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

DISPERSED BEHAVIOR AND PERCEPTIONS IN ASSORTATIVE SOCIETIES

By

Mira Frick, Ryota Iijima, and Yuhta Ishii

April 2018

COWLES FOUNDATION DISCUSSION PAPER NO. 2128

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

Motivated by the fact that people's perceptions of their societies are routinely incorrect, we study the possibility and implications of misperception in social interactions. We focus on coordination games with *assortative* interactions, where agents with higher types (e.g., wealth, political attitudes) are more likely than lower types to interact with other high types. Assortativity creates scope for misperception, because what agents observe in their local interactions need not be representative of society as a whole. To model this, we define a tractable solution concept, "local perception equilibrium" (LPE), that describes possible behavior and perceptions when agents' beliefs are derived only from their local interactions. We show that there is a unique form of misperception that can persist in *any* environment: This is *assortativity neglect*, where all agents believe the people they interact with to be a representative sample of society as a whole. Relative to the case with correct perceptions, assortativity neglect generates two mutually reinforcing departures: A "false consensus effect," whereby agents' perceptions of average characteristics in the population are increasing in their own type; and more "dispersed" behavior in society, which adversely affects welfare due to increased miscoordination. Finally, we propose a comparative notion of when one society is more assortative than another and show that more assortative societies are characterized precisely by greater action dispersion and a more severe false consensus effect, and as a result, greater assortativity has an ambiguous effect on welfare.

*This version: April 18, 2018. Frick: Yale University (mira.frick@yale.edu); Iijima: Yale University (ryota.iijima@yale.edu); Ishii: Centro de Investigación Económica, ITAM (yuhta.ishii@itam.mx). We thank Nageeb Ali, Attila Ambrus, Dirk Bergemann, Drew Fudenberg, Simone Galperti, Wayne Yuan Gao, Helena Garcia, Ben Golub, Andrei Gomberg, Sanjeev Goyal, Marina Halac, Matt Jackson, Navin Kartik, Daria Khromenkova, David Miller, Weicheng Min, Stephen Morris, Romans Pancs, Jacopo Perego, Larry Samuelson, Rani Spiegler, Philipp Strack, Linh To, as well as seminar audiences at ASU Economic Theory Conference, Berkeley, Cambridge, Cornell, Duke, Glasgow, ITAM, Kyoto, Michigan, NSF Conference on Network Science in Economics (Vanderbilt), Rochester, St. Andrews, UCLA, UT Austin, and Yale.

1 Introduction

1.1 Motivation and Overview

People's perceptions of the societies they live in are routinely and substantially incorrect. An active empirical literature spanning the social sciences documents systematic biases in individuals' perceived distributions of key population characteristics such as income or political attitudes.[1] Additionally, in social psychology and more recently economics, empirical research on "network cognition" finds systematic discrepancies between people's perceived and actual "interaction structures," i.e., the patterns of who interacts with whom.[2] While a large theoretical literature highlights the impact of social structures on individual behavior and welfare (for surveys, see Goyal, 2012; Jackson, 2010; Vega-Redondo, 2007), this literature has thus far assumed that agents possess a perfect understanding of their societies. In this paper, we take a first step toward analyzing the possibility and implications of misperceptions in social interactions.

Our starting point is the fact that many social interactions are *assortative*, in the sense that people interact more often with others with similar characteristics.[3] Thus, wealthier people are in general more likely than poorer people to have wealthy friends; and conservatives are more likely than liberals to interact with other conservatives. Assortativity creates scope for misperception, because what individuals observe in their local interactions need not be representative of society as a whole. Understanding the ramifications of this is made all the more relevant by evidence that societies are growing increasingly assortative.[4] Our analysis seeks to answer the following questions: What kinds of misperceptions can persist in assortative societies; what are the implications for behavior and welfare; and how are these affected by the extent of assortativity in society?

