3. Suppose that Assumptions B.1 & B.2 hold. Then the following two statements apply.

(i) For any two social groups $k_1, k_2 \in \{1, \ldots, K\}$, define the IV estimand $\beta^{IV}_{k_1,k_2}$ so that:

$$\beta^{IV}_{k_1,k_2} = \mathrm{E}(Z_n X_n')^{-1}\,\mathrm{E}(Z_n Y_n), \quad \text{where: } \begin{cases} Y_n &= F^{-1}_{\varepsilon|k_1}(\bar{\omega}^{k_1}_{n,a}) - F^{-1}_{\varepsilon|k_2}(\bar{\omega}^{k_2}_{n,a}) \\ X_n &= (1, \bar{\omega}^1_{n,a}, \ldots, \bar{\omega}^K_{n,a})' \\ Z_n &= (1, \bar{\omega}^1_{n,b}, \ldots, \bar{\omega}^K_{n,b})' \end{cases} \tag{24}$$

This quantity equals $\beta^{IV}_{k_1,k_2} = \big[h_{k_1} - h_{k_2} + \mathrm{E}(v_{nk_1,a}) - \mathrm{E}(v_{nk_2,a}), J_{k_1 1} - J_{k_2 1}, \ldots, J_{k_1 K} - J_{k_2 K}\big]'$.

(ii) Let $\alpha_n = W_n' d$ for some observed vector W_n, and define the IV estimand β^{IV}_k so that:

$$\beta^{IV}_k = \mathrm{E}(Z_n X_n')^{-1}\,\mathrm{E}(Z_n Y_n), \quad \text{where: } \begin{cases} Y_n &= F^{-1}_{\varepsilon|k}(\bar{\omega}^k_{n,a}) \\ X_n &= (1, W_n, \bar{\omega}^1_{n,a}, \ldots, \bar{\omega}^K_{n,a})' \\ Z_n &= (1, W_n, \bar{\omega}^1_{n,b}, \ldots, \bar{\omega}^K_{n,b})' \end{cases} \tag{25}$$

For any social group k, the estimand β^{IV}_k is equal to $[h_k + \mathrm{E}(v_{nk,a}), d', J_{k1}, \ldots, J_{kK}]'$.

By the sample analogue principle, I can construct estimators for $\beta^{IV}_{k_1,k_2}$ and β^{IV}_k that will converge in probability to the desired parameters as the number of networks grows large.[27] By standard arguments, these estimators are capable of inference, and they may be used for hypothesis testing. Importantly, this estimation strategy yields consistent estimates even when the size of each network remains small. Therefore, I do no rely on double asymptotics.

To evaluate the efficacy of the estimation method, I conduct Monte Carlo simulations. In Table 1, I compare the performance of the IV estimates as I vary the size of each network, as well as the number of networks, in the simulated data. I perform this analysis with both IV estimators: *(i)* computing $\{\hat{\beta}^{IV}_{k_1,k_2}\}_{k_1,k_2}$ when α_n is unknown, and then *(ii)* computing $\{\hat{\beta}^{IV}_k\}_k$ when $\alpha_n = W_n' d$ for some known vector W_n. Throughout all these simulations, I fix $K = 2$.

The simulation results in Table 1 illustrate two key properties of the IV estimators. First, the estimators perform better in settings with larger networks. Indeed, as the network sizes grow, the sample average choices $\{\bar{\omega}^k_{n,a}\}_{k,n}$ will better approximate the expectations $\{m^{k*}_n\}_{k,n}$, reducing the amount of noise in the model. Second, the estimators become more precise as

[26] Here, an implicit assumption is that there are at least two agents within any network n and social group k. Otherwise, it will be impossible to further partition the sample, which means that $\bar{\omega}^k_{n,a}$ and $\bar{\omega}^k_{n,b}$ are undefined.

[27] Specifically, the estimator $(\frac{1}{N}\sum_{n=1}^N Z_n X_n')^{-1}(\frac{1}{N}\sum_{n=1}^N Z_n Y_n)$ converges to $\mathrm{E}(Z_n X_n')^{-1}\,\mathrm{E}(Z_n Y_n)$ as $N \to \infty$.

the number of networks N increases. This pattern is implied by Theorem 3, which shows that, for any fixed network sizes, I can consistently estimate the social interaction effects.[28]

Table 1: Performance of IV Estimators for Different Network Sizes

		Mean Squared Error					
Number of	Agents per	IV Estimator *(i)*		IV Estimator *(ii)*			
Networks	Network	$\widehat{J_{11} - J_{12}}$	$\widehat{J_{22} - J_{21}}$	$\hat{J}_{11}$	$\hat{J}_{12}$	$\hat{J}_{21}$	$\hat{J}_{22}$
N = 50	100	0.249	0.922	2.740	4.264	1.858	2.226
	500	0.041	0.102	0.153	0.610	0.154	0.377
	1000	0.017	0.044	0.070	0.254	0.067	0.155
	2000	0.008	0.020	0.032	0.112	0.032	0.068
N = 500	100	0.066	0.122	0.188	0.989	0.155	0.859
	500	0.006	0.013	0.017	0.076	0.016	0.060
	1000	0.002	0.005	0.007	0.028	0.007	0.020
	2000	0.001	0.002	0.003	0.012	0.003	0.008
N = 5000	100	0.049	0.042	0.023	0.152	0.014	0.064
	500	0.003	0.003	0.003	0.022	0.002	0.025
	1000	0.001	0.001	0.001	0.006	0.001	0.006
	2000	0.000	0.000	0.000	0.002	0.000	0.002

Notes. MSE's are computed across $M = 5000$ simulation draws. For each specification, I assume equal network sizes, and I construct X_n using 2/3 of the network sample, while using the remaining 1/3 to define the instrument Z_n. I assume that agents have logistic preferences, i.e., $\varepsilon_i | k \overset{\text{i.i.d.}}{\sim} \text{Logistic}(0, 1)$ for $k \in \{1, 2\}$. For more details about the data generating process, see the Online Appendix and replication code.

VI. Application: Differences in Classroom Peer Effects by Gender

In this section, I present an empirical application of my model and identification strategy. This application uses data from Project STAR, a large-scale education experiment that randomly assigned students to classrooms of different sizes and subsequently measured their test scores. Within this context, I study the role of peer effects on academic performance, and I use the generalized interactions framework to assess how these spillovers differ by gender.

