Discrete Choice with Generalized Social Interactions

Oscar Volpe*

July 7, 2022

Abstract

This paper explores how identity affects individual behavior through social interactions. I study a discrete choice model where agents wish to conform to the actions of certain types of people in their network, while seeking to deviate from the actions of others. Under this generalized framework, I explore what equilibrium outcomes arise from noncooperative decision-making. I find conditions that allow me to extend the properties of games with strategic complementarities to models that have strategic substitutabilities. Using these conditions, I characterize the existence, multiplicity, and dynamic stability of equilibria. I also show how negative interactions lead to welfare trade-offs between agents in a network. Finally, I discuss how this model can be brought to data. I introduce a novel strategy for achieving (nonparametric) partial identification of the interaction terms. This strategy provides an economic framework for measuring polarization between social groups given data about individual choices.

Key Words: social interactions, identity, polarization, strategic complementarity, dynamic stability, non-negative matrix, nonparametric identification.

I give special thanks to Steven Durlauf, whose continued guidance has made this project possible. I am also grateful for valuable comments and advice from Stéphane Bonhomme, Ozan Candogan, Arun Chandrasekhar, Michael Dinerstein, Ali Hortaçsu, Robert Moffitt, Magne Mogstad, Philip Reny, and Alex Torgovitsky, as well as the participants of the Applied Micro working group, IO working group, and 3rd year research seminar at the University of Chicago. All errors in this paper are my own.

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

1 Introduction

The role of social interactions in individual decision-making has received widespread attention in economics. This work has largely been focused on settings with uniform strategic complementarities, where agents seek to conform to the average behavior of the rest of their network. Meanwhile, negative social interactions, which occur when agents wish to deviate from the behavior of others, have posed longstanding challenges for economic analysis.

In this paper, I provide a framework for studying generalized interactions, where individuals may experience both positive and negative social influences. I consider a context with binary choices in which agents are affected differently by different types of people in their network. I explore what aggregate behaviors can arise in equilibrium and discuss how this framework is empirically tractable. This work may be especially useful for analyzing the role of social identity (e.g., gender, race, or politics) on economic behavior, as well as the emergence of polarized and segregated societies.

Models of social interactions rest on the general idea that people, when forming decisions, are directly influenced by the choices of others. When these spillover effects are positive, interactions produce social multipliers, which can explain large differences in aggregate outcomes across populations. An active research program studies this phenomenon by modeling a network of agents who experience conformity effects; see Blume et al. (2011) and de Paula (2016) for discussions. These models can exhibit a variety of interesting properties, including the potential to have multiple, dynamically stable equilibria.

Yet, despite extensive research about settings with uniformly positive interactions, little attention has been given to the possibility of negative interaction effects. One issue is that the tools economists use to analyze models with complementarities do not readily transfer to cases with strategic substitutabilities, i.e. negative interactions. For example, models involving negative interactions 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. These characteristics, among others, make it challenging to study equilibrium behavior.

Nevertheless, negative social interactions appear in many real-world contexts. Consider, for example, the role of social identity in determining fashion trends. Men and women tend to dress differently amid pressure to conform to the styles that are associated with their own gender identities. Negative interaction effects likely exist in this context. For men, there is stigma about being perceived as feminine, while women face expectations not to look masculine. Hence, the distinction between two social categories (“men” and “women”) leads people to select different types of outfits. Note that fashion can also be used to signal someone's social class. In particular, wealthy people might wear expensive clothes to distinguish themselves from those who have less money. This inclination to deviate from the behavior of other people can be modeled as a negative interaction effect.

As a second example, consider parents' motivations for naming their children. As argued by Lieberson (2000), a child's first name typically reflects social influences related to cultural identity. Throughout the twentieth century, immigrant parents tended to give their children more American names as a way to assimilate themselves into US society.[1] Whatever the complex mechanisms driving parents to assimilate, this basic desire to conform to the social norms in a new country is a type of positive interaction effect. Meanwhile, the emergence of distinctively Black names in US history has been cited as evidence of African Americans seeking to reclaim a unique Black identity.[2] In this case, the incentive to distinguish oneself from other groups of people is a negative interaction effect.

Another way that negative interactions manifest themselves in society is through political polarization. Recent studies have documented how people's behaviors are becoming increasingly divided along partisan lines; see, for example, Bertrand & Kamenica (2020) and Boxell, Gentzkow, & Shapiro (2022). Today in the US, political identity is highly predictive of someone's decision to wear a mask or to become vaccinated against the COVID-19 virus.[3] Scholars have attributed these correlations to partisan animosity, as well as to misinformation that spreads easily in a divisive political climate.[4] Both of these explanations point to negative interactions. In the first case, agents act out of resentment toward members of a different political party. For example, a conservative may choose not to put on a mask as an act of defiance against liberal mask-wearers. In the second case, agents choose whether to believe information depending the source. For example, someone might be skeptical of a social media post about vaccines if it is written by someone with different political views.

The examples above are emblematic of the ways that identity shapes social interactions and, by implication, human behavior. As argued by Akerlof & Kranton (2000), social identity explains a wide range of economic phenomena, and it is relevant for most everyday decisions that people make. Moreover, the distinction between social categories leads to diverse types of interaction effects, which may be positive (e.g., admiration, trust) or negative (e.g., prejudice, distrust) depending on the context. Consequently, one might imagine that specific patterns arise where agents pool or separate their choices on the basis of identity.

Exactly what equilibrium outcomes occur when agents are affected positively by some

[1] For more on this discussion, see Carneiro, Lee, & Reis (2020) and Abramitzky, Boustan, & Eriksson (2020).
[2] For research on distinctively Black names, see Fryer & Levitt (2004) and Cook, Logan, & Parman (2022).
[3] See Bertrand et al. (2020) and the Kaiser Family Foundation's COVID-19 Vaccine Monitor (2022).
[4] See Allcott et al. (2020), Kerr, Panagopoulos, & van der Linden (2021) and Bursztyn et al. (2022).

and negatively by others? When are these outcomes stable in dynamic settings, and what do they imply about social welfare? To address these questions, I study a binary choice model, where agents are divided into different social groups. Following Brock & Durlauf (2001), I assume that each agent is influenced by the expected average behaviors in the population. Crucially, however, I break the uniform complementarities assumption, so that the social interactions vary, and may even be negative, across groups. Under this generalized framework, the interaction effects can be expressed in terms of a matrix $\mathbf{J}$, where each entry $J_{k\ell}$ indicates how people in group k are affected by people in group ℓ.

I explore the equilibrium properties of the model under noncooperative decision-making. In particular, I characterize what types of interaction effects, i.e. what restrictions on $\mathbf{J}$, lead to the existence, uniqueness, and dynamic stability of equilibria. I identify two conditions, one strictly weaker than the other, each guaranteeing the existence of a pure strategy Nash equilibrium. Both conditions have meaningful economic interpretations. In either a strict or weak sense, they require that agents are not negatively affected by the aggregate behavior in their own group. Using these two conditions, I show how key results in the complementarities literature extend to a much broader class of models that involve substitutabilities.

I start by establishing necessary and sufficient conditions for the existence of multiple equilibria. I demonstrate that multiplicity depends on one single statistic: the spectral radius of the network matrix. This number measures the intensity of the cycles in a network, thereby quantifying the collective strength of the interaction effects. I show that a unique equilibrium exists whenever the spectral radius lies below a certain threshold, whereas multiple equilibria arise when the spectral radius is above this threshold.

I then turn to the issue of dynamic stability. In practice, locally stable equilibria are the ones that a researcher observes, while unstable equilibria represent tipping points between equilibrium outcomes. Therefore, it is critical to understand when the model has a locally stable equilibrium versus when it is globally unstable. I prove that a dynamically stable equilibrium almost always exists as long as agents are not strongly repelled by members of their own group. In Section 5.3, I describe the implications of this result in the context of school choice, where agents care about the peer composition in a school.

Next, I compare the expected utility of agents at different equilibria. I find that negative interactions introduce welfare trade-offs. Specifically, if two agents who would otherwise prefer the same action are negatively influenced by one another, then the equilibrium that maximizes welfare for one must minimize welfare for the other. I discuss what this finding implies about social inefficiency in the case of COVID-19 vaccines and political polarization.

To obtain my results, I rely on very few parametric assumptions. For example, when analyzing the model, I assume that private utility varies across groups and that the error distributions are nonparametric. Moreover, I describe how my findings generalize to a much broader class of models where individuals are embedded in a network. I compare models of local interactions, where agents interact with a subset of the network, to models of global interactions, where agents interact with the entire network. In doing so, I am able to classify what types of network structures are associated with stable equilibrium outcomes.

Finally, I turn to the question of identification. That is, given data on individual choices, what can be said about the role of social interactions? I consider a context where agents are affected differently by different types of people, and where all the interactions take place within a local network (e.g., a neighborhood). I also assume that the researcher has data pertaining to multiple independent networks. Given this context, I introduce a new strategy to partially identify the social interaction effects, while also accounting for the issues of reflection and correlated unobservables. I show that, by differencing-out the network fixed effects for members of two groups residing in the same network, I am able to recover the differences between the interaction effects. I consider the benefits and limitations of this identification strategy, as well as the potential for self-selection into networks. I then explain how this approach can be used to measure the amount of polarization between social groups.

Related Literature

This paper is connected to three large literatures. First, it relates to a longstanding literature on strategic complementarity and substitutability in economics. The microeconomic foundations for models of complementarity are largely developed by Donald Topkis (see Topkis (1998) for a summary), as well as Vives (1990), and Milgrom & Roberts (1990). This work leverages key results from lattice theory, e.g., Tarski's fixed point theorem, to study equilibrium behavior in settings with uniform complementarities.[5] Subsequently, scholars have drawn from these ideas to advance theories of positive social interactions. Examples include Glaeser, Sacerdote, & Scheinkman (1996, 2003), Brock & Durlauf (2001, 2002), Calvo-Armengol, Patacchini, & Zenou (2009), and Cabrales, Calvo-Armengol, & Zenou (2011), among others.[6] The most relevant paper for my analysis is Brock & Durlauf (2001), which develops a binary choice model where all agents want to conform to the behaviors of others.

Strategic substitutabilities also appear in economics, albeit outside of the social interactions literature. One notable example is a public goods game, which often involves strict

[5] Cooper & John (1988) also provide an early discussion about the role of complementarity in economics. Bernheim (1994) describes how agents act under conformity effects even when their private preferences may vary. Further theoretical contributions have been made by Milgrom & Shannon (1994) and Athey (2001, 2002).

[6] In addition, Calvo-Armengol & Jackson (2004, 2007) study complementarities in labor-market networks.

substitutes. Recently, Bramoulle & Kranton (2007), Bramoulle, Kranton, & D'Amours (2014), and Elliott & Golub (2019) have studied network-based models of public goods. These authors have made important contributions to the field, while focusing primarily on models with continuous action spaces where network effects are weak enough to generate a unique equilibrium. However, as Jackson & Zenou (2015) explain, there are still unresolved questions about how and why models with substitutabilities behave differently from models with uniform complementarities. My findings in this paper should help to reconcile these differences by characterizing when models with negative interactions are "well-behaved".

Second, this paper contributes to an extensive literature about the identification of social interaction effects. Much of this work has focused on the linear-in-means setup; see, e.g., Kline & Tamer (2020) for a recent discussion. However, I consider a discrete choice model with incomplete information. This type of model is studied by Brock & Durlauf (2001, 2007), Aradillas-Lopez (2010, 2012), Bajari et al. (2010), and others. Similar models have also been studied in the treatment effects literature; see, for example, Sobel (2006), Manski (2013), and Lazzati (2015). In this paper, I propose a new identification strategy, which exploits within-network variation in individual-level characteristics, i.e. social identities. This strategy applies techniques of functional differencing (see, e.g., Bonhomme, 2012) to control for network-level determinants of choices. Although this approach is new to the social interactions literature, similar techniques have been proposed in other fields. 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.

Third, this paper relates to a large and growing literature on the economics of identity. Modern work on this topic originates from a collection of papers by Akerlof & Kranton (2000, 2002, 2005, 2008). This work provides rich discussions about the importance of identity for economic outcomes. Other theories of identity in economics include Dessi (2008), Shayo (2009), Benabou & Tirole (2011), and Bonomi, Gennaioli, & Tabellini (2021). Increasingly, applied work has highlighted the role of social identity in forming people's preferences; for example, consider Bertrand et al. (2015, 2020), Bursztyn et al. (2017a,b), and Dahl et al. (2020, 2021).[7] Although I do not explore the mechanisms that drive social interaction effects, I hope that the generalized framework presented in this paper will help to better integrate theories of identity into models of social interactions.

This paper proceeds as follows. Section 2 presents the binary choice model and discusses the implications of negative interaction effects for equilibria. Section 3 characterizes the key equilibrium properties of the model. Section 4 describes how these results apply in the

[7] I refer to Charness & Chen (2020) and Shayo (2020) for comprehensive discussions of this literature.

example of vaccines and political polarization. Section 5 considers alternative specifications and explores how the equilibrium properties generalize to different types of network-based models. Section 6 outlines the identification strategy. Finally, Section 7 concludes.

2 Model

2.1 Utility Maximization

I consider a binary choice model with social interaction effects that vary on the basis of group identity. This model generalizes Brock & Durlauf's (2001) binary choice framework by allowing individuals to be influenced differently, perhaps even negatively, by different types of people. For example, agents may wish to conform to the average behavior in certain groups, while also seeking to distinguish themselves from others. Unlike previous work, this model allows the error distributions to be nonparametric. This feature will ensure that the properties of the model are robust under relatively loose functional form assumptions.

Suppose there is a population of I individuals that is divided into K mutually-exclusive groups. Each agent i chooses a binary action ω_i from the set $\{-1, 1\}$ at some common time. Let $\bar{\omega}^k$ denote the average action in group k, and let $\bar{\omega}^k_{-i}$ be the average action among members of group k excluding i. The utility function for an agent i in group k is:

$$V^k(\omega_i) = u^k(\omega_i) + J_{kk}\omega_i E_i(\bar{\omega}^k_{-i}) + \sum_{\ell \neq k} J_{k\ell}\omega_i E_i(\bar{\omega}^\ell) + \varepsilon^k(\omega_i), \tag{1}$$

for $k = 1, \ldots, K$. Here, $u^k(\cdot)$ represents the private utility associated with a choice, and $\varepsilon^k(\cdot)$ is a random utility term that is independently distributed across agents. Notice that each of these terms is superscripted by k, as the functional forms may vary by group membership. Define $E_i(\bar{\omega}^k_{-i})$ and $E_i(\bar{\omega}^\ell)$ as agent i's subjective expectations about $\bar{\omega}^k_{-i}$ and $\bar{\omega}^\ell$, respectively. 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.[8] 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 replace $u^k(\cdot)$ with an affine function: $u^k(\omega_i) = h_k\omega_i + \eta_k$. Also, I write $\varepsilon^k(\omega_i) = \varepsilon_i\omega_i + \xi_i$ 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 preference for this action.

[8]For motivation behind the proportional spillovers assumption, I refer to Brock & Durlauf (2001).

Assume that the random payoff terms ε_i follow group-specific distributions:

$$P(\varepsilon_i \leq z|k) = F_{\varepsilon|k}(z), \tag{2}$$

for $k = 1, \ldots, K$, where each function $F_{\varepsilon|k}(\cdot)$ is continuously differentiable and symmetric about zero, taking positive density everywhere. To ensure a broad scope for the model, I do not make further parametric assumptions about these error distributions. So, this framework will apply under 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 collection of distribution functions $\{F_{\varepsilon|k}\}_{k=1}^{K}$, which describe how likely it is that any random payoff is realized in each group. Third, there is a matrix $\mathbf{J} \in \mathbb{R}^{K \times K}$, which contains all the social interaction effects:

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

Throughout this paper, I will refer to $\mathbf{J}$ as the *interaction matrix*. It can 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. 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, some of the interactions might be negative.

2.2 Equilibrium

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

$$P(\omega_i = 1|k) = F_{\varepsilon|k}\Big(h_k + J_{kk} E_i(\bar{\omega}_{-i}^{k}) + \sum_{\ell \neq k} J_{k\ell} E_i(\bar{\omega}^{\ell})\Big) \tag{4}$$

Since ω_i takes values in $\{-1, 1\}$, the expected action $E(\omega_i|k)$ is equal to $2P(\omega_i = 1|k) - 1$.