To address these questions, we introduce a tractable model of assortative local interactions. A society consists of a large population of agents with linearly ordered types (e.g., income or political attitudes), 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 are engaged in a coordination game, where players' best response functions (e.g., consumption decisions or political activities) are increasing in their own type, as well as in expected behavior among their matches and in society as a whole (e.g., due to peer effects or adherence to social norms). Two key modeling contributions allow us to investigate the effect of assortativity on behavior and perceptions: First, we introduce a solution concept, "local perception equilibrium," that describes possible behavior and (mis)perceptions of society when agents' beliefs are derived

[1] See, e.g., Cruces, Perez-Truglia, and Tetaz (2013); Norton and Ariely (2011) for income; and Ahler (2014); Westfall, Van Boven, Chambers, and Judd (2015) for political attitudes. We discuss concrete findings below.

[2] In economics, see Dessi, Gallo, and Goyal (2016); Breza, Chandrasekhar, and Tahbaz-Salehi (2018). Relevant work in social psychology includes Krackhardt (1987); Kumbasar, Rommey, and Batchelder (1994); Krackhardt and Kilduff (1999); Kilduff, Crossland, Tsai, and Krackhardt (2008); see Brands (2013) for a survey.

[3] See Pin and Rogers (2016) for a survey and discussion of potential sources of assortativity, such as institutional constraints determining meeting opportunities or socio-psychological factors (e.g., homophily).

[4] See, e.g., Jargowsky (1996); Reardon and Bischoff (2011) for studies of increased residential segregation by income in the US, and Bishop (2009); Lawrence, Sides, and Farrell (2010); Huber and Malhotra (2017) for evidence of growing online and offline segregation by political attitudes.

only from observing their matches' behavior. Second, we develop comparative notions that capture when one society is "more assortative" than another.

Our main findings are threefold. First, there is a unique form of misperception of the interaction structure that can persist in *any* environment: This is *assortativity neglect*, i.e., the belief that all agents interact with a representative sample of society. Second, assortativity neglect generates two mutually reinforcing departures from the correct perceptions benchmark: A "false consensus effect," whereby agents' perceptions of average characteristics in the population are increasing in their own type; and more "dispersed" behavior in society, which adversely affects welfare due to increased miscoordination. Finally, we show that more assortative societies are characterized precisely by greater action dispersion and a more severe false consensus effect, and as a result, greater assortativity has an ambiguous effect on welfare.

Our analysis proceeds in two steps: We first examine Nash equilibrium behavior, i.e., assuming correct perceptions. This serves as a benchmark and stepping stone for our analysis of local perception equilibrium, but also constitutes an independent contribution to the literature on network/local interaction games under correct perceptions, which has so far abstracted away from preference heterogeneity and assortativity and focused instead on implications of other societal features such as degree distributions and centrality measures (see the discussion of related literature in Section 1.2).

Our game admits a unique Nash equilibrium, which we characterize in terms of a particular monotone Markov process over the type space (see the penultimate paragraph of the introduction). This enables us to show that the extent of assortativity in society manifests itself in a simple way: More assortative societies correspond precisely to more dispersed behavior. To capture greater assortativity, we define an order over the *copulas* associated with any society (see Section 2.2) that reflects that people in top quantiles of the population (e.g., in the top 5% of the wealth distribution) are less likely to interact with bottom quantiles. Greater action dispersion, in the sense of mean-preserving spread, captures the idea that the gap between higher and lower types' actions increases; for example, if wealthier people step up their consumption of luxury goods or if left and right-leaning individuals each engage in more partisan political activities. Using tools from majorization theory (Marshall, Olkin, and Arnold, 2010), Theorem 1 shows that one society is more assortative than another if and only if it gives rise to more dispersed behavior regardless of the type distribution and coordination game. At the same time, we also highlight two alternative sources of action dispersion: Stronger coordination motives and increased type dispersion.

We next incorporate the possibility of misperception. A *local perception equilibrium* (LPE) in a given society consists of a true strategy profile along with a perceived society and perceived strategy profile for each agent, subject to three requirements: Each agent (i) correctly observes the distribution of actions among his matches; (ii) holds perceptions that "rationalize" these observations; and (iii) best-responds to his perceptions. Thus, unlike Nash equilibrium, where each agent best-responds to the correct beliefs about the underlying society and behavior, LPE allows agents to be wrong about both. However, by (i), these misperceptions must be *observationally consistent*, in the sense that they imply the correct distribution of matches' actions. Thus, LPE captures misperceptions that are persistent, because agents' beliefs are never contradicted by what they observe in their local interactions.