VI.A. Description of the Data and Specification Choices

The Project STAR experiment took place among students in 79 Tennessee public schools who entered kindergarten in 1985. Within each school, students and teachers were randomly assigned to three types of classrooms: small (13 to 17 students), large (22 to 25 students), and large with a teacher's aide. At the end of the school year, students took the *Stanford*

[28] When implementing this IV strategy, the researcher has flexibility in choosing how to partition the networks. Namely, one can select how much of the sample to use to define $\{\bar{\omega}^k_{n,a}\}_{k,n}$ and $\{\bar{\omega}^k_{n,b}\}_{k,n}$, as long as this selection process does not depend on the individual-level determinants of choices. Given this freedom, there may be a way to spit the sample in an "optimal way" in order to improve estimator performance. Also, if data on individual choices is available, then the researcher may obtain more precise estimates via maximum likelihood estimation. These issues are worth considering when estimating the model, and I leave them as a subject for future work.

Achievement Tests in Math and Reading, and the scores from these exams were recorded as part of the study. In total, the experiment encompassed 6,325 students across 325 classrooms.[29] For more details about the design and implementation of the program, see Word et al. (1990).

The data reports each student's raw test score, which I transform into a binary outcome to make the setting suitable for my model. Since there was no pass/fail threshold for these tests, I study a variety of outcomes, which include: scoring in the top 25%, 50%, and 75% among Tennessee kindergarten students on both the math and reading exams. I estimate the model separately for each of these outcomes to assess the impact of peer effects on each one.

Under the protocols of the experiment, students within a given school are unable to self-select into classrooms. So, Assumption B.1 holds conditional on a school. I account for this conditional independence by including school fixed effects when estimating the model. Following Graham (2008), I also give a version of the estimates where I restrict the sample to schools that have three classrooms (one of each type). This exercise addresses the unlikely possibility that there was nonrandom assignment of students to classrooms of the same type.

To evaluate how peer effects differ by gender, I focus on a version of the model with two social identities: male (m) and female (f). The Project STAR data is well-suited for studying these gender differences because male and female students are both well-represented within and across classrooms. Among participating students, there were 3,250 boys and 3,075 girls.

Student outcomes are specified according to equation (14). This framework allows for unknown gender effects h_f and h_m, which may reflect prior socialization or developmental differences between girls and boys. It also incorporates peer effects $(J_{ff}, J_{fm}, J_{mf}, J_{mm})$ that vary based on gender identity. Furthermore, the model allows for unobserved classroom characteristics α_n, which could include anything from teacher quality to the furnishings and layout of the room. Throughout my analysis, I make no assumptions about α_n—only that it is a continuously-distributed random variable. Therefore, my approach allows for a wide range of classroom-level determinants that are typically not accounted for in applied work.

When estimating the model, I assume that the idiosyncratic payoff terms follow logistic distributions, i.e., $\varepsilon_i|k \overset{\text{i.i.d.}}{\sim} \text{Logistic}(0,1)$ for $k \in \{f, m\}$. This assumption relieves the burden of recovering these distributions nonparametrically, a task that can be quite challenging in practice. In my setting, the data is simply not rich enough for nonparametric estimation.[30]

[29] The public-use data does not contain classroom identifiers. However, following Boozer & Cacciola (2001) and Graham (2008), I can uniquely assign students to classrooms by matching on observed classroom characteristics. I recover a sample of 6,248 students—among which 5,801 have non-missing test scores—across 321 classrooms.

[30] Recall from Theorem 2 that recovering $\{F_{\varepsilon|k}\}_k$ requires significant variation of individual-level covariates in each network. For small to moderate sample sizes, these functions may be poorly approximated, which causes the rest of the estimates to be imprecise. If $\{F_{\varepsilon|k}\}_k$ are known a priori, then this step is entirely avoided. Hence, there is always a trade-off between making fewer parametric assumptions and achieving more precise estimates.

VI.B. Empirical Results

Table 2 reports the IV estimates for $J_{ff} - J_{mf}$ and $J_{mm} - J_{fm}$ using the Project STAR data. To interpret these results, I first consider the descriptive evidence for gender differences in test scores. Averaging across all Project STAR classrooms, girls performed 1.01% better than boys on their combined math and reading scores. However, this difference does not reflect student outcomes at the classroom-level, where the gender gap favored boys nearly as many times as it favored girls. On average, the gender gap within a classroom was 2.83%—nearly three times higher than it was across classrooms. This disparity may suggest that certain factors at the classroom-level were contributing to gender differences in student achievement. By examining these patterns under the generalized interactions framework, I can isolate the role of differential classroom peer effects from other classroom factors that affect test scores.

The estimates in Table 2 indicate the presence of strong classroom peer effects that differ by gender. Both $J_{ff} - J_{mf}$ and $J_{mm} - J_{fm}$ are estimated to be positive, which implies that the pressure to conform is higher among peers of the same gender than it is for peers of opposite genders. In particular, a 1% increase in the expected fraction of girls scoring in the top 50% in math is estimated to raise the log-odds of achieving this outcome by 0.047 more for a female student than for a male student, on average. These estimates are roughly consistent across all six outcome variables. Moreover, there is no evidence to reject the hypothesis that $J_{ff} - J_{mf}$ equals $J_{mm} - J_{fm}$, which suggests that the peer effect differences may be symmetric.