Assume agents have rational expectations about the behavior in each group. That is, let $E_i(\omega_j|k) = E(\omega_j|k)$ for all agents i and j, and all groups k. So, while people might not directly observe the choices of others, they do correctly infer the group averages in expectation.

In equilibrium, $E(\omega_i|k) = E(\omega_j|k)$ for all i, j, and k by symmetry of the conditional expected choice equations. Therefore, the expectation $E(\omega_i|k)$ for any agent i equals the expected average action for any subset of the members of group k. Let m^{k*} denote the value of $E(\bar{\omega}^k)$ in equilibrium. The vector $m^* = (m^{1*}, \ldots, m^{K*})$ satisfies the following K equations:

$$m^{k*} = 2F_{\varepsilon|k}\left(h_k + \sum_{\ell=1}^{K} J_{k\ell} m^{\ell*}\right) - 1, \tag{5}$$

for $k = 1, \ldots, K$. An equilibrium is characterized by the expected group mean choices that solve this system. In other words, a fixed point solution to (5) corresponds to the expected average behaviors in each group that are consistent with individually optimal decisions.

Before proceeding, it will be useful to introduce some key terminology. I define the Jacobian matrix of the equilibrium system (5) evaluated at $m^* \in [-1, 1]^K$ to be:

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

where $\boldsymbol{\beta}(m^*) = \operatorname{diag}\big(2f_{\varepsilon|1}\big(h_1 + \sum_{\ell=1}^{K} J_{1\ell} m^{\ell*}\big), \ldots, 2f_{\varepsilon|K}\big(h_K + \sum_{\ell=1}^{K} J_{K\ell} m^{\ell*}\big)\big)$ is a diagonal matrix of scaled density functions. Each term $f_{\varepsilon|k}\big(h_k + \sum_{\ell=1}^{K} J_{k\ell} m^{\ell*}\big)$ can be interpreted as the relative likelihood that an agent in group k is close to indifferent between the two actions at a particular equilibrium m^*. Therefore, $\mathbf{D}(m^*)$ equals the interaction matrix $\mathbf{J}$ where each row k is weighted according to the expected fraction of group k that is near indifferent at m^*. By construction, all the density functions $f_{\varepsilon|k}(\cdot)$ are strictly positive, which means that every entry $D_{k\ell}(m^*)$ of the Jacobian matrix has the same sign as the interaction effect $J_{k\ell}$.

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

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

Note that $\mathbf{A}$ is convergent, in the sense that $\lim_{t\to\infty} \mathbf{A}^t = \mathbf{0}$, if and only if $\rho(\mathbf{A}) < 1$. Therefore, a larger spectral radius is associated with a more expansive matrix. When applied to the Jacobian $\mathbf{D}(m^*)$, the spectral radius provides a way to measure the collective strength of the social interaction effects within and across groups. In the context of this model, $\rho(\mathbf{D}(m^*))$ will be instrumental in characterizing the existence of multiple equilibria.

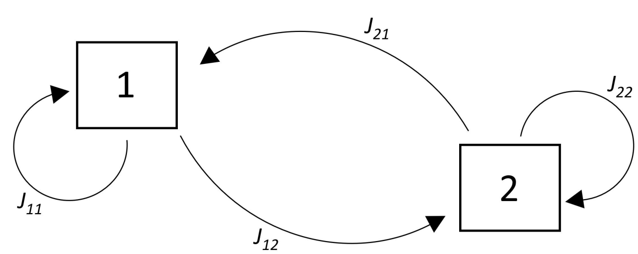
2.3 Potential for Negative Interactions

If all the entries of the interaction matrix were non-negative, then the model would exhibit strategic complementarities between all agents. In this case, each person's utility would be a supermodular function of individual choices, meaning that the marginal payoff from one's action (weakly) increases when anyone else takes that same action. These types of models tend to have important properties. For example, they possess equilibria that are well-ordered, in the sense that they form a complete lattice. They also typically involve dynamically stable equilibria, i.e. ones that can be solved for via fixed-point iteration on (5).[9]

When the interaction matrix contains negative entries, these equilibrium properties do not generally apply. This failure is not just a technical point, but it is driven by a deeper intuition. To understand why, suppose there is only one group in the population, i.e. $G = 1$, and assume, 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 $m^* = 2F_\varepsilon(Jm^*) - 1$ (group subscripts are removed for notational convenience). When $J < 0$, agents are repelled by the expected average action in the population. Therefore, if $E(\bar{\omega})$ is high, then people tend to prefer the low action. Conversely, if $E(\bar{\omega})$ is low, then most people choose the high action. When $J < 0$ is sufficiently large in magnitude, the equilibrium behavior in the model becomes unstable.[10] This instability carries an economic interpretation. That is, a population will never settle on one average action when everyone is always discontented with what that action will be.

When there are multiple groups, the issue of negative interactions becomes more complicated. To illustrate this point, suppose there are now two groups, i.e. $G = 2$. 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 might not exist if the within-group interaction effects J_{11} and J_{22} are very negative. Now, consider what happens when $J_{12} > 0$ and $J_{21} < 0$. Members of group 1 want to conform to the mean behavior in group 2, while members of group 2 seek to distinguish themselves from group 1. These social influences, when they are strong enough, lead to self-contradictory preferences. For example, if $E(\bar{\omega}^2)$ is high, then $E(\bar{\omega}^1)$ is also high, which means that $E(\bar{\omega}^2)$ is low, and so forth. Under noncooperative decision-making, this type of scenario can generate unstable equilibrium outcomes.

Figure 1: Social Interactions with Two Groups



[9] I refer to Milgrom & Roberts (1990) and Milgrom & Shannon (1994) for properties of supermodular games.

[10] Specifically, dynamic instability occurs when $J < -(2f_\varepsilon(0))^{-1}$. For an explanation, see Appendix C.

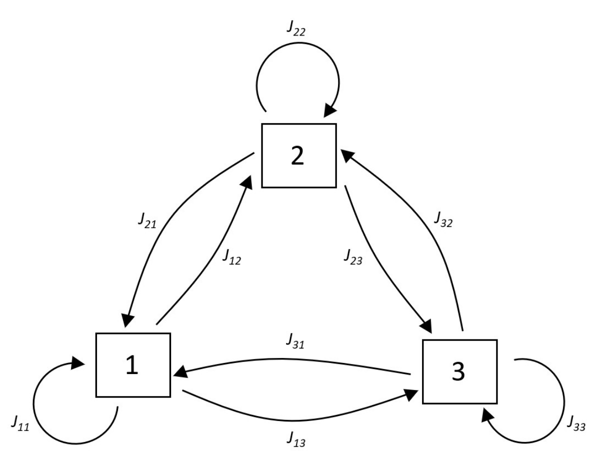
Going forward, it will be important to classify when models with negative interactions are "well-behaved". In other words, what types of matrices $\mathbf{J}$ are associated with stable behavioral outcomes? The examples above point to an issue that arises when preferences are self-contradictory in the sense that a population responds negatively to its own expected average choice. In general, if agents are sufficiently repelled by the behavior in their own group, then the model will exhibit *frustration*, a property where the interaction effects are incompatable with dynamically stable equilibria. So, it makes sense to rule out these cases.

Whenever there are multiple groups, within-group social utility arises from two channels: (1) a direct interaction with your own group and (2) an indirect interaction via the other groups that then interact with your group. Therefore, a natural condition might be that all within-group interactions, however indirect, should be non-negative. Formally, I write:

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

This condition requires that there is no channel through which individuals are affected negatively by their own group. For example, with two groups, *A.1* implies $J_{11} \geq 0$, $J_{22} \geq 0$, and $J_{12}J_{21} \geq 0$. Notice that $J_{12}J_{21} \geq 0$ does not preclude the possibility of negative between-group effects, but it does mean that J_{12} and J_{21} cannot have opposite signs. Therefore, *A.1* assumes that agents exhibit within-group strategic complementarities and that the between-group relations are mutual: they either "agree to agree" or "agree to disagree". When there are three groups, *A.1* also requires that the product of interaction effects for cycles of length three are non-negative, i.e. $J_{k\ell}J_{\ell j}J_{jk} \geq 0$ for all $k,\ell,j \in \mathcal{K}$. This case is depicted in Figure 2.

Figure 2: Social Interactions with Three Groups



Condition *A.1* helps to clarify the nature of relations between different social groups. Specifically, it allows someone to characterize when two groups are positively or negatively influenced by one another. To formalize these notions, consider the following definitions.

Definition 1. Group k is *connected* to group ℓ if there exist $j_1, j_2, \ldots, j_M \in \mathcal{K}$ where $J_{kj_1} J_{j_1 j_2} \ldots J_{j_M \ell} \neq 0$.

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

Whenever group k is positively (negatively) influenced by group ℓ, its members will seek to conform to (deviate from) the average behavior in ℓ. Suppose all groups are connected, i.e. that $\mathbf{J}$ is an irreducible matrix.[11] Then *A.1* implies: (1) every group is positively influenced by itself and (2) any two groups are either positively or negatively influenced by one another.

In certain contexts, it might be unrealistic to assume that *A.1* holds. In such cases, a much weaker condition will still ensure that the model of social interactions is "well-behaved". To understand why, recall that *A.1* restricts all sources of within-group social utility, when what really matters is that people should not wish to deviate from the mean behavior in their own group. For example, even if $J_{k\ell} J_{\ell k} < 0$ (violating *A.1*), it may be that agents are still positively affected overall by their own group when $J_{kk} > 0$ and $J_{\ell\ell} > 0$ are strong. Therefore, it makes sense to impose a looser requirement on the interaction matrix, which captures the idea that within-group interactions are cumulatively non-negative. This condition is stated below.

A.2. There exists an invertible matrix $\mathbf{B}$ such that $\mathbf{BJB}^{-1}$ has all non-negative entries.

Condition *A.2* requires that the interaction matrix $\mathbf{J}$ is similar to a non-negative matrix. Note that, when two matrices are similar, they represent the same linear operator under different bases. As I explain in Section 3, this condition allows me to prove that certain key properties associated with models of strategic complements can be extended to environments that involve strategic substitutes. In particular, *A.2* implies that the model has at least one pure strategy Nash equilibrium that is almost always dynamically stable.

To see how conditions *A.1* and *A.2* are related, consider the following result.

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

By this lemma, *A.1* is a special case of *A.2* where the change-of-basis matrix $\mathbf{B}$ is diagonal.

Since *A.2* is strictly weaker than *A.1*, it applies in a variety of settings where *A.1* is unlikely to hold. For example, suppose the interaction matrix $\mathbf{J}$ is symmetric, and assume:

$$J_{kk} \geq \sum_{\ell \neq k} |J_{k\ell}|, \tag{8}$$

for $k = 1, \ldots, K$. In this case, $\mathbf{J}$ is said to be *diagonally dominant*. This type of matrix always satisfies *A.2*, even though it need not satisfy *A.1*. To see why, note that every symmetric

[11] The matrix $\mathbf{J}$ is *irreducible* if, for any $k, \ell \in \mathcal{K}$, there is some $j_1, j_2, \ldots, j_M \in \mathcal{K}$ where $J_{kj_1} J_{j_1 j_2} \ldots J_{j_M \ell} \neq 0$.

matrix $\mathbf{J}$ can be diagonalized so that $\mathbf{J} = \mathbf{B}^{-1}\mathbf{\Lambda}\mathbf{B}$, where $\mathbf{B}$ is orthogonal and $\mathbf{\Lambda}$ is a diagonal matrix of eigenvalues of $\mathbf{J}$. By the Geršgorin circle theorem (see Appendix B for a formal statement), $\mathbf{J}$ is positive semidefinite, which implies that all the entries of $\mathbf{\Lambda}$ are non-negative.[12] This example is illustrative of the types of interactions that *A.2* allows but *A.1* does not. That is, if the level of cohesion in each group is sufficiently strong relative to the between-group interaction effects, then the model is "well-behaved" even when *A.1* fails.

Both *A.1* and *A.2* will be essential when interpreting the equilibrium behavior of the model. Under condition *A.2*, I prove results about the existence, multiplicity, and dynamic stability of equilibria. Under condition *A.1*, I obtain additional properties that allow me to characterize the welfare of different agents in equilibrium. As discussed in Section 5.2, both of these conditions apply to some of the most commonly-studied network structures. In addition, I give examples demonstrating that *A.2* is a much weaker restriction than *A.1*.[13]

3 Equilibrium Properties

3.1 Existence

There are two ways to prove the existence of an equilibrium in this model. First, since the error distributions $\{F_{\varepsilon|k}\}_{k=1}^{K}$ are continuous and the support of m^* is $[-1,1]^K$, Brouwer's fixed point theorem guarantees there is at least one solution to (5). I state this result below.

Property 1. There is at least one equilibrium $m^* \in [-1,1]^K$ in the binary choice model.

Note that this strategy for proving existence requires that the random payoff terms ε_i are continuously distributed. If these terms were to follow a discrete distribution or if the model had no random utility component, then Brouwer's fixed point theorem would not apply.

A second way to prove that equilibria exist is to invoke Tarski's fixed point theorem. This theorem ensures that there exists an equilibrium when the system (5) is non-decreasing, which happens only if $\mathbf{J}$ is a non-negative matrix. However, this result can be extended to cases where $\mathbf{J}$ satisfies condition *A.2*. Specifically, whenever *A.2* holds, the equilibrium system (5) maps to an alternate system of equations with a non-negative Jacobian matrix.[14] It follows that any equilibrium in the original model corresponds to an equilibrium in a different model where there are supermodular payoffs. Through this mapping, I can use

[12] More generally, every symmetric and positive semidefinite matrix satisfies condition *A.2*.

[13] Just as *A.1* and *A.2* allow me to generalize the properties of strategic complementarities, other matrix restrictions have been used to extend longstanding theories in economics. In the gross substitutes literature, examples include Metzler matrices and Morishima matrices. This work has been applied in numerous areas. For example, Tobin (1969) relied on gross substitutes restrictions to study the impacts of monetary policy.

[14] In Appendix B, I give a formal justification for this property, and I discuss its implications for equilibria.

Tarski's theorem to prove that equilibria exist under condition *A.2*. Crucially, this approach does not require that (5) is continuous. So, it applies to a much broader class of models. As I explain in Section 5.1, this argument is useful for proving the existence of pure strategy Nash equilibria in game-theoretic models that involve strategic substitutabilities.

3.2 Uniqueness

This model has the potential to exhibit multiple equilibria. In other words, for a given set of parameter values, there may be multiple solutions to (5). In this subsection, I examine what types of social environments will generate multiplicity versus uniqueness.

Unless the social interaction effects are sufficiently strong, the model always has a unique equilibrium. Recall that the overall intensity of social interactions can be measured by the spectral radius $\rho(\mathbf{D})$. When the entries of $\mathbf{J}$ grow larger in magnitude, the Jacobian $\mathbf{D}$ becomes more expansive, which means that $\rho(\mathbf{D}(m))$ is higher for all $m \in [-1,1]^K$. Whenever this spectral radius is less than 1 for every $m \in [-1,1]^K$, there exists a unique equilibrium. This property, which is stated below, is a consequence of the contraction mapping theorem.

Property 2. If $\sup_{m\in[-1,1]^K} \rho(\mathbf{D}(m)) < 1$, then the binary choice model has a unique equilibrium.

Having obtained a sufficient condition for uniqueness, the next step is to establish a necessary condition. That is, when exactly do multiple equilibria exist? I find that multiplicity occurs when agents experience a strong desire to conform to the average action in their own group. In these settings, aggregate behaviors are self-reinforcing, which can lead to multiple equilibrium outcomes from the same fundamentals. So, for there to be more than one equilibrium, it must be that agents are overall positively affected by their own group over some part of the support of m^*. This notion of net complementarity within groups is captured by condition *A.2*. If *A.2* is satisfied, and if the interaction effects are strong enough at some equilibrium m^*, then the model has multiple equilibria. Consider the following property.