What kinds of misperceptions can arise under LPE? Focusing first on perceived interaction structures, Theorem 2 shows that there is a particularly simple form of misperception that can be sustained in LPE in *any* society and strategic environment. This is *assortativity neglect*, i.e., the belief that all agents interact with a representative sample of society. Moreover, assortativity neglect is the *only* perception of the interaction structure that can arise in any environment; any other perceived interaction structure is inconsistent with LPE in some settings. A key observation driving this "robust sustainability" of assortativity neglect is the following fundamental feature (Lemma 3): Unlike any other misperception, assortativity neglect can explain *any* observed behavior; specifically, agents can rationalize others' behavior by attributing all action dispersion to type dispersion, ignoring the fact that variation across other agents' behavior may also be driven by differences in their matches' behavior. Lemma 3 can be viewed as providing a microfoundation for the "fundamental attribution error" (Ross, 1977), a central bias studied in social psychology that refers to people's tendency to attribute others' behavior to intrinsic characteristics rather than to external factors such as social influence.

We next show that in any given environment, being able to sustain assortativity neglect in LPE uniquely determines agents' behavior and their perceptions of population characteristics. As a result, assortativity neglect both offers a unified explanation for existing empirical findings and provides several new predictions:

First, we highlight two mutually reinforcing implications of assortativity neglect: Perceptions of population characteristics are subject to the aforementioned "false consensus effect" and behavior is more dispersed than under correct perceptions. Going back to Ross, Greene, and House (1977), the fact that people's perceptions of others' attributes are positively correlated with their own has been documented empirically in a wide variety of settings. For example, individuals' perceptions of the median income in their society are substantially increasing in own income, and supporters (opponents) of a particular political position tend to overestimate (underestimate) its support in the population.[5] In our setting, this effect arises in equilibrium due to the fact that higher types tend to observe higher actions and under assortativity neglect, falsely attribute this to a higher type mean in the population. Moreover, we show that this increases action dispersion, by exacerbating the difference between higher and lower types' coordination incentives. As a result, higher types observe an even higher distribution of actions, further amplifying the false consensus effect.

Second, we demonstrate how agents' misperceptions about population characteristics are shaped by the nature of their social interactions. In particular, while under assortativity neglect (as under Nash) increased assortativity and coordination motives both give rise to more dispersed behavior, they have opposing effects on perceptions: The false consensus effect is exacerbated in more assortative societies, but it is less pronounced when coordination motives are stronger. Moreover, depending on the relative strength of coordination motives and assortativity, agents can either

[5]See the references following Corollary 1.

under- or overestimate type dispersion in their society; this may help shed light on empirical evidence of both forms of misperception, e.g., widespread underestimation of income inequality but overestimation of political attitude polarization.[6]

Our analysis has important implications for welfare in assortative societies. First, misperception in the form of assortativity neglect Pareto decreases welfare relative to the Nash benchmark, both subjectively (when utilities are evaluated according to agents' perceptions) and objectively (according to correct perceptions). This is because the fact that assortativity neglect exacerbates action dispersion, increases miscoordination costs for all agents.[7] Second, increased assortativity can have an ambiguous effect on welfare: On the one hand, greater assortativity leads to less dispersed match distributions, facilitating coordination, but on the other hand, it increases action dispersion across types, which hinders coordination. We show that the latter effect dominates whenever coordination motives are sufficiently strong and is further exacerbated by assortativity neglect.

Finally, at a methodological level, a contribution of our paper is to import monotone Markov processes into the study of local interaction/network games. Even though our setting is static, every society can be viewed as inducing a discrete-time Markov process over its space of types: Starting with any type θ_0, this process first draws a match θ_1 for θ_0 according to the interaction structure, then draws type θ_1's match θ_2, and so on; crucially, assortativity corresponds to this process being *monotone* (Daley, 1968). We show that equilibrium strategies (under both Nash and LPE) can be represented as particular discounted sums of t-step ahead expectations of this process. This fact, along with monotonicity of the process, plays a key role throughout our proofs, while also helping build intuition for our results.