Table 2: IV Estimates for Math and Reading Test Scores

	Outcome Variable:					
	Math			Reading		
	Top 25%	Top 50%	Top 75%	Top 25%	Top 50%	Top 75%
$J_{ff} - J_{mf}$	4.219	4.710***	5.116	4.659***	4.886***	4.444***
	(3.109)	(0.866)	(3.530)	(0.230)	(0.920)	(0.151)
$J_{mm} - J_{fm}$	4.310	4.845***	5.509	4.556***	4.924***	4.674***
	(2.747)	(0.879)	(4.958)	(0.302)	(0.957)	(0.175)
Intercept	0.002	0.220	0.024	−0.008	0.059	0.166
	(0.605)	(0.313)	(0.280)	(0.085)	(0.301)	(0.108)
Number of Classrooms	321	321	321	321	321	321
School Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
$F_{(df1,df2)}$ 1st-Stage ($\bar{\omega}_n^m$)	$8.17_{(2,60)}$	$9.09_{(2,83)}$	$4.71_{(2,55)}$	$7.27_{(2,50)}$	$6.43_{(2,86)}$	$5.73_{(2,44)}$
$F_{(df1,df2)}$ 1st-Stage ($\bar{\omega}_n^f$)	$5.41_{(2,60)}$	$11.25_{(2,83)}$	$9.53_{(2,55)}$	$14.12_{(2,50)}$	$9.22_{(2,86)}$	$5.06_{(2,44)}$

Notes. Estimates are obtained by computing $\hat{\beta}_{f,m}^{IV}$, which corresponds to the estimand in equation (24). For implementation, I randomly split each classroom so that half the sample is used to form endogenous variables X_n, and the remaining half is used to form instruments Z_n. *p<0.1; **p<0.05; ***p<0.01.

To defend the validity of my approach, I perform two types of robustness checks. First, I test whether any of the gender-specific parameters $(h_f, h_m, J_{ff}, J_{fm}, J_{mf}, J_{mm})$ depend on

observed classroom characteristics, such as the share of minority students, the poverty rate, the location of a school (urban versus rural), as well as a teacher's education and experience. Such dependence would suggest that the model is misspecified, which could interfere with identification. Although I cannot test for variation along unobserved dimensions, I find no evidence to indicate that gender-specific effects differ by observed classroom characteristics.

Second, I investigate how sensitive the IV estimates are to the way that the classrooms are partitioned. Specifically, I produce histograms that show how each of the estimates differ across a variety of random classroom partitions. I show that, under alternative partitions, the parameter estimates would not deviate too much from the results reported in Table 2. So, the estimates reliably demonstrate strong evidence of differential peer effects by gender.

VII. Conclusion

The goal of this paper is to extend the theory of discrete choice and social interactions to more general settings, where agents are affected differently by different people. I analyze how aggregate outcomes depend on the features of a network, and I consider externalities that arise under both positive and negative spillovers. Finally, I show how data may be used to recover the social interaction effects in the model, and I propose a new estimation strategy that can be applied in any context with finite networks and unobserved contextual effects.

One contribution of this work is to classify when models with negative interactions have the same types of equilibrium properties as models with uniformly positive interactions. I present two conditions (Assumptions A and A.1), which both imply that agents are not repelled by their own actions. These conditions ensure that there is almost always a locally stable equilibrium, and they allow me to derive a sufficient condition for multiple equilibria. Under these conditions, I also draw conclusions about welfare that extend over a broad class of network-based models. To my knowledge, both these conditions are new to the literature.

In the second part of the paper, I explain how to tackle two key obstacles to identification. First, I show that the generalized interactions framework offers a panel structure, which may be leveraged to isolate the role of social interactions from unobserved network effects. Then, I demonstrate how to use internal instruments to correct for measurement error that enters the model when the expected average choices are not observed. This method assures that the model will be empirically tractable even in settings with small to moderate network sizes.

I implement my framework using data from Project STAR, and I find evidence that peer effects differ systematically on the basis of gender. These differences would not be captured by a model that assumes uniform interaction effects. Hence, this application highlights the advantages of the generalized interactions framework. Note that my analysis only scratches the surface on the economic implications of heterogeneous interactions. One could take the model further by allowing for full heterogeneity in spillover effects between all individuals. Therefore, while this paper contributes to the discussion about heterogeneous interactions in discrete choice environments, there is still much room for future research on this topic.

References

AKERLOF, G. A., AND R. E. KRANTON (2000): "Economics and Identity," *The 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.

ALLENDE, C. (2019): "Competition Under Social Interactions and the Design of Education Policies," Working Paper.

ARADILLAS-LOPEZ, A. (2010): "Semiparametric Estimation of a Simultaneous Game with Incomplete Information," *Journal of Econometrics*, 157, 409–431.

ATHEY, S. (2001): "Single Crossing Properties and the Existence of Pure Strategy Equilibria in Games of Incomplete Information," *Econometrica*, 69, 861–889.

——— (2002): "Monotone Comparative Statics under Uncertainty," *The Quarterly Journal of Economics*, 117, 187–223.

BALLESTER, C., A. CALVÓ-ARMENGOL, AND Y. ZENOU (2006): "Who's Who in Networks. Wanted: The Key Player," *Econometrica*, 74, 1403–1417.

BERNHEIM, B. D. (1994): "A Theory of Conformity," *Journal of Political Economy*, 102, 841–77.

BERRY, S. T., AND P. A. HAILE (2022): "Nonparametric Identification of Differentiated Products Demand Using Micro Data," working paper.

BERTRAND, M., AND E. KAMENICA (2023): "Coming Apart? Cultural Distances in the United States over Time," *American Economic Journal: Applied Economics*, 15, 100–141.

BHATTACHARYA, D., P. DUPAS, AND S. KANAYA (2023): "Demand and Welfare Analysis in Discrete Choice Models with Social Interactions," *The Review of Economic Studies*, rdad053.

BLUME, L., W. BROCK, S. DURLAUF, AND Y. IOANNIDES (2011): "Identification of Social Interactions," Volume 1 of Handbook of Social Economics: North-Holland, 853–964.

BLUME, L., W. BROCK, S. DURLAUF, AND R. JAYARAMAN (2015): "Linear Social Interactions Models," *Journal of Political Economy*, 123, 444–496.

BOOZER, M. A., AND S. E. CACCIOLA (2001): "Inside the 'Black Box' of Project STAR: Estimation of Peer Effects Using Experimental Data," Center Discussion Paper 832.

BOSTWICK, V. K., AND B. A. WEINBERG (2022): "Nevertheless She Persisted? Gender Peer Effects in Doctoral STEM Programs," *Journal of Labor Economics*, 40, 397–436.

BOXELL, L., M. GENTZKOW, AND J. M. SHAPIRO (2022): "Cross-Country Trends in Affective Polarization," *The Review of Economics and Statistics*, 1–60.

BRAMOULLÉ, Y., AND R. KRANTON (2007): "Risk Sharing Across Communities," *American Economic Review*, 97, 70–74.

BRAMOULLÉ, Y., R. KRANTON, AND M. D'AMOURS (2014): "Strategic Interaction and Networks," *American Economic Review*, 104, 898–930.

