Yale University

EliScholar – A Digital Platform for Scholarly Publishing at Yale

Cowles Foundation Discussion Papers | Cowles Foundation

4-1-2018

Dispersed Behavior and Perceptions in Assortative Societies

Mira Frick

Ryota Iijima

Yuhta Ishii

Follow this and additional works at: https://elischolar.library.yale.edu/cowles-discussion-paper-series

 Part of the Economics Commons

Recommended Citation

Frick, Mira; Iijima, Ryota; and Ishii, Yuhta, "Dispersed Behavior and Perceptions in Assortative Societies" (2018). *Cowles Foundation Discussion Papers*. 145.
https://elischolar.library.yale.edu/cowles-discussion-paper-series/145

This Discussion Paper is brought to you for free and open access by the Cowles Foundation at EliScholar – A Digital Platform for Scholarly Publishing at Yale. It has been accepted for inclusion in Cowles Foundation Discussion Papers by an authorized administrator of EliScholar – A Digital Platform for Scholarly Publishing at Yale. For more information, please contact elischolar@yale.edu.

DISPERSED BEHAVIOR AND PERCEPTIONS IN ASSORTATIVE SOCIETIES

By

Mira Frick, Ryota Iijima, and Yuhta Ishii

April 2018

Revised March 2019

COWLES FOUNDATION DISCUSSION PAPER NO. 2128R



COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

Dispersed Behavior and Perceptions in Assortative Societies*

Mira Frick Ryota Iijima Yuhta Ishii

Abstract

We take an equilibrium-based approach to study the interplay between behavior and misperceptions in coordination games with assortative interactions. Our focus is *assortativity neglect*, where agents fail to take into account the extent of assortativity in society. We show, first, that assortativity neglect amplifies action dispersion, both in fixed societies and by exacerbating the effect of social changes. Second, unlike other misperceptions, assortativity neglect is a misperception that agents can rationalize in any true environment. Finally, assortativity neglect provides a lens through which to understand how empirically documented misperceptions about distributions of population characteristics (e.g., income inequality) vary across societies.

*This version: March 27, 2019. Frick: Yale University (mira.frick@yale.edu); Iijima: Yale University (ryota.iijima@yale.edu); Ishii: Centro de Investigación Económica, ITAM (yuhta.ishii@itam.mx). This research was supported by National Science Foundation grant SES-1824324. We thank Nageeb Ali, Attila Ambrus, Roland Bénabou, Dirk Bergemann, Ron Berman, Kyle Chauvin, Drew Fudenberg, Simone Galperti, Wayne Yuan Gao, Helena Garcia, Ben Golub, Andrei Gomberg, Sanjeev Goyal, Marina Halac, Johannes Hörner, Matt Jackson, Philippe Jehiel, Navin Kartik, Daria Khromenkova, David Miller, Weicheng Min, Stephen Morris, Pietro Ortoleva, Romans Pancs, Jacopo Perego, Larry Samuelson, Joel Sobel, Rani Spiegler, Philipp Strack, Linh To, Leeat Yariv, as well as seminar audiences at ASU Economic Theory Conference, Bielefeld, Bonn, BRIC (Columbia), Cambridge, Cornell, Duke, East Coast Behavioral and Experimental Workshop (Maryland), Glasgow, ITAM, Kyoto, Marketing-IO Conference (Yale), Michigan, NSF Conference on Network Science in Economics (Vanderbilt), Princeton, Rochester, St. Andrews, UC Berkeley, UCLA, Tokyo, TSE, UT Austin, and Yale.

1 Introduction

In many social and economic settings, from consumption choices to political activities, individual incentives depend on behavior in society, for example due to network externalities or adherence to social norms. An important channel through which people learn about behavior in society is by interacting with and observing the behavior of their peers, such as neighbors, coworkers, or online acquaintances. Many of these interactions are *assortative*, in the sense that people interact more with others with similar characteristics—e.g., richer people are typically more likely than poorer people to have rich friends and conservatives are more likely than liberals to know other conservatives; indeed, evidence suggests that societies may be growing increasingly assortative.[1] Assortativity creates scope for *misperception*, because the behavior that individuals observe in their own interactions need not be representative of society as a whole. But if people misperceive behavior in society, this affects their own incentives and behavior, which in turn might influence others' observations, perceptions, and behavior.

In this paper, we take an equilibrium-based approach to study the interplay between behavior and misperceptions in assortative societies. Our main focus is on a natural form of misperception, *assortativity neglect*, where agents fail to take into account the extent of assortativity in society, mistakenly believing that everyone interacts with a representative (or more representative than actual) sample of the population.[2] Our analysis allows us to address how assortativity neglect affects agents' behavior and their perceived distributions of population characteristics (e.g., income or political attitudes); how this depends on the structure of society; and why assortativity neglect might be an especially "persistent" misperception.

We consider a class of coordination games with assortative interactions. A society consists of a large population of agents with linearly ordered types, along with an interaction structure that randomly matches pairs of types. Capturing assortativity, higher types' match distributions first-order stochastically dominate those of lower types. Agents' coordination incentives are modeled by best response functions that are linearly increasing in their own type, as well as in expected behavior among their matches and in society as a whole. To investigate the interplay between behavior and (mis)perceptions, we consider an equilibrium notion in the spirit of self-confirming equilibrium (Battigalli, 1987; Fudenberg and Levine, 1993), where agents best-respond to perceptions of behavior and society that need only be consistent with the behavior they observe among their matches. To capture which misperceptions might be especially persistent, we also formalize a *coherency* criterion that requires agents to be able to "rationalize" the behavior they observe based on their perceptions.

[1] E.g., Jargowsky (1996); Reardon and Bischoff (2011) document increased residential segregation by income in the US, and Bishop (2009); Lawrence, Sides, and Farrell (2010); Huber and Malhotra (2017) find growing offline and online segregation by political attitudes.

[2] Section 2.2 discusses empirical findings related to this misperception.

Our main findings are threefold. First, relative to the correct perceptions benchmark, assortativity neglect amplifies action dispersion, both by increasing the gap between higher and lower types' behavior in any fixed society and by acting as a "multiplier" that exacerbates the effect of social changes (e.g., increased assortativity) on action dispersion. Second, assortativity neglect is always a coherent misperception, regardless of the true level of assortativity in society; by contrast, other misperceptions of the interaction structure are not always coherent. Finally, assortativity neglect provides a lens through which to understand empirically documented misperceptions about distributions of key population characteristics (e.g., income inequality) and how they are shaped by the nature of agents' social interactions.

To illustrate the setup and main findings, we present a simple example in the context of education investment and residential sorting by income:

1.1 Illustrative Example

Consider a continuum population of agents, each of whom is endowed with one of three equally likely types, representing socioeconomic status/income: "rich" (θ_h), "middle-class" (θ_m), or "poor" (θ_l), with $\theta_h > \theta_m = 0 > \theta_l$. For simplicity, focus on the symmetric case, $\theta_h = -\theta_l$. Each agent knows her own type, but does not directly observe other agents' types. Due to neighborhood sorting by income, the richer an agent the more likely she is to interact with other high income agents. Specifically, pairwise interaction probabilities are summarized by

P	θ_h	θ_m	θ_l
θ_h	ρ	$\frac{1}{3}$	$\frac{2}{3}-\rho$
θ_m	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
θ_l	$\frac{2}{3}-\rho$	$\frac{1}{3}$	ρ

where the parameter $\rho \in (\frac{1}{3}, \frac{2}{3})$ measures the degree of assortativity.

Each agent θ_i chooses a level $s(\theta_i) \in \mathbb{R}$ of monetary/time investment in education.[3] Assume that θ_i's best-response against strategy profile s takes the form

$$\mathrm{BR}_{\theta_i}(s) = \frac{1}{3}\theta_i + \frac{1}{3}\underbrace{\mathbb{E}[s(\theta)|\theta_i]}_{:=\sum_j P(\theta_i,\theta_j)s(\theta_j)} + \frac{1}{3}\underbrace{\mathbb{E}[s(\theta)]}_{:=\sum_j \frac{1}{3}s(\theta_j)}.$$

That is, agents of higher socioeconomic status have an intrinsic tendency to invest more in education; but additionally, θ_i's optimal education investment is increasing in local average investment among the agents she interacts with (e.g., due to peer effects) and in global average

[3] We consider education investment to include decisions such as expenditures on educational materials or tutors or the amount of effort exerted at school, but assume school choice (and other decisions that might endogenously affect sorting) to be exogenous (e.g., because everyone enrolls in their neighborhood school).

investment in society as a whole (e.g., because future returns to education are boosted by economy-wide knowledge spillovers).[4] For simplicity, we assume an equal weight $\frac{1}{3}$ on each component. There is a unique (Bayes) Nash equilibrium: $s^{NE}(\theta_h) = \frac{\theta_h}{3-2(\rho-\frac{1}{3})} = -s^{NE}(\theta_l)$ and $s^{NE}(\theta_m) = 0$.

While Nash equilibrium assumes that agents best-respond to correct perceptions of behavior in society, we formalize a notion of *assortativity neglect* equilibrium: Here agents are only able to observe the local action distributions in their neighborhoods, and draw inferences from these based on the misperception that there is no sorting, i.e., they perceive the sorting parameter to be $\hat{\rho} = \frac{1}{3}$.[5] We show that this leads to the following misspecified best response function:

$$\mathrm{BR}^{AN}_{\theta_i}(s) = \frac{1}{3}\theta_i + \frac{2}{3}\mathbb{E}[s(\theta)|\theta_i],$$

where θ_i best-responds to the correct local action average $\mathbb{E}[s(\theta)|\theta_i]$ but mistakenly believes this to coincide with the global action average. Equilibrium actions are uniquely given by $s^{AN}(\theta_h) = \frac{\theta_h}{3-4(\rho-\frac{1}{3})} = -s^{AN}(\theta_l)$ and $s^{AN}(\theta_m) = 0$.

Increased action dispersion. Our first main finding is that assortativity neglect increases action dispersion relative to Nash, through two channels. First, in any given society, assortativity neglect leads to more polarized education investment than Nash:

$$s^{AN}(\theta_h) > s^{NE}(\theta_h) > s^{NE}(\theta_l) > s^{AN}(\theta_l).$$

The intuition is simple and reflects a mutually reinforcing interplay between agents' misperceptions and behavior: Since the rich are more likely than the poor to interact with other rich agents, they tend to observe higher education investment among their peers. Under assortativity neglect, this gives rise to a "false consensus effect" (Ross, Greene, and House, 1977), where perceptions of average education investment in society, and hence of returns to education, are increasing in agents' socioeconomic status (and thus in their own investment). Relative to correct perceptions, this increases the education investment gap between the rich and poor, which in turn, through observation of their peers' behavior, feeds into the false consensus effect.

Second, assortativity neglect acts as a *multiplier* of social changes that increase action

[4] See, e.g., Bénabou (1993, 1996a,b); Fernández and Rogerson (1996, 2001); Durlauf (1996) for related (correctly specified) models of education investment with sorting and local and/or global complementarities.

[5] This is an extreme notion of assortativity neglect, where agents fully neglect sorting. Section 5.1.2 considers weaker forms of assortativity neglect that allow for any perceived sorting parameter $\hat{\rho} \in [\frac{1}{3}, \rho]$.

dispersion. For instance, the effect of an increase in the degree ρ of neighborhood sorting is

$$\underbrace{\frac{\partial}{\partial\rho}\left(s^{AN}(\theta_h)-s^{AN}(\theta_l)\right)}_{=\frac{8\theta_h}{\left(3-4(\rho-\frac{1}{3})\right)^2}} > \underbrace{\frac{\partial}{\partial\rho}\left(s^{NE}(\theta_h)-s^{NE}(\theta_l)\right)}_{=\frac{4\theta_h}{\left(3-2(\rho-\frac{1}{3})\right)^2}} > 0.$$

That is, the rich–poor education gap rises under Nash, but even more so under assortativity neglect. Intuitively, greater sorting has a direct effect on the education gap under Nash, by increasing differences in local peer effects between the rich and poor. However, under assortativity neglect this additionally magnifies the false consensus effect, because the rich (poor) mistakenly attribute their increased (decreased) local education investment levels to a global trend in society, further polarizing their responses. An increase in income inequality $d := \theta_h - \theta_l$ has an analogous effect.

Coherency of assortativity neglect. So far, we have assumed that agents neglect sorting but do not question whether the behavior they observe "makes sense" given this perception. Suppose next that each agent seeks to rationalize what she sees, in the sense that, given her perception of society, the behavior she observes must be consistent with everyone best-responding to others' behavior, a requirement we refer to as *coherency*. Our second main finding is that assortativity neglect can always be coherently sustained, no matter how assortative the actual society.

Agents' rationalization procedure under assortativity neglect relates to the "fundamental attribution error" studied in social psychology (Ross, 1977), i.e., a tendency to attribute others' behavior to intrinsic characteristics rather than external factors such as social influence. For instance, a middle-class agent θ_m observes three different investment levels $a_i := s^{AN}(\theta_i)$ $(i = h, m, l)$, but believes everyone to face the same peer effects, as she perceives local average behavior in each neighborhood to equal the global action average $\bar{a} := \frac{1}{3}\sum_i a_i$. Nevertheless, θ_m can rationalize observed differences in education investment by attributing them purely to differences in income: If she perceives rich and poor types to be $\hat{\theta}_i = \frac{3\theta_i}{3-4(\rho-\frac{1}{3})}$ $(i = h, l)$, then $a_i = \frac{1}{3}\hat{\theta}_i + \frac{2}{3}\bar{a}$; i.e., a_i is a best response for $\hat{\theta}_i$ given θ_m's perceptions.

While assortativity neglect is always coherent, we will show that any misperception of interaction patterns other than assortativity neglect is *not* coherent in some environments. For instance, in an asymmetric version of the present example where $\theta_h \neq -\theta_l$, we will derive an upper bound $\bar{\rho} \in [\rho, \frac{2}{3})$ on the degree $\hat{\rho}$ of sorting that θ_m can coherently perceive. Thus, θ_m can arbitrarily underestimate assortativity, but there are limits to how much she can overestimate assortativity.

Misperceptions of type distributions. Note that despite the fact that middle-class agents interact with a representative sample of the population, θ_m misperceives the income

distribution in society. This is because instead of being able to form perceptions of the income distribution by directly observing peers' income levels, θ_m draws inferences by rationalizing the *behavior* she observes among her peers. In the present example, assortativity neglect leads θ_m to overestimate income inequality, i.e., $\hat{\theta}_h > \theta_h > \theta_l > \hat{\theta}_l$, as she compensates for neglecting socioeconomic differences in peer effects by attributing the rich–poor education gap entirely to income. In general societies, a countervailing force is that an agent's peers may be less heterogeneous than the overall population, so that assortativity neglect might lead her to underestimate type dispersion. Section 4.2 shows that whether agents over- or underestimate type dispersion depends on the relative strength of strategic complementarities and assortativity, and we relate this to empirical findings on cross-country differences in perceptions of income inequality. We also analyze agents' misperceptions of their own position relative to the mean type and show that this has implications for a society's demand for redistribution.

1.2 Overview

The paper proceeds as follows. Section 2 defines assortativity neglect equilibrium in general assortative societies. Moving beyond the simple illustrative example above, equilibrium behavior does not typically admit a closed-form solution. Instead, Section 3 analyzes and contrasts behavior under Nash and assortativity neglect by recasting assortative societies as monotone Markov processes over their space of types. Moreover, by representing interaction patterns in terms of the copula associated with a society, we identify an appropriate non-parametric measure of assortativity under which more assortative societies are characterized precisely by greater action dispersion, but more so under assortativity neglect than Nash. Section 4 analyzes the coherency of assortativity neglect and its implications for perceived type distributions in general environments. Section 5 discusses several extensions and related literature, and Section 6 offers concluding remarks.

2 Model

2.1 Society and Coordination Game

There is a continuum of agents with mass normalized to 1. Each agent is identified with a type $\theta \in \mathbb{R}$. An agent's type is her private information. Agents interact according to a random matching technology. A society P specifies the probability with which any pair of types θ and θ' are matched:[6]

[6]As is also common in the network game literature, we view society P as exogenous and focus on analyzing equilibrium behavior (and perceptions). See Pin and Rogers (2016) for a survey of potential sources of assortativity, such as institutional constraints determining meeting opportunities or socio-psychological factors (e.g.,

Definition 1. A ***society*** is a joint cdf P over $\mathbb{R} \times \mathbb{R}$ that is:

1. *symmetric*: $P(\theta, \theta') = P(\theta', \theta)$ for all θ, θ'
2. *assortative*: $P(\cdot|\theta)$ first-order stochastically dominates $P(\cdot|\theta')$ if $\theta \geq \theta'$.[7]

Symmetry is a consistency condition required to describe a random matching in a population. Assortativity captures the idea that higher types are more likely than lower types to interact with other high types. Note that a society P jointly summarizes an underlying ***type distribution***, described by the marginal distribution $F := \operatorname{marg} P$, and a ***matching technology***, which for every type θ specifies the distribution $P(\cdot|\theta)$ of θ's matches. We assume that the distribution F is absolutely continuous, L^1 and has a connected support, denoted by Θ. Let $\mathcal{F}$ denote the set of all cdfs with these properties.

Society is engaged in an incomplete-information coordination game. Agents have symmetric action sets $A = \mathbb{R}$, and a strategy profile is a Borel-measurable function $s : \Theta \to A$ that specifies an action $s(\theta)$ for each type θ. To model coordination motives, we focus on linear best response functions, which are widely used in the literature on network games and incomplete-information coordination games;[8] Section 5.1.3 discusses more general best response functions. Specifically, there exist coefficients $\gamma, \beta \geq 0$ with $\gamma + \beta < 1$ such that each type θ's best response against any strategy profile s in society P is given by

$$\mathrm{BR}_\theta(s, P) = (1 - \gamma - \beta)\theta + \gamma \mathbb{E}_P[s(\theta')|\theta] + \beta \mathbb{E}_F[s(\theta')]. \tag{1}$$

The first term captures that higher types have an intrinsic tendency to take higher actions. The second term represents a local coordination motive, whereby each type θ's best response is increasing in her matches' expected behavior. Finally, reflecting a global coordination motive, θ's best response is also increasing in the average action in society as a whole.