Property 3. Suppose *A.2* holds. If there is an equilibrium $m^* \in [-1,1]^K$ where $\rho(\mathbf{D}(m^*)) > 1$, then there are two more equilibria $\underline{m}^*, \overline{m}^* \in [-1,1]^K$ where $\rho(\mathbf{D}(\underline{m}^*)) \leq 1$ and $\rho(\mathbf{D}(\overline{m}^*)) \leq 1$.

This property has two implications. First, it provides a sufficient condition for multiplicity. That is, if $\rho(\mathbf{D}(m^*))$ is above unity at some equilibrium m^*, then the model has at least three distinct equilibria. Recall that $\mathbf{D}(m^*)$ equals $\boldsymbol{\beta}(m^*)\mathbf{J}$, where $\mathbf{J}$ is the interaction matrix and $\boldsymbol{\beta}(m^*)$ measures how likely agents are to be near indifferent between the two actions at the equilibrium m^*. Therefore, multiple equilibria are more likely to occur in settings where social interactions are strong and/or where agents are relatively indifferent between choices.

The second implication of Property 3 is the restriction it places on the value of $\rho(\mathbf{D}(m^*))$

across different equilibria. In particular, if m^* is a unique equilibrium, then $\rho(\mathbf{D}(m^*)) \leq 1$. If there are multiple equilibria, then at least two of them satisfy $\rho(\mathbf{D}(m^*)) \leq 1$. This restriction lays the groundwork for Property 5, which addresses the dynamic stability of equilibria.

To justify Property 3, I first give a proof for the case where $\mathbf{J}$ is non-negative. This proof relies on key mathematical results that are associated with non-negative matrices. Namely, I apply the Perron-Frobenius theorem, which asserts that the largest real eigenvalue of any square, non-negative, irreducible matrix corresponds to a strictly positive eigenvector.[15] I then show how this proof may be extended to any setting where condition *A.2* is satisfied, i.e. where $\mathbf{J}$ is similar to a non-negative matrix. Importantly, this similarity transformation implies that the equilibrium system (5) shares many of the same characteristics as systems that have non-negative Jacobian matrices. These arguments are laid out in Appendix B.

In certain contexts, Property 2 and Property 3 coincide, leading to one condition that is both necessary and sufficient for multiple equilibria to exist. This concurrence takes place when *A.2* holds and $\rho(\mathbf{D}(m))$ is maximized at some equilibrium m^*. In this case, uniqueness arises if $\rho(\mathbf{D}(m^*)) < 1$ (Property 2) and multiplicity arises if $\rho(\mathbf{D}(m^*)) > 1$ (Property 3).[16] One notable example where this reasoning applies is when $0 \in \operatorname{argmax}_{x\in\mathbb{R}} f_{\varepsilon|k}(x)$ for $k \in \mathcal{K}$. This restriction ensures that, without deterministic private utility bias (i.e. when $h = \mathbf{0}_K$), agents are most likely to be indifferent between choices at $m^* = \mathbf{0}_K$. If this restriction is satisfied and if $h = \mathbf{0}_K$, then the spectral radius $\rho(\mathbf{D}(m))$ is maximized at $m^* = \mathbf{0}_K$, which is also an equilibrium. Letting $\boldsymbol{\beta}_0 = \text{diag}(2f_{\varepsilon|1}(0), \ldots, 2f_{\varepsilon|G}(0))$, I arrive at the following result.

Property 4. Suppose *A.2* holds and assume $0 \in \operatorname{argmax}_{x\in\mathbb{R}} f_{\varepsilon|k}(x)$ for all $k \in \mathcal{K}$.

(i) Let $h = \mathbf{0}_K$. Whenever $\rho(\boldsymbol{\beta}_0\mathbf{J}) > 1$, the model has at least three equilibria. One of these equilibria is $m^* = \mathbf{0}_K$ and the other two equilibria are symmetric about zero. Conversely, whenever $\rho(\boldsymbol{\beta}_0\mathbf{J}) < 1$, the model will have a unique equilibrium $m^* = \mathbf{0}_K$.

(ii) Let $h \neq \mathbf{0}_K$. Then there is a set $\mathcal{H}$, which is defined in terms of $\boldsymbol{\beta}$ and $\mathbf{J}$, such that the model possesses multiple equilibria if and only if $\rho(\boldsymbol{\beta}_0\mathbf{J}) > 1$ and h is an element of $\mathcal{H}$.

Property 4 helps to clarify how multiplicity of equilibria depends on the relative strengths of h and $\mathbf{J}$. For any interaction matrix $\mathbf{J}$, there is always some h that guarantees a unique equilibrium. In general, when private utility causes agents to be more indifferent between choices, multiple equilibria are more likely to exist. The case where $0 \in \operatorname{argmax}_{x\in\mathbb{R}} f_{\varepsilon|k}(x)$ for all $k \in \mathcal{K}$ is particularly relevant since it encompasses any model where the error distributions $\{F_{\varepsilon|k}\}_{k=1}^{K}$ are (weakly) concave over $\mathbb{R}_+$. Therefore, it applies under some of the most

[15] The spectral properties of non-negative matrices are well-studied in mathematics (see Meyer (2000), sec. 8).

[16] When $\rho(\mathbf{D}(m^*)) = 1$, equilibrium analysis becomes more ambiguous. This case is not especially relevant for this model since it occurs with probability measure zero. See Appendix A for further discussion.

common parametric specifications of $\{F_{\varepsilon|k}\}_{k=1}^K$, such as logistic or Gaussian errors.

3.3 Dynamic Stability

To examine the stability of different equilibria over time, I consider a dynamic analogue of the binary choice model. Here, an equilibrium is deemed to be locally stable whenever it can be solved for through fixed point iteration on (5). Formally, I define $m^* \in [-1,1]^K$ to be a *locally stable equilibrium* if it is a limiting solution to:

$$m_t^k = 2F_{\varepsilon|k}\left(h_k + \sum_{\ell=1}^K J_{k\ell} m_{t-1}^\ell\right) - 1, \tag{9}$$

for $k = 1, \ldots, K$, where m_0 resides in some sufficiently small neighborhood of m^*.

An equilibrium $m^* \in [-1,1]^K$ is locally stable whenever the system (5) is a contraction at the vector m^*, which happens if $\rho(\mathbf{D}(m^*)) < 1$ (see Appendix C for the proof). So, by Property 2, a unique, locally stable equilibrium always exists when the social interactions are weak enough to ensure that $\rho(\mathbf{D}(m)) < 1$ for all $m \in [-1,1]^K$. As the interaction effects grow stronger, however, unstable equilibria can emerge. In these settings, it is worth understanding what sorts of interactions will still lead to some dynamically stable outcomes.

As a general rule, the model cannot have any stable equilibria when agents are strongly repelled by the expected average choice in their own group. One way to rule out this possibility is to impose condition *A.2*, which implies that the within-group interactions are cumulatively non-negative. Indeed, whenever *A.2* holds, the model almost always has at least one locally stable equilibrium.[17] This result, which is stated below, follows from Property 3.

Property 5. Suppose *A.2* holds. For almost all distribution functions $\{F_{\varepsilon|k}\}_{k=1}^K$:

(i) If there is a unique equilibrium, then it is locally stable.

(ii) If there are multiple equilibria, then at least two of them are locally stable.

Consider any equilibrium m^* that is locally stable. If group k is positively influenced by group ℓ, then $\lim_{h_\ell \to \infty} m^{k*} = 1$ and $\lim_{h_\ell \to -\infty} m^{k*} = -1$. Conversely, if group k is negatively influenced by group ℓ, then $\lim_{h_\ell \to \infty} m^{k*} = -1$ and $\lim_{h_\ell \to -\infty} m^{k*} = 1$. Now, suppose that k_1 is positively influenced by ℓ and that k_2 is negatively influenced by ℓ. Then, for any collection

[17] More generally, *A.2* ensures that the model has more locally stable equilibria than unstable equilibria. Also, if $\mathbf{J}$ is a positive semi-definite matrix (a special case of *A.2*), then there is an odd number of equilibria, where exactly $d+1$ are locally stable and d are unstable for some $d \in \mathbb{N}$. See Appendices A and B for justification. Note that this finding is consistent with a longstanding literature about odd numbers of equilibria in game theory; see, e.g., Wilson (1971) Harsanyi (1973), Kohlberg & Mertens (1986), and Govindan & Wilson (2001).

of groups $j_1, j_2, \ldots, j_M \in \mathcal{K}$, $\lim_{J_{\ell j_1} J_{j_1 j_2} \cdots J_{j_M \ell} \to \infty} (m^{k_1 *}, m^{k_2 *})$ equals either $(-1, 1)$ or $(1, -1)$.

3.4 Welfare Trade-offs

In this model, no strict Pareto ranking exists across equilibria, since extreme realizations of the random payoff term ε_i can dominate an individual's utility function. Therefore, to compare the welfare of agents at different equilibria, I consider the expected utility of a "typical" person in each group, i.e. prior to observing ε_i and ξ_i. In doing so, I am able to evaluate which equilibrium makes members of different groups better off on average. For an individual in group k and an equilibrium m^*, this expected utility is:

$$\begin{aligned} E(\max_{\omega_i} V^k(\omega_i)|m^*) &= E\left(\max_{\omega_i} \left\{ h_k \omega_i + \eta_k + \sum_{\ell=1}^{K} J_{k\ell} \omega_i m^{\ell *} + \varepsilon_i \omega_i + \xi_i \right\} \right) \\ &= E\left(\left| h_k + \sum_{\ell=1}^{K} J_{k\ell} m^{\ell *} + \varepsilon_i \right| \right) + \eta_k + E(\xi_i) \end{aligned} \tag{10}$$

When $E(\max_{\omega_i} V^k(\omega_i)|m^*)$ grows larger, agents in group k receive greater utility, on average, at m^*. So, this quantity provides a way to measure the aggregate welfare in group k.

To assess which equilibrium generates the highest expected utility in each group, it will be necessary to characterize how the equilibria are ordered on $[-1, 1]^K$. If there are strategic complementarities between all agents, i.e. if $\mathbf{J}$ is non-negative, then Tarski's fixed point theorem implies that the set of equilibria forms a complete lattice. In this case, there is always a maximal and a minimal equilibrium. In other words, there exists one equilibrium where $E(\bar{\omega}^k)$ is highest for all k and one equilibrium where $E(\bar{\omega}^k)$ is lowest for all k.

If there are negative interactions, then the equilibria in the model do not generally form a lattice structure. Consequently, there may not be any equilibrium that simultaneously maximizes (or minimizes) the expected mean choice in every group. This feature makes it challenging to determine the order of equilibria. Fortunately, when condition *A.1* holds, the equilibria are still ordered in a distinctive way. To see how, consider the following property.

Property 6. Suppose *A.1* holds. For any two groups k and ℓ:

(i) If k and ℓ are positively influenced by one another, then the equilibrium where $E(\bar{\omega}^k)$ is maximal (or minimal) is the same equilibrium where $E(\bar{\omega}^\ell)$ is maximal (or minimal).

(ii) If k and ℓ are negatively influenced by one another, then the equilibrium where $E(\bar{\omega}^k)$ is maximal must be the same equilibrium where $E(\bar{\omega}^\ell)$ is minimal, 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 context, there exist two *extremal equilibria* (call them $\underline{m}^*$ and $\overline{m}^*$) at which $E(\bar{\omega}^k)$ is either maximized or minimized for all groups k. As I prove in Appendix A, both $\underline{m}^*$ and $\overline{m}^*$ are always locally stable. So, under appropriate initial conditions, iteration on best-response dynamics will converge to an extremal equilibrium.

As seen in the next result, $E(\max_{\omega_i} V^k(\omega_i)|m^*)$ is always maximized at an extremal equilibrium. In particular, if agents in group k tend to prefer the positive (negative) action, then they will maximize their expected utility at the equilibrium where $E(\bar{\omega}^k)$ is highest (lowest).

Property 7. Suppose *A.1* holds. Let $\underline{m}^*$ and $\overline{m}^*$ denote the two extremal equilibria.

(i) If $h = \mathbf{0}_K$, then $\underline{m}^*$ and $\overline{m}^*$ are symmetric about zero and both maximize $E(\max_{\omega_i} V^k(\omega_i)|m^*), \forall k$.

(ii) For any group k, there is a threshold T_k such that:

- When $h_k > T_k$, the equilibrium where $E(\bar{\omega}^k)$ is highest maximizes $E(\max_{\omega_i} V^k(\omega_i)|m^*)$.
- When $h_k < T_k$, the equilibrium where $E(\bar{\omega}^k)$ is lowest maximizes $E(\max_{\omega_i} V^k(\omega_i)|m^*)$.

Why is the threshold T_k not equal to zero? In any group k, an agent's deterministic preference over the choices depends on h_k, as well as on h_ℓ for every ℓ to which k is connected. For example, even if $h_k > 0$, agents in group k may tend to prefer the negative action if (1) $h_\ell < 0$ for some ℓ that positively influences k or if (2) $h_\ell > 0$ for some ℓ that negatively influences k. Only if h_k is strong enough to overcome these external influences, i.e. if h_k lies above some threshold T_k, will $E(\max_{\omega_i} V^k(\omega_i)|m^*)$ be maximized at the highest equilibrium.

Taken together, Properties 6 and 7 imply that negative interactions introduce welfare trade-offs. To see why, consider two groups k and ℓ, and assume $h_k > T_k$ and $h_\ell > T_\ell$. By Property 7(ii), agents in k and ℓ tend to be better off at the equilibria where $E(\bar{\omega}^k)$ and $E(\bar{\omega}^\ell)$ are highest, respectively. By Property 6, these "best" equilibria coincide when k and ℓ are positively influenced by one another, and they fail to coincide when k and ℓ are negatively influenced by one another. So, there is a trade-off. If two groups are biased toward the same action, and if they are sufficiently repelled by one another, then they maximize their expected utility at different equilibria. A similar trade-off arises whenever two groups prefer different actions, i.e. $h_k > T_k$ and $h_\ell < T_\ell$, and they are positively influenced by one another.

These welfare trade-offs become even more striking upon considering the next property.

Property 8. Fix any group k, and let $\underline{m}^*$ ($\overline{m}^*$) be the equilibrium where $E(\bar{\omega}^k)$ is lowest (highest).

(i) If $\underline{m}^{k*}, \overline{m}^{k*} \geq 0$, then $\underline{m}^*$ minimizes $E(\max_{\omega_i} V^k(\omega_i)|m^*)$ and $\overline{m}^*$ maximizes $E(\max_{\omega_i} V^k(\omega_i)|m^*)$.

(ii) If $\underline{m}^{k*}, \overline{m}^{k*} \leq 0$, then $\underline{m}^*$ maximizes $E(\max_{\omega_i} V^k(\omega_i)|m^*)$ and $\overline{m}^*$ minimizes $E(\max_{\omega_i} V^k(\omega_i)|m^*)$.

Suppose that more than half of the people in groups k and ℓ choose the positive action at every equilibrium. This event occurs when both groups are strongly biased toward $\omega_i = 1$. If k and ℓ are positively influenced by one another, then they share the same best and worst equilibria. If they are negatively influenced by one another, then the equilibrium that is best for one group is also worst for the other group. In other words, it is impossible to maximize the aggregate welfare in group k without minimizing the aggregate welfare in group ℓ.

4 Example: Vaccines and Political Polarization

To illustrate how Properties 1-8 can be used to study real-world behavior, I return to the example about COVID-19 vaccination rates in the US. Consider a stylized model where there are two political identities: Democrats (D) and Republicans (R). Members of both parties must decide whether or not to receive a vaccine. Their choices depend on privately-held preferences, as well as social interaction effects that vary based on political partisanship.

Assume that agents prefer to resemble people in their own party, i.e. that $J_{DD}, J_{RR} \geq 0$. Moreover, suppose that agents want to distinguish themselves from members of the other party, i.e. that $J_{DR}, J_{RD} \leq 0$. In this context, the interaction matrix $\mathbf{J}$ satisfies condition *A.1*, which means that there almost always exists at least one dynamically stable equilibrium.

As of April 2022, the self-reported vaccination rate is 92% among Democrats and 55% among Republicans.[18] Under the binary choice framework, the difference between these vaccination rates can be explained in two ways. First, the outcomes could be driven primarily by private preferences, rather than by social interactions. In this case, Democrats are inherently more inclined to receive a vaccine than Republicans, i.e. $h_D > h_R$, and partisan animosity does not play a significant role. This explanation suggests that there is a unique equilibrium and that the interaction effects are not strong enough to generate multiplicity.