BROCK, W., AND S. DURLAUF (2001): "Discrete Choice with Social Interactions," *The Review of Economic Studies*, 68, 235–260.

——— (2007): "Identification of Binary Choice Models with Social Interactions," *Journal of Econometrics*, 140, 52–75.

CABRALES, A., A. CALVO-ARMENGOL, AND Y. ZENOU (2011): "Social Interactions and Spillovers," *Games and Economic Behavior*, 72, 339–360.

CALVO-ARMENGOL, A., E. PATACCHINI, AND Y. ZENOU (2009): "Peer Effects and Social Networks in Education," *The Review of Economic Studies*, 76, 1239–1267.

CARTWRIGHT, D., AND F. HARARY (1956): "Structural balance: A Generalization of Heider's Theory," *Psychological Review*, 63, 277–293.

CHARNESS, G., AND Y. CHEN (2020): "Social Identity, Group Behavior, and Teams," *Annual Review of Economics*, 12, 691–713.

COOPER, R., AND A. JOHN (1988): "Coordinating Coordination Failures in Keynesian Models," *The Quarterly Journal of Economics*, 103, 441–463.

ELLIOTT, M., AND B. GOLUB (2019): "A Network Approach to Public Goods," *Journal of Political Economy*, 127, 730–776.

ELLISON, G. (1993): "Learning, Local Interaction, and Coordination," *Econometrica*, 61, 1047–1071.

GALEOTTI, A., S. GOYAL, M. O. JACKSON, F. VEGA-REDONDO, AND L. YARIV (2010): "Network Games," *The Review of Economic Studies*, 77, 218–244.

GLAESER, E. L., B. SACERDOTE, AND J. A. SCHEINKMAN (1996): "Crime and Social Interactions," *The Quarterly Journal of Economics*, 111, 507–548.

——— (2003): "The Social Multiplier," *Journal of the European Economic Association*, 345–353.

GRAHAM, B. S. (2008): "Identifying Social Interactions through Conditional Variance Restrictions," *Econometrica*, 76, 643–660.

HARSANYI, J. C. (1973): "Oddness of the number of equilibrium points: A new proof," *International Journal of Game Theory*, 2, 235–250.

HENRY, D. (1981): *Geometric Theory of Semilinear Parabolic Equations.* Volume 840 of Lecture Notes in Mathematics, Berlin: Springer-Verlag.

HOROWITZ, J. (2009): *Semiparametric and Nonparametric Methods in Econometrics.* Volume 692.

HOXBY, C. M. (2000): "The Effects of Class Size on Student Achievement: New Evidence from Population Variation," *The Quarterly Journal of Economics*, 115, 1239–1285.

JACKSON, M. O., AND Y. ZENOU (2015): "Games on Networks," Volume 4 of Handbook of Game Theory with Economic Applications: Elsevier, 95–163.

KLINE, B., AND E. TAMER (2020): "Econometric Analysis of Models with Social Interactions," in *The Econometric Analysis of Network Data* ed. by Graham, B., and de Paula, A.: Academic Press, 149–181.

KOHLBERG, E., AND J.-F. MERTENS (1986): "On the Strategic Stability of Equilibria," *Econometrica*, 54, 1003–1037.

LAVY, V., AND A. SCHLOSSER (2011): "Mechanisms and Impacts of Gender Peer Effects at School," *American Economic Journal: Applied Economics*, 3, 1–33.

MANSKI, C. F. (1985): "Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator," *Journal of Econometrics*, 27, 313–333.

——— (1988): "Identification of Binary Response Models," *Journal of the American Statistical Association*, 83, 729–738.

——— (1993): "Identification of Endogenous Social Effects: The Reflection Problem," *The Review of Economic Studies*, 60, 531–542.

MILGROM, P., AND J. ROBERTS (1990): "The Economics of Modern Manufacturing: Technology, Strategy, and Organization," *The American Economic Review*, 80, 511–528.

——— (1994): "Comparing Equilibria," *The American Economic Review*, 84, 441–459.

MILGROM, P., AND C. SHANNON (1994): "Monotone Comparative Statics," *Econometrica*, 62, 157–180.

MILNOR, J. (1965): *Topology from the Differentiable Viewpoint*: The University Press of Virginia.

PAULA, A. D. (2017): *Econometrics of Network Models* Volume 1 of Econometric Society Monographs, 268–323: Cambridge University Press.

SHAYO, M. (2020): "Social Identity and Economic Policy," *Annual Review of Economics*, 12, 355–389.

TAMER, E. (2003): "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria," *The Review of Economic Studies*, 70, 147–165.

TOPKIS, D. (1998): *Supermodularity and Complementarity*, Frontiers of Economic Research: Princeton University Press.

VIVES, X. (1990): "Nash equilibrium with strategic complementarities," *Journal of Mathematical Economics*, 19, 305–321.

WILSON, R. (1971): "Computing Equilibria of N-Person Games," *SIAM Journal on Applied Mathematics*, 21, 80–87.

WOOLDRIDGE, J. (2013): *Introductory Econometrics: A Modern Approach*: Cengage Learning.

WORD, E. R., J. JOHNSTON, H. P. BAIN ET AL. (1990): "The State of Tennessee's Student/Teacher Achievement Ratio (STAR) Project: Technical Report (1985-1990)."

Appendix: Proofs

This appendix contains proofs of the main results in the paper. Additional discussion and findings (e.g., justification for remarks that are made in the footnotes and robustness analyses for the empirical application) are provided in a supplementary appendix, which is intended only for online publication.