While our analysis takes best response function (1) as its primitive, a utility function that gives rise to (1) is

$$U_P(\theta, a, s) = -\int \int \big(a - (1 - \gamma - \beta)\theta - \gamma s(\theta') - \beta s(\theta'')\big)^2 dP(\theta'|\theta) dF(\theta''). \tag{2}$$

That is, θ faces a quadratic miscoordination cost that reflects the gap between her action and a weighted sum of her type and realized actions among her matches and in society.

homophily).

[7] $P(\cdot|\theta)$ denotes the conditional distribution given that one of the two types in the match is realized to be θ. By the symmetry assumption, it is irrelevant which of the two types' realizations we condition on.

[8] See, e.g., Jackson and Zenou (2013) for a survey of the former and Morris and Shin (2002); Angeletos and Pavan (2007); Bergemann and Morris (2013) for the latter.

In addition to education investment (Section 1.1), other economic examples include many consumption decisions that depend on income either positively (e.g., luxury good consumption) or negatively (e.g., smoking), but may also exhibit both direct peer-to-peer consumption complementarities and material or socio-psychological global payoff complementarities (e.g., economy-wide technological complementarities or a desire to adhere to a social norm). Likewise, types might represent political attitudes on a left-right spectrum and actions the extent to which agents manifest support for particular positions (e.g., by posting political content on social media), where related local and global coordination/conformity motives may be at play.

2.2 Copula Representation and Assortativity Neglect

As noted, any society P jointly summarizes an underlying type distribution $F = \text{marg}_\Theta P$ and a matching technology $(P(\cdot|\theta))_{\theta\in\Theta}$. In general, varying the matching technology of P also changes the type distribution, and vice versa. To be able to disentangle these two dimensions, we also make use of the following equivalent representation of societies, which expresses who interacts with whom not in terms of types $\theta \in \Theta$ (e.g., a particular wealth level), but in terms of type *quantiles* $x \in [0,1]$ (e.g., the 5th wealth percentile).

Formally, for any society P with type distribution $F \in \mathcal{F}$, define $C(x,x')$ to be the probability that two agents whose type quantiles are below x and x' are matched; that is,

$$C(x,x') := P(F^{-1}(x), F^{-1}(x')) \text{ for all } x,x' \in (0,1), \tag{3}$$

and $C(x,0) = C(0,x) := 0$, $C(x,1) = C(1,x) := x$ for all $x \in [0,1]$.[9] Note that C is a (two-dimensional) ***copula***, i.e., a joint cdf over $[0,1]^2$ with uniform marginals. Moreover, copula C inherits symmetry and assortativity from P.[10] We refer to symmetric and assortative copulas as ***interaction structures***.

Any society induces an interaction structure via (3). Conversely, given any interaction structure C and type distribution $F \in \mathcal{F}$, defining $P(\theta,\theta') := C(F(\theta), F(\theta'))$ for all θ,θ' yields a society. Thus, pairs (F,C) of type distributions $F \in \mathcal{F}$ and interaction structures C yield an equivalent representation of societies. From now on, we focus on societies whose interaction structure C admits a density function $c(\cdot,\cdot)$ that is positive and absolutely continuous on $(0,1)^2$; let $\mathcal{C}$ denote the class of all interaction structures with these properties.[11]

Example 1 (Gaussian societies)**.** As a simple parametric example, let P be given by a sym-

[9] For any cdf G, its inverse is defined by $G^{-1}(x) := \inf\{v \in \mathbb{R} : G(v) \geq x\}$ for each $x \in (0,1)$.

[10] That is, $C(x,x') = C(x',x)$ for all x,x'; and $C(\cdot|x)$ first-order stochastically dominates $C(\cdot|x')$ if $x \geq x'$.

[11] Assuming a positive and absolutely continuous density is not essential for most results, but it simplifies the exposition.

metric bivariate Gaussian distribution, where

$$(\theta_1, \theta_2) \sim \mathcal{N}\left(\begin{pmatrix} \mu \\ \mu \end{pmatrix}, \begin{pmatrix} \sigma^2 & \sigma_{1,2} \\ \sigma_{1,2} & \sigma^2 \end{pmatrix}\right)$$

and the correlation coefficient $\rho := \frac{\sigma_{1,2}}{\sigma^2} \in [0, 1)$ is nonnegative to ensure assortativity. The corresponding type distribution F is normally distributed and fully parametrized by its mean μ and variance σ^2. Interaction structure C is fully parametrized by the correlation coefficient ρ: $C(x, x') := \Phi_\rho(\Phi^{-1}(x), \Phi^{-1}(x'))$, where Φ is the standard normal cdf and Φ_ρ is the cdf of the joint normal distribution with mean vector $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and covariance matrix $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. ▲

The next subsection will introduce the idea that instead of knowing the actual society $P = (F, C)$, each type θ has in mind a perceived society $\hat{P}_\theta$ that may differ from P. We will be especially interested in a natural form of misperception, where agents fail to take into account the fact that society is assortative. By decomposing θ's perceived society $\hat{P}_\theta$ into a perceived type distribution $\hat{F}_\theta$ and perceived interaction structure $\hat{C}_\theta$,[12] we can formalize this misperception as follows. Let the ***independent interaction structure*** $C_I \in \mathcal{C}$ be given by $C_I(x, x') = xx'$ for all x, x'; that is, regardless of an agent's own quantile x, the distribution $C_I(\cdot|x)$ of her matches' quantiles is uniform on $[0, 1]$.

We say that θ suffers from ***assortativity neglect*** if $\hat{C}_\theta = C_I$. Thus, under assortativity neglect, θ fails to take into account that agents' match distributions depend on their position in the type distribution and instead perceives everyone to interact with a representative sample of society as a whole. Note that $\hat{C}_\theta = C_I$ is equivalent to the requirement that $\hat{P}_\theta(\cdot|\theta') = \hat{F}_\theta$ for all θ'; that is, assortativity neglect imposes no restrictions on θ's perception $\hat{F}_\theta$ of the type distribution in society, as long as she perceives all agents' match distributions to coincide with $\hat{F}_\theta$. However, Section 4 will show that assortativity neglect uniquely pins down $\hat{F}_\theta$ in (coherent) *equilibrium*.

Assortativity neglect can be seen as a form of "projection bias," whereby θ projects her own (perceived) match distribution onto all other agents' match distributions (i.e., $\hat{P}_\theta(\cdot|\theta) = \hat{P}_\theta(\cdot|\theta')$); this is broadly in line with findings in the empirical literature on "network cognition."[13] Assortativity neglect also relates to empirically documented information-processing biases, notably "selection neglect" (Enke, 2017; Esponda and Vespa, 2014, 2018; Barron, Huck, and Jehiel, 2019), where agents fail to take into account that the information they see may be

[12] That is, $\hat{F}_\theta = \text{marg}_\Theta \hat{P}_\theta$ and $\hat{C}_\theta(x, x') = \hat{P}_\theta\left(\hat{F}_\theta^{-1}(x), \hat{F}_\theta^{-1}(x')\right)$ for all $x, x' \in (0, 1)$.

[13] E.g., Dessi, Gallo, and Goyal (2016) elicit subjects' assessments of degree distributions on a network and document a projection bias, where subjects project their own number of neighbors (i.e., their degree) onto other agents in the network. More broadly, a literature in social psychology documents related "location effects" in individuals' perceptions of their interaction structures (for a survey, see Brands, 2013).

subject to selection effects. Section 5.2 discusses the relationship with theoretical work that incorporates these and related biases.

Finally, we note that the above notion of assortativity neglect is quite extreme, as θ perceives everyone to interact with a perfectly representative sample of society. Section 5.1.2 will discuss weaker forms of assortativity neglect.

2.3 Assortativity Neglect Equilibrium

To study the equilibrium implications of assortativity neglect, we move beyond the standard equilibrium notion of (Bayes) Nash which assumes that agents best-respond to correct perceptions about the underlying strategy profile s and society P.

Instead, we employ an equilibrium notion in the spirit of self-confirming equilibrium (Battigalli, 1987; Fudenberg and Levine, 1993) that formalizes the following idea. We think of our coordination game as the steady state of a setting where successive generations of agents choose actions based only on observing the behavior in their *local* neighborhoods. That is, before a new generation of types take actions, each new type θ first observes the current action distribution among types from her match distribution $P(\cdot|\theta)$ (e.g., the current education investment distribution in her neighborhood). In the steady state of such a setting, we might expect agents to possess a good understanding of the action distributions among their matches. However, as this is only a very partial snapshot of society, this information need not be enough for them to correctly identify s and P.

Capturing this, in any given society P, our equilibrium notion summarizes which true behavior s and (mis)perceptions of s and P can jointly arise, subject to two requirements: (1) Each agent best-responds to her perceptions of s and P; (2) these perceptions need not be correct, but are consistent with the true action distribution among her matches.

Formally, we first define each type θ's ***local action distribution***, i.e., the distribution of actions among θ's matches. When the true strategy profile and society are s and P, this is given by the distribution $H_\theta^{s,P}$ over actions that arises when θ's matches are drawn from $P(\cdot|\theta)$ and behave according to strategy profile s; that is,[14]

$$H_\theta^{s,P}(a) = \int_\Theta \mathbb{1}_{\{s(\theta')\leq a\}}\, dP(\theta'|\theta) \text{ for all } a \in A. \tag{4}$$

Definition 2. Given any (P, γ, β), an ***equilibrium*** is an L^1 strategy profile s[15] together with a perceived society $\hat{P}_\theta$ and perceived strategy profile $\hat{s}_\theta$ for each type θ satisfying:

1. *Best response*: $s(\theta) = \mathrm{BR}_\theta(\hat{s}_\theta, \hat{P}_\theta)$.

[14] Here $\mathbb{1}_{\{s(\theta')\leq a\}}$ denotes the indicator random variable on the event $\{\theta' : s(\theta') \leq a\}$.
[15] That is, $\int |s(\theta)| dF(\theta) < \infty$.

2. *Observational consistency*: $H_\theta^{s,P} = H_\theta^{\hat{s}_\theta, \hat{P}_\theta}$.

An ***assortativity neglect equilibrium (ANE)*** is an equilibrium $(s, (\hat{s}_\theta, \hat{P}_\theta)_\theta)$ such that $\hat{C}_\theta = C_I$ for all θ.

Underlying any equilibrium is a true society P and strategy profile s. While Nash equilibrium assumes that players best-respond to correct perceptions about both, Definition 2 allows more generally that each type θ best-responds to some possibly incorrect perceived society $\hat{P}_\theta$ and strategy profile $\hat{s}_\theta$. However, by observational consistency, these perceptions are disciplined by the requirement that θ's perceived local action distribution $H_\theta^{\hat{s}_\theta, \hat{P}_\theta}$ coincide with the true local action distribution $H_\theta^{s,P}$.

Our main focus is on the case of assortativity neglect equilibrium, where all types θ suffer from assortativity neglect. In Section 3, we will show that ANE always exists and behavior under ANE is uniquely pinned down, so that Definition 2 is suitable for understanding and contrasting the behavioral implications of assortativity neglect and Nash.

At the same time, Definition 2 is quite weak in that it captures a "mechanical" view of other agents' behavior: As long as θ's perceptions predict the correct local action distribution, she can hold arbitrary perceptions $\hat{s}_\theta$ of behavior in society, without asking whether this behavior $\hat{s}_\theta$ "makes sense" under her perceived society $\hat{P}_\theta$. This renders Definition 2 very permissive: Indeed, we will see that essentially any perceived strategy profiles (including $\hat{s}_\theta$ that are decreasing in types) and any perceived type distributions can arise under ANE; more generally, any perceived society $\hat{P}_\theta$ can be sustained as part of some equilibrium (Appendix E.1).

Given this, we introduce a refinement that imposes the following coherency requirement on each agent's perceptions of behavior and society: We require that the behavior $\hat{s}_\theta(\theta')$ that θ attributes to any other type θ' should be a best response given $\hat{s}_\theta$ and $\hat{P}_\theta$.

Definition 3. An equilibrium $(s, (\hat{s}_\theta, \hat{P}_\theta)_\theta)$ is ***coherent*** if each type θ's perceptions satisfy $\hat{s}_\theta(\theta') = \mathrm{BR}_{\theta'}(\hat{s}_\theta, \hat{P}_\theta)$ for each θ'.

Coherent equilibria capture misperceptions that might be thought of as especially "persistent:" Indeed, each agent's perceptions $\hat{s}_\theta$ and $\hat{P}_\theta$ not only correctly predict her observations (by observational consistency), but additionally can rationalize these observations (by coherency). Thus, based on introspection alone and absent additional external information, there is a sense in which agents might never realize that they hold misperceptions. Section 4 will show that assortativity neglect is persistent in this sense, as in any society there exists a coherent ANE; moreover, the coherent ANE is unique, yielding sharp predictions for agents' perceived type distributions under assortativity neglect. We will also see that other misperceptions of society need not be coherent.

Remark 1. We briefly comment on some features of our equilibrium notion:

Connection with SCE. In general games, self-confirming equilibrium (SCE) captures players who best-respond to beliefs about others' strategies that need not be correct, but need only be consistent with some limited feedback about opponents' behavior.[16] Definition 2 applies this idea to a setting where agents' feedback is limited to observing the action distribution among their matches, which are determined by an underlying assortative society P. In addition to beliefs about strategies, we also treat beliefs about P as endogenous equilibrium objects. Moving beyond SCE, Definition 3 imposes a rationalization requirement that relates to rationalizable conjectural equilibrium (see the third paragraph below).

No feedback from payoffs. As in other important special cases of SCE (e.g., Eyster and Rabin, 2005; Jehiel, 2005; Jehiel and Koessler, 2008), we implicitly assume that agents do not receive additional feedback from their payoffs. In the above interpretation in terms of successive generations of short-lived agents, this captures that each new agent can base her action choice only on current behavior but not payoffs in her neighborhood, as fits settings (e.g., education investment) where payoffs are realized only in the more distant future.[17] An alternative interpretation is that global payoffs (e.g., to smoking) represent "psychological" utilities without direct material feedback (e.g., a desire to conform to a perceived social norm). Finally, note that even if agents observe their payoffs, this information need not be enough to identify s and P;[18] thus, non-Nash behavior can continue to arise.

Nash rationalization. Coherency requires each agent θ to be able to use her *own* perceptions $\hat{s}_\theta, \hat{P}_\theta$ to rationalize the behavior $\hat{s}_\theta(\theta')$ she attributes to any other agent θ'. We refer to this requirement as Nash rationalization, as it implies that θ's perceived strategy profile $\hat{s}_\theta$ is a Nash equilibrium given her perceived society $\hat{P}_\theta$. Section 5.1.4 discusses a more permissive coherency notion in the spirit of rationalizable conjectural equilibrium (Rubinstein and Wolinsky, 1994; Esponda, 2013), where θ (despite being confident in her own perceptions $\hat{s}_\theta$ and $\hat{P}_\theta$) entertains the possibility that θ' best-responds to different perceptions.

Homogeneous misperceptions. Under ANE, all agents perceive the same interaction structure C_I. Section 5.1.1 presents an extension to a hybrid model where only some fraction of agents of each type suffer from assortativity neglect and the remaining fraction have correct perceptions. Additionally, our weaker notions of assortativity neglect (Section 5.1.2) allow perceived interaction structures $\hat{C}_\theta$ to vary across types θ. ▲

[16] Section 5.2 discusses the SCE literature in more detail.

[17] Concretely, global payoffs might represent an adolescent's future employment prospects (which depend on current education investment in society as a whole, due to the global complementarities highlighted in Section 1.1), but at the time of investing in her education she might only have access to information about current education investment (and not the long-run employment outcomes) among adolescents in her neighborhood.

[18] E.g., under quadratic-loss utilities (2), realized payoffs $U_P(\theta, a, s)$ do not uniquely identify local and global average behavior.

3 Behavior under Assortativity Neglect

3.1 Increased Action Dispersion under Assortativity Neglect

We now proceed to analyze and contrast behavior under Nash and assortativity neglect. A first key observation is the following: Even though our model is static, any society P induces a discrete-time Markov process over its space of types Θ, with initial distribution given by the type distribution $F = \text{marg}_\Theta P$ and transition kernel represented by the matching technology $(P(\cdot|\theta))_{\theta\in\Theta}$. That is, this process first draws an initial type $\theta_0 \in \Theta$ according to F, then draws type θ_0's match θ_1 according to $P(\cdot|\theta_0)$, type θ_1's match θ_2 according to $P(\cdot|\theta_1)$, and so on. We refer to this Markov process as the process of ***t-step ahead matches*** in society and also denote it by P. Note that F is a stationary distribution of the process. Moreover, assortativity of P corresponds precisely to this process being *monotone* (Daley, 1968); this feature plays an important role throughout our analysis.[19]

The process of t-step ahead matches yields a simple description of the Nash strategy profile s^{NE} of our game. Indeed, iterating the best response condition (1), we have

$$\begin{aligned} s^{NE}(\theta) &= (1-\gamma-\beta)\theta + \gamma\mathbb{E}_P\left[s(\theta_1) \mid \theta_0 = \theta\right] + \beta\mathbb{E}_F[s(\theta')] = \dots \\ &= (1-\gamma-\beta)\sum_{t=0}^{\tau}\gamma^t\left(\mathbb{E}_P\left[\theta_t \mid \theta_0 = \theta\right] + \beta\mathbb{E}_F[s(\theta')]\right) + \gamma^{\tau+1}\mathbb{E}_P[s(\theta_{\tau+1}) \mid \theta_0 = \theta] \end{aligned}$$

for all θ and $\tau \in \mathbb{N}$. In Appendix B.1, we verify that the higher-order term $\gamma^{\tau+1}\mathbb{E}_P[s(\theta_{\tau+1})|\theta_0 = \theta]$ vanishes as $\tau \to \infty$, yielding the following result:

Lemma 1 (Nash equilibrium)**.** *For any (P, γ, β), there exists a unique Nash equilibrium. Nash strategies are strictly increasing in types with average action $\mathbb{E}_F[s^{NE}(\theta)] = \mathbb{E}_F[\theta]$ and*

$$s^{NE}(\theta) = (1-\gamma-\beta)\sum_{t=0}^{\infty}\gamma^t\mathbb{E}_P[\theta_t|\theta_0 = \theta] + \frac{\beta\mathbb{E}_F[\theta']}{1-\gamma} \quad \textit{for all } \theta. \tag{5}$$

Thus, θ's Nash action is a weighted average of a γ-discounted sum of her expected t-step ahead matches and a constant that depends only on γ, β, and the type mean $\mathbb{E}_F[\theta']$ in society. Her behavior is increasing in her type for two reasons: First, because of the direct effect that higher types prefer higher actions. Second, because higher types are more likely to meet other high types; this is reflected by the fact that, due to the monotonicity of the Markov process P, all t-step ahead expectations $\mathbb{E}_P[\theta_t|\theta_0 = \theta]$ are (weakly) increasing in θ.