The remainder of the paper is structured as follows. Section 2 defines our coordination game with assortative interactions along with our comparative notion of assortativity. Section 3 studies the Nash equilibrium benchmark. Section 4 defines LPE. Section 5 analyzes LPE misperceptions and behavior, while Section 6 considers welfare implications. Section 7 discusses additional directions. Section 8 is a conclusion.

1.2 Related Literature

Our paper relates most closely to the literature on network/local interaction games (Ballester, Calvo-Armengol, and Zenou, 2006; Jackson and Yariv, 2007; Galeotti, Goyal, Jackson, Vega-Redondo, and Yariv, 2010; Bramoulle, Kranton, and DíAmours, 2014). As is common in this literature, we study coordination games with linear best-response functions. We make two main contributions:

First, we introduce a framework that allows agents to hold persistent misperceptions about interaction structures—an important departure from the standard assumption in this literature that the underlying network is common knowledge. While systematic misperceptions about interaction structures have been widely documented (see the first paragraph of the introduction), to the best of our knowledge, we are the first to provide an economic model of their origins and implications.

[6] For references, see the discussion following Proposition 4.

[7] To formalize this, we employ the widely used quadratic-loss utility specification.

We provide specific predictions for what forms of misperceptions can be sustained, in particular establishing the robust sustainability of assortativity neglect. This is consistent with experimental findings in Dessi, Gallo, and Goyal (2016) and gives rise to further well-documented misperceptions about population characteristics. Our results also show that such misperceptions can have substantial implications for behavior and welfare.

A recent paper by Jackson (2018) studies implications of the "friendship paradox,"[8] showing that this leads equilibrium actions under local interactions to be higher than under uniform global interactions. He also analyzes "naive" agents who are assumed to behave as in the local interaction case but receive utilities according to the uniform global interaction case. This form of naiveté bears some resemblance to assortativity neglect in our setting. However, beyond modeling differences, an important distinction is that while naiveté in Jackson (2018) is imposed exogenously, our approach models agents' perceptions as endogenous equilibrium objects and derives assortativity neglect as the unique robustly sustainable perception of the interaction structure. Additionally, our papers focus on different phenomena: the effect of the friendship paradox is absent in our setting, because we abstract away from degree heterogeneity; instead we focus on assortativity, which is absent in Jackson's framework.

Second, we characterize the impact of assortativity and type heterogeneity on equilibrium outcomes under both Nash and LPE. While assortativity is a significant phenomenon in many social interactions, its implications for equilibrium behavior have not been well explored; instead, the existing literature tends to study the impact of other societal features, such as degree distributions and centrality measures, under homogeneous preferences.[9] Our modeling approach, in particular our representation of interaction structures in terms of copulas, allows us to define non-parametric comparative notions of assortativity and to establish a tight connection between greater assortativity and increased action dispersion. A key observation in deriving this result is to recast the problem as a comparison of monotone Markov processes.[10] This approach should prove useful in studying network games beyond our specific context.

The idea behind LPE builds on the literature on learning-based equilibrium concepts, notably self-confirming equilibrium and its variations (Battigalli, 1987; Fudenberg and Levine, 1993; Dekel, Fudenberg, and Levine, 2004; Esponda and Pouzo, 2016). One difference is that while this literature typically focuses on incorrect beliefs about others' strategies,[11] LPE additionally treats agents' beliefs about interaction structures and type distributions as explicit equilibrium objects whose properties and comparative statics we investigate. A second key departure is that in order to be able to draw inferences from matches' observed behavior to properties of the interaction structure

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

[9] Calvó-Armengol, Patacchini, and Zenou (2009) study a model with heterogeneous agents by generalizing Ballester, Calvo-Armengol, and Zenou (2006) but do not consider assortativity or comparative statics. Golub and Jackson (2012) consider DeGroot learning dynamics in which random networks are generated according to homophily; however, agents in their setting have homogeneous preferences.