Proof of Property 1

Proof. Define $\mathcal{Q} : [0,1]^K \to [0,1]^K$ so that $\mathcal{Q}_k(m) = F_{\varepsilon|k}\big(h_k + \sum_{\ell=1}^K J_{k\ell} m^\ell\big)$ for $k = 1, \ldots, K$. Since $\mathcal{Q}$ is a continuous, self-mapping function on a non-empty, compact, convex set in $\mathbb{R}^K$, Brouwer's fixed point theorem ensures that there exists a vector m^* that solves $m^* = \mathcal{Q}(m^*)$. □

Proof of Property 2

Proof. Suppose $\rho(\mathbf{D}(m^*)) < 1$. For any $\epsilon > 0$, there exists a matrix norm $||\cdot||$ for which $||\mathbf{D}(m^*)|| \leq \rho(\mathbf{D}(m^*)) + \epsilon$. By defining ϵ so that $\epsilon < 1 - \rho(\mathbf{D}(m^*))$, it follows that $||\mathbf{D}(m^*)|| < 1$ for some matrix norm. The system (5) is a contraction at m^* under this norm, which implies:

$$||m_t - m^*|| = ||\mathcal{Q}(m_{t-1}) - \mathcal{Q}(m^*)|| \leq \kappa ||m_{t-1} - m^*||,$$

where $\kappa = [0, 1)$ for any vector m_{t-1} that lies in some sufficiently small neighborhood of m^*. Iterating on the inequality above ensures that $||m_t - m^*|| \leq \kappa^t ||m_0 - m^*||$, where $\lim_{t\to\infty} \kappa^t = 0$.

Next, suppose that $\rho(\mathbf{D}(m^*)) > 1$. Since $||\mathbf{D}(m^*)|| \geq \rho(\mathbf{D}(m^*))$ for any matrix norm $||\cdot||$, it must be that $||\mathbf{D}(m^*)|| > 1$. By Henry (1981), Theorem 5.1.5., there exists some $u > 0$ such that, for any $\delta > 0$, there is an initial iterate m_0 where $||m_0 - m^*|| < \delta$ for which some future iterate m_t, where $t \geq 1$, satisfies $||m_t - m^*|| \geq u$. It follows that m^* is an unstable equilibrium. □

Proof of Property 3

Proof. Define the mapping $\mathcal{H} : [0,1]^K \to \mathbb{R}^K$ so that $\mathcal{H}_k(m) = m^k - \mathcal{Q}_k(m)$ for $k = 1, \ldots, K$. By definition, m^* is an equilibrium if and only if $\mathcal{H}(m^*) = \mathbf{0}_K$. Since no equilibrium lies on the boundary of $[0,1]^K$, I restrict attention to the interior $(0,1)^K$.[31] Let $\mathbf{D}_\mathcal{H}$ denote the Jacobian matrix of $\mathcal{H}$, and define a set $\mathcal{C} = \{m \in (0,1)^K : \det(\mathbf{D}_\mathcal{H}(m)) = 0\}$. By Sard's Theorem, $\mathcal{H}(\mathcal{C})$ has Lebesgue measure zero. Therefore, for almost all $\mathcal{H}(m)$ evaluated on the domain $(0,1)^K$, the matrix $\mathbf{D}_\mathcal{H}(m)$ is invertible. Moreover, for any fixed $y = \mathcal{H}(m)$, almost all distributions $\{F_{\varepsilon|k}\}_{k=1}^K$ would satisfy $y \notin \mathcal{H}(\mathcal{C})$. So, when applied to the case where $\mathcal{H}(m) = \mathbf{0}_K$, Sard's Theorem ensures that $\mathbf{D}_\mathcal{H}(m^*)$ is almost always invertible over the full set of equilibria m^*. By the inverse function theorem, it is further guaranteed that each equilibrium is locally unique. Finally, since the set of equilibria is compact, it must have a finitely-many elements.

To prove that there is an odd number of equilibria, I rely on a version of the Poincaré-Hopf Index Theorem that is proven in Milnor (1965), Ch. 6. This theorem is presented below.

Poincaré-Hopf Theorem. Let v be a smooth vector field on the disk D^K that points outward on the boundary and has a finitely many isolated zeros $\{x^*_{(j)}\}_{j=1}^M$ satisfying $\det(\mathbf{D}_v(x^*_{(j)})) \neq 0$. Then $\sum_{j=1}^M \text{index}_{x^*_{(j)}}(v) = 1$ where $\text{index}_{x^*_{(j)}}(v)$ equals 1 if $\det(\mathbf{D}_v(x^*_{(j)})) > 0$ and -1 otherwise.

[31] Since each $F_{\varepsilon|k}$ has positive density everywhere, $\mathcal{H}_k(m) < 0$ when $m_k = 0$ and $\mathcal{H}_k(m) > 0$ when $m_k = 1$.

This theorem applies within this context because the set of equilibria is defined on $(0,1)^K$, which is diffeomorphic to an open disk. In addition, $\mathcal{H}$ is a smooth vector field pointing outward at all boundary points, since $\lim_{m_k \to 0} \mathcal{H}(m) < 0$ and $\lim_{m_k \to 1} \mathcal{H}(m) > 0$ for all k. Finally, for almost all $\{F_{\varepsilon|k}\}_{k=1}^K$, it is the case that $\mathcal{H}$ has finitely-many isolated zeros and that, at each of these zeros, the Jacobian has a nonzero determinant. Therefore, the index theorem ensures that $\sum_{j=1}^M \text{index}_{m^*_{(j)}}(\mathcal{H}) = 1$, which implies that there is an odd number of equilibria.

Finally, suppose there are d_s locally stable equilibria. At each one of these equilibria m^*, the eigenvalues of $\mathbf{D}(m^*)$ all lie below unity in absolute value with probability one. Hence:

$$\det\big(\mathbf{D}_{\mathcal{H}}(m^*)\big) = \det\big(I - \mathbf{D}(m^*)\big) = \prod_{k=1}^K \big(1 - \lambda_k(m^*)\big) > 0,$$

where $\{\lambda_k(m^*)\}_{k=1}^K$ denote the eigenvalues of $\mathbf{D}(m^*)$. By the index theorem, there must also be at least $d_s - 1$ equilibria m^* at which $\det(\mathbf{D}_{\mathcal{H}}(m^*)) < 0$. Each of these equilibria have at least one eigenvalue that exceeds unity. Therefore, there are at least $d_s - 1$ unstable equilibria. □