By contrast, consider any assortativity neglect equilibrium $(s, (\hat{P}_\theta, \hat{s}_\theta)_\theta)$. For each θ, the best-

[19] Higher-order expectations of others' types (Samet, 1998) without monotonicity play an important role in other papers on coordination games (e.g., Morris and Shin, 2002; Golub and Morris, 2017).

response requirement means that $s(\theta) = (1-\gamma-\beta)\theta + \gamma\mathbb{E}_{\hat{P}_\theta}[\hat{s}_\theta(\theta')|\theta] + \beta\mathbb{E}_{\hat{F}_\theta}[\hat{s}_\theta(\theta')]$. Moreover, since θ suffers from assortativity neglect (i.e., $\hat{C}_\theta = C_I$), she perceives her local action mean and global action mean to coincide, i.e., $\mathbb{E}_{\hat{P}_\theta}[\hat{s}_\theta(\theta')|\theta] = \mathbb{E}_{\hat{F}_\theta}[\hat{s}_\theta(\theta')]$. Finally, observational consistency implies that θ is correct about her local action action mean, i.e., $E_{\hat{P}_\theta}[\hat{s}_\theta(\theta')|\theta] = E_P[s(\theta')|\theta]$. Combining these observations yields the misspecified best response

$$s(\theta) = (1-\gamma-\beta)\theta + (\gamma+\beta)\mathbb{E}_P[s(\theta')|\theta] =: \mathrm{BR}_\theta^{AN}(s, P). \tag{6}$$

As under Nash, iterating (6) implies that in any ANE, behavior is uniquely given by

$$s^{AN}(\theta) = (1-\gamma-\beta)\sum_{t=0}^{\infty}(\gamma+\beta)^t\mathbb{E}_P[\theta_t|\theta_0=\theta] \text{ for all } \theta, \tag{7}$$

i.e., θ's action is a $(\gamma+\beta)$-discounted sum of her expected t-step ahead matches under P.[20]

Generalizing the illustrative example in Section 1.1, a first key implication is that assortativity neglect increases action dispersion relative to Nash. Formally, comparing (5) and (7), the fact that t-step ahead expectations $\mathbb{E}_P[\theta_t|\theta_0=\theta]$ are increasing in θ implies

$$s^{AN}(\theta) - s^{AN}(\theta') \geq s^{NE}(\theta) - s^{NE}(\theta') \text{ for all } \theta > \theta',$$

with strict inequality if $\beta > 0$ and $C \neq C_I$. Equivalently, in terms of the induced cdfs over actions, the ANE action distribution $H^{AN} = F \circ s^{AN^{-1}}$ is ***more dispersive*** than the Nash action distribution $H^{NE} = F \circ s^{NE^{-1}}$; that is, $H^{AN^{-1}}(x) - H^{AN^{-1}}(y) \geq H^{NE^{-1}}(x) - H^{NE^{-1}}(y)$ for all $0 < y \leq x < 1$. Since the average actions under (5) and (7) coincide, this implies that H^{AN} is a mean-preserving spread of H^{NE}.[21]

Proposition 1 (Assortativity neglect equilibrium)**.** *For any (P, β, γ), there exists an assortativity neglect equilibrium. In any assortativity neglect equilibrium, behavior s^{AN} is given by (7). The action distribution H^{AN} is more dispersive than the Nash action distribution H^{NE}.*

While Proposition 1 uniquely pins down behavior s^{AN}, ANE are far from unique: As the proof shows (Appendix B.2), ANE allows for essentially arbitrary perceived type distributions $\hat{F}_\theta$ and strategy profiles $\hat{s}_\theta$. However, Section 4 will show that there is a unique *coherent* ANE.

Proposition 1 reflects a "false consensus effect:" Under any monotonic strategy profile, assortativity neglect leads higher types to perceive first-order stochastically higher global action

[20] Both (5) and (7) can be interpreted as steady states of the following adjustment/learning process: Starting with any monotone strategy profile s^0, if agents repeatedly best-respond to previous period behavior according to (1) and (6), play converges to (5) and (7), respectively.

[21] Recall that cdf H_1 is a mean-preserving spread of H_2 if $\int \phi(a)\, dH_1(a) \geq \int \phi(a)\, dH_2(a)$ for any convex function $\phi : \mathbb{R} \to \mathbb{R}$ for which the integrals are well-defined. For cdfs that share the same mean, the dispersive order is stronger than the mean-preserving spread order (e.g., Shaked and Shanthikumar, 2007).

distributions. Relative to correct perceptions, this drives up the gap between higher and lower types' best responses and in turn, through observation of matches' behavior, reinforces the gap in their perceptions of global behavior. Under coherent ANE, we will see that the same effect applies to agents' perceived type distributions. These predictions are in line with the eponymous finding in social psychology (Ross, Greene, and House, 1977; Marks and Miller, 1987) that people's perceptions of others' attributes and behaviors tend to be positively correlated with their own attributes and behaviors. This has been documented in a wide variety of settings, from perceptions of income levels or political attitudes to the perceived prevalence of smoking in society.[22]

3.2 Multiplier Effect of Assortativity Neglect

Proposition 1 shows that assortativity neglect increases action dispersion relative to Nash in any fixed society. We now highlight a second channel through which assortativity neglect can lead to more dispersed behavior: By amplifying the effect of two key social changes—increased assortativity and increased type heterogeneity. Remark 2 briefly points to some implications of these findings in specific economic examples.

While the illustrative example in Section 1.1 suggests an intuitive connection between increased assortativity and action dispersion, formalizing this requires an appropriate non-parametric notion of when one society is more assortative than another. We capture this by means of the following partial order over interaction structures: Call C_1 ***more assortative*** than C_2 (denoted $C_1 \succsim_{MA} C_2$) if $C_1(x^*, y^*) \geq C_2(x^*, y^*)$ for any $x^*, y^* \in (0, 1)$. That is, as illustrated in Figure 1, C_1 assigns higher probability than C_2 to low–low matches (between agents below quantiles x^* and y^*), or equivalently, assigns higher probability to high–high matches (above quantiles x^* and y^*).[23] Note that the independent interaction structure C_I is the least assortative interaction structure, while the most assortative one is the perfectly assortative interaction structure where each quantile is matched only with types of the same quantile.[24] Moreover, for Gaussian societies (Example 1) it can be shown that $C_{\rho_1} \succsim_{MA} C_{\rho_2}$ if and only if $\rho_1 \geq \rho_2$; that is, assortativity is parametrized precisely by the correlation coefficient.

To consider the effect of increased assortativity, we compare Nash and ANE action distributions and strategies H_i^{NE}, s_i^{NE}, H_i^{AN}, s_i^{AN} across environments (F, C_i, γ, β) that differ only

[22] E.g., Dawtry, Sutton, and Sibley (2015) find that individuals' perceptions of mean income are increasing in own income and Bauman and Geher (2002) that perceived support for particular policies (e.g., abortion) is higher among supporters than opponents. Botvin, Botvin, Baker, Dusenbury, and Goldberg (1992) document that perceived smoking prevalence is higher for smokers than non-smokers and positively related to the number of friends who smoke (in line with misinference from local observations).

[23] In the statistics literature (e.g., Joe, 1997), this ordering is referred to as PQD (positive quadrant dependence) or concordance order and is used more generally to compare any two-dimensional cdfs. For other uses in economics, see e.g., Meyer and Strulovici (2012).

[24] Formally, $C_{Per}(x^*, y^*) = \min\{x^*, y^*\}$.

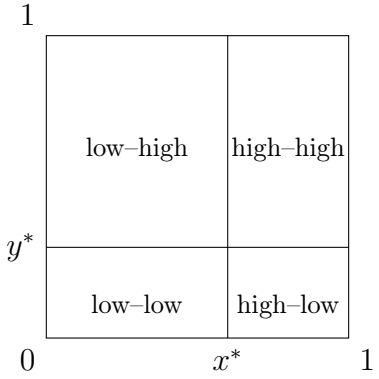


Figure 1: $C_1 \succsim_{MA} C_2$ if and only if C_1 assigns higher probability than C_2 to the high–high region and low–low region and lower probability to the low–high and high–low region.

in their interaction structures C_i:[25]

Proposition 2 (Effect of Assortativity)**.** *For any $C_1, C_2 \in \mathcal{C}$, the following are equivalent:*

1. *C_1 is more assortative than C_2.*
2. *H_1^{NE} is a mean-preserving spread of H_2^{NE} for all (F, γ, β).*
3. *H_1^{AN} is a mean-preserving spread of H_2^{AN} for all (F, γ, β).*
4. *For all $\theta^* \in \Theta$ and (F, γ, β),*

$$\mathbb{E}_F[s_1^{AN}(\theta) - s_2^{AN}(\theta) \mid \theta \geq \theta^*] \geq \mathbb{E}_F[s_1^{NE}(\theta) - s_2^{NE}(\theta) \mid \theta \geq \theta^*] \geq 0.$$

The equivalence of parts 1–3 shows that $\succsim_{MA}$ is an appropriate measure of assortativity to formalize the tight connection between increased assortativity and action dispersion: Not only does more assortativity lead to greater action dispersion (by "$1 \Rightarrow 2, 3$"), but greater action dispersion is indeed a defining feature of more assortative societies (by "$2, 3 \Rightarrow 1$"). However, while this is true under both Nash and ANE, part 4 highlights that any given rise in assortativity has a *stronger* effect on action dispersion under assortativity neglect: High types' actions increase more on average (and equivalently, low types' actions decrease more) than under Nash.[26]

[25] An analog of Proposition 2 can be provided by replacing mean-preserving spread with the dispersive order and $\succsim_{MA}$ with a stronger more-assortative order (see Appendix D.2.2). Likewise, an analog of Proposition 3 holds by replacing dispersive order with mean-preserving spread.

[26] Note that for $i \in \{\text{NE}, \text{AN}\}$, H_1^i is a mean-preserving spread of H_2^i iff $\mathbb{E}_F[s_1^i(\theta) - s_2^i(\theta) \mid \theta \geq \theta^*] \geq 0$ for all θ^*. Thus, part 4 captures that the mean-preserving spread increase from H_2^i to H_1^i is greater under ANE than Nash.

Section 1.1 discussed the basic intuition for the forward direction: A rise in assortativity increases differences in *local* coordination incentives across types (under Nash and ANE), but under ANE it additionally increases differences in perceived *global* coordination incentives by magnifying the false consensus effect. To establish the formal equivalence result, Appendix B.3 exploits the Markov process representation of Nash and ANE in (5) and (7), reducing the problem to a comparison of t-step ahead expectations across different societies.

Specifically, for any society $P = (F, C)$ and $t \geq 1$, let D_P^t denote the distribution of $\mathbb{E}_P[\theta_t|\theta_0]$ where θ_0 is distributed according to F; that is, D_P^t is the distribution of expected t-step ahead matches in society. Monotonicity of the Markov process P allows us to establish a "duality" between the more assortative order over interaction structures and the mean-preserving spread order over expected t-step ahead match distributions D_{F,C_i}^t; that is, C_1 is more assortative than C_2 if and only if D_{F,C_1}^t is a mean-preserving spread of D_{F,C_2}^t for all F and t.[27] We then complete the proof by combining the representations of Nash and ANE as discounted sums of expected t-step ahead matches with the fact that the mean-preserving spread order is linear and continuous.

Finally, comparing environments (F_i, C, γ, β) that differ only in their type distributions, we obtain an analogous result: Increased type dispersion corresponds to increased action dispersion under both Nash and ANE, but the effect is stronger under ANE.

Proposition 3 (Effect of Type Dispersion)**.** *For any $F_1, F_2 \in \mathcal{F}$, the following are equivalent:*

1. *F_1 is more dispersive than F_2.*
2. *H_1^{NE} is more dispersive than H_2^{NE} for any (C, γ, β).*
3. *H_1^{AN} is more dispersive than H_2^{AN} for any (C, γ, β).*
4. *For all (C, γ, β) and $x, y \in (0, 1)$ with $x > y$,*

$$\Delta_{x,y} H_1^{AN} - \Delta_{x,y} H_2^{AN} \geq \Delta_{x,y} H_1^{NE} - \Delta_{x,y} H_2^{NE} \geq 0,$$

where $\Delta_{x,y} H := H^{-1}(x) - H^{-1}(y)$ for any cdf H.

Remark 2. We briefly discuss some economic implications of the multiplier effect:

Education inequality. A growing literature studies the possibility of decreasing the socioeconomic education gap by reducing income-based residential segregation, e.g., through programs such as "Moving to Opportunity."[28] In line with the Nash prediction of Proposition 2, much

[27] This result follows from Lemma B.1, which shows that $C_1 \gtrsim_{MA} C_2$ if and only if D_{F,C_1}^1 is a mean-preserving spread of D_{F,C_2}^1 for all F, combined with Lemma A.3, which establishes that the mean-preserving spread order is "isotone."

[28] E.g., Chetty, Hendren, and Katz (2016). Surveys of the literature on neighborhood/peer effects in education include Durlauf (2004); Sacerdote (2011).

theoretical and empirical work emphasizes the role of direct local complementarities, either socio-psychological (e.g., peer effects) or material (e.g., local provision of educational facilities). Under ANE, we additionally highlight an inferential channel: Decreasing the segregation of disadvantaged individuals might reduce their misinferences from local observations about the returns to education (which are subject to global complementarities).[29] Moreover, this inferential channel suggests that further reduction in the education gap might be achieved through informational interventions that correct misperceptions.

Smoking. Many developed countries have seen rising socioeconomic disparities in cigarette consumption.[30] One explanation is that growing information about the risks of smoking may have been absorbed more readily by individuals of higher socioeconomic status, but there is a discussion in the empirical literature about whether socioeconomic variations in knowledge of smoking risks are indeed significant enough to account for large behavior differences.[31] Proposition 3 suggests that the behavioral impact of even a moderate rise in the socioeconomic knowledge gap might be exacerbated under assortativity neglect:[32] The direct effect on behavior could be amplified by the indirect effect of an increased gap in the perceived societal prevalence of smoking. See the discussion following Proposition 1 for evidence of a false consensus effect in the context of smoking. ▲

Example 2 (Example 1 continued)**.** In a Gaussian society P parametrized by (μ, σ^2, ρ), each type θ_0's distribution $P(\cdot|\theta_0)$ of matches is also normal with mean $\mathbb{E}_P[\theta_1|\theta_0] = (1-\rho)\mu + \rho\theta_0$; inductively, $\mathbb{E}_P[\theta_t|\theta_0] = (1-\rho^t)\mu + \rho^t\theta_0$ for all t. Applying (5) and (7) yields strategy profiles

$$s^{NE}(\theta) = \frac{1-\beta-\gamma}{1-\gamma\rho}\theta + \frac{\beta+\gamma(1-\rho)}{1-\gamma\rho}\mu, \quad s^{AN}(\theta) = \frac{1-\beta-\gamma}{1-(\beta+\gamma)\rho}\theta + \frac{(\beta+\gamma)(1-\rho)}{1-(\beta+\gamma)\rho}\mu,$$

with corresponding action variances $\left(\frac{1-\beta-\gamma}{1-\gamma\rho}\right)^2\sigma^2$ and $\left(\frac{1-\beta-\gamma}{1-(\beta+\gamma)\rho}\right)^2\sigma^2$, respectively. In line with Proposition 1, ANE features a higher action variance than Nash. Moreover, both variances are increasing in ρ and σ^2, but the derivative is higher under ANE than Nash, reflecting the multiplier effect in Propositions 2–3. ▲

[29] Streufert (2000) proposes a non-equilibrium model where poor individuals assess the returns to schooling by exogenously sampling outcomes in their neighborhood, neglecting the fact that successful individuals do not appear in their observations as they have left the neighborhood.

[30] E.g., US smoking rates from the mid-1960s to 2000 declined by 62% among the top income quintile but by only 9% among the lowest income quintile (Wan, 2017).

[31] See, e.g., Siahpush, McNeill, Hammond, and Fong (2006) for evidence of significant knowledge differences and Pampel, Krueger, and Denney (2010); Cutler and Lleras-Muney (2010) (p. 357 & p. 9) for a counterpoint.

[32] To apply Proposition 3, we interpret types as capturing ignorance about smoking risks, which are decreasing in income, and we assume that sorting is based on income. An increase in the socioeconomic knowledge gap can then be captured by increased dispersion of F, while holding fixed interaction structure C.

4 Coherency of Assortativity Neglect

4.1 Rationalizing Assortativity Neglect

We now show that in any environment, there is a coherent ANE. Thus, even when agents seek to rationalize observed behavior, assortativity neglect is a misperception that can persist regardless of how assortative the actual society is. Moreover, whereas Proposition 1 only pinned down ANE behavior, we show that the coherent ANE is unique. In particular, for any (P, γ, β), it fully determines agents' misperceptions $\hat{F}_\theta$ about the type distribution, paving the way for comparative statics analysis of how these are shaped by the nature of social interactions.

Proposition 4 (Coherency of Assortativity Neglect)**.** *For any (P, γ, β), there exists a unique coherent assortativity neglect equilibrium.*

The existence of coherent ANE is due to the following defining feature of assortativity neglect: If an agent's perceived interaction structure is C_I, she can rationalize *any* behavior. Formally, Lemma B.2 in the appendix shows that for any action distribution $\hat{H}$ and any γ, β, there is an appropriate type distribution $\hat{F}$ such that $\hat{H}$ is the Nash action distribution in society $(\hat{F}, C_I)$; moreover, C_I is the *only* interaction structure with this feature.