A second way to explain the difference in vaccination rates is through social interactions. Under this interpretation, agents in both parties might share similar private preferences, but partisan resentment causes them to make different choices. Therefore, negative interaction effects J_{DR} and J_{RD} would boost the vaccination rate among Democrats, while reducing the rate among Republicans. By Property 3, this environment may have multiple equilibria if the interaction effects are sufficiently strong. While the equilibrium that has been realized involves a higher vaccination rate for Democrats, a second stable equilibrium might involve a higher vaccination rate for Republicans. In this context, there would also be an unstable equilibrium, which acts as a tipping point between these two stable equilibrium outcomes.

[18] These disaggregated statistics come from Kaiser Family Foundation's COVID-19 Vaccine Monitor (2022).

If there are multiple equilibria, then the equilibrium that has been realized is more favorable to Democrats than it is to Republicans. To see why, note that more than half of the people in each party claim to be vaccinated. Based on this observation, the model indicates that both Democrats and Republicans tend to be biased toward receiving a vaccine. Therefore, agents tend to maximize their utility at the equilibrium where their own political party has the highest vaccination rate. As it stands, the current equilibrium would be superior for most Democrats and inferior for most Republicans. Properties 6 and 8 formalize this idea.

From a policy perspective, negative interactions introduce social inefficiencies. To illustrate this point, consider any intervention that makes vaccines less costly to Democrats, i.e. one that increases h_D. Whenever $J_{RD} < 0$, this intervention has the externality of reducing the vaccination rate among Republicans. In a highly divisive political climate, increasing h_D can even lead to a reduction in the overall vaccination rate. Therefore, policies can sometimes have adverse consequences if there are negative interaction effects in the network.

5 Extensions

5.1 Games on Networks

The properties in Section 3 are not unique to the binary choice framework. They carry implications for a more general class of models in which individuals are connected through a network structure. As an illustration of this point, consider a game where each player k makes a choice a_k from a compact action space A_k in $\mathbb{R}$. A player k's best response, given a profile of actions $a \in \{A_1, \ldots, A_K\}$, is:

$$a_k^* = q_k\Big(\sum_{\ell=1}^{K} J_{k\ell} a_\ell\Big), \tag{11}$$

where $q_k(\cdot)$ is some non-decreasing function. The interdependence of the players' payoffs is captured by the directed graph $(\mathcal{K}, \mathbf{J})$, where the nodes $\mathcal{K} = \{1, \ldots, K\}$ represent players and the matrix $\mathbf{J}$ contains the interaction effects.

This game encompasses a variety of economic models. For example, the action a_k could represent a player's contribution toward a public good, with everyone benefiting from how much their neighbors contribute. Bramoulle, Kranton, & D'Amours (2014) study this type of game, where there are strategic substitutabilities between agents.[19] In their paper, players choose actions from the interval $[0, 1]$ and best responses are $a_k^* = \max\big\{0, 1 - \delta \sum_{\ell=1}^{K} g_{k\ell} a_\ell\big\}$

[19] More recent work on public goods games in networks includes Allouch (2017) and Elliott & Golub (2019).

for $\delta \in [0, 1]$, where $g_{k\ell} \in \{0, 1\}$ indicates whether players k and ℓ are linked. This setup can be expressed as (11) by setting $q_k(x) = \max\{0, 1+x\}$ and $J_{k\ell} = -\delta g_{k\ell}$. Alternatively, consider a public goods game where agents choose from a finite action space. This specification may also be written in terms of (11) where the functions $q_k(\cdot)$ are discontinuous.

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 take a collective action. 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 the action profiles a^* at which no player k wishes to deviate from a^*_k. In games of strategic complements, where each $J_{k\ell}$ is non-negative, an equilibrium always exists. However, if there are strategic substitutes, then existence is not guaranteed without further assumptions (e.g., continuous best responses and compact, convex action spaces). One of the key findings in this paper is that any game with interaction effects that satisfy *A.2* will possess an equilibrium.

Aside from ensuring existence, condition *A.2* is also useful for studying the uniqueness and dynamic stability of equilibria. Specifically, if the system of best responses (11) is continuously differentiable at the equilibria, then the spectral radius of its Jacobian matrix serves as a sufficient statistic for determining multiplicity. Additionally, condition *A.2* guarantees that this type of game almost always has at least one locally stable equilibrium.

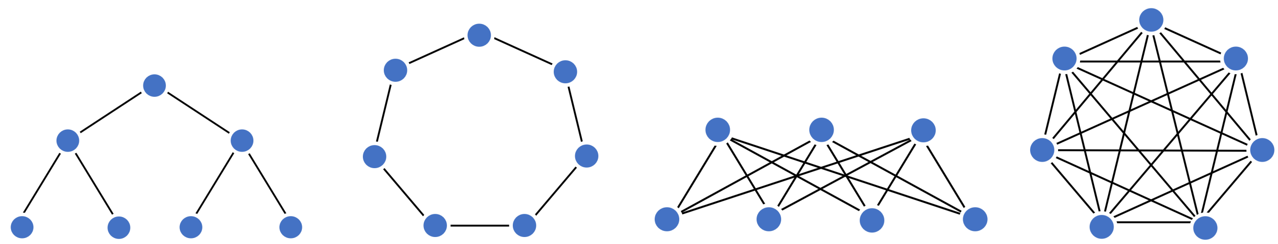
Even when *A.2* holds, the equilibria in this game may not be well-ordered. In particular, they might not form a lattice structure, which is useful for drawing welfare comparisons. As seen through Property 6, ensuring well-ordered equilibria comes from imposing the stronger condition *A.1*. This condition guarantees that there 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 characteristic of equation (11) is that every player k's best response (weakly) increases or decreases monotonically in a_ℓ, for all $\ell \in \mathcal{K}$. Consequently, if (11) is continuously differentiable, then the entries of the Jacobian matrix retain the same sign for every value of a. However, this monotonicity assumption can be weakened. Suppose, for example, that whether players conform to, or deviate from, one another depends on their current state of behaviors. Then, conditions *A.1* and *A.2* might only apply for some action profiles a, but not for others. In these cases, proving equilibrium properties would amount to arguing that *A.1* and/or *A.2* are satisfied locally over certain regions of the support of a.

5.2 Stable Network Structures

Given the implications of *A.1* and *A.2* for equilibrium behavior, it is worth understanding what network structures will satisfy these two conditions. Namely, what types of exclusion restrictions, i.e. 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 3: Graphs with Seven Nodes



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

Example 1 (Trees): Consider a tree with K nodes. This type of network may be especially relevant for studying interactions within social hierarchies. Note that trees encompass two of the most 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 condition *A.1* is satisfied for any tree. This example demonstrates how exclusion restrictions are useful for arguing that *A.1* holds. That is, when the links are sparse such that there are few or no cycles, *A.1* may be likely to hold even in very large networks. 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* cannot hold. However, if the game is adapted so that J_{kk} and $J_{\ell\ell}$ are both positive and large, then condition *A.2* may still apply, and the equilibrium behavior will be stable.[20]

Example 2 (Circles): Consider a circle with K nodes. This type of network is often used to illustrate domino effects that can arise in models of local interactions; see, for example, Ellison (1993). In such models, agents only interact with a few close contacts, rather than with

[20]The case where $J_{k\ell}J_{\ell k} < 0$ and $J_{kk}, J_{\ell\ell} = 0$ is strategically similar to the matching pennies game, which has no pure strategy Nash equilibrium. However, equilibria may exist in this game if $J_{kk} > 0$ and $J_{\ell\ell} > 0$ are strong relative to $J_{k\ell}$ and $J_{\ell k}$. For example, consider the matrices $\mathbf{J}$ and $\mathbf{B}$, where:

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

If $\delta \geq 3$, then $\mathbf{BJB}^{-1}$ is a non-negative matrix. Since $\mathbf{J}$ satisfies *A.2*, the corresponding game has an equilibium.

the entire population. For *A.1* or *A.2* to hold, this network would need to satisfy conditions that may seem unnaturally restrictive. To see why, suppose that the social interactions are symmetric, and let n_{edge} denote the number of edges involving negative interactions. Whenever n_{edge} is even, *A.1* always applies. However, if n_{edge} is odd, then every player is negatively influenced by itself, which means that neither *A.1* nor *A.2* can hold. Therefore, the stability of equilibria for a circle depends on the parity of negative interactions.

Example 3 (Bipartite Graphs): Consider a game with strategic substitutabilities between all players, i.e. where $J_{k\ell} \leq 0$ for any k, ℓ. This game satisfies *A.1* if and only if the corresponding graph is bipartite. As a general rule, *A.1* holds whenever the players can be divided into two teams, such that negative interactions only exist between members of different teams. Recall that, even when *A.1* fails, the weaker condition *A.2* would still guarantee the existence and dynamic stability of equilibria in models of strategic substitutes.

Example 4 (Complete Graphs): Consider a network where all K players are linked. This type of structure applies in contexts with global interactions, where agents are influenced by the aggregate behavior of the entire population. If the interactions are symmetric, then the fraction of complete graphs for which **J** satisfies *A.1* is $1/2^{\gamma_K}$, where $\gamma_K = (K-1)(K-2)/2$ for $K \geq 2$. So, when there are three players, this fraction is $1/2$. When there are seven players (depicted in Figure 3d), it is $1/32,768$. Hence, condition *A.1* is far more likely to apply in games where there are very few players and/or where the interaction effects are fairly uniform. For example, in the binary choice model, social interactions differ at the group-level. In a large population with few social groups, it is often realistic to assume *A.1*. Moreover, a complete graph may still satisfy *A.2* in cases where *A.1* does not apply. Recall, for example, that any symmetric, diagonally dominant matrix **J** satisfies *A.2*.

5.3 Preferences over Network Composition

So far, I have assumed that utility depends on the expected average choice in each group, regardless of which group comprises a bigger share of the population. I now consider an alternative setup, where utility depends on the expected composition of people who make a choice. That is, suppose agents care about $E(k|\omega_i)$ instead of $E(\omega_i|k)$.

This reformulation applies when modeling how social interactions drive selection into networks. For example, suppose agents are choosing whether to enter a new environment, such as a school, and they care about what groups of people they are likely to encounter. This scenario naturally leads to negative interaction effects, since a preference that one group is in the majority is equivalent to a preference that other groups are in the minority. Such models can be analyzed using the same techniques that I outlined in the previous sections.

Consider a stylized model where there are two social groups: a and b.[21] For notational convenience, I re-scale the choices so that $\omega_i \in \{1,2\}$. Let λ_a denote the share of the population in group a, and let $s_a(\omega_i)$ be the conditional probability of being in group a given the choice ω_i. Assume that an agent's payoff from selecting ω_i depends on the expected composition of people who also select ω_i. Therefore, the utility function can be written as:

$$V^k(\omega_i) = u^k(\omega_i) + J_k s_a(\omega_i) + \varepsilon^k(\omega_i), \tag{12}$$

for $k \in \{a,b\}$, where $u^k(\cdot)$ and $\varepsilon^k(\cdot)$ are specified exactly as before. Under this framework, the term J_k indicates how much agents in group k value being around people in group a.

An equilibrium is defined to be the values of $s_a(1)$ and $s_a(2)$ that are consistent with individually optimal choices. By Bayes' rule, equilibria are fixed point solutions to:

$$s_a(\omega_i) = \frac{\lambda_a P(\omega_i|a)}{\lambda_a P(\omega_i|a) + (1-\lambda_a)P(\omega_i|b)}, \tag{13}$$

for $\omega_i \in \{1,2\}$, where $P(\omega_i = 2|k) = F_{\varepsilon|k}\big(h_k + J_k(s_a(2) - s_a(1))\big)$. Just as before, the existence of an equilibrium is guaranteed by Brouwer's fixed point theorem.

Uniqueness and dynamic stability of equilibria depend on the Jacobian matrix $\mathbf{D} \in \mathbb{R}^{2\times 2}$ of the system (13). Letting $\beta_k = f_{\varepsilon|k}\big(h_k + J_k(s_a(2) - s_a(1))\big)$, this matrix has diagonal entries:

$$D_{jj} = s_a(j)[1 - s_a(j)] \left(\frac{J_a \beta_a}{P(\omega_i = j|a)} - \frac{J_b \beta_b}{P(\omega_i = j|b)} \right), \tag{14}$$

for $j \in \{1,2\}$, while the off-diagonal entries are $D_{12} = -D_{11}$ and $D_{21} = -D_{22}$. Notice that each row j of $\mathbf{D}$ is scaled by $s_a(j)[1-s_a(j)]$, which increases as $s_a(j)$ approaches $1/2$. Hence, the Jacobian matrix is more expansive when the agents making each choice are more diverse. Conversely, the Jacobian becomes less expansive when $s_a(j)$ approaches 0 or 1 for some j, which implies that one group makes up the majority of people choosing $\omega_i = j$.

Suppose that people prefer to associate with members of their own group, i.e. let $J_a \geq 0$ and $J_b \leq 0$. In this case, the Jacobian matrix always satisfies condition *A.1*. Therefore, the model almost always has a dynamically stable equilibrium even when the interaction effects are strong enough to generate multiplicity. However, if people prefer not to associate with members of their own group, i.e. if $J_a < 0$ or $J_b > 0$, then the model may not have any stable equilibria. As discussed Section 2.3, this instability arises when agents are repelled by their own group to such an extent that their preferences become self-contradictory.

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

Consider a context where one group is more desirable to everybody; e.g., let J_a and J_b both be positive. This type of scenario appears in numerous applications. For example, in Allende (2019), all parents prefer to send their children to schools that are made up of wealthy, high-achieving peers. In such cases, stable equilibria exist when J_b is small, while global instability occurs when J_b is large. So, the existence of stable equilibria requires that members of group b do not overwhelmingly seek to associate with people in group a. Note that this analysis applies to a static model, where group membership is fixed. In the example of school choice, certain characteristics (e.g., student achievement) are likely to change over time, while other traits (e.g., racial identity) are likely to remain constant. One area for future work is to study how the dynamics of identity drive the formation of networks.

6 Econometrics

6.1 Empirical Setting

I now show how the binary choice model can be brought to data. In particular, I explain how a researcher can measure the amount of polarization between two social groups by examining individual behavior. For any two groups k_1 and k_2, I define the level of polarization as $\delta_{k_1 k_2} = J_{k_1 k_1} + J_{k_2 k_2} - J_{k_1 k_2} - J_{k_2 k_1}$. This term quantifies how much people want to resemble their own group plus how much they want to distinguish themselves from the other group. In this section, I prove $\delta_{k_1 k_2}$ is identified for all $k_1, k_2 \in \mathcal{K}$ under few parametric assumptions.

I focus on a setting where all the social interactions take place in a local network. In practice, this network may refer to a neighborhood, a school, or a workplace. Agents are only influenced by the people in their network, and these influences are allowed to differ on the basis of social identity. Assume that there is data about multiple networks and that each individual i is drawn at random from a network $n \in \{1, \ldots, N\}$. Moreover, suppose that the researcher observes both the social identity k and network membership n of every agent.

For empirical implementation, I make two slight adjustments to the model. First, I rescale the choices so that ω_i takes values of 0 or 1. Note that this modification does not affect any of the equilibrium properties of the model, and it is made purely for notational convenience. Second, I relax the assumption that the private utility terms $h_1, \ldots, h_K$ are constant. By doing so, I can incorporate covariates and network fixed effects into the model.