Proof of Lemma 1

Proof. "⇒" Suppose that $\mathbf{J}$ is similar to a non-negative matrix $\mathbf{A}$ by way of a diagonal change-of-basis matrix $\mathbf{B}$. That is, there exists some diagonal $\mathbf{B}$ for which $\mathbf{A} = \mathbf{B}\mathbf{J}\mathbf{B}^{-1} \geq \mathbf{0}$. Since $\mathbf{B}$ is diagonal, the elements of $\mathbf{A}$ can be expressed as $A_{k\ell} = B_{kk}J_{k\ell}/B_{\ell\ell}$, for all k, ℓ. Thus, for any selection of indices k and $\ell_1, \ldots, \ell_M$ in the set $\mathcal{K} = \{1, \ldots, K\}$, it must be that:

$$\begin{aligned} 0 &\leq A_{k\ell_1}A_{\ell_1\ell_2}\cdots A_{\ell_M k} \\ &= \Big(\frac{B_{kk}}{B_{\ell_1\ell_1}}J_{k\ell_1}\Big)\Big(\frac{B_{\ell_1\ell_1}}{B_{\ell_2\ell_2}}J_{\ell_1\ell_2}\Big)\cdots\Big(\frac{B_{\ell_M\ell_M}}{B_{kk}}J_{\ell_M k}\Big) \\ &= J_{k\ell_1}J_{\ell_1\ell_2}\cdots J_{\ell_M k} \end{aligned}$$

"⇐" Suppose that *A.1* holds. I first restrict attention to the case where $\mathbf{J}$ is irreducible. For any k, I define $\{\gamma_\ell^k\}_{\ell=1}^K$ so that $\gamma_\ell^k = 1$ if k is positively influenced (see Definition 2) by ℓ, and $\gamma_\ell^k = -1$ otherwise. Next, fixing some index $k_0 \in \mathcal{K}$, I construct the matrix $\mathbf{B}$ such that:

$$\mathbf{B} = \text{diag}\begin{bmatrix}\gamma_1^{k_0} \\ \vdots \\ \gamma_K^{k_0}\end{bmatrix}$$

Notice that $\mathbf{B}$ is involutory, i.e. $\mathbf{B}^{-1} = \mathbf{B}$. Thus, for all (g, ℓ), Assumption A.1 ensures that:

$$[\mathbf{B}\mathbf{J}\mathbf{B}^{-1}]_{k,\ell} = [\mathbf{B}\mathbf{J}\mathbf{B}]_{k,\ell} = \gamma_k^{k_0}\gamma_\ell^{k_0}J_{k\ell} = \gamma_{k_0}^{k}\gamma_\ell^{k_0}J_{k\ell} = \gamma_\ell^k J_{k\ell} = |J_{k\ell}|$$

It follows that $\mathbf{B}\mathbf{J}\mathbf{B}^{-1}$ equals the absolute value of $\mathbf{J}$. Therefore, $\mathbf{B}\mathbf{J}\mathbf{B}^{-1}$ is non-negative. Finally, if $\mathbf{J}$ is not irreducible, then this same reasoning applies to all irreducible blocks of $\mathbf{J}$. So, it must always be true that $\mathbf{B}\mathbf{J}\mathbf{B}^{-1}$ is non-negative for some diagonal matrix $\mathbf{B} \in \mathbb{R}^{K\times K}$. □

Proof of Property 4

Proof. First, suppose that $\mathbf{J}$—and therefore $\mathbf{D}(m)$ for every $m \in [0,1]^K$—is a non-negative matrix. In addition, assume there exists an equilibrium $m^* \in (0,1)^K$ satisfying $\rho(\mathbf{D}(m^*)) > 1$.

If $\mathbf{D}(m^*)$ is an irreducible matrix, then the Perron-Frobenius theorem guarantees that:

$$\mathbf{D}(m^*)x = \rho(\mathbf{D}(m^*))x > x,$$

for some strictly positive vector $x \in \mathbb{R}^K_{++}$. It follows that $\mathbf{D}(m^*)\delta x > \delta x$ for any scalar $\delta > 0$. By taking the first-order Taylor approximation of $\mathcal{Q}(m^*+\delta x)$ about m^*, I obtain the following:

$$\mathcal{Q}(m^* + \delta x) = \underbrace{\mathcal{Q}(x^*)}_{=x^*} + \underbrace{\mathbf{D}(x^*)\delta x}_{>\delta x} + \mathrm{h}_1(m^* + \delta x)\delta x, \quad \text{where } \lim_{\delta \to 0} \mathrm{h}_1(m^* + \delta x) = 0$$

For sufficiently small δ, the vector $a = m^* + \delta x$, where $a \in (0,1)^K$, will satisfy $\mathcal{Q}(a) > a > m^*$. By an analogous argument, there also exists some $b = m^* - \delta x$, which satisfies $\mathcal{Q}(b) < b < m^*$. Since no equilibrium lies on the boundary of $[0,1]^K$, the following inequalities are satisfied:

$$\mathbf{0}_K < \mathcal{Q}(\mathbf{0}_K) < \mathcal{Q}(b) < b < m^* < a < \mathcal{Q}(a) < \mathcal{Q}(\mathbf{1}_K) < \mathbf{1}_K$$

Brouwer's fixed point theorem ensures that $\mathcal{Q}$ has two more fixed points $\underline{m}^*$ and $\overline{m}^*$, such that $\mathbf{0}_K < \underline{m}^* < m^*$ and $m^* < \overline{m}^* < \mathbf{1}_K$. Moreover, if either of these equilibria is unstable, i.e., if either $\rho(\mathbf{D}(\underline{m}^*)) > 1$ or $\rho(\mathbf{D}(\overline{m}^*)) > 1$, then these same arguments can be used to show that there exist two more equilibria: one that lies between the two unstable equilibria and another that lies between the unstable equilibrium and the boundary. Since there is a finite number of equilibria with probability one, I may conclude that there are almost always more stable equilibria than unstable equilibria. Taken together with Property 3, this result implies that there is almost always exactly one more stable equilibrium than there are unstable equilibria.

Next, consider the case where $\mathbf{D}(m^*)$ is a reducible matrix. If $\rho(\mathbf{D}(m^*)) > 1$, then the same must hold for some irreducible block of $\mathbf{D}(m^*)$. Let $\mathcal{B}$ denote the set of indices within this block. Applying the Perron-Frobenius theorem to that block, there exists some vector x, satisfying $x_\ell > 0$ for $\ell \in \mathcal{B}$ and $x_\ell = 0$ otherwise, so that $\mathcal{Q}(m^* + \delta x) > m^* + \delta x$ for $\delta > 0$ sufficiently small. Setting $a = m^* + \delta x$, it follows that $m^* < a < \mathcal{Q}(a) < \mathcal{Q}(\mathbf{1}_K) < \mathbf{1}_K$. By Brouwer's theorem, there must exist a fixed point of $\mathcal{Q}$ between m^* and $\mathbf{1}_K$, and (by analogous arguments) a fixed point of $\mathcal{Q}$ between $\mathbf{0}_K$ and m^*. So, just as in the previous case, there is always exactly one more stable equilibrium than there are unstable equilibria.