To see the idea, note that in order for $\hat{H}$ to be the Nash action distribution in society $(\hat{F}, \hat{C})$, it must be possible to decompose the action difference between any two quantiles $x > y$ into two terms—the corresponding type difference and the difference in local coordination incentives, where the latter results from differences in x and y's matches' behavior and is greater under more assortative $\hat{C}$ (as highlighted by Proposition 2):

$$\underbrace{\hat{H}^{-1}(x) - \hat{H}^{-1}(y)}_{\Delta(\text{actions})} = (1-\gamma-\beta)\underbrace{(\hat{F}^{-1}(x) - \hat{F}^{-1}(y))}_{\Delta(\text{types})} + \underbrace{\gamma \int \hat{H}^{-1}(z)\, d\left(\hat{C}(z|x) - \hat{C}(z|y)\right)}_{\Delta(\text{local coordination incentives})}. \quad (8)$$

The key feature of the independent interaction structure $\hat{C} = C_I$ is that there is no difference between x and y's coordination incentives, as all agents face the *same* distribution of matches. This makes it possible to rationalize $\hat{H}$ by simply attributing all action dispersion to type dispersion. Moreover, for any $\hat{H}$ there is a unique type distribution $\hat{F}$ that achieves this while also yielding the correct action mean $\mathbb{E}_{\hat{H}}[a] = \mathbb{E}_{\hat{F}}[\theta]$. To prove Proposition 4, Appendix B.4 applies this idea to each type's local action distribution $H_\theta^{s^{AN},P}$ under ANE, which uniquely determines θ's perceived type distribution $\hat{F}_\theta$ and strategy profile $\hat{s}_\theta$.

As discussed in the illustrative example, this rationalization procedure is reminiscent of the "fundamental attribution error" (Ross, 1977), as it attributes variation in others' behavior entirely to intrinsic characteristics (types) and neglects that external factors (differences in peer effects) might also be at play. While this procedure is always succesful under assortativity

neglect, Section 4.3 will show that it does not apply under other perceived interaction structures, as ignoring differences in local coordination incentives is not feasible when $\hat{C} \neq C_I$. We will use this to establish a converse of Proposition 4 and to derive limits on agents' coherent misperceptions in any given environment.

4.2 Perceived Type Distributions

Since coherent ANE uniquely pins down $\hat{F}_\theta$, it provides a lens through which to study how agents' misperceptions of population characteristics are shaped by the nature of their social interactions. As we discuss below, this may be helpful in understanding empirical findings on misperceptions of income distributions and has implications for models of demand for redistribution.

Throughout this section, given any environment (P, γ, β), let $\hat{F}_\theta$ denote θ's perceived type distribution in the coherent ANE at (P, γ, β). We begin by noting that under coherent ANE, the "false consensus effect" we observed for perceived behavior following Proposition 1 extends to $\hat{F}_\theta$: The only way for higher types to rationalize their higher perceived action distributions is by perceiving higher type distributions.

Corollary 1 (False Consensus Effect)**.** *For any (P, γ, β), $\hat{F}_\theta$ is increasing in θ with respect to first-order stochastic dominance.*

The false consensus effect would also emerge if agents were able to directly observe their local match distributions $P(\cdot|\theta)$ and formed perceptions of the global type distribution by projecting $P(\cdot|\theta)$ onto society as a whole.[33] Indeed, this setting corresponds to coherent ANE when coordination motives $\gamma + \beta$ are zero, as in this case $s^{AN}(\theta) = \theta$ for all types. However, we now exhibit two ways in which the fact that under coherent ANE, agents form perceptions by rationalizing observed local *behavior* leads to more nuanced predictions, where the strength of coordination motives $\gamma + \beta$ relative to other societal features such as assortativity plays a key role.

First, we show that coordination motives drive up agents' perceptions of type dispersion (e.g., income inequality) in society. Moreover, coordination motives matter not only quantitatively, but affect the direction of agents' misperceptions: Depending on the strength of $\gamma + \beta$, agents may over- or underestimate global type dispersion. Notably, overestimation is possible even when agents' local match distributions are less dispersed than the global type distribution, which is a natural feature that holds on average in any society.[34]

[33] In the context of income distributions, the idea that agents directly observe peers' types and project them onto society underlies Cruces, Perez-Truglia, and Tetaz (2013) and Windsteiger (2018), who emphasize that this may lead to underestimation of income inequality.

[34] Formally, by the law of iterated variance, any society P satisfies $\mathbb{E}_F[\mathrm{Var}_P[\theta'|\theta]] \leq \mathrm{Var}_F[\theta]$.

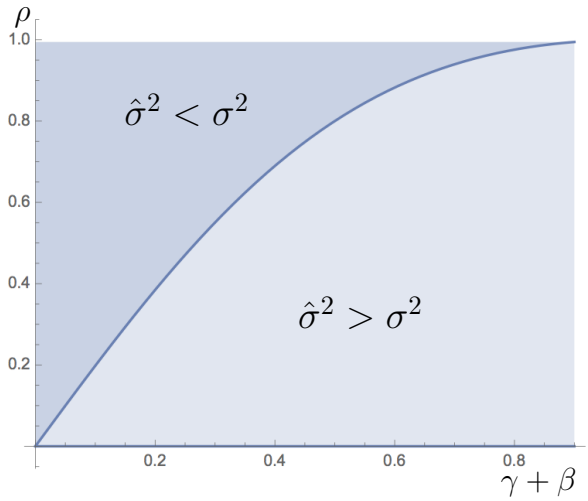


Figure 2: Over-/underestimation of type variance in Example 3.

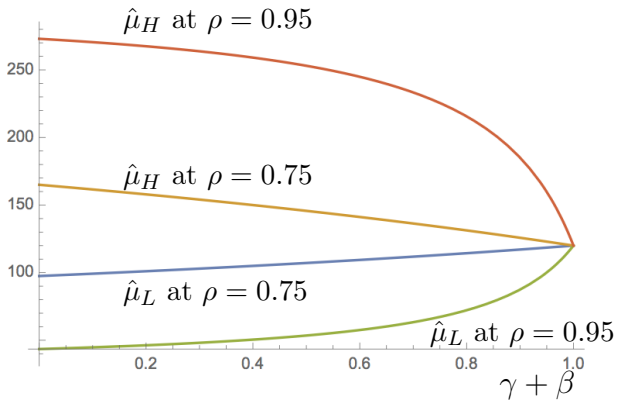


Figure 3: Dispersion of perceived type means in Example 4.

Proposition 5 (Perceived type dispersion)**.** *Fix any P and $\beta_1 + \gamma_1 \leq \beta_2 + \gamma_2$. Let $\hat{F}^i_\theta$ denote type θ's perceived type distribution under the coherent ANE at (P, β_i, γ_i). Then $\hat{F}^2_\theta$ is more dispersive than $\hat{F}^1_\theta$ for all θ.*

Example 3 (Over-/underestimation of type dispersion)**.** Consider a Gaussian society $P = (\mu, \sigma^2, \rho)$ with $\rho > 0$. By Example 2, ANE behavior satisfies $s^{AN}(\theta') = \frac{1-\beta-\gamma}{1-(\beta+\gamma)\rho}\theta' + \frac{(\beta+\gamma)(1-\rho)}{1-(\beta+\gamma)\rho}\mu$. Hence, θ's local action distribution is normally distributed with mean $\mu + (\theta - \mu)\rho\frac{1-\beta-\gamma}{1-(\beta+\gamma)\rho}$ and variance $\sigma^2\frac{(1-\gamma-\beta)^2(1-\rho^2)}{(1-(\beta+\gamma)\rho)^2}$. Thus, coherency implies that $\hat{F}_\theta$ is normal with mean $\hat{\mu}_\theta = \mu + (\theta - \mu)\rho\frac{1-\beta-\gamma}{1-(\beta+\gamma)\rho}$ and variance $\hat{\sigma}^2_\theta = \sigma^2\frac{1-\rho^2}{(1-(\beta+\gamma)\rho)^2}$. Note that θ's match distribution has variance $\sigma^2(1-\rho^2)$, which is less than the global type variance σ^2. If $\gamma + \beta = 0$, then θ's perceived type distribution is equal to her match distribution, so that $\hat{\sigma}^2_\theta < \sigma^2$. However, $\hat{\sigma}^2_\theta$ is increasing in $\gamma + \beta$, and $\hat{\sigma}^2_\theta > \sigma^2$ whenever $\frac{2(\gamma+\beta)}{1+(\gamma+\beta)^2} > \rho$. Thus, θ overestimates (resp. underestimates) type dispersion when $\gamma + \beta$ is large (resp. small) relative to the assortativity parameter ρ. Figure 2 illustrates. ▲

To understand Proposition 5, it is important to note that stronger coordination motives in fact lead to *less* dispersed ANE *behavior* (Appendix C.5): This is because as $\gamma + \beta$ increases, θ's behavior $s^{AN}(\theta) = (1-\gamma-\beta)\sum_{t=0}^{\infty}(\gamma+\beta)^t\mathbb{E}_P[\theta_t|\theta_0 = \theta]$ is influenced less by her own type and more by her matches' and higher-order matches' characteristics, and the latter exhibit less variation across agents than their own types. However, under assortativity neglect, agents mistakenly believe that matches' characteristics do not feature *any* variation across agents, which other things equal, would lead them to expect even less local action dispersion than actual. To maintain coherency and observational consistency, they compensate by perceiving higher type dispersion.

Remark 3 (Perceptions of income inequality)**.** As Hauser and Norton (2017) summarize in a recent survey, a growing empirical literature documents that "people around the world hold incorrect perceptions of [income] inequality in their country—but with variation. In the US and United Kingdom, for example, underestimation of inequality is relatively common, while overestimation occurs in other countries, such as France and Germany." Cross-country differences in perceived inequality are often at odds with actual levels of inequality; e.g., Niehues (2014) finds higher perceived levels of inequality in France and Germany than in the US and UK, despite the fact that actual inequality is greater in the latter two countries.

Based on the premise that people draw inferences about inequality by observing their peers' consumption patterns and neglecting assortativity, our analysis provides a theoretical lens for understanding how cross-country variations in perceived inequality might be influenced by societal features other than actual inequality. For instance, our above findings suggest that ceteris paribus, perceived inequality might be higher in countries with stronger social conformity tendencies and/or less socioeconomic segregation; that both under- and overestimation of inequality is possible depending on the relative strength of these two features;[35] and that cross-country differences in these features can lead perceived inequality to be higher in more equal countries than less equal ones.[36] ▲

Next, we show that while the false consensus effect implies that perceived type means $\hat{\mu}_\theta := \mathbb{E}_{\hat{F}_\theta}[\theta']$ are increasing in own type θ, coordination motives weaken this effect: The distribution of $\hat{\mu}_\theta$ in the population (i.e., when θ is distributed according to F) is less dispersed the higher $\gamma + \beta$, and under a regularity condition, perceived type means converge to the true mean as $\gamma + \beta \to 1$. Again, this result matters not only quantitatively, but has qualitative implications for how agents perceive their own position relative to the mean, which we show might affect society's demand for redistribution.

Proposition 6 (Dispersion of perceived type means)**.** *Fix any P and $\beta_1 + \gamma_1 \leq \beta_2 + \gamma_2$. Let $\hat{M}^i$ denote the population distribution of perceived type means $\hat{\mu}_\theta$ under the coherent ANE at*

[35] This finding might also help understand empirically documented misperceptions in other contexts, e.g., why it might be possible for people to overestimate political attitude polarization in society (as documented by Ahler, 2014; Westfall, Van Boven, Chambers, and Judd, 2015) even if political attitudes among their peers are more homogenous than in society as a whole.

[36] Careful empirical analysis of these predictions is beyond the scope of this paper, but we note that in the case of the US/UK vs. France/Germany they may enjoy some empirical support: First, several studies document higher socioeconomic segregation in the US (and to a lesser extent the UK) than in continental Europe (e.g., Musterd, 2005; Quillian and Lagrange, 2016). Second, an anti-conformism measure that has been employed in cross-country comparisons in economics (e.g., Gorodnichenko and Roland, 2011) and is known to correlate highly with anti-conformist behavior in lab experiments (Bond and Smith, 1996) is Hofstede's "individualism index" (Hofstede, 2001); this ranks the US (91 out of 100) and UK (89) among the least conformist countries and France (71) and Germany (67) as somewhat more conformist (cf. https://www.hofstede-insights.com/product/compare-countries).

(P, β_i, γ_i). *Then* $\hat{M}^1$ *is a mean-preserving spread of* $\hat{M}^2$. *Moreover, if* $\lim_{t\to\infty} \mathbb{E}_P[\theta_t|\theta_0 = \theta] = \mathbb{E}_F[\theta']$, *then* $\lim_{\gamma+\beta\to 1} \hat{\mu}_\theta = \mathbb{E}_F[\theta']$.

Under coherent ANE, θ's perceived type mean is equal to her local action average, $\hat{\mu}_\theta = \mathbb{E}_P[s^{AN}(\theta')|\theta]$. Hence, Proposition 6 follows from the preceding observation that stronger coordination motives lead to less dispersed ANE behavior. As noted, the higher $\gamma + \beta$, the more is each agent's behavior influenced by her more distant t-step ahead matches. Thus, if expected t-step ahead matches approximate the population mean as $t \to \infty$, then θ's perceived mean converges to the truth as $\gamma + \beta \to 1$. The requirement that $\lim_{t\to\infty} \mathbb{E}_P[\theta_t|\theta_0 = \theta] \to \mathbb{E}_F[\theta']$ holds in all examples in this paper; Appendix C.6.1 provides a simple sufficient condition, capturing that the matching technology is not too local and society is well connected.

In Example 3, symmetry of the Gaussian environment yields perceived means that are a convex combination of the true mean and agents' own types, so that (i) all agents correctly perceive which side of the mean they are on; (ii) perceived positions $\theta - \hat{\mu}_\theta$ relative to the mean are increasing in θ. Example 4 shows that in asymmetrically segregated societies both (i) and (ii) can be violated. However, by Proposition 6, the stronger coordination motives are relative to other fundamentals such as assortativity, the less severe are these violations.

Example 4 (Perceived position relative to the mean)**.** Consider the following partitional society: With probability $\frac{2}{3}$, θ is drawn from a "lower class" of types distributed uniformly on $\Theta_L = [0, 60]$, and with probability $\frac{1}{3}$, θ belongs to the "upper class" distributed uniformly on $\Theta_H = [60, 540]$. For each $\theta \in \Theta_i$, θ's match is drawn uniformly from her own class Θ_i with probability ρ_i, and is drawn uniformly from the other class with probability $1 - \rho_i$, where $\rho_L =: \rho \in [\frac{2}{3}, 1]$ and $\rho_H = 2\rho - 1$ to ensure assortativity and symmetry.

The true type mean is 120. In the coherent ANE,[37] all lower class types perceive the mean to be $\hat{\mu}_L \in [30, 120]$ and upper class types perceive $\hat{\mu}_H \in [120, 300]$, where Proposition 6 implies that $\hat{\mu}_L$ is increasing and $\hat{\mu}_H$ is decreasing in $\gamma+\beta$; see Figure 3.[38] Whenever $\gamma+\beta$ is sufficiently small and the sorting parameter ρ is sufficiently large, then $\hat{\mu}_L < 60$. Thus, for all $\theta_L \in (\hat{\mu}_L, 60)$ and $\theta_H \in (60, \hat{\mu}_H)$, we have $\theta_L - \hat{\mu}_L > 0 > \theta_H - \hat{\mu}_H$; that is, the upper range of lower class types misperceive themselves to be above average, while the lower range of upper class types perceive themselves to be below average (incorrectly for $\theta_H \in (120, \hat{\mu}_H)$). By contrast, as $\gamma+\beta \to 1$, both $\hat{\mu}_L$ and $\hat{\mu}_H$ tend to the true mean 120, so that agents have more and more accurate perceptions of their position relative to mean. ▲

[37] ANE actions can be calculated as $s^{AN}(\theta) = \frac{30((\gamma+\beta)(12\rho-11)-1)}{(\gamma+\beta)(3\rho-2)-1} + (1-\gamma-\beta)(\theta-30)$ for each $\theta \in \Theta_L$ and $s^{AN}(\theta) = \frac{60((\gamma+\beta)(6\rho-1)-5)}{(\gamma+\beta)(3\rho-2)-1} + (1-\gamma-\beta)(\theta-300)$ for each $\theta \in \Theta_H$. The corresponding perceptions are given by $\hat{\mu}_L = \frac{30((\gamma+\beta)(3\rho-2)+9\rho-10)}{(\gamma+\beta)(3\rho-2)-1}$ and $\hat{\mu}_H = \frac{60((\gamma+\beta)(15\rho-10)-9\rho+4)}{(\gamma+\beta)(3\rho-2)-1}$.

[38] Note that $\hat{\mu}_L$ ($\hat{\mu}_H$) is also decreasing (increasing) in ρ. In general, analogous to Proposition 6, it can be shown that the distribution $\hat{M}$ of perceived type means is mean-preserving spread increasing in the $\gtrsim_{MA}$-order over interaction structures C.

Remark 4 (Preference for redistribution)**.** Several recent empirical and theoretical studies emphasize that people's misperceptions of income distributions might affect a society's demand for redistribution.[39] Using a simple version of Meltzer and Richard's (1981) majority voting model as an illustration, Appendix E.2 suggests two ways in which our findings add to this point. First, assortativity neglect can affect the identity of the decisive voter: While under correct perceptions the implemented redistribution scheme is the one preferred by the median income type, this need not hold when agents' perceived positions relative to the mean are nonmonotonic (as in Example 4). Second, both the identity of the decisive voter and her demand for redistribution are shaped not only by the true income distribution, but also depend systematically on coordination motives and assortativity. In Example 4, assortativity neglect reduces demand for redistribution relative to correct perceptions, and completely eliminates demand for redistribution as $\gamma + \beta \to 0$ and $\rho \to 1$. ▲

4.3 Limits on Coherent Misperceptions

Proposition 4 showed that in any environment, there is a coherent equilibrium in which all agents suffer from assortativity neglect (i.e., hold perception $\hat{C}_\theta = C_I$). The following result provides a converse: For any interaction structure $\hat{C}$ other than C_I and any type θ, there are environments in which θ *cannot* sustain perception $\hat{C}$ in any coherent equilibrium. For this result, we impose the following mild regularity requirement: We call an interaction structure $\hat{C}$ ***regular*** if there exists some $y \in (0,1)$ such that $|\{x \in (0,1) : \hat{C}(x|x) = y\}| = 1$.[40]

Proposition 7. *Fix any regular $\hat{C} \neq C_I$ and any θ. There exists (P, γ, β) such that all coherent equilibria satisfy $\hat{C}_\theta \neq \hat{C}$.*

To see the idea, just as for Proposition 4, first consider rationalizing an arbitrary action distribution $\hat{H}$ as Nash. As summarized by (8), in order for $\hat{H}$ to be Nash at some perceived society $(\hat{F}, \hat{C})$, observed action dispersion must be consistent with the type dispersion and variation in local coordination incentives implied by $(\hat{F}, \hat{C})$. The key difference with Proposition 4 is that if $\hat{C} \neq C_I$, then coordination incentives do vary across agents. Indeed, for some action distributions $\hat{H}$, coordination incentives across agents will differ by more than their action

[39] Cruces, Perez-Truglia, and Tetaz (2013); Engelhardt and Wagener (2015); Gimpelson and Treisman (2018); Niehues (2014) find empirically that people's perceptions of income inequality or of their own income quantile better predict their attitudes to redistribution than the actual income distribution. In an endogenous sorting model, Windsteiger (2018) assumes agents underestimate differences across income classes and shows that this reduces demand for redistribution; in contrast with Appendix E.2, the median income voter is decisive in her setting and coordination motives play no role as agents draw inferences from observed types rather than actions.