[10] Outside network economics, some previous work (e.g., Müller and Stoyan, 2002) compares monotone Markov processes in terms of first-order dominance; by contrast, our comparison results are about dispersion of higher-order expectations.

[11] Some exceptions are Fudenberg and Levine (2006) and Esponda (2008).

and type distribution, agents in LPE rationalize this observed behavior by finding perceptions under which it is Nash. This is unlike self-confirming equilibrium, where agents simply best-respond to opponents' observed behavior without forming a model that rationalizes their actions, and is closer in spirit to rationalizable conjectural equilibrium (RCE) (Rubinstein and Wolinsky, 1994; Esponda, 2013). We discuss the relationship in more detail in Section 7.1, where we show that LPE can be viewed as a tractable refinement of RCE, in particular differing from the latter in that it provides unique predictions for agents' perceived type distributions under assortativity neglect.

Finally, as mentioned in the introduction, a large empirical literature across the social sciences (that we survey throughout Section 5) documents systematic misperceptions about distributions of key population characteristics. Our theoretical framework offers sharp predictions for what kinds of misperceptions can persist and how they vary with the underlying environment: We show that the robust sustainability of assortativity neglect can provide a unified explanation of several well-documented misperceptions in this literature (e.g., the false consensus effect and over-/underestimation of type dispersion), while also yielding new testable predictions for how such misperceptions are shaped by interaction patterns and coordination motives.

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 $\theta \in \mathbb{R}$, representing, e.g., wealth levels or political attitudes on a left-right spectrum. An agent's type is his 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:[12]

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 $\theta \geq \theta'$.[13]

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. As discussed in the introduction, this property is widely documented in many social interactions, and a large literature provides foundations for assortativity based on a variety of factors including socio-psychological motivations (e.g., homophily) or institutional constraints.

Note that a society P jointly summarizes an underlying ***population***, described by the marginal type distribution $F := \text{marg}\, P$, and a ***matching technology***, which for every type θ specifies the

[12]This can be seen as a reduced form representation of a network that focuses on type/preference heterogeneity and abstracts away from degree heterogeneity, as in Morris and Shin (2005). Galeotti, Goyal, Jackson, Vega-Redondo, and Yariv (2010) and Jackson and Yariv (2007) use similar approaches, but they focus on degree heterogeneity and abstract away from type heterogeneity.

[13]$P(\cdot|\theta)$ 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.

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 measurable L^1 function[14] $s : \Theta \to A$ that specifies an action $s(\theta)$ for each type θ. When types represent wealth, actions might capture brand or quality choice in consumption decisions, such as what kind of car to drive or clothes to wear; while in the case of political attitudes, actions could represent the extent to which agents manifest support for particular positions or candidates on a day-to-day basis, e.g., by displaying yard signs or bumper stickers or posting political content on social media.

To model coordination motives, we follow much of the literature on network/local interaction games[15] by considering linear best response functions: 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

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

The first term captures that higher types have an intrinsic tendency to take higher actions; e.g., wealthier people have a greater propensity to consume luxury goods, and more fervent conservatives/liberals derive higher utility from manifesting support for conservative/liberal positions. The second term represents a local coordination motive, whereby each type θ's best response is increasing in his matches' expected behavior; this captures well-documented peer effects in consumption decisions or political activities. Finally, reflecting a global coordination motive, θ's best response is also increasing in the average action in society; this could represent social status concerns in consumption or a desire to adhere to a social norm (e.g., refraining from manifestation of extreme political positions that are considered taboo).

Note that (1) assumes that θ best responds to a correct belief about s and P, as is the case under (Bayes) Nash equilibrium, which we analyze in Section 3. In Section 4, we will introduce a solution concept, "local perception equilibrium," that allows for the possibility of misperception of s and P.[16]

With the sole exception of the welfare analysis in Section 6, the best response function (1) is all that matters for our results and the exact specification of agents' utilities is irrelevant. For concreteness, a widely used utility specification (e.g., Morris and Shin, 2002; Angeletos and Pavan, 2007; Bergemann and Morris, 2013)[17] that gives rise to (1) is the quadratic-loss utility

$$u_P(a, \theta, s) = -\mathbb{E}_P[(a - \theta - \gamma s(\theta') - \beta \mathbb{E}_F[s(\theta')])^2 \mid \theta]. \tag{2}$$

That is, agents wish to minimize the expected square loss relative to their bliss-point action $\theta +$

[14] That is, $\int |s(\theta)| dF(\theta) < \infty$.