For an agent i with social identity k who resides in network n, the choice ω_i depends on:

(*a*) individual-level factors, both observable ($X_i \in \mathbb{R}^r$) and unobservable ($\varepsilon_i \in \mathbb{R}$)

(*b*) network-level factors, both observable ($W_n \in \mathbb{R}^s$) and unobservable ($\alpha_n \in \mathbb{R}$)

(*c*) rational expectations about $\bar{\omega}^k_{n,-i}$ and $\{\bar{\omega}^\ell_n\}_{\ell \neq k}$, denoted by m^k_n and $\{m^\ell_n\}_{\ell \neq k}$ (respectively)

Given these considerations, I replace each term h_k with $\alpha_k + \alpha_n + X_i'c + W_n'd$, which is a linear function of group fixed effects α_k, network fixed effects α_n, individual-level observables X_i, and network-level observables W_n. Hence, an agent's choice ω_i is defined to be:

$$\omega_i = \mathbb{1}\left\{\alpha_k + \alpha_n + X_i'c + W_n'd + \sum_{\ell=1}^{K} J_{k\ell} m^\ell_n + \varepsilon_i \geq 0\right\} \tag{15}$$

Going forward, I make three simplifying assumptions about the error structure. First, I assume that ε_i, conditional on group membership k, is independent of individual-level observables and network-level factors. That is, let $P(\varepsilon_i \leq z | \{X_j\}_{j \in n}, W_n, \alpha_n, \alpha_k) = F_{\varepsilon|k}(z)$, where the group-specific error distributions $F_{\varepsilon|k}$ are specified exactly as in Section 2.1.[22] Second, I assume that X_i, conditional on group membership k, is independent of network-level factors. Therefore, I can write $P(X_i \leq x | W_n, \alpha_n, \alpha_k) = F_{X|k}(x)$, where $F_{X|k}$ denotes the distribution of X_i in group k. Taken together, these first two assumptions imply that there is no self-selection into networks based on X_i and ε_i. Although these restrictions may be weakened to allow for selection on observables, my analysis abstracts away from the possibility of selection into networks due to unobservable factors.[23] Finally, I assume that the errors ε_i and ε_j are independent for any two agents i and j within and across networks.

Just as before, agents act with incomplete information and rational expectations. Although they do not directly observe the realized behaviors of everybody in their network, they do infer the expected average choice $E(\bar{\omega}^k_n)$ for every group $k \in \mathcal{K}$. This setup is likely to apply whenever the size of the networks is large. In such contexts, agents might not see X_j and ε_j for every individual j, but they would likely know about W_n, α_n, and $\{\alpha_k\}_{k=1}^K$, as well as the distribution functions $\{F_{\varepsilon|k}\}_{k=1}^K$ and $\{F_{X|k}\}_{k=1}^K$. In addition, for large networks, the average choice $\bar{\omega}^k_n$ is will be closely approximated by the expectation $E(\bar{\omega}^k_n) = E(\omega_i | W_n, \alpha_n, \alpha_k)$.

An equilibrium in a network n is defined by the values of $(m^1_n, \ldots, m^K_n)$ that arise from individually optimal behaviors. For any k, I write $m^k_n = \int E(\omega_i | X_i, W_n, \alpha_n, \alpha_k) dF_{X|k}$, where:

$$E(\omega_i | X_i, W_n, \alpha_n, \alpha_k) = F_{\varepsilon|k}\left(\alpha_k + \alpha_n + X_i'c + W_n'd + \sum_{\ell=1}^{K} J_{k\ell} m^\ell_n\right) \tag{16}$$

As shown in Section 3.2, multiple equilibria may arise if the social interaction effects $\{J_{k\ell}\}_{k,\ell}$

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

[23] Suppose there is selection on observables, i.e. $P(X_i \leq x | W_n, \alpha_n, \alpha_k) = F_{X|W_n,k}$. Then, even if $F_{X|W_n,k}$ does not equal $F_{X|k}$, identification follows by comparing networks with similar observable characteristics W_n.

are sufficiently large in magnitude. When multiple equilibria are present, assume that agents know which one is realized, so there is no coordination involved in selecting an equilibrium.

6.2 Identification Results

The binary choice model is fully identified if there is only one set of parameter values that is consistent with the data under the modeling assumptions. Even when there is no self-selection into networks based on individual-level factors ε_i and X_i, two key barriers to identification remain. First, there is the *reflection problem* (see Manski, 1993), which takes place when $\{m_n^\ell\}_{\ell=1}^K$ and W_n are functionally dependent. This issue makes it impossible to disentangle the role of social interactions from contextual network effects. Second, there is the possibility of network-level unobservables, which occurs whenever $\alpha_n \neq 0$ for some n. If there are unobserved network factors, then the researcher cannot isolate the effect of social interactions on aggregate outcomes. Hence, identification of $\{J_{k\ell}\}_{k,\ell}$ tends to break down.

I propose a new strategy to partially identify the social interaction effects in the model. Crucially, this approach imposes no restrictions on the network-level determinants of choices, i.e. W_n and α_n. Therefore, it overcomes both of the identification challenges mentioned in the previous paragraph. The intuition behind 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 Claim.)* For any agent i in group k_1 and any agent j in group k_2, where i and j reside in network n: $E(\omega_i|X_i, W_n, \alpha_n, \alpha_{k_1}) = E(\omega_i|X_i, X_j, \alpha_{k_1}, \alpha_{k_2}, E(\omega_j|X_j, W_n, \alpha_n, \alpha_{k_2}))$.

Proof. The conditional expected choices $E(\omega_i|X_i, W_n, \alpha_n, \alpha_{k_1})$ and $E(\omega_j|X_j, W_n, \alpha_n, \alpha_{k_2})$ for agents i and j, respectively, are both defined according to equation (16). Since $F_{\varepsilon|k_2}$ is strictly increasing, its inverse $F_{\varepsilon|k_2}^{-1}$ exists. So, the equation for agent j may be re-written as:

$$\alpha_n = F_{\varepsilon|k_2}^{-1}(E(\omega_j|X_j, W_n, \alpha_n, \alpha_{k_2})) - \alpha_{k_2} - X_j'c - W_n'd - \sum_{\ell=1}^{K} J_{k_2\ell} m_n^\ell \tag{17}$$

Plugging this expression for α_n into the equation for agent i, I obtain:

$$\begin{aligned} E(\omega_i|X_i, W_n, \alpha_n, \alpha_{k_1}) = F_{\varepsilon|k_1}\bigg(&(\alpha_{k_1} - \alpha_{k_2}) + (X_i - X_j)'c \\ &+ \sum_{\ell=1}^{K} (J_{k_1\ell} - J_{k_2\ell}) m_n^\ell + F_{\varepsilon|k_2}^{-1}(E(\omega_j|X_j, W_n, \alpha_n, \alpha_{k_2}))\bigg) \end{aligned} \tag{18}$$

Note that all the network-level factors cancel out. Now, $E(\omega_i|X_i, W_n, \alpha_n, \alpha_{k_1})$ is expressed in terms of X_i, X_j, $\{m_n^\ell\}_{\ell=1}^K$, and $E(\omega_j|X_j, W_n, \alpha_n, \alpha_{k_2})$, which are all observed by the researcher. □

This lemma demonstrates how heterogeneity in social identity can be used to account for all contextual network effects. Consider any two individuals i and j with different social identities (k_1 and k_2, respectively) who both reside in the same network n. All the network-level determinants of $E(\omega_i|X_i, W_n, \alpha_n, \alpha_{k_1})$ are captured by $E(\omega_j|X_j, W_n, \alpha_n, \alpha_{k_2})$. Moreover, any difference between these conditional expectations must be driven by individual-level observables (i.e. X_i versus X_j), as well as social factors related to identity (i.e. k_1 versus k_2). This framework provides a panel structure, allowing me to control for the network effects in the model by comparing outcomes among different social groups within the same network.

As a consequence of Lemma 2, I can recover the terms c, $\alpha_{k_1} - \alpha_{k_2}$, and $\{J_{k_1\ell} - J_{k_2\ell}\}_{\ell=1}^K$ for any two groups $k_1, k_2 \in \mathcal{K}$. I give two versions of this identification result. First, I provide conditions for semiparametric identification, where the error distribution functions $\{F_{\varepsilon|k}\}_{k=1}^K$ are known by the researcher. Then, I offer conditions for nonparametric identification, where the functions $\{F_{\varepsilon|k}\}_{k=1}^K$ are unknown. While the nonparametric version allows for greater model flexibility, it also requires that there is a large amount of variability in the data. In both versions, identification is achieved without making any assumptions about W_n and α_n.

Semiparametric Identification

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

C.1. For some group $k \in \mathcal{K}$, supp$(X|k)$ is not contained in a proper linear subspace of $\mathbb{R}^r$.

C.2. Let $N > K$, and assume there is variation in $(m_n^1, \ldots, m_n^K)$ across networks $n \in \{1, \ldots, N\}$.

Assumption *C.1*, which corresponds to *X1* in Manski (1988), implies that there is no linear dependence in the individual-level observables for at least one group k. Assumption *C.2* requires that there is some amount of variation in the equilibrium outcomes across networks. Importantly, this variation across networks will allow me to isolate the role of social interactions from other group-specific factors that may affect people's choices. Taken together, *C.1* and *C.2* are sufficient to guarantee identification for any known $\{F_{\varepsilon|k}\}_{k=1}^K$.

Property 9.1. Suppose *C.1* and *C.2* hold, and assume the error distributions $\{F_{\varepsilon|k}\}_{k=1}^K$ are known. The parameters c, $\{\alpha_{k_1} - \alpha_{k_2}\}_{k_1,k_2}$, and $\{J_{k_1\ell} - J_{k_2\ell}\}_{k_1,k_2,\ell}$ in the model are point identified.

In practice, the error distributions are often only known up to a scale parameter (e.g., the variance), in which case $\{J_{k_1\ell} - J_{k_2\ell}\}_{k_1,k_2,\ell}$ are identified up to scale. However, if the distributions are exactly known, then—under appropriate normalizations—all the interaction effects $\{J_{k\ell}\}_{k,\ell}$ can be recovered. To see how, let $J_{\ell\ell} = 1$ for some group ℓ. Since $\{J_{k\ell} - J_{\ell\ell}\}_{k=1}^K$ are point identified, each term $J_{k\ell}$ may be determined relative to $J_{\ell\ell}$. Hence, by normalizing the diagonal entries of $\mathbf{J}$ to one, all the social interaction effects can be recovered from data.

Nonparametric Identification

Nonparametric identification is achieved by replacing *C.1* with a stronger assumption.

C.3. For all groups $k \in \mathcal{K}$, $\text{supp}(X|k)$ is not contained in a proper linear subspace of $\mathbb{R}^r$; also, there exists some component x_j of X such that, for almost every value of $x_{-j|k}$, the conditional distribution of $x_{j|k}$ given $x_{-j|k}$ has a positive density everywhere on $\mathbb{R}$.

This condition states that some component of X_i varies continuously within each group k over an unbounded support. Whenever *C.2* and *C.3* hold, there will be sufficient variation in the data to recover the parameters even when the error distributions $\{F_{\varepsilon|k}\}_{k=1}^K$ are unknown.

Property 9.2. If *C.2* and *C.3* hold, then $(c, \{\alpha_{k_1} - \alpha_{k_2}\}_{k_1,k_2}, \{J_{k_1\ell} - J_{k_2\ell}\}_{k_1,k_2,\ell})$ is identified up to scale.

Measuring Polarization

Three types of parameters are identified in this model. First, there is the vector of coefficients c on the individual-level observables X_i. Second, there are the differences in group fixed effects $\alpha_{k_1} - \alpha_{k_2}$ for any two groups $k_1, k_2 \in \mathcal{K}$. Third, there is the set $\{J_{k_1\ell} - J_{k_2\ell}\}_{k_1,k_2,\ell}$, where each $J_{k_1\ell} - J_{k_2\ell}$ indicates how agents in k_1 and k_2 differ in their preferences to conform to the behavior in group ℓ. By subtracting $J_{k_1k_2} - J_{k_2k_2}$ from $J_{k_1k_1} - J_{k_2k_1}$, I obtain:

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

which specifies how agents in k_1 and k_2 tend to prefer their own group over the other group.

From these parameters $\{\delta_{k_1k_2}\}_{k_1,k_2}$, one can measure the amount of polarization between any two social groups in a way that is motivated by the economic model. Moreover, this model allows there to be network fixed effects, group fixed effects, and individual-level factors, while the residual variation in behavioral outcomes is attributed to social interactions. In economics, polarization is a relatively abstract concept, and the terms $\{\delta_{k_1k_2}\}_{k_1,k_2}$ provide just one way to interpret it. Under this interpretation, polarization may be defined with respect to a particular choice. For example, if identity is more salient for some choices than others, then the terms $\{\delta_{k_1k_2}\}_{k_1,k_2}$ would depend on the decision that agents are making.

Lastly, I emphasize that this discussion is preliminary and by no means exhaustive. The identification strategy I propose is useful because it does not require instrumental variables. Rather, it exploits a panel structure within the network to control for contextual effects. However, my approach does not account for the possibility of self-selection into networks, which can pose an additional threat to identification. In practice, this issue is often addressed through the use of instruments, or by a selection correction procedure that involves explicitly modeling entry into networks; see, e.g., Blume et al. (2011) for further discussion. When self-selection is likely, my approach may be used in conjunction with these other methods.

7 Conclusion

The primary goal of this paper is to extend the theory of discrete choice and social interactions to more general settings, where agents experience a wide array of social influences. I analyze how aggregate behavioral outcomes depend on the features of a network, and I characterize externalities that arise in the presence of negative interaction effects.

One contribution of this work is to classify when models with negative interactions have the same types of properties as models with uniformly positive interactions. I obtain two conditions (*A.1* and *A.2*), which both require that agents are not repelled by their own behavior. These conditions allow me to prove a sufficiency result for the existence of multiple equilibria. In addition, they imply that the model almost always has a dynamically stable equilibrium. To my knowledge, both of these conditions are new to the literature.

Although I focus on social interactions, the findings in this paper may be used to study other types of network-based models where agents choose from a compact action space. In Section 5, I describe how my results may be broadly applicable, and I give special attention to four of the most commonly-studied network structures: trees, circles, bipartite graphs, and complete graphs. Note that my analysis only scratches the surface of the economic implications of negative interactions. There is far more research to be done on this topic, such as incorporating market prices or allowing the interaction effects to change over time.

In Section 6 of the paper, I consider how data can be used to recover the social interaction effects in the model. I show how the generalized interactions framework offers a panel structure, which can be exploited to isolate the role of social interactions from contextual network factors. By controlling for the network effects, my approach overcomes one of the major barriers to identification in models of social interactions. Nevertheless, it does not address the issue of self-selection into networks. One important area for future work is to better understand and account for the possibility of selection bias.

References

ABRAMITZKY, R., L. BOUSTAN, AND K. ERIKSSON (2020): "Do Immigrants Assimilate More Slowly Today Than in the Past?" *American Economic Review: Insights*, 2, 125–41.

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.

——— (2005): "Identity and the Economics of Organizations," *Journal of Economic Perspectives*, 19, 9–32.

——— (2008): "Identity, Supervision, and Work Groups," *American Economic Review*, 98, 212–17.

ALLCOTT, H., L. BOXELL, J. CONWAY, M. GENTZKOW, M. THALER, AND D. YANG (2020): "Polarization and public health: Partisan differences in social distancing during the coronavirus pandemic," *Journal of Public Economics*, 191, 104254.

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

ALLOUCH, N. (2017): "The cost of segregation in (social) networks," *Games and Economic Behavior*, 106, 329–342.

ARADILLAS-LOPEZ, A. (2010): "Semiparametric estimation of a simultaneous game with incomplete information," *Journal of Econometrics*, 157, 409–431.

——— (2012): "Pairwise-difference estimation of incomplete information games," *Journal of Econometrics*, 168, 120–140.

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.

AUERBACH, E. (2022): "Identification and Estimation of a Partially Linear Regression Model Using Network Data," *Econometrica*, 90, 347–365.

BAJARI, P., H. HONG, J. KRAINER, AND D. NEKIPELOV (2010): "Estimating Static Models of Strategic Interactions," *Journal of Business and Economic Statistics*, 28, 469–482.

BAXTER, R. (2007): *Exactly Solved Models in Statistical Mechanics*, Dover Books on Physics: Dover Publications.

BAYER, P., AND C. TIMMINS (2005): "On the equilibrium properties of locational sorting models," *Journal of Urban Economics*, 57, 462–477.

BENABOU, R., AND J. TIROLE (2011): "Identity, Morals, and Taboos: Beliefs as Assets," *The Quarterly Journal of Economics*, 126, 805–855.

BENJAMIN, D. J., J. J. CHOI, AND A. J. STRICKLAND (2010): "Social Identity and Preferences," *American Economic Review*, 100, 1913–28.