[40] Function $x \mapsto \hat{C}(x|x)$ maps each quantile x to its "local quantile" under $\hat{C}$, i.e., to the fraction of its matches with quantile below its own. Regularity rules out that this function oscillates arbitrarily, requiring there to be at least one value y that is achieved exactly once. This is weaker than the requirement that the map is monotone, which is satisfied under both Gaussian interactions (Example 1) and under $\hat{C} = C_I$.

differences, making it impossible to rationalize $\hat{H}$ regardless of the perceived type distribution $\hat{F}$.[41] The proof of Proposition 7 builds on this observation, exhibiting environments (P, γ, β) in which in any coherent equilibrium, θ cannot rationalize her perceived global action distribution under perception $\hat{C}$.

Propositions 4 and 7 capture a sense in which assortativity neglect is a misperception that is especially simple to sustain: While assortativity neglect allows agents to only reason about others' intrinsic characteristics, any other misperception requires taking into account variations in coordination incentives, and in some environments this will lead to violations of coherency. Of course, in any *fixed* environment (P, γ, β), there may be scope for many coherent perceptions other than assortativity neglect (notably, the correct perception $\hat{C} = C$). Nevertheless, as the following example illustrates, the same logic as above can imply upper bounds on how much assortativity agents can perceive in a given environment:

Example 5 (Upper bounds on perceived assortativity)**.** Consider the same setting as in Section 1.1 with three types $\theta_h > \theta_m = 0 > \theta_l$, but drop the symmetry assumption to allow for $\theta_h \neq -\theta_l$. For simplicity, assume there are only local coordination incentives, with best-response function $\text{BR}_{\theta_i}(s) = \frac{1}{2}\theta_i + \frac{1}{2}\mathbb{E}[s(\theta)|\theta_i]$.[42]

Since there are no global coordination incentives, it is immediate from Definition 2 that true behavior in any equilibrium is the same as Nash. Fixing a true sorting parameter $\rho \in [\frac{1}{3}, \frac{2}{3}]$, we focus on which perceptions $\hat{\rho} \in [\frac{1}{3}, \frac{2}{3}]$ about ρ the middle-class type can sustain in a coherent equilibrium. By observational consistency, θ_m knows the three equilibrium action levels $a_i = s^{NE}(\theta_i)$. Moreover, θ_m knows her own type. However, θ_m might hold misperceptions about rich and poor types, perceiving them to be $\hat{\theta}_h$ and $\hat{\theta}_l$, as long as perceived types continue to be ordered by

$$\hat{\theta}_h > \theta_m = 0 > \hat{\theta}_l.^{43} \tag{9}$$

Consider the actual and perceived action differences between the rich and middle-class. By the best-response condition, the actual difference is given by

$$a_h - a_m = \frac{1}{2}\underbrace{(\theta_h - \theta_m)}_{\Delta(\text{types})} + \underbrace{\frac{1}{2}(\rho - \frac{1}{3})(a_h - a_l)}_{\Delta(\text{local coordination incentives})}. \tag{10}$$

[41] Concretely, pick $x > y$ such that the distribution $\hat{C}(\cdot|x)$ of x's matches strictly stochastically dominates that of y. Then consider an action distribution $\hat{H}$ whose xth and yth quantiles are very similar ($\hat{H}^{-1}(x) - \hat{H}^{-1}(y) \approx 0$) but such that x's matches on average take substantially higher actions than y's matches. In this case, it is impossible to satisfy (8) regardless of $\hat{F}$, because x and y's coordination incentives already differ by more than their actions.

[42] The same logic extends to general γ, β; see Appendix D.2.2.

[43] (9) must hold because $\hat{\theta}_i$ is the type to whom θ_m attributes action a_i (i.e., $a_i = \hat{s}_{\theta_m}(\hat{\theta}_i)$), and under coherency θ_m's perceived strategy profile $\hat{s}_{\theta_m}$ is increasing in types.

At the same time, the coherency condition requires θ_m to believe that a_h is a best response for $\hat{\theta}_h$ under sorting parameter $\hat{\rho}$, which implies a perceived gap of

$$a_h - a_m = \frac{1}{2}(\hat{\theta}_h - \theta_m) + \frac{1}{2}(\hat{\rho} - \frac{1}{3})(a_h - a_l). \tag{11}$$

Combining (10) and (11) and the symmetric conditions for action differences between the poor and middle-class yields

$$\hat{\theta}_h - \theta_h = (\rho - \hat{\rho})(a_h - a_l) = \theta_l - \hat{\theta}_l. \tag{12}$$

If θ_m underestimates sorting, i.e., $\hat{\rho} \leq \rho$, then the fact that $a_h > a_l$ implies that θ_m overestimates θ_h and underestimates θ_l (i.e., $\hat{\theta}_h \geq \theta_h$ and $\hat{\theta}_l \leq \theta_l$). This is always consistent with the ordering restriction (9). By contrast, if $\hat{\rho} > \rho$, then θ_m underestimates θ_h and overestimates θ_l, and if $\hat{\rho}$ is too great, her perceived types $\hat{\theta}_h$ and $\hat{\theta}_l$ might violate (9). This typically implies an upper bound $\bar{\rho}$ on θ_m's sustainable perceptions $\hat{\rho} \in [\frac{1}{3}, \frac{2}{3}]$. Specifically, assuming $\theta_h \geq -\theta_l$, the sustainable perceptions must satisfy $\hat{\rho} < \bar{\rho} := \rho - \theta_l \frac{2(\frac{4}{3}-\rho)}{\theta_h - \theta_l}$.[44] For $\rho = \frac{1}{2}$, $\theta_l = -10$, and $\theta_h = 190$, the upper bound is $\bar{\rho} = \frac{7}{12} < \frac{2}{3}$. As $\theta_h \to \infty$, $\bar{\rho}$ shrinks to ρ, so that underestimating ρ is always sustainable but limits on coherent overestimation can be arbitrarily tight. ▲

The above rationalization procedure for $\hat{\rho} < \rho$ reflects a weaker form of "fundamental attribution error:" θ_m underestimates (though need not fully neglect) differences in external circumstances across agents, and compensates by overattributing behavior differences to intrinsic characteristics. The example suggests that while this rationalization procedure is always internally consistent (i.e., coherent), the same need not be true for the opposite bias of overestimating environmental differences. Supplementary Appendix D.2.2 shows how this logic extends to general settings.

5 Discussion

We briefly discuss several extensions of our analysis and review related literature.

5.1 Extensions

5.1.1 Hybrid model and negative externalities of assortativity neglect

Under ANE, all agents suffer from assortativity neglect. More realistically, some agents might be less prone to misperception than others, for example due to having access to information about global (rather than just local) action distributions. As a simple illustration, Appendix D.1

[44] Indeed, note that $a_h - a_l = \frac{1}{2}(\theta_h - \theta_l) + (\rho - \frac{1}{3})(a_h - a_l)$, yielding $a_h - a_l = \frac{\theta_h - \theta_l}{2(\frac{4}{3}-\rho)}$. Combining (12), (9) and $\theta_h \geq -\theta_l$ then yields the upper bound $\bar{\rho}$. An analogous bound can be obtained if $\theta_h < -\theta_l$.

considers the following hybrid model: Independently of types, only fraction $\alpha \in [0,1]$ of agents suffer from assortativity neglect (subject to the best-response and observational consistency requirements in Definition 2), while the remaining share of agents best-respond to *correct* perceptions about the strategy profile and society. We show that equilibrium strategies of both groups again admit Markov process representations that generalize those for ANE and Nash, based on which all main results extend.

Notably, while behavior among assortativity neglect agents is more dispersed than among correct agents, a greater fraction of assortativity neglect agents exacerbates action dispersion among *both* groups: Just as in Proposition 1, best responses against any fixed monotone strategy are more dispersed among assortativity neglect agents, because the false consensus effect leads higher types to perceive stronger global coordination incentives; but even for correct agents, the resulting increase in equilibrium action dispersion translates into more dispersed local coordination incentives, and more so the greater α. This captures a sense in which assortativity neglect agents exert a negative externality on society, as they drive up miscoordination among all agents.[45]

5.1.2 Weaker forms of assortativity neglect

For simplicity, we have focused on a rather extreme notion of assortativity neglect, where agents believe that everyone interacts with a perfectly representative sample of society. Appendix D.2 shows that several of our main insights extend to weaker forms of assortativity neglect. First, Appendix D.2.1 focuses on the finding that assortativity neglect increases action dispersion relative to Nash. Whereas under ANE agents' perceptions of their own match quantile distributions $\hat{C}_\theta(\cdot|\hat{F}_\theta(\theta))$ are the same for all θ (namely uniform on $[0,1]$), we show that this finding generalizes to a weaker notion of assortativity neglect that only requires perceived match distributions to be "less increasing" in θ than the actual match distributions $C(\cdot|F(\theta))$. Behavior in this case continues to admit a Markov process representation that nests Nash and ANE. In Gaussian societies, this weaker form of assortativity neglect is satisfied by a class of coherent equilibria where agents commonly perceive the correlation coefficient to be any $\hat{\rho}$ that is less than the true ρ.

Second, Appendix D.2.2 focuses on coherency of assortativity neglect. Extending our finding that assortativity neglect is especially easy to rationalize, as it is the only perception of the interaction structure that agents can coherently sustain across *all* environments (Propositions 4 and 7), we formalize a sense in which in any *fixed* environment, it is easier to coherently perceive lower levels of assortativity than higher levels. The result implies that in any environment, θ can coherently perceive a continuum range of interaction structures $\hat{C}$ that are less assortative than

[45] This observation contrasts with Jehiel's (2018) model of investment under selection neglect, where the effect of misperception is weakened the greater the share of agents who suffer from selection neglect.

the true interaction structure C. As in Example 5, θ's rationalization procedure overattributes differences in actions to differences in types, reflecting a weaker form of "fundamental attribution error." By contrast, as Example 5 illustrated, some environments feature limits on how much θ can overestimate assortativity.

5.1.3 More general best response functions

As in many papers on network games and incomplete information coordination games, we have focused on linear best response functions. This enabled us to obtain Markov process representations of Nash and ANE strategies, which played an important role in deriving the comparative statics results in Sections 3.2 and 4.2. At the same time, several of our main insights do not rely on the linearity of best responses. Indeed, consider a more general best response function of the form $\mathrm{BR}_\theta(s,P) = b(\theta, H_\theta^{s,P}, H^{s,P})$, where $H^{s,P}$ denotes the global action distribution at (s,P), map $b : \mathbb{R}\times\Delta(\mathbb{R})\times\Delta(\mathbb{R})\to\mathbb{R}$ is continuously increasing in θ and satisfies $\lim_{\theta\to\infty} b(\theta,H,H') = \infty$ and $\lim_{\theta\to-\infty} b(\theta,H,H') = -\infty$ for any cdfs H, H', and we assume suitable regularity conditions that ensure existence of Nash and ANE. Note that b may be first-order stochastic dominance (FOSD)-increasing or decreasing in local and global action distributions, allowing for positive coordination (as in our baseline model) or anticoordination.

Based on the same "fundamental attribution error"-like rationalization procedure as in Proposition 4, our finding that assortativity neglect is a coherent misperception generalizes: For any ANE, there is a coherent ANE with the same behavior.[46] In comparing Nash and ANE behavior, a complication is that general best response functions may feature equilibrium multiplicity. One setting that allows one to side-step this issue are additively separable best-responses with only global coordination incentives, $\mathrm{BR}_\theta(s,P) = \theta + \psi(H^{s,P})$. Note that in this case all Nash equilibria are equally dispersive (in terms of the dispersive order). Moreover, if ψ is FOSD-increasing (positive coordination incentives), then generalizing the logic in Proposition 1, it is straightforward to show that every monotone ANE is more dispersive than any Nash equilibrium, as higher types perceive higher global coordination motives. By contrast, if ψ is FOSD-decreasing (anticoordination), then assortativity neglect has the opposite implication and every monotone ANE is less dispersive than any Nash.

[46] To see this, let s denote the strategy profile at some ANE. For each θ,θ', set $\hat{s}_\theta(\theta') := b(\theta', H_\theta^{s,P}, H_\theta^{s,P})$ and $\hat{F}_\theta(\theta') := H_\theta^{s,P}(\hat{s}_\theta(\theta'))$. Then $H_\theta^{\hat{s}_\theta,\hat{P}_\theta}(a) = \hat{F}_\theta(\hat{s}_\theta^{-1}(a)) = H_\theta^{s,P}(a)$, ensuring observational consistency. Under $\hat{C}_\theta = C_I$, this implies that θ's perceived local and global action distributions both equal $H_\theta^{s,P}$. Thus, by construction $\hat{s}_\theta(\theta') = \mathrm{BR}_{\theta'}(\hat{s}_\theta,\hat{P}_\theta)$, ensuring coherency. Finally, s satisfies the best-response requirement, as the original ANE construction implies $s(\theta) = b(\theta, H_\theta^{s,P}, H_\theta^{s,P})$.

5.1.4 Weaker rationalization/coherency notions

Our coherency notion (Definition 3) requires each agent θ to be able to use her *own* perceptions $\hat{s}_\theta, \hat{P}_\theta$ to rationalize the behavior $\hat{s}_\theta(\theta')$ she attributes to any other agent θ'. This suggests a simple and dogmatic worldview whereby θ believes that all other types θ' share her perceptions $\hat{s}_\theta$ and $\hat{P}_\theta$ and best-respond to them. One might consider weakening this requirement to allow θ to hold higher-order beliefs about other agents' perceptions. To accommodate this, Appendix D.3 adapts rationalizable conjectural equilibrium (RCE) (Rubinstein and Wolinsky, 1994; Esponda, 2013) to our setting: Here θ believes that θ' best-responds to possibly different perceptions than her own, as long as these perceptions can in turn rationalize the behavior that θ believes θ' to observe, where θ' may again believe her matches to hold different perceptions from her own, and so on.[47]

Several of our main insights extend to RCE. First, as RCE is a special case of Definition 2 but more permissive than coherent equilibrium, assortativity neglect trivially remains sustainable and leads to the same behavior. Second, we show that the finding that other misperceptions of C are not sustainable in all environments (Proposition 7) and that any given environment may feature upper bounds on perceived assortativity (Example 5) also remains valid. However, whereas the coherent ANE is unique, RCE with assortativity neglect need not pin down agents' perceptions. Thus, coherency can be viewed as a refinement of RCE that captures misperceptions that are especially simple and internally consistent (i.e., don't require agents to contemplate higher-order belief disagreement), while also having the virtue of yielding sharp predictions and comparative statics for agents' perceived type distributions.

5.2 Related Literature

This paper contributes to a growing literature in behavioral game theory that employs variants of self-confirming equilibrium to capture agents who, due to misinferences from limited observational feedback, hold incorrect beliefs about opponents' behavior.[48] While in our setting agents neglect selection effects arising from assortativity in society, a number of papers consider agents who neglect selection due to missing feedback about non-implemented projects/transactions, e.g., in settings of adverse selection (Esponda, 2008), voting (Esponda and Pouzo, 2017), or

[47] Fudenberg and Kamada (2015) introduce related solution concepts for extensive form games. Lipnowski and Sadler (2018) consider RCE in which the observation structure is parametrized by a network; in their setting the network does not affect game payoffs, and they assume that the true network structure and agents' types are common knowledge.

[48] Esponda and Pouzo's (2016) Berk-Nash equilibrium captures steady-state behavior in general games with misspecified players. While Berk-Nash allows agents' beliefs not to perfectly match observed feedback, we follow self-confirming equilibrium in requiring observational consistency. See also Spiegler (2016, 2017).

investment (Jehiel, 2018).[49] A related inferential bias, "correlation neglect," underlies cursed equilibrium (Eyster and Rabin, 2005) and analogy-based expectation equilibrium (Jehiel, 2005; Jehiel and Koessler, 2008), where agents are correct about the type distribution in their population but misperceive the correlation between others' types and actions; these solution concepts reduce to Bayes-Nash equilibrium in the static private-value environment of this paper.[50] Chauvin (2018) proposes an equilibrium model of discrimination: Agents belong to observable groups and each group generates an outcome distribution that depends jointly on members' individual traits and on population beliefs about the group, but others' beliefs about each group are based on misinferences that attribute observed outcomes purely to members' traits. This misinference procedure relates to the "fundamental attribution error" that arises in our setting as an implication of agents' rationalization of assortativity neglect.[51] None of the aforementioned papers consider the question of coherency/rationalization of agents' misperceptions.

While we analyze the equilibrium implications of assortativity neglect for coordination games in fixed societies, other recent papers consider the effect of related selection biases on endogenous sorting. Levy and Razin (2017) study the coevolution of sorting into different school types and beliefs about school quality: agents' beliefs are shaped by communicating with school peers while ignoring selection into schools. They characterize when polarized beliefs about school quality are sustained in the long run. Windsteiger (2018) considers steady-state sorting into income classes when agents directly observe their peers' incomes but underestimate income differences across classes; she shows that this misperception reduces demand for redistribution. Our result that assortativity neglect is a coherent misperception in all societies can be viewed as a conceptual foundation for examining its implications in the aforementioned models.

As noted, assortativity neglect can be seen as a form of "projection bias," where agents project their own (perceived) match distributions onto other agents' match distributions. Other work considers agents who project their tastes onto others, e.g., in the context of auctions (Breitmoser, 2018) and social learning (Gagnon-Bartsch, 2017; Bohren and Hauser, 2018). Madarász (2012, 2016) considers information projection, where agents overestimate the similarity of others' information. Unlike our paper, these models do not impose observational consistency and do not consider the coherency of agents' misperceptions.