Finally, I explain how this fixed point property, which is specific to monotone mappings, i.e., where $\mathbf{J}$ is non-negative, can be extended to a certain class of non-monotone mappings. Suppose that Assumption A is satisfied, i.e., let there be an invertible matrix $\mathbf{B}$ such that $\mathbf{B}\mathbf{J}\mathbf{B}^{-1}$ is non-negative. For $\mathcal{I} = [0,1]^K$, define the mapping $\hat{\mathcal{Q}} : \mathbf{B}\mathcal{I} \to \mathbf{B}\mathcal{I}$ in such a way that $\hat{\mathcal{Q}}(m) = \mathbf{B}\mathcal{Q}(\mathbf{B}^{-1}m)$. This mapping has a Jacobian matrix of $\mathbf{D}_{\hat{\mathcal{Q}}}(m) = \mathbf{B}\mathbf{D}(\mathbf{B}^{-1}m)\mathbf{B}^{-1}$. Note that $\phi : \mathbb{R}^K \to \mathbb{R}^K$, where $\phi(m) = \mathbf{B}m$, is a bijective linear map. Therefore, $\phi(\cdot)$ is a *homeomorphism*, and it preserves interior points.[32] It follows that $\mathbf{B}\mathbf{D}(m)\mathbf{B}^{-1}$ is non-negative on $\mathrm{int}(\mathcal{I})$ if and only if $\mathbf{D}_{\hat{\mathcal{Q}}}(m) = \mathbf{B}\mathbf{D}(\mathbf{B}^{-1}m)\mathbf{B}^{-1}$ is non-negative on $\mathbf{B}\mathrm{int}(\mathcal{I})$, which equals $\mathrm{int}(\mathbf{B}\mathcal{I})$. So, under Assumption A, the matrix $\mathbf{D}_{\hat{\mathcal{Q}}}(m)$ is non-negative for every $m \in \mathrm{int}(\mathbf{B}\mathcal{I})$.

To summarize, Assumption A implies that $\hat{\mathcal{Q}}(x) = \mathbf{B}\mathcal{Q}(\mathbf{B}^{-1}x)$ is monotonic on $\mathrm{int}(\mathbf{B}\mathcal{I})$. Note also that m^* is a fixed point of $\mathcal{Q}$ if and only if $\mathbf{B}m^*$ is a fixed point of $\hat{\mathcal{Q}}$. To see why, write $\hat{\mathcal{Q}}(\mathbf{B}m^*) = \mathbf{B}\mathcal{Q}(m^*) = \mathbf{B}m^*$. In addition, note that $\mathbf{D}(m)$ and $\mathbf{D}_{\hat{\mathcal{Q}}}(\mathbf{B}m)$ are similar

[32] A *homeomorphism* is a continuous bijection between two topological spaces that has a continuous inverse.

matrices, since $\mathbf{D}_{\hat{\mathcal{Q}}}(\mathbf{B}m) = \mathbf{B}\mathbf{D}(m)\mathbf{B}^{-1}$, which implies that they share the same eigenvalues. In particular, their spectral radii are equivalent: $\rho(\mathbf{D}_{\mathcal{Q}}(m)) = \rho(\mathbf{D}_{\hat{\mathcal{Q}}}(\mathbf{B}m))$ for all $m \in \text{int}(\mathcal{I})$.

By this relationship, I can assess the existence, uniqueness, and local stability of fixed points of the mapping $\mathcal{Q}$ by focusing on the (non-decreasing) mapping $\hat{\mathcal{Q}}$. While the fixed points of $\mathcal{Q}$ and $\hat{\mathcal{Q}}$ may be ordered differently on their respective domains, their number and local stability are the same. So, this property extends to models where Assumption A holds. □

Proof of Property 5

Proof. Since the random payoffs ε_i are distributed symmetrically about zero, the expected utility $\mathrm{E}\left(\max_{\omega_i} U_i(\omega_i|k)|m^*\right)$ defined in equation (10) strictly increases in $|h_k + \sum_{\ell=1}^K J_{k\ell} m^{\ell *}|$. In addition, since $|\mathrm{E}(\bar{\omega}^k)|$ is a strictly increasing function of $|h_k + \sum_{\ell=1}^K J_{k\ell} m^{\ell *}|$ in equilibrium, it must be that $\mathrm{E}\left(\max_{\omega_i} U_i(\omega_i|k)|m^*\right)$ is highest at the equilibrium where $|\mathrm{E}(\bar{\omega}^k)|$ is highest. Parts *(i)* & *(ii)* follow directly from the fact that $\lim_{h_k \to \infty} \mathrm{E}(\bar{\omega}^k) = 1$ and $\lim_{h_k \to -\infty} \mathrm{E}(\bar{\omega}^k) = -1$. □

Proof of Property 6

Proof. Suppose $\mathbf{J}$ satisfies A.1. By Lemma 1, there exists some diagonal matrix $\mathbf{B}$ for which $\mathbf{B}\mathbf{J}\mathbf{B}^{-1}$ is non-negative. Without loss of generality, choose $\mathbf{B}$ to be the matrix that is constructed in the "$\Leftarrow$" part of the proof of Lemma 1. That is, fix some group k, and define $\mathbf{B} = \text{diag}[\gamma_1^k, \ldots, \gamma_K^k]$, where $\gamma_\ell^k = 1$ if k is positively influenced by ℓ, and $\gamma_\ell^k = -1$ otherwise.