[15] See, e.g., Jackson and Zenou (2013) for a survey.

[16] See Hopkins and Kornienko (2004), Ghiglino and Goyal (2010), Immorlica, Kranton, Manea, and Stoddard (2017) for related models of peer effects or status concerns in consumption, which however do not allow for the possibility of misperception.

[17] These papers use a continuum population global interactions framework and focus on Gaussian type distributions.

$\gamma s(\theta') + \beta \mathbb{E}_F[s(\theta')]$, which is influenced by own type, the action of their realized match, and the global average action in society.

2.2 Copula Representation and Comparing Assortativity

Recall that any society P jointly summarizes an underlying population, given by the 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 population, and vice versa. To be able to disentangle differences along these two dimensions, we will frequently make use of the following equivalent representation of societies.

The idea is to express 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). 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,

$$C(x, x') := P(F^{-1}(x), F^{-1}(x')) \tag{3}$$

for all $x, x' \in (0, 1)$, $C(x, 0) = C(0, x) := 0$ and $C(x, 1) = C(1, x) := x$ for all $x \in [0, 1]$.[18] Note that C is a (two-dimensional) ***copula***, i.e., a joint cdf over $[0, 1]^2$ with *uniform* marginals; moreover, C inherits symmetry and assortativity from P.

Definition 2. An ***interaction structure*** is a two dimensional copula C that is (i) *symmetric*: $C(x, x') = C(x', x)$ for all $x, x' \in (0, 1)$; and (ii) *assortative*: $C(\cdot|x)$ first-order stochastically dominates $C(\cdot|x')$ if $x \geq x'$.

As noted, any society induces an interaction structure via (3). Conversely, given any interaction structure C and population $F \in \mathcal{F}$, defining $P(\theta, \theta') := C(F(\theta), F(\theta'))$ for all θ, θ' yields a society. Thus, pairs (F, C) of populations $F \in \mathcal{F}$ and interaction structures C yield an equivalent representation of societies. Using the (F, C) decomposition, we can compare societies with different interaction structures but the same population, or with different populations but the same interaction structure. In the following we will move freely between the two representations.

In particular, by comparing societies in terms of their interaction structures, we can define a comparative notion that captures when one society is more assortative than another:

Definition 3. Given interaction structures C_1 and C_2, we say that C_1 is ***more assortative*** than C_2, denoted $C_1 \succsim_{MA} C_2$, if $C_1(\cdot|x \geq x^*)$ first-order stochastically dominates $C_2(\cdot|x \geq x^*)$ for all $x^* \in (0, 1)$.

Recall that assortativity of C means that the distribution $C(\cdot|x)$ of matches' quantiles is increasing in own quantile x with respect to first-order stochastic dominance. Definition 3 says that C_1 is more assortative than C_2 if this effect is stronger under C_1, in the sense that conditional on

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

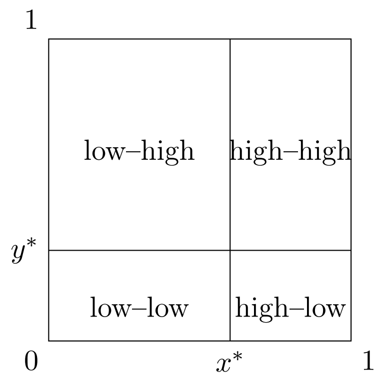

Figure 1: In the space $[0,1] \times [0,1]$ of pairs of quantiles, any interaction structure C assigns probability $(1-x^*)C(y^*|x \geq x^*)$ to the high–low region corresponding to all matches between a quantile above x^* and a quantile below y^*. Thus, $C_1 \succsim_{MA} C_2$ if and only if for all x^* and y^*, C_1 assigns lower probability than C_2 to the high–low region; equivalently, C_1 assigns higher probability than C_2 to the high–high region and low–low region and lower probability to the low–high region.