Our analysis of linear best response coordination games in a fixed society relates to the

[49] Osborne and Rubinstein (1998, 2003) consider agents who observe finite but unbiased samples of others' actions and draw misinferences due to neglecting sampling noise, while our focus is on agents whose samples are biased but infinite.

[50] Non-strategic models of selection/correlation neglect include, e.g., Streufert (2000); DeMarzo, Vayanos, and Zwiebel (2003); Glaeser and Sunstein (2009); Ortoleva and Snowberg (2015); Levy and Razin (2015); Ellis and Piccione (2017).

[51] See Kaneko and Matsui (1999) for a related model of discrimination based on inductive game theory. Ettinger and Jehiel (2010) formalize a notion of "fundamental attribution error" in a bargaining setting.

literature on static network games.[52] Several papers consider network games with random matching/incomplete information about the network structure, employing Bayes-Nash equilibrium or dynamic adjustment processes, e.g., Jackson and Yariv (2007); Galeotti, Goyal, Jackson, Vega-Redondo, and Yariv (2010) (for a survey, see Section 3.5 of Jackson and Zenou, 2013).[53] Two recent papers relate more closely to our focus on agents' misperceptions of interaction patterns: Jackson (2018) studies implications of the "friendship paradox,"[54] showing that this leads Nash equilibrium actions under local interactions to be higher than under uniform global interactions. He also analyzes naive agents who behave as in the local interaction case even though utilities depend on uniform global interactions. The effect of the friendship paradox is absent in our setting, because we do not model degree heterogeneity; instead we focus on type heterogeneity and assortativity, which is absent in Jackson's model. Battigalli, Panebianco, and Pin (2018) study self-confirming equilibrium in network games with a focus on learning dynamics and perceived centrality. Our analysis of the coherency/rationalization of agents' misperceptions and its implications for agents' perceived type distributions has no counterpart in either Jackson (2018) or Battigalli, Panebianco, and Pin (2018).

6 Concluding Remarks

This paper analyzes the behavioral implications of assortativity neglect, based on a relaxation of Nash equilibrium that captures agents whose perceptions are derived only from their assortative local interactions. We show that ssortativity neglect exacerbates action dispersion in coordination games, both in any fixed society and by amplifying the impact of social changes (increased assortativity and type heterogeneity). We also highlight why assortativity neglect might be an especially persistent form of misperception: A rationalization procedure reminiscent of the fundamental attribution error makes it possible for agents to coherently sustain assortativity neglect in any environment, whereas other misperceptions are not always coherent. Finally, the model provides a lens through which to understand how empirically documented misperceptions of population characteristics (e.g., income or political attitude distributions) may be shaped by the nature of agents' social interactions.

Our results have implications for agents' welfare under misperception: Appendix E.3 shows that whether or not assortativity neglect makes agents worse off depends on whether θ's payoffs are interpreted *objectively* (i.e., according to the true behavior in society under ANE), as

[52] In dynamic local interaction games, it has been highlighted that the effect of the network structure can be exaggerated when agents are myopic rather than forward-looking (e.g., Section 3.5 of Özgür, 2011). While reminiscent of our multiplier effect, the mechanism is quite different (forward-looking agents are less sensitive to local behavioral changes as they anticipate that others will have a chance to change their actions).

[53] The literature on learning in networks often employs updating heuristics in the style of DeGroot (1974) that do not require agents to form beliefs about the network structure; see Golub and Sadler (2016) for a survey.

[54] This refers to the mathematical fact that people's neighbors on average have higher degrees than themselves.

might be appropriate if utilities are viewed as material payoffs; or *subjectively* (according to θ's perception of behavior), as might fit a psychological interpretation of utilities. Objectively, assortativity neglect always Pareto-decreases welfare relative to Nash, but subjective welfare can be lower or higher than Nash depending on fundamentals.

Finally, while the paper has focused on linear best-response coordination games (and their generalizations in Section 5.1.3), we note that assortativity neglect might be of relevance in other population games, both static (e.g., discrete-action games such as political protests) and dynamic (e.g., social learning; see Section 7.2 of Frick, Iijima, and Ishii, 2019).

Appendix: Main Proofs

This appendix presents our main proofs (Appendix A–B). The supplementary online appendix presents omitted proofs and additional results (Appendix C–E).

A Preliminaries

A.1 Operator T_C induced by interaction structure C

Many of our proofs make use of a particular operator T_C over the the space of inverse cdfs that is induced by any interaction structure C. To define this, let L^1 be the space of all measurable functions $f : (0,1) \to \mathbb{R}$ such that $\int_0^1 |f(x)|dx < \infty$, endowed with the L^1 norm. Let $\mathcal{I} \subseteq L^1$ denote the subset consisting of absolutely continuous[55] and weakly increasing functions. For each cdf $F \in \mathcal{F}$, we have that F^{-1} is strictly increasing, absolutely continuous and that $\int_0^1 |F^{-1}(x)|dx = \int |\theta| dF(\theta) < \infty$, so that $F^{-1} \in \mathcal{I}$. Conversely, for any $f \in \mathcal{I}$ that is strictly increasing, we have $f^{-1} \in \mathcal{F}$.

Given any interaction structure C, define the operator T_C over L^1 by

$$T_C f(x) = \int_0^1 f(y)\, dC(y|x)$$

for all $f \in L^1$. If $C \in \mathcal{C}$ with density c, then we can write $T_C f(x) = \int_0^1 c(y,x) f(y) dy$ for all $f \in L^1$. The following lemma records some basic properties of T_C that we invoke without reference from now on.

Lemma A.1. *Fix any $C \in \mathcal{C}$. Then T_C is a continuous operator from L^1 to L^1 with the following properties:*

1. *$\|T_C f\| \leq \|f\|$ for each $f \in L^1$.*

[55] That is, for any $x, x' \in (0,1)$, there is an integrable function f' such that $f(x) = f(x') + \int_{x'}^{x} f'(y)dy$.

2. $T_C f \in \mathcal{I}$ *for any* $f \in \mathcal{I}$.

3. *For any* $\gamma \in [0,1)$ *and* $f \in L^1$,

$$\lim_{\tau \to \infty} \sum_{t=0}^{\tau} \gamma^t (T_C)^t f = \sum_{t=0}^{\infty} \gamma^t (T_C)^t f \in L^1,$$

where $(T_C)^t$ *is defined by* $(T_C)^0(f) := f$ *and* $(T_C)^{t+1}(f) := (T_C)^t(T_C f)$ *for all* f *and* t.

Proof. See Online Appendix C.1. □

A.2 Properties of the mean-preserving spread order over $\mathcal{I}$

Define a binary relation $\succsim_m$ over $\mathcal{I}$ by setting $f \succsim_m g$ if and only if $\int_0^1 \phi(f(x))dx \geq \int_0^1 \phi(g(x))dx$ for all convex functions ϕ such that $\phi \circ f, \phi \circ g \in L^1$. Note that for F, $G \in \mathcal{F}$, F is a mean-preserving spread of G if and only if $F^{-1} \succsim_m G^{-1}$. The following characterization of $\succsim_m$ is standard (e.g., Section 3.A.1 in Shaked and Shanthikumar, 2007):

Lemma A.2. *Let* $f, g \in \mathcal{I}$. *Then the following are equivalent:*

1. $f \succsim_m g$

2. $\int_y^1 f(x)dx \geq \int_y^1 g(x)dx$ *for all* $y \in (0,1)$, *with equality for* $y = 0$.

We say that a preorder (i.e., reflexive and transitive binary relation) $\succsim$ over $\mathcal{I}$ is ***linear*** if for any $f, g, h \in \mathcal{I}$ and $\alpha_1, \alpha_2 > 0$, we have $f \succsim g$ if and only if $\alpha_1 f + \alpha_2 h \succsim \alpha_1 g + \alpha_2 h$; ***continuous*** if for any $f_n \to f \in \mathcal{I}, g_n \to g \in \mathcal{I}$ with $f_n \succsim g_n$ for each n, we have $f \succsim g$; and ***isotone*** if $f \succsim g$ implies $T_C f \succsim T_C g$ for any $C \in \mathcal{C}$.

The following lemma verifies that order $\succsim_m$ over $\mathcal{I}$ satisfies these three basic properties:

Lemma A.3. $\succsim_m$ *is a preorder over* $\mathcal{I}$ *that is linear, continuous, and isotone.*

Proof. See Online Appendix C.1. □

Finally, we show that $(T_C)^t f$ is $\succsim_m$-decreasing in t:

Lemma A.4. $(T_C)^t f \succsim_m (T_C)^{t+1} f$ *for all* $t \geq 0$, $C \in \mathcal{C}$ *and* $f \in \mathcal{I}$.

Proof. See Online Appendix C.1. □

B Proofs

B.1 Proof of Lemma 1

Fix any (P, γ, β) with $P = (F, C)$ and let $\mu := \mathbb{E}_F[\theta]$. Since $F \in \mathcal{F}$, $F^{-1} \in \mathcal{I}$ with F^{-1} strictly increasing. Define

$$h(x) := (1-\gamma-\beta)\sum_{t\geq 0}\gamma^t (T_C)^t F^{-1}(x) + \frac{\beta\mu}{1-\gamma}$$

for each $x \in (0,1)$. Note that by construction, $h = (1-\gamma-\beta)F^{-1} + \gamma T_C h + \beta T_{C_I} h$. Moreover, h is strictly increasing, since $(T_C)^t F^{-1}$ is weakly increasing for each $t \geq 0$ and strictly increasing for $t = 0$. Note also that for each t, $(T_C)^t F^{-1} \in \mathcal{I}$ and hence there exists $(T_C^t F^{-1})' : (0,1) \to \mathbb{R}_+$ such that $(T_C)^t F^{-1}(x) - (T_C)^t F^{-1}(x') = \int_{x'}^{x} (T_C^t F^{-1})'(y)dy$ for all $x > x'$. Thus, h is absolutely continuous as

$$h(x) - h(x') = \lim_{\tau\to\infty} \sum_{t=0}^{\tau} \int_{x'}^{x} (1-\gamma-\beta)\gamma^t (T_C^t F^{-1})'(y)dy$$
$$= \lim_{\tau\to\infty} \int_{x'}^{x} (1-\gamma-\beta) \sum_{t=0}^{\tau} \gamma^t (T_C^t F^{-1})'(y)dy = \int_{x'}^{x} (1-\gamma-\beta) \sum_{t\geq 0} \gamma^t (T_C^t F^{-1})'(y)dy$$

where the last equality holds by the monotone convergence theorem.

Let $s(\theta) := h(F(\theta))$ for each $\theta \in \Theta$. Since $h \in L^1$, we have $\int |s(\theta)|dF(\theta) = \int |h(x)|dx < \infty$, so that s is L^1. Moreover, s inherits strict monotonicity and absolute continuity (by the change of variable theorem) from h and F. Finally, s is a Nash equilibrium because for each type θ and $x = F(\theta)$, we have

$$s(\theta) = h(x) = (1-\gamma-\beta)F^{-1}(x) + \gamma T_C h(x) + \beta T_{C_I} h(x) = (1-\gamma-\beta)\theta + \gamma \mathbb{E}_P[s(\theta') \mid \theta] + \beta \mathbb{E}_F[s(\theta')].$$

To show uniqueness of equilibrium, consider any Nash equilibrium $\hat{s}$. Define $\hat{h}(x) := \hat{s}(F^{-1}(x))$ for each x. By the best-response condition for $\hat{s}$, we have

$$\hat{h} = (1-\gamma-\beta)F^{-1} + \gamma T_C \hat{h} + \beta T_{C_I} \hat{h} \tag{13}$$

Iterating (13) yields

$$\hat{h} = (1-\gamma-\beta)F^{-1} + \beta T_{C_I}\hat{h} + \gamma T_C\left((1-\gamma-\beta)F^{-1} + \beta T_{C_I}\hat{h}\right) + \gamma^2 (T_C)^2 \hat{h} = \dots$$
$$= (1-\gamma-\beta)\sum_{t=0}^{\tau} \gamma^t (T_C)^t \left(F^{-1} + \beta T_{C_I}\hat{h}\right) + \gamma^{\tau+1}(T_C)^{\tau+1}\hat{h}$$

for all $\tau \in \mathbb{N}$. Since we also have $h = (1-\gamma-\beta)F^{-1} + \gamma T_C h + \beta T_{C_I} h$, the analogous iteration

holds for h. Thus,

$$\|\hat{h}-h\| \leq (1-\gamma-\beta)\|\sum_{t=0}^{\tau}\gamma^t(T_C)^t\left(F^{-1}-\beta T_{C_I}\hat{h}-F^{-1}+\beta T_{C_I}h\right)\|+\gamma^{\tau+1}\|(T_C)^{\tau+1}(\hat{h}-h)\|$$
$$\leq (1-\gamma-\beta)\|\sum_{t=0}^{\tau}\gamma^t(T_C)^t\left(\beta T_{C_I}(h-\hat{h})\right)\|+\gamma^{\tau+1}\|\hat{h}-h\|,$$

which converges to $(1-\gamma-\beta)\|\sum_{t=0}^{\infty}\gamma^t(T_C)^t\left(\beta T_{C_I}(h-\hat{h})\right)\|$ as $\tau\to\infty$. But integrating both sides of (13) with respect to x, we obtain $\int_0^1\hat{h}(x)dx = T_{C_I}\hat{h}(y)=\mu$ for each y, and analogously $T_{C_I}h(y)=\mu$ from the best-response condition for h. Thus, $\|\hat{h}-h\|=0$, whence $\hat{s}=s$. □

B.2 Proof of Proposition 1

Fix any (P,γ,β). As shown in the main text, the strategy profile s^{AN} in any ANE must satisfy equation (6). This implies that s^{AN} is the Nash equilibrium at (P,γ',β'), where $\gamma'=\gamma+\beta$ and $\beta'=0$. Thus, s^{AN} must satisfy (7) by Lemma 1.

To verify the existence of ANE, for each θ, first take an arbitrary $\hat{F}_\theta\in\mathcal{F}$. Then choose any $\hat{s}_\theta(\cdot)$ such that the observational consistency condition $H_\theta^{s^{AN},P}=\hat{H}_\theta^{\hat{s}_\theta,\hat{P}_\theta}$ holds. For example, this can be achieved by setting $\hat{s}_\theta(\theta')=(H_\theta^{s^{AN},P})^{-1}(1-\hat{F}_\theta(\theta'))$ for each θ'.[56] Given observational consistency, (6) ensures that the best response condition is satisfied.

To show that action distribution H^{AN} is more dispersive than the Nash action distribution H^{NE}, it suffices to show that $s^{AN}(\theta)-s^{AN}(\theta')\geq s^{NE}(\theta)-s^{NE}(\theta')$ for all $\theta>\theta'$. But observe that for all τ, the monotonicity of process P implies $\mathbb{E}_P[\theta_t|\theta_0=\theta]\geq\mathbb{E}_P[\theta_t|\theta_0=\theta']$, whence

$$0\leq(1-\beta-\gamma)\sum_{t=0}^{\tau}((\gamma+\beta)^t-\gamma^t)\left(\mathbb{E}_P[\theta_t|\theta_0=\theta]-\mathbb{E}_P[\theta_t|\theta_0=\theta']\right).$$

By (5) and (7), the RHS converges to $(s^{AN}(\theta)-s^{AN}(\theta'))-(s^{NE}(\theta)-s^{NE}(\theta'))$ as $\tau\to\infty$, proving the desired claim. □

B.3 Proof of Proposition 2

We first show that the more-assortative order $\succsim_{MA}$ is the "dual order" of the mean-preserving spread order $\succsim_m$:

Lemma B.1. *Fix any $C_1,C_2\in\mathcal{C}$. Then $C_1\succsim_{MA}C_2$ if and only if $T_{C_1}F^{-1}\succsim_m T_{C_2}F^{-1}$ for all $F\in\mathcal{F}$.*

[56] To highlight that ANE puts little restriction on perceived strategy profiles, this construction defines a $\hat{s}_\theta$ that is decreasing. In general, any perceived strategy profile $\hat{s}_\theta$ such that defining $\hat{F}_\theta(\theta'):=H_\theta^{s^{AN},P}(\hat{s}_\theta((-\infty,\theta']))$ for all θ' yields a perceived type distribution $\hat{F}_\theta\in\mathcal{F}$, can be sustained as part of an ANE.

Proof. First observe that $C_1 \succsim_{MA} C_2$ if and only if $C_1(\cdot|x \geq y)$ first-order stochastically dominates $C_2(\cdot \mid x \geq y)$ for any $y \in (0,1)$. This is because

$$\begin{aligned} C_1 \succsim_{MA} C_2 &\iff \int_0^z \int_0^y c_1(x',x)dxdx' \geq \int_0^z \int_0^y c_2(x',x)dxdx' \ \forall y,z \in (0,1) \\ &\iff \int_0^z \int_y^1 c_1(x',x)dxdx' \leq \int_0^z \int_y^1 c_2(x',x)dxdx' \ \forall y,z \in (0,1) \\ &\iff C_1(z \mid x \geq y) \leq C_2(z \mid x \geq y) \ \forall y,z \in (0,1) \end{aligned}$$

where the second line uses $\int_0^z \int_0^1 c_i(x',x)dxdx' = \int_0^1 \int_0^z c_i(x',x)dx'dx = z$ for each $i = 1,2$.