As shown in the proof of Property 4, the mapping $\hat{\mathcal{Q}}(m) = \mathbf{B}\mathcal{Q}(\mathbf{B}^{-1}m)$ is non-decreasing on $\text{int}(\mathbf{B}\mathcal{I})$. Also, because $(\mathbf{B}\mathcal{I}, \leq)$ is a complete lattice, Tarski's fixed point theorem ensures that the set of fixed points of $\hat{\mathcal{Q}}$ forms a complete lattice. In particular, $\hat{\mathcal{Q}}$ has greatest and least fixed points. Let $\mathbf{B}\overline{m}^*$ denote the greatest fixed point of $\hat{\mathcal{Q}}$. Then $\mathrm{E}(\bar{\omega}^\ell)$ is maximal at the equilibrium $\overline{m}^*$ if k is positively influenced by ℓ, and $\mathrm{E}(\bar{\omega}^\ell)$ is minimal at $\overline{m}^*$ otherwise. Let $\mathbf{B}\underline{m}^*$ be the lowest fixed point of $\hat{\mathcal{Q}}$. Then $\mathrm{E}(\bar{\omega}^\ell)$ is minimal at $\underline{m}^*$ whenever k is positively influenced by ℓ, and $\mathrm{E}(\bar{\omega}^\ell)$ is maximal at $\underline{m}^*$ otherwise. This argument holds for any $k, \ell \in \mathcal{K}$. □

Proof of Theorem 1

Proof. Pick any two social groups k_1 and k_2, and define the function $\nu_{k_1,k_2} : \mathbb{R}^K \to \mathbb{R}$ such that $\nu_{k_1,k_2}(m) = (h_{k_1} - h_{k_2}) + \sum_{\ell=1}^K (J_{k_1\ell} - J_{k_2\ell}) m^\ell + F^{-1}_{\varepsilon|k_2}(m^{k_2})$. By equation (18), the expected average action $m_n^{k_1 *}$ equals $F_{\varepsilon|k_1}(\nu_{k_1,k_2}(m_n^*))$. Additionally, since the cumulative distribution function $F_{\varepsilon|k_1} : \mathbb{R} \to [0,1]$ is strictly increasing over $\mathbb{R}$, the following equality will be satisfied:

$$\begin{aligned} m_n^{k_1 *} &= F_{\varepsilon|k_1}\big(h_{k_1} - h_{k_2} + \sum_{\ell=1}^K (J_{k_1\ell} - J_{k_2\ell}) m_n^{\ell *} + F^{-1}_{\varepsilon|k_2}(m_n^{k_2 *})\big) \\ &= F_{\varepsilon|k_1}\big(\widehat{h_{k_1} - h_{k_2}} + \sum_{\ell=1}^K (\widehat{J_{k_1\ell} - J_{k_2\ell}}) m_n^{\ell *} + F^{-1}_{\varepsilon|k_2}(m_n^{k_2 *})\big) \end{aligned}$$

if and only if $(h_{k_1} - h_{k_2}) - (\widehat{h_{k_1} - h_{k_2}}) = \sum_{\ell=1}^K \left[(\widehat{J_{k_1\ell} - J_{k_2\ell}}) - (J_{k_1\ell} - J_{k_2\ell})\right] m_n^{\ell *}$. This property holds for all networks n. Also, since each of the functions $\{F_{\varepsilon|k}\}_{k=1}^K$ is nonlinear, $\{m_n^{k*}\}_{k=1}^K$ are nonlinear functions of one another. Sufficient variation in $\{m_n^{k*}\}_{k=1}^K$ across networks ensures:

$$h_{k_1} - h_{k_2} = \widehat{h_{k_1} - h_{k_2}} \quad \text{and} \quad J_{k_1\ell} - J_{k_2\ell} = \widehat{J_{k_1\ell} - J_{k_2\ell}},$$

for every $\ell \in \mathcal{K}$. Also, since k_1 and k_2 are chosen arbitrarily, this result holds for all $k_1, k_2 \in \mathcal{K}$. The proof in the case where $\alpha_n = W_n' d$ relies on an analogous argument, and it also closely follows the proof given by Brock & Durlauf (2007), Proposition 1. I exclude it for this reason. □

Proof of Theorem 2

Proof. To demonstrate that c, $\{h_{k_1} - h_{k_2}\}_{k_1,k_2}$, and $\{J_{k_1\ell} - J_{k_2\ell}\}_{k_1,k_2,\ell}$ are identified, I apply Corollary 5 of Proposition 2 in Manski (1988). This result not only allows me to recover the parameters of interest, but it also guarantees that the error distributions $\{F_{\varepsilon|k}\}_{k=1}^K$ are identified up to scale. Ordinarily, recovering the error distributions would be unnecessary. However, in this setting, the presence of endogenous interaction effects means that $\{m_n^{\ell*}\}_{\ell=1}^K$ is functionally dependent on $\{F_{\varepsilon|k}\}_{k=1}^K$. To handle this issue, I take a two-step approach to identification: first I recover c and $\{F_{\varepsilon|k}\}_{k=1}^K$, then I use these quantities to recover the rest.

To start, I show that c and $\{F_{\varepsilon|k}\}_{k=1}^K$ are identified up to scale. Consider any group $k \in \mathcal{K}$. By Assumption B.3, there is some element x_j of X that varies continuously across $\mathbb{R}$. Without loss of generality, let $x_j = x_1$, and normalize the coefficient c_1 to one. In addition, fix some network n, and define the quantity $\zeta_n^k = h_k + \alpha_n + \sum_{\ell=1}^K J_{k\ell} m_n^{\ell*}$. For any agent in group k who resides in network n and has individual-level observables X_i, the expected choice ω_i is:

$$\mathrm{E}(\omega_i | X_i, k, \alpha_n, \{m_n^{\ell*}\}_{\ell=1}^K) = F_{\varepsilon|k}(\zeta_n^k + X_i' c)$$

To recover $\{c, F_{\varepsilon|k}\}$, I must show $F_{\varepsilon|k}(\zeta_n^k + X_i' c) = \hat{F}_{\varepsilon|k}(\hat{\zeta}_n^k + X_i' \hat{c})$ implies $c = \hat{c}$ and $F_{\varepsilon|k} = \hat{F}_{\varepsilon|k}$ for any $X_i \in \text{supp}(X|k)$. This property holds by Manski's (1988) corollary. Moreover, since this argument holds for all $k \in \mathcal{K}$, I conclude that c and $\{F_{\varepsilon|k}\}_{k=1}^K$ are identified up to scale.

Having shown that c and $\{F_{\varepsilon|k}\}_{k=1}^K$ can be recovered, the rest of the proof follows by the same arguments used to prove Theorem 1. So, the rest of the parameters are also identified. □

Proof of Theorem 3

Proof. Updates in progress.