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., P. CORTES, C. OLIVETTI, AND J. PAN (2020): "Social Norms, Labour Market Opportunities, and the Marriage Gap Between Skilled and Unskilled Women," *The Review of Economic Studies*, 88, 1936–1978.

BERTRAND, M., AND E. KAMENICA (2020): "Coming Apart? Cultural Distances in the United States Over Time," working paper.

BERTRAND, M., E. KAMENICA, AND J. PAN (2015): "Gender Identity and Relative Income within Households," *The Quarterly Journal of Economics*, 130, 571–614.

BISIN, A., A. MORO, AND G. TOPA (2011): "The Empirical Content of Models with Multiple Equilibria in Economies with Social Interactions," Working Paper 17196, National Bureau of Economic Research.

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

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

BONHOMME, S. (2012): "Functional Differencing," *Econometrica*, 80, 1337–1385.

BONOMI, G., N. GENNAIOLI, AND G. TABELLINI (2021): "Identity, Beliefs, and Political Conflict," *The Quarterly Journal of Economics*, 136, 2371–2411.

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

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

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

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

——— (2002): "A Multinomial-Choice Model of Neighborhood Effects," *American Economic Review*, 92, 298–303.

——— (2007): "Identification of binary choice models with social interactions," *Journal of Econometrics*, 140, 52–75.

BURSZTYN, L., B. FERMAN, S. FIORIN, M. KANZ, AND G. RAO (2017a): "Status Goods: Experimental Evidence from Platinum Credit Cards," *The Quarterly Journal of Economics*, 133, 1561–1595.

BURSZTYN, L., T. FUJIWARA, AND A. PALLAIS (2017b): "'Acting Wife': Marriage Market Incentives and Labor Market Investments," *American Economic Review*, 107, 3288–3319.

BURSZTYN, L., A. RAO, C. ROTH, AND D. YANAGIZAWA-DROTT (2022): "Opinions as Facts," ECONtribute Discussion Papers Series 159, University of Bonn and University of Cologne, Germany.

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

CALVO-ARMENGOL, A., AND M. JACKSON (2007): "Networks in labor markets: Wage and employment dynamics and inequality," *Journal of Economic Theory*, 132, 27–46.

CALVO-ARMENGOL, A., AND M. O. JACKSON (2004): "The Effects of Social Networks on Employment and Inequality," *American Economic Review*, 94, 426–454.

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

CARNEIRO, P., S. LEE, AND H. REIS (2020): "Please call me John: Name choice and the assimilation of immigrants in the United States, 1900–1930," *Labour Economics*, 62, 101778.

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

COOK, L. D., T. LOGAN, AND J. PARMAN (2022): "The antebellum roots of distinctively black names," *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 55, 1–11.

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

DAHL, G. B., A. R. KOSTOL, AND M. MOGSTAD (2014): "Family Welfare Cultures," *The Quarterly Journal of Economics*, 129, 1711–1752.

DAHL, G. B., A. KOTSADAM, AND D.-O. ROOTH (2020): "Does Integration Change Gender Attitudes? The Effect of Randomly Assigning Women to Traditionally Male Teams," *The Quarterly Journal of Economics*, 136, 987–1030.

DAHL, G., C. FELFE, P. FRIJTERS, AND H. RAINER (2021): "Caught between Cultures: Unintended Consequences of Improving Opportunity for Immigrant Girls," *The Review of Economic Studies*.

DESSI, R. (2008): "Collective Memory, Cultural Transmission, and Investments," *American Economic Review*, 98, 534–60.

DURLAUF, S. N. (2001): "A Framework for the Study of Individual Behavior and Social Interactions," *Sociological Methodology*, 31, 47–87.

DURLAUF, S. N., AND Y. M. IOANNIDES (2010): "Social Interactions," *Annual Review of Economics*, 2, 451–478.

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.

FRICK, M., R. IIJIMA, AND Y. ISHII (2018): "Dispersed Behavior and Perceptions in Assortative Societies," Cowles Foundation Discussion Papers 2128, Cowles Foundation for Research in Economics, Yale University.

FRYER, R. G., AND S. D. LEVITT (2004): "The Causes and Consequences of Distinctively Black Names," *The Quarterly Journal of Economics*, 119, 767–805.

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*, 1, 345–353.

GOVINDAN, S., AND R. WILSON (2001): "Direct Proofs of Generic Finiteness of Nash Equilibrium Outcomes," *Econometrica*, 69, 765–769.

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

HATCHER, A., C. U. PRESS, AND C. U. D. O. MATHEMATICS (2002): *Algebraic Topology*, Algebraic Topology: Cambridge University Press.

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

HORST, U., AND J. A. SCHEINKMAN (2006): "Equilibria in Systems of Social Interactions," *Journal of Economic Theory*, 130, 44–77.

JACKSON, M. O., B. W. ROGERS, AND Y. ZENOU (2017): "The Economic Consequences of Social-Network Structure," *Journal of Economic Literature*, 55, 49–95.

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

JOHNSON, C. R., AND P. TARAZAGA (2004): "On Matrices with Perron–Frobenius Properties and Some Negative Entries," *Positivity*, 8, 327–338.

KENNAN, J. (2001): "Uniqueness of Positive Fixed Points for Increasing Concave Functions on Rn: An Elementary Result," *Review of Economic Dynamics*, 4, 893–899.

KERR, J., C. PANAGOPOULOS, AND S. VAN DER LINDEN (2021): "Political polarization on COVID-19 pandemic response in the United States," *Personality and Individual Differences*, 179, 110892.

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.

LAZZATI, N. (2015): "Treatment response with social interactions: Partial identification via monotone comparative statics," *Quantitative Economics*, 6, 49–83.

LEVY, R. (2021): "Social Media, News Consumption, and Polarization: Evidence from a Field Experiment," *American Economic Review*, 111, 831–70.

LIEBERSON, S. (2000): *A Matter of Taste: How Names, Fashions, and Culture Change*, New Haven, Connecticut: Yale University Press, New Haven, Connecticut: Yale University Press.

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.

——— (2000): "Economic Analysis of Social Interactions," *Journal of Economic Perspectives*, 14, 115–136.

——— (2013): "Identification of treatment response with social interactions," *The Econometrics Journal*, 16, S1–S23.

MCLENNAN, A. (2018): *Advanced Fixed Point Theory for Economics* in , Springer Books (978-981-13-0710-2): Springer.

MEYER, C. (2000): *Matrix Analysis and Applied Linear Algebra*, Other Titles in Applied Mathematics: Society for Industrial and Applied Mathematics (SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104).

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

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

MORRIS, S. (2000): "Contagion," *The Review of Economic Studies*, 67, 57–78.

MUNKRES, J. (2000): *Topology*, Featured Titles for Topology: Prentice Hall, Incorporated.

NAKAJIMA, R. (2007): "Measuring Peer Effects on Youth Smoking Behaviour," *The Review of Economic Studies*, 74, 897–935.

ORTEGA, J. M. (1972): *Numerical Analysis: A Second Course*.

DE PAULA, A. (2016): "Econometrics of network models," CeMMAP working papers CWP06/16, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

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

PESKI, M., AND B. SZENTES (2013): "Spontaneous Discrimination," *American Economic Review*, 103, 2412–36.

PLISCHKE, M., AND B. BERGERSEN (2005): *Equilibrium Statistical Physics (3rd Edition).*: World Scientific Publishing Company.

SHAYO, M. (2009): "A Model of Social Identity with an Application to Political Economy: Nation, Class, and Redistribution," *The American Political Science Review*, 103, 147–174.

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

SOBEL, M. E. (2006): "What Do Randomized Studies of Housing Mobility Demonstrate?: Causal Inference in the Face of Interference," *Journal of the American Statistical Association*, 101, 1398–1407.

STEIN, D., AND C. NEWMAN (2013): *Spin Glasses and Complexity*: Princeton University Press.

TOBIN, J. (1969): "A General Equilibrium Approach To Monetary Theory," *Journal of Money, Credit and Banking*, 1, 15–29.

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.

YOUNG, H. P. (2009): "Innovation Diffusion in Heterogeneous Populations: Contagion, Social Influence, and Social Learning," *American Economic Review*, 99, 1899–1924.

Appendix A

Fixed Point Properties of Monotone Mappings

Let $\mathcal{I} = [\underline{x}, \overline{x}]$ be an interval in $\mathbb{R}^n$, and let $\mathcal{Q} : \mathcal{I} \to \mathcal{I}$ be a mapping that is non-decreasing in each of its inputs. A fixed point x^* of $\mathcal{Q}$ is defined as a solution to $\mathcal{Q}(x^*) = x^*$. In this appendix section, I characterize the existence and multiplicity of fixed points of $\mathcal{Q}$.

Existence

Arguing that a fixed point exists is straightforward. Because $(\mathcal{I}, \leq)$ is a complete lattice and $\mathcal{Q} : \mathcal{I} \to \mathcal{I}$ is monotone increasing, Tarski's fixed point theorem ensures that $\mathcal{Q}$ has a fixed point somewhere on $\mathcal{I}$. In addition, the set of fixed points forms a complete lattice.

Multiplicity

I now examine when $\mathcal{Q}$ has multiple fixed points. To do so, I make two extra assumptions. First, suppose $\underline{x} < \mathcal{Q}(\underline{x})$ and $\mathcal{Q}(\overline{x}) < \overline{x}$, so there are no fixed points on the boundary. Second, suppose $\mathcal{Q}$ is continuously differentiable at every fixed point x^*. Under these two assumptions, I can define $\mathbf{D}_{\mathcal{Q}}(x^*)$ as the Jacobian matrix of $\mathcal{Q}$ evaluated at $x^* \in \text{int}(\mathcal{I})$. I also define the spectral radius $\rho(\mathbf{D}_{\mathcal{Q}}(x^*))$ as the largest eigenvalue of $\mathbf{D}_{\mathcal{Q}}(x^*)$ in absolute value.

Multiplicity Result. If there exists a fixed point x^* of $\mathcal{Q}$ where $\rho(\mathbf{D}_{\mathcal{Q}}(x^*)) > 1$, then there exist two more fixed points $\underline{x}^*, \overline{x}^*$ of $\mathcal{Q}$ where $\underline{x}^* < x^* < \overline{x}^*$, $\rho(\mathbf{D}_{\mathcal{Q}}(\underline{x}^*)) \leq 1$, and $\rho(\mathbf{D}_{\mathcal{Q}}(\overline{x}^*)) \leq 1$.

Proof. Suppose $\mathcal{Q}(x^*) = x^*$ and $\rho(\mathbf{D}_{\mathcal{Q}}(x^*)) > 1$ for some $x^* \in \text{int}(\mathcal{I})$. I consider two cases.

First, assume $\mathbf{D}_{\mathcal{Q}}(x^*)$ is an irreducible matrix. As $\mathbf{D}_{\mathcal{Q}}(x^*)$ is also non-negative, the Perron-Frobenius theorem implies:

$$\mathbf{D}_{\mathcal{Q}}(x^*)\hat{x} = \rho(\mathbf{D}_{\mathcal{Q}}(x^*))\hat{x} > \hat{x},$$

for some strictly positive vector $\hat{x} \in \mathbb{R}^n_{++}$. It follows that $\mathbf{D}_{\mathcal{Q}}(x^*)\lambda\hat{x} > \lambda\hat{x}$ for any $\lambda > 0$.

Taking the first-order Taylor approximation of $\mathcal{Q}(x^* + \lambda\hat{x})$ about x^*, I obtain:

$$\mathcal{Q}(x^* + \lambda\hat{x}) \approx \underbrace{\mathcal{Q}(x^*)}_{=x^*} + \underbrace{\mathbf{D}_{\mathcal{Q}}(x^*)\lambda\hat{x}}_{>\lambda\hat{x}}$$

For some sufficiently small λ, the vector $a = x^* + \lambda\hat{x}$, where $a \in \text{int}(\mathcal{I})$, satisfies $\mathcal{Q}(a) > a$. Therefore, letting $b = \overline{x}$, it must be that $a < \mathcal{Q}(a) < \mathcal{Q}(b) < b$. Tarski's fixed point theorem ensures that $\mathcal{Q}$ has a fixed point $\overline{x}^*$ somewhere between a and b. By construction, $\overline{x}^* > x^*$.

By Tarski's theorem, it must be that $\mathcal{Q}$ has greatest and least fixed points. Without loss of generality, suppose $\overline{x}^*$ is the greatest fixed point of $\mathcal{Q}$ on $\mathcal{I}$. If $\rho(\mathbf{D}_{\mathcal{Q}}(\overline{x}^*)) > 1$, then (by the same reasoning as before) $\mathcal{Q}$ has a fixed point that is strictly greater than $\overline{x}^*$. Arriving at a contradiction in this case, I conclude that $\rho(\mathbf{D}_{\mathcal{Q}}(\overline{x}^*)) \leq 1$. Using an analogous argument, I can locate another fixed point $\underline{x}^*$ of $\mathcal{Q}$, which satisfies $\underline{x}^* < x^*$ and $\rho(\mathbf{D}_{\mathcal{Q}}(\underline{x}^*)) \leq 1$.

Lastly, consider the case where $\mathbf{D}_{\mathcal{Q}}(x^*)$ is a reducible matrix. If $\rho(\mathbf{D}_{\mathcal{Q}}(x^*)) > 1$, then the same must hold for some irreducible block of $\mathbf{D}_{\mathcal{Q}}(x^*)$. Let $\mathcal{B}$ denote the set of indices within this block. Applying the Perron-Frobenius theorem to that block, there exists some vector $\hat{x}$, satisfying $\hat{x}_\ell > 0$ for $\ell \in \mathcal{B}$ and $\hat{x}_\ell = 0$ otherwise, so that $\mathcal{Q}(x^* + \lambda\hat{x}) > x^* + \lambda\hat{x}$ for $\lambda > 0$

sufficiently small. Setting $a = x^* + \lambda\hat{x}$ and $b = \overline{x}$, it follows that $a < \mathcal{Q}(a) < \mathcal{Q}(b) < b$. By Tarski's theorem, there is a fixed point of $\mathcal{Q}$ between a and b. So, just as in the previous case, there are fixed points $\underline{x}^*, \overline{x}^* \in \text{int}(\mathcal{I})$ where $\underline{x}^* < x^* < \overline{x}^*$, $\rho(\mathbf{D}_\mathcal{Q}(\underline{x}^*)) \leq 1$ and $\rho(\mathbf{D}_\mathcal{Q}(\overline{x}^*)) \leq 1$. □

One important special case occurs when each component of $\mathcal{Q}$ depends on a single input. Namely, suppose that $\mathcal{Q} = (\mathcal{Q}_1(x_{j_1}), \ldots, \mathcal{Q}_n(x_{j_n}))$ for some choices of $j_1, \ldots, j_n \in \{1, \ldots, n\}$. In this case, the matrix $\mathbf{D}_\mathcal{Q}$ has at most one nonzero entry in each row. These systems are especially convenient to study because they can be inverted in such a way that the inverse mapping $\mathcal{Q}^{-1}$ is also non-decreasing.[24] As a consequence of this property, I can write down an approximate converse to the *Multiplicity Result*. Consider the following corollary.

Corollary. (Multiplicity) Assume $\mathcal{Q} = (\mathcal{Q}_1(x_{j_1}), \ldots, \mathcal{Q}_n(x_{j_n}))$ for some $j_1, \ldots, j_n \in \{1, \ldots, n\}$, and assume there are two fixed points $\underline{x}^*, \overline{x}^*$ of $\mathcal{Q}$ where $\rho(\mathbf{D}_\mathcal{Q}(\underline{x}^*)) < 1$ and $\rho(\mathbf{D}_\mathcal{Q}(\overline{x}^*)) < 1$. Then there must be another fixed point x^* of $\mathcal{Q}$ where $\underline{x}^* < x^* < \overline{x}^*$ and where $\rho(\mathbf{D}_\mathcal{Q}(x^*)) \geq 1$.