Next note that for any $F \in \mathcal{F}$, $T_{C_1}F^{-1} \succsim_m T_{C_2}F^{-1}$ holds if and only if $\int_y^1 T_{C_1}f(x)dx \geq \int_y^1 T_{C_2}f(x)dx$ for all $y \in (0,1)$ with equality if $y = 0$. But

$$\begin{aligned} &\int_y^1 T_{C_1}f(x)dx \geq \int_y^1 T_{C_2}f(x)dx, \ \ \forall y \in (0,1) \\ \iff &\int_y^1 \int_0^1 c_1(x',x)f(x')dx'dx \geq \int_y^1 \int_0^1 c_2(x',x)f(x')dx'dx, \ \ \forall y \in (0,1) \\ \iff &\int_0^1 \int_y^1 \tfrac{1}{1-y} c_1(x',x)dx f(x')dx' \geq \int_0^1 \int_y^1 \tfrac{1}{1-y} c_2(x',x)dx f(x')dx', \ \ \forall y \in (0,1). \end{aligned}$$

Since the set of all F^{-1} with $F \in \mathcal{F}$ consists of all L^1, strictly increasing and absolutely continuous functions on $(0,1)$, this implies that $T_{C_1}F^{-1} \succsim_m T_{C_2}F^{-1}$ holds for all $F \in \mathcal{F}$ if and only if $C_1(\cdot \mid x \geq y)$ first-order stochastically dominates $C_2(\cdot \mid x \geq y)$ for any $y \in (0,1)$. By the first paragraph, this is equivalent to $C_1 \succsim_{MA} C_2$. □

Proof of Proposition 2. **(1) ⇒ (2)**: Suppose that $C_1 \succsim_{MA} C_2$ and consider any F, γ, β. Let $f := F^{-1}$, which is in $\mathcal{I}$ since $F \in \mathcal{F}$. We first show by induction that $(T_{C_1})^t f \succsim_m (T_{C_2})^t f$ for all t. For $t = 1$ this is true by Lemma B.1. Suppose the claim holds for some $t \geq 1$. Then

$$(T_{C_1})^{t+1} f = T_{C_1} (T_{C_1})^t f \succsim_m T_{C_1} (T_{C_2})^t f \succsim_m T_{C_2} (T_{C_2})^t f = (T_{C_2})^{t+1} f,$$

where the first comparison follows from the inductive hypothesis by isotonicity of $\succsim_m$, and the second one holds by Lemma B.1. Thus, by transitivity of $\succsim_m$, we have $(T_{C_1})^{t+1} f \succsim_m (T_{C_2})^{t+1} f$, as required.

Next, note that linearity of $\succsim_m$ and $C_1 \succsim_{MA} C_2$ implies

$$\begin{aligned} &\sum_{t=0}^{\tau} \gamma^t (T_{C_1})^t F^{-1} \succsim_m \left(\gamma^\tau (T_{C_2})^\tau + \sum_{t=0}^{\tau-1} \gamma^t (T_{C_1})^t \right) F^{-1} \succsim_m \\ \succsim_m &\left(\sum_{t=\tau-1}^{\tau} \gamma^t (T_{C_2})^t + \sum_{t=0}^{\tau-2} \gamma^t (T_{C_1})^t \right) F^{-1} \succsim_m \cdots \succsim_m \sum_{t=0}^{\tau} \gamma^t (T_{C_2})^t F^{-1} \end{aligned}$$

for any $\tau \in \mathbb{N}$. Moreover, by Lemma A.1, as $\tau \to \infty$, we have

$$\sum_{t=0}^{\tau} \gamma^t (T_{C_1})^t F^{-1} \to \sum_{t=0}^{\infty} \gamma^t (T_{C_1})^t F^{-1}, \quad \sum_{t=0}^{\tau} \gamma^t (T_{C_2})^t F^{-1} \to \sum_{t=0}^{\infty} \gamma^t (T_{C_2})^t F^{-1}.$$

Thus, by continuity and linearity of $\succsim_m$, we have

$$(1-\gamma-\beta) \sum_{t=0}^{\infty} \gamma^t (T_{C_1})^t F^{-1} + \frac{\beta\mu}{1-\gamma} \succsim_m (1-\gamma-\beta) \sum_{t=0}^{\infty} \gamma^t (T_{C_2})^t F^{-1} + \frac{\beta\mu}{1-\gamma},$$

where $\mu = \mathbb{E}_F[\theta]$. Thus, H_1^{NE} is a mean-preserving spread of H_2^{NE} at (F, γ, β).
(2) ⇒ (3): Immediate from the fact that the ANE action distribution at (P, γ, β) coincides with the Nash action distribution at $(P, \gamma + \beta, 0)$.
(3) ⇒ (1): Let $h_{F,C_i,\gamma,\beta}$ denote the inverse of the ANE action distribution at (F, C_i, γ, β). Suppose $h_{F,C_1,\gamma,\beta} \succsim_m h_{F,C_2,\gamma,\beta}$ for all (F, γ, β). Setting $f := F^{-1}$ and $\delta := \gamma + \beta$, we have

$$(1-\delta) \sum_{t\geq 0} \delta^t T_{C_1} f = h_{F,C_1,\gamma,\beta} \succsim_m h_{F,C_2,\gamma,\beta} = (1-\delta) \sum_{t\geq 0} \delta^t T_{C_2} f.$$

By linearity of $\succsim_m$ and since $(T_{C_i})^0(f) = f$ for $i = 1, 2$, this implies

$$T_{C_1} f + \sum_{t\geq 2} \delta^t (T_{C_1})^t f \succsim_m T_{C_2} f + \sum_{t\geq 2} \delta^t (T_{C_2})^t f. \tag{14}$$

Note that for each $i = 1, 2$,

$$\|T_{C_i} f + \sum_{t\geq 2} \delta^t (T_{C_i})^t f - T_{C_i} f\| \leq \sum_{t\geq 2} \delta^t \|(T_{C_i})^t f\| \leq \sum_{t\geq 2} \delta^t \|f\|$$

so that, as $\delta \to 0$, $T_{C_i} f + \sum_{t\geq 2} \delta^t (T_{C_i})^t f \to T_{C_i} f$. Thus, by continuity of $\succsim_m$, (14) yields $T_{C_1} f \succsim T_{C_2} f$. As this is true for all $f = F^{-1}$, we have $C_1 \succsim_{MA} C_2$ by Lemma B.1.
(1) ⇔ (4): We first show that (1) implies (4). By the proof of "(1) ⇒ (2)," we have $(T_{C_1})^t F^{-1} \succsim_m (T_{C_2})^t F^{-1}$ for all t. Thus,

$$\left((\gamma+\beta)^t (T_{C_1})^t + \gamma^t (T_{C_2})^t\right) F^{-1} \succsim_m \left(\gamma^t (T_{C_1})^t + (\gamma+\beta)^t (T_{C_2})^t\right) F^{-1},$$

as $(\gamma+\beta)^t \geq \gamma^t \geq 0$ and by linearity of $\succsim_m$. Then linearity and continuity of $\succsim_m$ also imply

$$\begin{aligned} &(1-\gamma-\beta) \textstyle\sum_{t=0}^{\infty} \left((\gamma+\beta)^t (T_{C_1})^t + \gamma^t (T_{C_2})^t\right) F^{-1} + \frac{\beta \mathbb{E}_F[\theta]}{1-\gamma} \\ &\succsim_m (1-\gamma-\beta) \textstyle\sum_{t=0}^{\infty} \left(\gamma^t (T_{C_1})^t + (\gamma+\beta)^t (T_{C_2})^t\right) F^{-1} + \frac{\beta \mathbb{E}_F[\theta]}{1-\gamma}. \end{aligned}$$

By monotonicity of equilibrium strategies, this yields for all θ^* that

$$\mathbb{E}_F[s_1^{AN}(\theta) + s_2^{NE}(\theta)|\theta \geq \theta^*] \geq \mathbb{E}_F[s_1^{NE}(\theta) + s_2^{AN}(\theta)|\theta \geq \theta^*],$$

which is equivalent to the first inequality in (4). The second inequality follows from (2), which is implied by (1) as we have seen above.

To see that (4) implies (1), note that the second inequality in (4) implies (2). Thus (1) follows from the above proofs. □

B.4 Proof of Proposition 4

The following characterization of C_I is used in the proofs of both Propositions 4 and 7.

Lemma B.2. *For any $\hat{C} \in \mathcal{C}$, $\gamma > 0$ and $\beta \geq 0$, the following are equivalent:*

1. *$\hat{C} = C_I$.*
2. *For any action distribution $\hat{H} \in \mathcal{F}$, there exists a type distribution $\hat{F}$ such that $\hat{H}$ is the Nash action distribution at $(\hat{F}, \hat{C}, \gamma, \beta)$.*

Moreover, if (2) holds then for every $\hat{H}$ the corresponding $\hat{F}$ is unique.

Proof. Fix any $\hat{C} \in \mathcal{C}$, $\gamma > 0$ and $\beta \geq 0$. Note first that by the proof of Lemma 1, for any action distribution $\hat{H} \in \mathcal{F}$ and type distribution $\hat{F} \in \mathcal{F}$, $\hat{H}$ is the Nash action distribution at $(\hat{F}, \hat{C}, \gamma, \beta)$ if and only if

$$\hat{H}^{-1}(x) = (1-\gamma-\beta)\hat{F}^{-1}(x) + \gamma \int \hat{H}^{-1}(z) d\hat{C}(z|x) + \beta \int \hat{H}^{-1}(z)dz \text{ for all } x \in (0,1). \quad (15)$$

(1) ⇒ (2): Suppose $\hat{C} = C_I$ and consider any action distribution $\hat{H} \in \mathcal{F}$. Define a type distribution $\hat{F}$ by $\hat{F}^{-1}(x) := \frac{1}{1-\gamma-\beta}\left(\hat{H}^{-1}(x) - (\gamma+\beta)\int \hat{H}^{-1}(z)dz\right)$ for each x. Since $\hat{H} \in \mathcal{F}$, it follows that $\hat{F}^{-1}$ is L^1, strictly increasing, and absolutely continuous, so that $\hat{F} \in \mathcal{F}$. Moreover, $\hat{C} = C_I$ implies $\int \hat{H}^{-1}(z)d\hat{C}(z|x) = \int \hat{H}^{-1}(z)dz$, so by (15), $\hat{H}$ is the Nash action distribution at $(\hat{F}, \hat{C}, \gamma, \beta)$.

(2) ⇒ (1): We prove the contrapositive. Suppose that $\hat{C} \neq C_I$. We will construct an action distribution $\hat{H}$ for which there is no type distribution $\hat{F}$ such that $\hat{H}$ is the Nash action distribution at $(\hat{F}, \hat{C}, \gamma, \beta)$. To do so, note first that since $\hat{C}$ is assortative and $\hat{C} \neq C_I$, there exists $x \in (0,1)$ such that $\hat{C}(x \mid z)$ is weakly decreasing and non-constant in z. Thus, there exist $y, y' \in (0,1)$ with $y > y'$ such that $\hat{C}(x \mid y) < \hat{C}(x \mid y')$. Moreover, we can assume that either (i) $y > y' \geq x$ or (ii) $x > y > y'$.[57]

[57] Indeed, either there exists $y > x$ such that $\hat{C}(x|y) > \hat{C}(x|x)$, in which case (i) holds setting $y' = x$. Or $\hat{C}(x|z) \leq \hat{C}(x|x)$ for all z, in which case the fact that $\hat{C}(x|z)$ is non-constant in z yields some $y' < x$ such that $\hat{C}(x|y') < \hat{C}(x|x)$ and continuity yields $y \in (y', x)$ such that $\hat{C}(x|y') < \hat{C}(x|y)$; thus, (ii) is satisfied.

Let $\mathbb{1}_{[x,1)}$ denote the indicator function on $[x,1)$, and let $(\hat{H}_n)_n$ be a sequence of action distributions in $\mathcal{F}$ such that $(\hat{H}_n^{-1})_n$ is uniformly bounded and $\hat{H}_n^{-1}(z) \to \mathbb{1}_{[x,1)}(z)$ for all $z \in (0,1)$.[58] Then, under both (i) and (ii), as $n \to \infty$, we have that $\hat{H}_n^{-1}(y) - \hat{H}_n^{-1}(y') \to 0$ and $\int \hat{H}_n^{-1}(z)\, d\left(\hat{C}(z|y) - \hat{C}(z|y')\right) \to \hat{C}(x|y') - \hat{C}(x|y) > 0$. Hence, setting $\hat{H} = \hat{H}_n$ for some sufficiently large n, we have

$$\hat{H}^{-1}(y) - \hat{H}^{-1}(y') < \gamma \int \hat{H}^{-1}(z)\, d\left(\hat{C}(z|y) - \hat{C}(z|y')\right). \tag{16}$$

But then, if $\hat{H}$ were the Nash action distribution under $(\hat{F}, \hat{C}, \gamma, \beta)$ for some type distribution $\hat{F}$, the best-response condition (15) would imply that

$$(1-\gamma-\beta)\left(\hat{F}^{-1}(y) - \hat{F}^{-1}(y')\right) = \hat{H}^{-1}(y) - \hat{H}^{-1}(y') - \gamma \int \hat{H}^{-1}(z)\, d\left(\hat{C}(z|y) - \hat{C}(z|y')\right) < 0.$$

This is impossible since $y > y'$.

For the moreover part, if $\hat{H}$ is the Nash action distribution at $(\hat{F}, C_I, \gamma, \beta)$, then by (15), $\hat{H}^{-1}(x) = (1-\gamma-\beta)\hat{F}^{-1}(x) + (\gamma+\beta)\int \hat{H}^{-1}(z)\, dz$ for all x. This uniquely pins down $\hat{F}$. □

Proof of Proposition 4. Observe that $(s, (\hat{s}_\theta, \hat{F}_\theta, C_I)_{\theta\in\Theta})$ is a coherent ANE if and only if (i) s is given by (7) (by Proposition 1), (ii) for each θ, $H_\theta^{s,P}$ is the Nash equilibrium action distribution under $(\hat{F}_\theta, C_I)$ and $\hat{s}_\theta$ is the corresponding strategy profile.

Therefore, by Lemma B.2, it suffices to show that $H_\theta^{s^{AN},P} \in \mathcal{F}$ for each θ where s^{AN} is defined by (7). Note that s^{AN} coincides with the Nash equilibrium at (P, γ', β'), where $\gamma' = \gamma + \beta$ and $\beta' = 0$. Thus, $(s^{AN})^{-1}$ is strictly increasing and absolutely continuous. By monotonicity of s^{AN}, $H_\theta^{s^{AN},P}(a) = P((s^{AN})^{-1}(a)|\theta)$ for each $a \in s^{AN}(\Theta)$. Since $P(\theta'|\theta)$ is absolutely continuous and strictly increasing in θ' on Θ, $H_\theta^{s^{AN},P}(a)$ is absolutely continuous (by the change of variable theorem) and strictly increasing in $a \in s^{AN}(\Theta)$.

Since s^{AN} is L^1 with respect to F, $\int\int |s^{AN}(\theta')| dP(\theta'|\theta) dF(\theta) = \int |s^{AN}(\theta)| dF(\theta) < \infty$. Thus, there exists $\Theta^* \subseteq \Theta$ such that $\Theta \setminus \Theta^*$ has Lebesgue measure zero and for every $\theta \in \Theta^*$, $\int |s^{AN}(\theta')| dP(\theta'|\theta) < \infty$. Hence, $H_\theta^{s^{AN},P}$ is L^1 for all $\theta \in \Theta^*$. As $H_\theta^{s^{AN},P}$ is FOSD-monotonic in θ, this implies that $H_\theta^{s^{AN},P}$ is L^1 for every $\theta \in \Theta$.[59] □

[58] Concretely, we can define $\hat{H}_n$ by $\hat{H}_n^{-1}(z) := \begin{cases} \frac{n-1}{n}\left(\frac{z}{x}\right)^n & \text{if } z \le x \\ \frac{1}{n(1-x)}(z-1)+1 & \text{if } z > x. \end{cases}$

[59] Indeed, take any $\theta \in \Theta \setminus \Theta^*$. If $\theta \in (\inf\Theta, \sup\Theta)$, pick $\theta', \theta'' \in \Theta^*$ with $\theta' < \theta < \theta''$. Then $\int |a| dH_\theta^{s^{AN},P}(a) = -\int_{-\infty}^0 a dH_\theta^{s^{AN},P}(a) + \int_0^\infty a dH_\theta^{s^{AN},P}(a) \le -\int_{-\infty}^0 a dH_{\theta'}^{s^{AN},P}(a) + \int_0^\infty a dH_{\theta''}^{s^{AN},P}(a) < \infty$. If $\theta = \sup\Theta$ (the case $\theta = \inf\Theta$ is analogous), then $\mathrm{supp} H_\theta^{s^{AN},P}$ is bounded above. Thus, $\int |a| dH_\theta^{s^{AN},P}(a) = -\int_{-\infty}^0 a dH_\theta^{s^{AN},P}(a) + \int_0^\infty a dH_\theta^{s^{AN},P}(a) \le -\int_{-\infty}^0 a dH_{\theta'}^{s^{AN},P}(a) + \int_0^\infty a dH_\theta^{s^{AN},P}(a) < \infty$ for any $\theta' \in \Theta^*$.

B.5 Proof of Proposition 7

Fix any θ and any regular $\hat{C} \in \mathcal{C}$ with $\hat{C} \neq C_I$. Fix an arbitrary $\gamma > 0$ and let $\beta = 0$. We will find a society (F, C) such that no coherent equilibrium at (F, C, γ, β) satisfies $\hat{C}_\theta = \hat{C}$.

Note first that by regularity of $\hat{C}$, there exists $z \in (0,1)$ such that $\{x \in (0,1) : \hat{C}(x|x) = z\} = \{\hat{x}\}$ for some $\hat{x}$. Moreover, since $\hat{C}$ admits a positive density on $(0,1)^2$, we have that $\hat{C}(\cdot \mid \hat{x})$ is a bijection from $(0,1)$ to $(0,1)$. Let $\hat{C}_{\hat{x}}^{-1}$ denote its inverse. As in the proof of the "(2) $\implies$ (1)" direction of Lemma B.2, we can find an action distribution $\hat{H} \in \mathcal{F}$ and some $y > y'$ such that

$$\hat{H}^{-1}(y) - \hat{H}^{-1}(y') < \gamma \int \hat{H}^{-1}(w)\, d\left(\hat{C}(w|y) - \hat{C}(w|y')\right). \tag{17}$$

Note that setting $\tilde{H}^{-1}(x) := \hat{H}^{-1}(\hat{C}_{\hat{x}}^{-1}(x))$ for all $x \in (0,1)$ defines another action cdf $\tilde{H} \in \mathcal{F}$. Moreover, by the "(1) $\implies$ (2)" direction of Lemma B.2, there exists a type distribution $\tilde{F} \in \mathcal{F}$ such that $\tilde{H}$ is the Nash action distribution under $(\tilde{F}, C_I, \gamma, \beta)$. Since $\tilde{F} \in \mathcal{F}$, there exists some $\nu \in \mathbb{R}$ such that $\tilde{F}(\theta + \nu) = z$. Define a new type distribution F by $F(\theta') := \tilde{F}(\theta' + \nu)$ for each θ'. Then the Nash action distribution H at (F, C_I, γ, β) satisfies

$$H^{-1}(x) = \tilde{H}^{-1}(x) + \nu \text{ for all } x \in (0,1). \tag{18}$$

We claim that in environment (F, C_I, γ, β), there exists no coherent equilibrium with $\hat{C}_\theta = \hat{C}$. Indeed, suppose for a contradiction that there is a coherent equilibrium with $\hat{C}_\theta = \hat{C}$, where the corresponding true strategy profile is s and θ's perceived type distribution and perceived strategy profile are $\hat{F}_\theta$ and $\hat{s}_\theta$. Since $\beta = 0$, the strategy profile s at this equilibrium coincides with the Nash equilibrium profile (see Lemma D.4). Hence, the true action distribution is given by H. Let $\hat{G}_\theta = \hat{F}_\theta \circ \hat{s}_\theta^{-1}$ denote θ's perceived global action distribution. We will derive a contradiction from the coherency requirement that $\hat{G}_\theta$ is the Nash action distribution at $(\hat{F}_\theta, \hat{C}, \gamma, \beta)$.