To prove this corollary, I can reverse the steps taken to prove the *Multiplicity Result*. Namely, start by assuming that $\rho(\mathbf{D}_\mathcal{Q}(\underline{x}^*)) < 1$ and $\rho(\mathbf{D}_\mathcal{Q}(\overline{x}^*)) < 1$ for fixed points $\underline{x}^*, \overline{x}^* \in \mathcal{I}$. Use the Perron-Frobenius theorem to show that $\underline{x}^* < Q(a) < a < b < Q(b) < \overline{x}^*$ for some appropriate choices of a and b. Since $\mathcal{Q}^{-1}$ is non-decreasing, Tarski's theorem guarantees a fixed point x^* of $\mathcal{Q}^{-1}$ (and $\mathcal{Q}$, by implication) between a and b. Additionally, $\rho(\mathbf{D}_\mathcal{Q}(x^*)) \geq 1$ is satisfied without loss of generality (otherwise this argument may be repeated indefinitely).

This corollary almost provides a converse to the *Multiplicity Result*. In fact, excluding the case where $\rho(\mathbf{D}_\mathcal{Q}) = 1$ at fixed points of $\mathcal{Q}$, it gives an exact converse. Moreover, note that almost every mapping $\mathcal{Q}$ does not have fixed points where $\rho(\mathbf{D}_\mathcal{Q}) = 1$. Therefore, for almost every mapping of the form $\mathcal{Q} = (\mathcal{Q}_1(x_{j_1}), \ldots, \mathcal{Q}_n(x_{j_n}))$, every fixed point where $\rho(\mathbf{D}_\mathcal{Q}) > 1$ is bounded by two fixed points where $\rho(\mathbf{D}_\mathcal{Q}) < 1$. It follows that there is an odd number of fixed points: d where $\rho(\mathbf{D}_\mathcal{Q}) > 1$ and $d + 1$ where $\rho(\mathbf{D}_\mathcal{Q}) < 1$ for some $d \in \mathbb{N}$.

What happens when $\rho(\mathbf{D}_\mathcal{Q}(x^*)) = 1$ at a fixed point x^* of $\mathcal{Q}$? Without imposing additional functional form restrictions on the mapping $\mathcal{Q}$, it is ambiguous what this type of fixed point would imply about multiplicity. For example, if $\mathcal{Q}$ is concave over $\mathcal{I}$, then there must always be a unique fixed point. This property follows directly from Kennan's (2001) paper on the "Uniqueness of Positive Fixed Points for Increasing Concave Functions in $\mathbb{R}^n$". Alternatively, if $\mathcal{Q}$ is strictly convex in some neighborhood of x^* where $\rho(\mathbf{D}_\mathcal{Q}(x^*)) = 1$, then Tarski's fixed point theorem can be used to argue that multiple fixed points exist.

Appendix B

Matrix Similarity and Homeomorphisms

In this appendix section, I discuss how the fixed point properties of monotone mappings can be extended to a particular class of non-monotone mappings. Following the setup from Appendix A, let $\mathcal{I} = [\underline{x}, \overline{x}]$ be an interval in $\mathbb{R}^n$. I restrict my attention to mappings $\mathcal{Q} : \mathcal{I} \to \mathcal{I}$

[24] Note that a non-negative matrix will have a non-negative inverse if and only if it is a monomial matrix. So, the Jacobian matrix of $\mathcal{Q}^{-1}$ will be non-negative as long as each component of $\mathcal{Q}$ depends on a different input. In addition, whenever $\mathbf{D}_\mathcal{Q}$ does not have full rank, this same argument applies to any invertible block of $\mathbf{D}_\mathcal{Q}$.

that are continuously differentiable such that $\mathcal{Q}(x)$ lies within the interior of $\mathcal{I}$ for all $x \in \mathcal{I}$.[25] Just as before, let $\mathbf{D}_{\mathcal{Q}}(x)$ denote the Jacobian matrix of $\mathcal{Q}$ evaluated at the vector x. Suppose that $\mathbf{D}_{\mathcal{Q}}(x)$ satisfies condition *A.2* for all $x \in \text{int}(x)$. That is, assume:

A.2. There is an invertible matrix $\mathbf{B}$ such that $\mathbf{B}\mathbf{D}_{\mathcal{Q}}(x)\mathbf{B}^{-1}$ is non-negative for all $x \in \text{int}(\mathcal{I})$.

To understand the implications of *A.2*, define the mapping $\hat{\mathcal{Q}} : \mathbf{B}\mathcal{I} \to \mathbf{B}\mathcal{I}$ in such a way that $\hat{\mathcal{Q}}(x) = \mathbf{B}\mathcal{Q}(\mathbf{B}^{-1}x)$. This mapping has a Jacobian matrix of $\mathbf{D}_{\hat{\mathcal{Q}}}(x) = \mathbf{B}\mathbf{D}_{\mathcal{Q}}(\mathbf{B}^{-1}x)\mathbf{B}^{-1}$. Note that $\phi : \mathbb{R}^n \to \mathbb{R}^n$, where $\phi(x) = \mathbf{B}x$, is a bijective linear map. Therefore, $\phi(\cdot)$ is a *homeomorphism*, and it preserves interior points.[26] It follows that $\mathbf{B}\mathbf{D}_{\mathcal{Q}}(x)\mathbf{B}^{-1}$ is non-negative on $\text{int}(\mathcal{I})$ if and only if the matrix $\mathbf{D}_{\hat{\mathcal{Q}}}(x) = \mathbf{B}\mathbf{D}_{\mathcal{Q}}(\mathbf{B}^{-1}x)\mathbf{B}^{-1}$ is non-negative on $\mathbf{B}\text{int}(\mathcal{I})$, which equals $\text{int}(\mathbf{B}\mathcal{I})$. So, under *A.2*, the matrix $\mathbf{D}_{\hat{\mathcal{Q}}}(x)$ is non-negative for all $x \in \text{int}(\mathbf{B}\mathcal{I})$.

To summarize, *A.2* implies that $\hat{\mathcal{Q}}(x) = \mathbf{B}\mathcal{Q}(\mathbf{B}^{-1}x)$ is non-decreasing on the space $\text{int}(\mathbf{B}\mathcal{I})$. Note also that x^* is a fixed point of $\mathcal{Q}$ if and only if $\mathbf{B}x^*$ is a fixed point of $\hat{\mathcal{Q}}$. To see why, write $\hat{\mathcal{Q}}(\mathbf{B}x^*) = \mathbf{B}\mathcal{Q}(x^*) = \mathbf{B}x^*$. In addition, note that $\mathbf{D}_{\mathcal{Q}}(x)$ and $\mathbf{D}_{\hat{\mathcal{Q}}}(\mathbf{B}x)$ are similar matrices, since $\mathbf{D}_{\hat{\mathcal{Q}}}(\mathbf{B}x) = \mathbf{B}\mathbf{D}_{\mathcal{H}}(x)\mathbf{B}^{-1}$, which implies that they share the same eigenvalues. In particular, their spectral radii are equivalent: $\rho(\mathbf{D}_{\mathcal{Q}}(x)) = \rho(\mathbf{D}_{\hat{\mathcal{Q}}}(\mathbf{B}x))$ for all $x \in \text{int}(\mathcal{I})$.

Given this relationship, I can assess the existence, uniqueness, and dynamic stability of fixed points of the mapping $\mathcal{Q}$ by focusing on the (non-decreasing) mapping $\hat{\mathcal{Q}}$. It follows that each of the results from Appendix A will also hold for any $\mathcal{Q}$ with a Jacobian matrix satisfying *A.2*. Although the fixed points of $\mathcal{Q}$ and $\hat{\mathcal{Q}}$ may be ordered differently on their respective domains, the number and local stability of fixed points will be the same.

Special Case 1: Condition A.1

One important special case of condition *A.2* is condition *A.1*. Suppose $\mathbf{D}_{\mathcal{Q}}(x)$ satisfies condition *A.1* for all $x \in \text{int}(x)$. Then, *A.2* must hold for some diagonal change-of-basis matrix $\mathbf{B}$. This property is stated in Section 2.3 as Lemma 1 and it is proven in Appendix C.

If *A.1* applies, then the mapping $\mathcal{Q}$ always exhibits extremal equilibria. To see why, recall that any fixed point $\mathbf{B}x^*$ of $\hat{\mathcal{Q}}$ corresponds to a fixed point x^* of $\mathcal{Q}$. Moreover, by Tarski's theorem, the set of fixed points of $\hat{\mathcal{Q}}$ form a complete lattice. So, when pre-multiplied by $\mathbf{B}$, the fixed points of $\mathcal{Q}$ also form a complete lattice. If $\mathbf{B}$ is diagonal, then $\mathbf{B}x^*$ can be written as $(B_{11}x_1^*, \ldots, B_{nn}x_n^*)$, where $B_{11}, \ldots, B_{nn}$ are the diagonal entries of $\mathbf{B}$. Therefore, when the vector $\mathbf{B}x^*$ is maximal or minimal, each component of x^* must also be maximal or minimal.

Special Case 2: Positive Semidefinite Matrices

Symmetric, positive semidefinite matrices also satisfy condition *A.2*, since they can be diagonalized, and all their eigenvalues are non-negative. Therefore, these matrices are similar to non-negative, diagonal matrices. This special case carries important implications for the fixed points of $\mathcal{Q}$. In particular, if $\mathbf{D}_{\mathcal{Q}}$ is similar to a matrix that is non-negative and diagonal, then the corollary from Appendix A applies. Namely, $\mathcal{Q}$ almost always has an odd

[25] These restrictions may be weakened substantially; e.g., full continuity is not necessary to impose. However, since the equilibrium system (5) satisfies these requirements, I will focus my attention on this special case.

[26] Formally, a *homeomorphism* is a continuous bijection between two topological spaces that has a continuous inverse. For overviews about the importance of homeomorphisms, see Munkres (2000) and Hatcher et al. (2002).

number of fixed points: d where $\rho(\mathbf{D}_\mathcal{Q}) > 1$ and $d+1$ where $\rho(\mathbf{D}_\mathcal{Q}) < 1$ for some $d \in \mathbb{N}$.

One special type of positive semidefinite matrix is a diagonally dominant matrix with real non-negative diagonal entries. This type of matrix is defined in equation (8). To show that this matrix is indeed positive semidefinite, I invoke the Geršgorin circle theorem.

Geršgorin Circle Theorem. Consider a complex-valued matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$, and define $\{r_i\}_i$ so that $r_i(\mathbf{A}) = \sum_{i \neq j} |a_{ij}|$. Let $D(a_{ii}, r_i(\mathbf{A}))$ be the closed disc centered at a_{ii} with radius $r_i(\mathbf{A})$. Then every eigenvalue of $\mathbf{A}$ must lie within at least one of these discs $D(a_{ii}, r_i(\mathbf{A}))$.

This foundational theorem of linear algebra is useful for bounding the eigenvalues of any square matrix. In particular, it guarantees that all the real eigenvalues of a diagonally dominant matrix must be non-negative. So, if the matrix is also symmetric, then *A.2* holds.[27]

Appendix C

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 $\mathbf{A} = \mathbf{BJB}^{-1} \geq \mathbf{0}$. 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 choice of k and $\ell_1, \ldots, \ell_M$ in $\{1, \ldots, K\}$, it must be that:

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

Given the inequality above, I may conclude that the matrix $\mathbf{J}$ satisfies condition *A.1*.

"⇐" Suppose that *A.1* holds. I first consider the case where $\mathbf{J}$ is irreducible. For any group k, define the terms $\{\gamma_\ell^k\}_{\ell=1}^K$ so that $\gamma_\ell^k = 1$ if k is positively influenced by ℓ, and $\gamma_\ell^k = -1$ otherwise. Next, fixing some group k_0, I construct the matrix $\mathbf{B}$ so 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. $\mathbf{B}^{-1} = \mathbf{B}$. For all (g, ℓ), condition *A.1* guarantees that:

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

It follows that $\mathbf{BJB}^{-1}$ equals the absolute value of $\mathbf{J}$. Therefore, $\mathbf{BJB}^{-1}$ 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 $\mathbf{BJB}^{-1}$ is non-negative for some diagonal matrix $\mathbf{B} \in \mathbb{R}^{K \times K}$. □

[27] Note that Stieltjes matrices are also symmetric and positive definite, which means that they satisfy *A.2*.

Proof of Property 1

Proof. (Brouwer's FPT) The set $[-1,1]^K \subset \mathbb{R}^K$ is non-empty, compact, and convex. The equilibrium system (5) is a continuous, self-mapping function $\mathcal{Q} : [-1,1]^K \to [-1,1]^K$. By Brouwer's fixed point theorem, $\mathcal{Q}$ must have a fixed point m^*, i.e. an equilibrium exists. □

Proof. (Tarski's FPT) Suppose *A.2* holds. Then, for all m, the matrix $\mathbf{BD}(m)\mathbf{B}^{-1}$ is non-negative for some $\mathbf{B}$. Define the function $\hat{\mathcal{Q}}(m) = \mathbf{B}\mathcal{Q}(\mathbf{B}^{-1}m)$, which (by construction) is non-decreasing and takes values on a complete lattice. Applying Tarski's fixed point theorem, $\hat{\mathcal{Q}}$ must have a fixed point $\mathbf{B}m^*$. By the arguments presented in Appendix B, it follows that m^* is a fixed point of $\mathcal{Q}$. Therefore, the model has at least one equilibrium. □

Proof of Property 2

Proof. To begin, I claim that (5) is a contraction at $m \in [-1,1]^K$ when $\rho(\mathbf{D}(m)) < 1$. To see why, suppose $\rho(\mathbf{D}(m)) < 1$. Then $\mathbf{D}$ is a convergent matrix. As proven in Ortega (1972), for any $\varepsilon > 0$, there exists some matrix norm $||\cdot||$ for which $\rho(\mathbf{D}(m)) \leq ||\mathbf{D}(m)|| < \rho(\mathbf{D}(m)) + \varepsilon$. By choosing $\varepsilon = 1 - \rho(\mathbf{D}(m))$, it follows that $||\mathbf{D}(m)|| < 1$ for some appropriate matrix norm. Therefore, the system (5) is a contraction at m under this norm. That is, I can write:

$$||\mathcal{Q}(m) - \mathcal{Q}(\hat{m})|| \leq \kappa ||m - \hat{m}||,$$

for $\kappa \in [0,1)$, where $\hat{m}$ lies within some sufficiently small neighborhood of the vector m.

Now, suppose $\sup_{m\in[-1,1]^K} \rho(\mathbf{D}(m)) < 1$. So, $\rho(\mathbf{D}(m)) < 1$ at every $m \in [-1,1]^K$. By the first part of this proof, it must be that the system (5) is a contraction everywhere on $[-1,1]^K$, which implies that it is a contraction mapping. Applying the Banach contraction mapping theorem, I conclude that (5) has only one fixed point, i.e. there is a unique equilibrium. □

Proof of Property 3

Proof. To prove this property, I refer to the *Multiplicity Result* in Appendix A, as well as the discussion in Appendix B. Namely, under *A.2*, there is a homeomorphism relating (5) to some non-decreasing system of equations. The fixed points of (5) map directly to fixed points of this monotone mapping. So, Property 3 follows directly from the *Multiplicity Result.* □

Proof of Property 4

Proof. Assume $0 \in \text{argmax}_{x\in\mathbb{R}} f_{\varepsilon|k}(x)$ for all $k \in \mathcal{K}$, and let $h = \mathbf{0}_K$. Then $\boldsymbol{\beta}(\mathbf{0}_K) \geq \boldsymbol{\beta}(m)$ for every $m \in [-1,1]^K$. It follows that the Jacobian matrix $\mathbf{D}(\mathbf{0}_K) = \boldsymbol{\beta}(\mathbf{0}_K)\mathbf{J}$ is larger in magnitude than $\mathbf{D}(m) = \boldsymbol{\beta}(m)\mathbf{J}$ for every $m \in [-1,1]^K$. So, I can write:

$$\rho(\mathbf{D}(m)) = \lim_{j\to\infty} ||\mathbf{D}^j(m)||^{1/j} \leq \lim_{j\to\infty} ||\mathbf{D}^j(\mathbf{0}_K)||^{1/j} = \rho(\mathbf{D}(\mathbf{0}_K)),$$

where the first and last equalities apply Gelfand's formula, and $||\cdot||$ may denote any matrix norm (e.g., the induced p-norm). Hence, $\rho(\mathbf{D}(m))$ is maximized at the vector $m = \mathbf{0}_K$.