To this end, note first that θ's perceived type quantile $\hat{F}_\theta(\theta)$ is given by $\hat{x}$. Indeed, θ's true quantile is $F(\theta) = z$. Hence, under the true interaction structure C_I, the fraction of θ's matches that play an action below $s(\theta)$ is $C_I(z|z) = z$. But by observational consistency, θ must be correct about this fraction; that is, θ must believe fraction z of her matches to have types below θ, i.e., $\hat{C}(\hat{F}_\theta(\theta)|\hat{F}_\theta(\theta)) = z$. By choice of z, this implies $\hat{F}_\theta(\theta) = \hat{x}$.

Given this, we have the following relationship between θ's perceived local and global action distributions $H_\theta^{\hat{s}_\theta, \hat{P}_\theta}$ and $\hat{G}_\theta$:

$$H_\theta^{\hat{s}_\theta, \hat{P}_\theta}{}^{-1}(x) = \hat{G}_\theta^{-1}(\hat{C}_{\hat{x}}^{-1}(x)) \text{ for all } x. \tag{19}$$

Indeed, since θ perceives her own quantile to be $\hat{x}$ and the interaction structure to be $\hat{C}$, θ

perceives the xth quantile among her matches to correspond to quantile $\hat{C}_{\hat{x}}^{-1}(x)$ in the overall population. Hence, θ's perception of the xth action quantile among θ's matches (i.e, the LHS) is the same as θ's perception of the $\hat{C}_{\hat{x}}^{-1}(x)$th action quantile in the overall population (i.e., the RHS).

Moreover, since the true interaction structure is C_I, θ's true local action distribution $H_\theta^{s,P}$ is given by the global action distribution H. Hence, by observational consistency, we have $H_\theta^{\hat{s}_\theta,\hat{P}_\theta} = H$. Combining this with (19) yields that $\hat{G}_\theta^{-1}(\hat{C}_{\hat{x}}^{-1}(x)) = H^{-1}(x)$ for all x; or equivalently (substituting $x = \hat{C}(q|\hat{x})$) that

$$\hat{G}_\theta^{-1}(q) = H^{-1}(\hat{C}(q|\hat{x})) = \tilde{H}^{-1}(\hat{C}(q|\hat{x})) + \nu = \hat{H}^{-1}(q) + \nu \text{ for all } q \in (0,1),$$

where the final two equalities follow from (18) and the definition of $\tilde{H}$ above. Combining this with equation (17), we have that $\hat{G}_\theta^{-1}(y) - \hat{G}_\theta^{-1}(y')$ is equal to

$$\hat{H}^{-1}(y) - \hat{H}^{-1}(y') < \gamma \int \hat{H}^{-1}(w)\, d\left(\hat{C}(w|y) - \hat{C}(w|y')\right) = \gamma \int \hat{G}_\theta^{-1}(w)\, d\left(\hat{C}(w|y) - \hat{C}(w|y')\right).$$

Since $\hat{G}_\theta$ is the Nash action distribution at $(\hat{F}_\theta, \hat{C}, \gamma, \beta)$, the best-response condition (15) then implies that $\hat{F}_\theta^{-1}(y) < \hat{F}^{-1}(y')$. This contradicts the fact that $y > y'$. □

References

Ahler, D. J. (2014): "Self-fulfilling misperceptions of public polarization," *The Journal of Politics*, 76(3), 607–620.

Alesina, A., and P. Giuliano (2011): "Preferences for redistribution," in *Handbook of social economics*, vol. 1, pp. 93–131. Elsevier.

Aliprantis, C. D., and K. Border (2006): *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer.

Angeletos, G.-M., and A. Pavan (2007): "Efficient use of information and social value of information," *Econometrica*, 75(4), 1103–1142.

Barron, K., S. Huck, and P. Jehiel (2019): "Everyday econometricians: Selection neglect and overoptimism when learning from others," Discussion paper, Working Paper.

Battigalli, P. (1987): "Comportamento razionale ed equilibrio nei giochi e nelle situazioni sociali," *unpublished undergraduate dissertation, Bocconi University, Milano*.

Battigalli, P., F. Panebianco, and P. Pin (2018): "Learning and Selfconfirming Equilibria in Network Games," working paper.

Bauman, K. P., and G. Geher (2002): "We think you agree: The detrimental impact of the false consensus effect on behavior," *Current Psychology*, 21(4), 293–318.

Bénabou, R. (1993): "Workings of a city: location, education, and production," *The Quarterly Journal of Economics*, 108(3), 619–652.

——— (1996a): "Equity and efficiency in human capital investment: the local connection," *The Review of Economic Studies*, 63(2), 237–264.

——— (1996b): "Heterogeneity, stratification, and growth: macroeconomic implications of community structure and school finance," *The American Economic Review*, pp. 584–609.

Bergemann, D., and S. Morris (2013): "Robust Predictions in Games with Incomplete Information," *Econometrica*, 81, 1251–1308.

Bishop, B. (2009): *The big sort: Why the clustering of like-minded America is tearing us apart*. Houghton Mifflin Harcourt.

Bohren, J. A., and D. Hauser (2018): "Social Learning with Model Misspecification: A Framework and a Robustness Result," working paper.

Bond, R., and P. B. Smith (1996): "Culture and conformity: A meta-analysis of studies using Asch's (1952b, 1956) line judgment task.," *Psychological bulletin*, 119(1), 111.

Botvin, G. J., E. M. Botvin, E. Baker, L. Dusenbury, and C. J. Goldberg (1992): "The false consensus effect: Predicting adolescents' tobacco use from normative expectations," *Psychological reports*, 70(1), 171–178.

Brands, R. A. (2013): "Cognitive social structures in social network research: A review," *Journal of Organizational Behavior*, 34(S1).

Breitmoser, Y. (2018): "Knowing me, imagining you: Projection and overbidding in auctions," *Games and Economic Behavior*.

Chauvin, K. (2018): "A Misattribution Theory of Discrimination," working paper.

Chetty, R., N. Hendren, and L. F. Katz (2016): "The effects of exposure to better neighborhoods on children: New evidence from the Moving to Opportunity experiment," *American Economic Review*, 106(4), 855–902.

Cruces, G., R. Perez-Truglia, and M. Tetaz (2013): "Biased perceptions of income distribution and preferences for redistribution: Evidence from a survey experiment," *Journal of Public Economics*, 98, 100–112.

Cutler, D. M., and A. Lleras-Muney (2010): "Understanding differences in health behaviors by education," *Journal of health economics*, 29(1), 1–28.

Daley, D. (1968): "Stochastically monotone Markov chains," *Probability Theory and Related Fields*, 10(4), 305–317.

Dawtry, R. J., R. M. Sutton, and C. G. Sibley (2015): "Why wealthier people think people are wealthier, and why it matters: From social sampling to attitudes to redistribution," *Psychological science*, 26(9), 1389–1400.

DeGroot, M. H. (1974): "Reaching a consensus," *Journal of the American Statistical Association*, 69(345), 118–121.

DeMarzo, P. M., D. Vayanos, and J. Zwiebel (2003): "Persuasion bias, social influence, and unidimensional opinions," *The Quarterly journal of economics*, 118(3), 909–968.

Dessi, R., E. Gallo, and S. Goyal (2016): "Network cognition," *Journal of Economic Behavior & Organization*, 123, 78–96.

Durlauf, S. N. (1996): "A theory of persistent income inequality," *Journal of Economic growth*, 1(1), 75–93.

——— (2004): "Neighborhood effects," in *Handbook of regional and urban economics*, vol. 4, pp. 2173–2242. Elsevier.

Ellis, A., and M. Piccione (2017): "Correlation misperception in choice," *American Economic Review*, 107(4), 1264–92.

Engelhardt, C., and A. Wagener (2015): "Biased perceptions of income inequality and redistribution," mimeo.

Enke, B. (2017): "What You See is All There Is," mimeo.

Esponda, I. (2008): "Behavioral equilibrium in economies with adverse selection," *The American Economic Review*, 98(4), 1269–1291.

——— (2013): "Rationalizable conjectural equilibrium: A framework for robust predictions," *Theoretical Economics*, 8(2), 467–501.

Esponda, I., and D. Pouzo (2016): "Berk–Nash Equilibrium: A Framework for Modeling Agents With Misspecified Models," *Econometrica*, 84(3), 1093–1130.

——— (2017): "Conditional Retrospective Voting in Large Elections," *American Economic Journal: Microeconomics*, 9(2), 54–75.

Esponda, I., and E. Vespa (2014): "Hypothetical thinking and information extraction in the laboratory," *American Economic Journal: Microeconomics*, 6(4), 180–202.

——— (2018): "Endogenous sample selection: A laboratory study," *Quantitative Economics*, 9(1), 183–216.

Ettinger, D., and P. Jehiel (2010): "A theory of deception," *American Economic Journal: Microeconomics*, 2(1), 1–20.

Eyster, E., and M. Rabin (2005): "Cursed equilibrium," *Econometrica*, 73(5), 1623–1672.

Fernández, R., and R. Rogerson (1996): "Income distribution, communities, and the quality of public education," *The Quarterly Journal of Economics*, 111(1), 135–164.

——— (2001): "Sorting and long-run inequality," *The Quarterly Journal of Economics*, 116(4), 1305–1341.

Frick, M., R. Iijima, and Y. Ishii (2019): "Misinterpreting Others and the Fragility of Social Learning," Cowles Foundation Discussion Paper # 2160.

Fudenberg, D., and Y. Kamada (2015): "Rationalizable partition-confirmed equilibrium," *Theoretical Economics*, 10(3), 775–806.

Fudenberg, D., and D. K. Levine (1993): "Self-confirming equilibrium," *Econometrica*, pp. 523–545.

Gagnon-Bartsch, T. (2017): "Taste Projection in Models of Social Learning," Discussion paper, working paper.

Galeotti, A., S. Goyal, M. O. Jackson, F. Vega-Redondo, and L. Yariv (2010): "Network games," *The review of economic studies*, 77(1), 218–244.

Gimpelson, V., and D. Treisman (2018): "Misperceiving inequality," *Economics & Politics*, 30(1), 27–54.

Glaeser, E. L., and C. R. Sunstein (2009): "Extremism and social learning," *Journal of Legal Analysis*, 1(1), 263–324.

Golub, B., and S. Morris (2017): "Expectations, Networks, and Conventions," working paper.

Golub, B., and E. Sadler (2016): "Learning in Social Networks," in *The Oxford Handbook of the Economics of Networks*.

Gorodnichenko, Y., and G. Roland (2011): "Which dimensions of culture matter for long-run growth?," *American Economic Review*, 101(3), 492–98.

Hairer, M., and J. C. Mattingly (2011): "Yet Another Look at Harris' Ergodic Theorem for Markov Chains," in *Seminar on Stochastic Analysis, Random Fields and Applications VI*, pp. 109–117. Springer.

Hauser, O. P., and M. I. Norton (2017): "(Mis) perceptions of inequality," *Current opinion in psychology*, 18, 21–25.

Hofstede, G. (2001): *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations*. Sage publications.

Huber, G. A., and N. Malhotra (2017): "Political homophily in social relationships: Evidence from online dating behavior," *The Journal of Politics*, 79(1), 269–283.

Jackson, M. O. (2018): "The friendship paradox and systematic biases in perceptions and social norms," *Journal of Political Economy*, forthcoming.

Jackson, M. O., and L. Yariv (2007): "Diffusion of behavior and equilibrium properties in network games," *The American economic review P&P*, 97(2), 92–98.

Jackson, M. O., and Y. Zenou (2013): "Games on Networks," in *Handbook of Game Theory*, ed. by P. H. Young, and S. Zamir, vol. 4. Elsevier Science.

Jargowsky, P. A. (1996): "Take the money and run: Economic segregation in US metropolitan areas," *American sociological review*, pp. 984–998.

Jehiel, P. (2005): "Analogy-based expectation equilibrium," *Journal of Economic Theory*, 123(2), 81–104.

——— (2018): "Investment strategy and selection bias: An equilibrium perspective on overoptimism," *American Economic Review*, 108(6), 1582–97.

Jehiel, P., and F. Koessler (2008): "Revisiting games of incomplete information with analogy-based expectations," *Games and Economic Behavior*, 62(2), 533–557.

Joe, H. (1997): *Multivariate models and multivariate dependence concepts*. CRC Press.

Kaneko, M., and A. Matsui (1999): "Inductive game theory: discrimination and prejudices," *Journal of Public Economic Theory*, 1(1), 101–137.

Lawrence, E., J. Sides, and H. Farrell (2010): "Self-segregation or deliberation? Blog readership, participation, and polarization in American politics," *Perspectives on Politics*, 8(1), 141–157.

Levy, G., and R. Razin (2015): "Correlation neglect, voting behavior, and information aggregation," *American Economic Review*, 105(4), 1634–45.

——— (2017): "The Coevolution of Segregation, Polarized Beliefs, and Discrimination: The Case of Private versus State Education," *American Economic Journal: Microeconomics*, 9(4), 141–70.

Lipnowski, E., and E. D. Sadler (2018): "Peer-Confirming Equilibrium," *Econometrica*, forthcoming.

Madarász, K. (2012): "Information projection: Model and applications," *The Review of Economic Studies*, 79(3), 961–985.

——— (2016): "Projection equilibrium: Definition and applications to social investment and persuasion," .

Marks, G., and N. Miller (1987): "Ten years of research on the false-consensus effect: An empirical and theoretical review.," *Psychological bulletin*, 102(1), 72.

Meltzer, A. H., and S. F. Richard (1981): "A rational theory of the size of government," *Journal of Political Economy*, 89(5), 914–927.

Meyer, M., and B. Strulovici (2012): "Increasing interdependence of multivariate distributions," *Journal of Economic Theory*, 147(4), 1460–1489.

Morris, S., and H. S. Shin (2002): "Social value of public information," *American Economic Review*, 92(5), 1521–1534.

Musterd, S. (2005): "Social and ethnic segregation in Europe: levels, causes, and effects," *Journal of Urban Affairs*, 27(3), 331–348.

Niehues, J. (2014): "Subjective perceptions of inequality and redistributive preferences: An international comparison," working paper.

Ortoleva, P., and E. Snowberg (2015): "Overconfidence in political behavior," *American Economic Review*, 105(2), 504–35.

Osborne, M. J., and A. Rubinstein (1998): "Games with procedurally rational players," *American Economic Review*, pp. 834–847.

——— (2003): "Sampling equilibrium, with an application to strategic voting," *Games and Economic Behavior*, 45(2), 434–441.

Özgür, O. (2011): "Local interactions," in *Handbook of Social Economics*, vol. 1, pp. 587–644. Elsevier.

Pampel, F. C., P. M. Krueger, and J. T. Denney (2010): "Socioeconomic disparities in health behaviors," *Annual review of sociology*, 36, 349–370.

Pin, P., and B. Rogers (2016): "Stochastic Network Formation and Homophily," in *The Oxford Handbook of the Economics of Networks*.

Quillian, L., and H. Lagrange (2016): "Socioeconomic segregation in large cities in France and the United States," *Demography*, 53(4), 1051–1084.

Reardon, S. F., and K. Bischoff (2011): "Income inequality and income segregation," *American Journal of Sociology*, 116(4), 1092–1153.

Ross, L. (1977): "The Intuitive Psychologist And His Shortcomings: Distortions in the Attribution Process1," in *Advances in experimental social psychology*, vol. 10, pp. 173–220. Elsevier.

Ross, L., D. Greene, and P. House (1977): "The "false consensus effect:" An egocentric bias in social perception and attribution processes," *Journal of experimental social psychology*, 13(3), 279–301.

Rubinstein, A., and A. Wolinsky (1994): "Rationalizable conjectural equilibrium: between Nash and rationalizability," *Games and Economic Behavior*, 6(2), 299–311.

Ryff, J. (1963): "On the representation of doubly stochastic operators," *Pacific Journal of Mathematics*, 13(4), 1379–1386.

Sacerdote, B. (2011): "Peer effects in education: How might they work, how big are they and how much do we know thus far?," in *Handbook of the Economics of Education*, vol. 3, pp. 249–277. Elsevier.

Samet, D. (1998): "Iterated expectations and common priors," *Games and economic Behavior*, 24(1-2), 131–141.

Shaked, M., and G. Shanthikumar (2007): *Stochastic orders*. Springer.

Siahpush, M., A. McNeill, D. Hammond, and G. Fong (2006): "Socioeconomic and country variations in knowledge of health risks of tobacco smoking and toxic constituents of smoke: Results from the 2002 International Tobacco Control (ITC) Four Country Survey," *Tobacco control*, 15(suppl 3), iii65–iii70.

Spiegler, R. (2016): "Bayesian networks and boundedly rational expectations," *The Quarterly Journal of Economics*, 131(3), 1243–1290.

——— (2017): ""Data Monkeys": A Procedural Model of Extrapolation from Partial Statistics," *The Review of Economic Studies*, 84(4), 1818–1841.

Streufert, P. (2000): "The effect of underclass social isolation on schooling choice," *Journal of Public Economic Theory*, 2(4), 461–482.

Wan, W. (2017): "America's new tobacco crisis: The rich stopped smoking, the poor didn't," *The Washington Post*, June 13, 2017.

Westfall, J., L. Van Boven, J. R. Chambers, and C. M. Judd (2015): "Perceiving political polarization in the United States: Party identity strength and attitude extremity exacerbate the perceived partisan divide," *Perspectives on Psychological Science*, 10(2), 145–158.

Windsteiger, L. (2018): "The Redistributive Consequences of Segregation and Misperceptions," working paper.