By Property 2, there is a unique equilibrium whenever $\rho(\mathbf{D}(\mathbf{0}_K)) < 1$. This equilibrium is located at $m^* = \mathbf{0}_K$, since the vector $\mathbf{0}_K$ always solves (5) when $h = \mathbf{0}_K$. By Property 3, there exist two extra equilibria when $\rho(\mathbf{D}(\mathbf{0}_K)) > 1$. Moreover, by symmetry of the distribution functions, this pair of equilibria is symmetric about $\mathbf{0}_K$. Therefore, Property 4(i) must hold.

Next, suppose that $h \neq \mathbf{0}_K$. In this case, $\rho(\boldsymbol{\beta}_0\mathbf{J}) < 1$ is no longer a sufficient condition for the existence of multiple equilibria. With nonzero private utility bias, the social interaction effects must be even stronger than they were before to guarantee multiplicity. To formalize this idea, let $\hat{\mathcal{Q}}(m) = \mathbf{B}\mathcal{Q}(\mathbf{B}^{-1}m)$ denote a non-decreasing mapping to which (5) is homeomorphic. Let $\mathcal{H}$ be the set of private utility vectors h for which there exist $a, b \in [-1, 1]^K$ satisfying $\hat{\mathcal{Q}}(a) < a < b < \hat{\mathcal{Q}}(b)$. I claim that there are multiple equilibria if and only if $h \in \mathcal{H}$.

To see why, suppose $h \in \mathcal{H}$. Then, Tarski's fixed point theorem guarantees the existence of fixed points $\mathbf{B}\underline{m}^*$ and $\mathbf{B}\overline{m}^*$ of $\hat{\mathcal{Q}}$, where $\mathbf{B}\underline{m}^* < a$ and $\mathbf{B}\overline{m}^* > b$. Hence, $\underline{m}^*$ and $\overline{m}^*$ are both equilibria. To show the other direction, start by supposing there are multiple equilibria. Then, $\hat{\mathcal{Q}}$ has multiple fixed points, which form a complete lattice. Let $\mathbf{B}\underline{m}^*$ and $\mathbf{B}\overline{m}^*$ be the fixed points of $\hat{\mathcal{Q}}$ that are minimal and maximal, respectively. As shown in Appendix A, the Jacobian matrix of $\hat{\mathcal{Q}}$ is convergent at these extremal fixed points. Applying the Perron-Frobenius theorem, it follows that $\hat{\mathcal{Q}}(a) < a < b < \hat{\mathcal{Q}}(b)$ for some appropriate choices of a and b where $\mathbf{B}\underline{m}^* < a$ and $\mathbf{B}\overline{m}^* > b$. Therefore, Property 4(ii) must hold.

□

Proof of Property 5

Proof. Fix any vector h and interaction matrix $\mathbf{J}$. For almost all distributions $\{F_{\varepsilon|k}\}_{k=1}^K$, there is no equilibrium m^* at which $\rho(\mathbf{D}(m^*)) = 1$. Now, suppose $\rho(\mathbf{D}(m^*)) < 1$. As demonstrated in the proof of Property 2, the system (5) must be a contraction at m^*. So, I can write:

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

where $\kappa = [0, 1)$ for any m_{t-1} in some sufficiently small neighborhood of m^*. Iterating on the inequality above ensures that $||m_t - m^*|| \leq \kappa^t||m_0 - m^*||$, where $\lim_{t\to\infty} \kappa^t = 0$. Hence, the vector m^* is a locally stable equilibrium, and Property 5 follows directly from Property 3.

□

Proof of Property 6

Proof. Suppose $\mathbf{J}$ satisfies *A.1*. By Lemma 1, there exists some diagonal matrix $\mathbf{B}$ for which $\mathbf{BJB}^{-1}$ 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:

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

where $\gamma_\ell^k = 1$ if group k is positively influenced by group ℓ, and $\gamma_\ell^k = -1$ otherwise.

As shown in Appendices A and B, the fixed points of the mapping $\hat{\mathcal{Q}}(m) = \mathbf{B}\mathcal{Q}(\mathbf{B}^{-1}m)$ form a complete lattice. Let $\mathbf{B}\overline{m}^*$ denote the greatest fixed point of $\hat{\mathcal{Q}}$. Then $E(\bar{\omega}^\ell)$ is maximal at the equilibrium $\overline{m}^*$ if k is positively influenced by ℓ, and $E(\bar{\omega}^\ell)$ is minimal at $\overline{m}^*$ otherwise. Let $\mathbf{B}\underline{m}^*$ be the lowest fixed point of $\hat{\mathcal{Q}}$. Then $E(\bar{\omega}^\ell)$ is minimal at the equilibrium $\underline{m}^*$ if k is positively influenced by ℓ, and $E(\bar{\omega}^\ell)$ is maximal at $\underline{m}^*$ otherwise. Since this argument holds for any two groups k and ℓ in $\mathcal{K}$, I may conclude that Property 6 holds.

□

Proof of Property 7

Proof. Consider the expected utility of a "typical" agent in group k. I write:

$$\begin{aligned} E(\max_{\omega_i} V^k(\omega_i)|m^*) &= E\Big(\Big|h_k + \sum_{\ell=1}^K J_{k\ell}m^{\ell*} + \varepsilon_i\Big|\Big) + \eta_k + E(\xi_i) \\ &= E\left(\max\Big\{h_k + \sum_{\ell=1}^K J_{k\ell}m^{\ell*} + \varepsilon_i, 0\Big\}\right) + E\left(\max\Big\{-h_k - \sum_{\ell=1}^K J_{kl}m^{\ell*} + \varepsilon_i, 0\Big\}\right) + \eta_k + E(\xi_i), \end{aligned}$$

where the last equality follows because ε_i and $-\varepsilon_i$ have the same distribution. Observe that this quantity is increasing in the absolute value of $h_k + \sum_{\ell=1}^K J_{k\ell}m^{\ell*}$ for all k. In other words, for any two equilibria $\hat{m}^*$ and $\tilde{m}^*$, agents in group k tend to prefer $\hat{m}^*$ over $\tilde{m}^*$ whenever:

$$\Big|h_k + \sum_{\ell=1}^K J_{k\ell}\hat{m}^{\ell*}\Big| \geq \Big|h_k + \sum_{\ell=1}^K J_{k\ell}\tilde{m}^{\ell*}\Big|$$

To prove Property 7(i), suppose $h = \mathbf{0}_K$. In this case, the equilibrium m^* that produces the highest aggregate welfare for each group k is the one that maximizes $\big|\sum_{\ell=1}^K J_{k\ell}m^{\ell*}\big|$ for every k. If $m^* = \mathbf{0}_K$ is the unique equilibrium, then $E(\max_{\omega_i} V^k(\omega_i)|m^*) = E(|\varepsilon_i|) + \eta_k + E(\xi_i)$. Note that, in this case, aggregate welfare increases in the variance of the random payoff ε_i.

Now assume that there are multiple equilibria. Since condition *A.1* holds, every equilibrium $m^* \in [-1,1]^K$ corresponds to a fixed point $x^* \in [-1,1]^K$ of the system:

$$x^{k*} = 2F_{\varepsilon|k}\Big(\sum_{\ell=1}^K |J_{k\ell}|x^{\ell*}\Big) - 1,$$

for $k \in \mathcal{K}$.[28] By Tarski's fixed point theorem, there is a (positive) greatest fixed point $\overline{x}^*$ and a (negative) least fixed point $\underline{x}^*$, which are symmetric about zero. For any group k, it must be that either $|J_{k\ell}|x^{\ell*} = J_{k\ell}m^{\ell*}$, for all ℓ, or $|J_{k\ell}|x^{\ell*} = -J_{k\ell}m^{\ell*}$, for all ℓ. In either case, the equilibria $\underline{m}^*$ and $\overline{m}^*$ corresponding to $\overline{x}^*$ and $\underline{x}^*$ satisfy $\sum_{\ell=1}^K J_{k\ell}\underline{m}^{\ell*} = -\sum_{\ell=1}^K J_{k\ell}\overline{m}^{\ell*}$ and:

$$\min\left\{\sum_{\ell=1}^K J_{k\ell}\underline{m}^{\ell*}, \sum_{\ell=1}^K J_{k\ell}\overline{m}^{\ell*}\right\} \leq \sum_{\ell=1}^K J_{k\ell}m^{\ell*} \leq \max\left\{\sum_{\ell=1}^K J_{k\ell}\underline{m}^{\ell*}, \sum_{\ell=1}^K J_{k\ell}\overline{m}^{\ell*}\right\}$$

for every equilibrium m^*. So, the quantity $\big|\sum_{\ell=1}^K J_{k\ell}m^{\ell*}\big|$ is maximized when $m^* \in \{\underline{m}^*, \overline{m}^*\}$, which implies that aggregate welfare in group k is greatest at either equilibrium $\underline{m}^*$ or $\overline{m}^*$.

[28] To see why, choose $\mathbf{B}$ to be the matrix that is constructed in the "$\Leftarrow$" part of the proof of Lemma 1.

To prove Property 7(ii), fix some group k. Let $\underline{m}^*$ ($\overline{m}^*$) be the equilibrium where m^k is minimal (maximal). Then $\sum_{\ell=1}^K J_{k\ell}\underline{m}^{\ell*} < \sum_{\ell=1}^K J_{k\ell}\overline{m}^{\ell*}$ holds as a consequence of Lemma 1. Define the term T_k so that $T_k = -\frac{1}{2}\sum_{\ell=1}^K J_{k\ell}(\underline{m}^{\ell*} + \overline{m}^{\ell*})$. If $h_k > T_k$, then:

$$h_k + \sum_{\ell=1}^K J_{k\ell}\overline{m}^{\ell*} > -h_k - \sum_{\ell=1}^K J_{k\ell}\underline{m}^{\ell*},$$

which implies $\left|h_k + \sum_{\ell=1}^K J_{k\ell}\overline{m}^{\ell*}\right| > \left|h_k + \sum_{\ell=1}^K J_{k\ell}\underline{m}^{\ell*}\right|$. Conversely, if $h_k < T_k$, then:

$$h_k + \sum_{\ell=1}^K J_{k\ell}\overline{m}^{\ell*} < -h_k - \sum_{\ell=1}^K J_{k\ell}\underline{m}^{\ell*},$$

which implies $\left|h_k + \sum_{\ell=1}^K J_{k\ell}\overline{m}^{\ell*}\right| < \left|h_k + \sum_{\ell=1}^K J_{k\ell}\underline{m}^{\ell*}\right|$. I conclude that Property 7(ii) holds.

□

Proof of Property 8

Proof. Suppose that $\underline{m}^{k*}, \overline{m}^{k*} \geq 0$. Then $m^{k*} \geq 0$ at every equilibrium m^*. It follows that:

$$h_k + \sum_{\ell=1}^K J_{k\ell}m^{\ell*} \geq 0$$

at every equilibrium m^*, which means that $E(\max_{\omega_i} V^k(\omega_i)|m^*)$ is increasing in $\sum_{\ell=1}^K J_{k\ell}m^{\ell*}$. Note that $\sum_{\ell=1}^K J_{k\ell}m^{\ell*}$ is maximal (or minimal) at the equilibrium where m^{k*} is maximal (or minimal). So, $\underline{m}^*$ minimizes $E(\max_{\omega_i} V^k(\omega_i)|m^*)$ and $\overline{m}^*$ maximizes $E(\max_{\omega_i} V^k(\omega_i)|m^*)$. Since the same reasoning applies in the case where $\underline{m}^{k*}, \overline{m}^{k*} \leq 0$, both (i) and (ii) must hold.

□

Proof of Property 9.1

Proof. By assumption *C.1*, there is some group k for which $\text{supp}(X|k)$ is not contained in a proper linear subspace of $\mathbb{R}^r$. Fix some network n, and let $\zeta_n^k = \alpha_k + \alpha_n + W_n'd + \sum_{\ell=1}^K J_{k\ell}m_n^\ell$. The conditional expected choice ω_i given $(X_i, W_n, \alpha_n, \alpha_k)$ may be written as:

$$E(\omega_i|X_i, W_n, \alpha_n, \alpha_k) = F_{\varepsilon|k}(\zeta_n^k + X_i'c)$$

To recover c, I must show that $F_{\varepsilon|k}(\zeta_n^k + X_i'c) = F_{\varepsilon|k}(\hat{\zeta}_n^k + X_i'\hat{c})$ implies $c = \hat{c}$. Since $F_{\varepsilon|k}$ is known, this property holds by Proposition 5 of Manski (1988). It follows that c is identified.

Having demonstrated that c can be recovered, I now focus on the other parameters. Consider any two social groups k_1 and k_2, and define the mapping $\phi(\cdot)$ that satisfies:

$$\phi(\nu) = \int F_{\varepsilon|k_1}\big((X_i - X_j)'c + \nu\big)dF_{X|k_1},$$

where X_j is chosen from $\text{supp}(X|k_2)$ and where the integral is evaluated over the conditional support of X_i given k_1. By construction, $\phi(\cdot)$ is nonlinear and monotonically increasing in ν.

For any network n, define $\nu_n = (\alpha_{k_1} - \alpha_{k_2}) + \sum_{\ell=1}^{K}(J_{k_1\ell} - J_{k_2\ell})m_n^\ell + F^{-1}_{\varepsilon|k_2}(E(\omega_j|X_j, W_n, \alpha_n, \alpha_{k_2}))$. By equation (18), the expected mean choice $m_n^{k_1}$ is equal to $\phi(\nu_n)$. Since $\phi(\cdot)$ is monotonic,

$$\begin{aligned} m_n^{k_1} &= \int F_{\varepsilon|k_1}\Big((\alpha_{k_1} - \alpha_{k_2}) + (X_i - X_j)'c + \sum_{\ell=1}^{K}(J_{k_1\ell} - J_{k_2\ell})m_n^\ell + F^{-1}_{\varepsilon|k_2}(E(\omega_j|X_j, W_n, \alpha_n, \alpha_{k_2}))\Big)dF_{X|k_1} \\ &= \int F_{\varepsilon|k_1}\Big((\widehat{\alpha_{k_1} - \alpha_{k_2}}) + (X_i - X_j)'c + \sum_{\ell=1}^{K}(\widehat{J_{k_1\ell} - J_{k_2\ell}})m_n^\ell + F^{-1}_{\varepsilon|k_2}(E(\omega_j|X_j, W_n, \alpha_n, \alpha_{k_2}))\Big)dF_{X|k_1} \end{aligned}$$

is satisfied if and only if $\sum_{\ell=1}^{K}\big[(J_{k_1\ell} - J_{k_2\ell}) - (\widehat{J_{k_1\ell} - J_{k_2\ell}})\big]m_n^\ell = (\alpha_{k_1} - \alpha_{k_2}) - (\widehat{\alpha_{k_1} - \alpha_{k_2}})$ for all networks $n \in \{1, \ldots, N\}$. Since the equilibrium mean choice levels are nonlinear functions of one another, sufficient variation in $(m_n^1, \ldots, m_n^K)$ across networks implies that:

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

for every $\ell \in \mathcal{K}$. Also, since k_1 and k_2 are chosen arbitrarily, this result holds for all $k_1, k_2 \in \mathcal{K}$. □

Proof of Property 9.2

Proof. To show that c, $\{\alpha_{k_1} - \alpha_{k_2}\}_{k_1,k_2}$, 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 *C.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 term $\zeta_n^k = \alpha_k + \alpha_n + W_n'd + \sum_{\ell=1}^{K} J_{k\ell}m_n^\ell$. For anyone in group k who resides in network n and has individual-level observables X_i, the expected choice ω_i is:

$$E(\omega_i|X_i, W_n, \alpha_n, \alpha_k) = F_{\varepsilon|k}(\zeta_n^k + X_i'c)$$

To recover $\{c, F_{\varepsilon|k}\}$, I must show $F_{\varepsilon|k}(\zeta_{n_0}^k + X_i'c) = \hat{F}_{\varepsilon|k}(\hat{\zeta}_{n_0}^k + X_i'\hat{c})$ implies $c = \hat{c}$ and $F_{\varepsilon|k} = \hat{F}_{\varepsilon|k}$ for all $X_i \in \text{supp}(X|k, n_0)$. This property holds by Manski's (1988) corollary. Also, since this argument applies for any $k \in \mathcal{K}$, 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 approach used to prove Property 9.1. So, the rest of the parameters are also identified. □