own quantile exceeding any given threshold x^*, matches' quantiles are stochastically higher under C_1 than C_2. For example, in more assortative societies, people in the top five percent of the wealth distribution are less likely to interact with those in the bottom 20%. Figure 1 provides further illustration. In the statistics literature (e.g., Joe, 1997),[19] $\succsim_{MA}$ is known as the PQD (positive quadrant dependence) or concordance order and is used more generally to compare any two-dimensional cdfs. To the best of our knowledge, we are the first to use this notion as a measure of assortativity. In Section 7.2, we introduce a second, stronger assortativity order and show that natural analogs of our main results remain valid.

The more assortative order has natural minimal and maximal elements. In particular, the $\succsim_{MA}$-least assortative interaction structure is the ***independent interaction structure*** C_I under which $C_I(x,x') = xx'$ for all x, x'; that is, the distribution $C_I(\cdot|x)$ of matches' quantiles is uniform on $[0,1]$ regardless of own quantile x, or equivalently, in any society $P = (F, C_I)$, matches' types are drawn from the unconditional type distribution F. The $\succsim_{MA}$-most assortative interaction structure is the ***perfectly assortative interaction structure*** C_{Per} given by $C_{Per}(x,x') = \min\{x,x'\}$; that is, each quantile x is matched only with types of the same quantile.

From now on, we will only consider 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.[20] We conclude the description of the model with a simple parametric example:

Example 1 (Gaussian societies)**.** Symmetric bivariate Gaussian distributions with positive corre-

[19] See also Meyer and Strulovici (2012).

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

lation represent a simple class of societies. Let P be given by the distribution of the form

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

where the correlation coefficient $\rho := \frac{\sigma_{1,2}}{\sigma^2} \in [0, 1)$ is assumed weakly positive to ensure assortativity. The corresponding population F is normally distributed with mean μ and variance σ^2. The corresponding interaction structure C is given by

$$C_\rho(x, x') := \Phi_\rho(\Phi^{-1}(x), \Phi^{-1}(x')),$$

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 $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. Note that whereas F is parametrized by μ and σ^2, C_ρ is fully parametrized by the correlation coefficient ρ. Moreover, it can be shown that $C_{\rho_1} \succsim_{MA} C_{\rho_2}$ if and only if $\rho_1 \geq \rho_2$, so that more assortativity is characterized precisely by greater correlation. □

3 Behavior in Assortative Societies: Nash Benchmark

As a stepping stone toward analyzing perceptions and behavior in assortative societies, this section abstracts away from the possibility of misperception and studies the Nash equilibrium of our game. We first observe that any society P can be associated with a particular monotone Markov process over its space of types and derive a representation of Nash strategies as a discounted sum of the t-step ahead expectations of this process. Using this observation, we establish a tight connection between assortativity and action dispersion: More assortative societies correspond precisely to more dispersed Nash behavior.

3.1 Markov Process Representation of Nash Equilibrium

While our model is static, a key observation for our analysis is the following: Any society P induces a discrete-time Markov process over its space of types Θ whose initial distribution is the type distribution $F = \text{marg}_\Theta P$ and whose transition kernel is 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 the 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 and that assortativity of P corresponds to the process being *monotone* (Daley, 1968); the latter feature will play a central role throughout the paper.

The process of t-step ahead matches yields a simple description of the (Bayes) Nash equilibrium strategy profile s of our game. Indeed, iterating the best response condition (1), s must satisfy

$$s(\theta_0) = \theta_0 + \gamma \mathbb{E}_P [s(\theta_1) \mid \theta_0] + \beta \mathbb{E}_F[s(\theta)] = \dots$$
$$= \sum_{t=0}^{\tau} \gamma^t \left(\mathbb{E}_P [\theta_t \mid \theta_0] + \beta \mathbb{E}_F[s(\theta)]\right) + \gamma^{\tau+1} \mathbb{E}_P[s(\theta_{\tau+1}) \mid \theta_0]$$

for all θ_0 and $\tau \in \mathbb{N}$. In Appendix A.2, we verify that the higher-order term $\gamma^{\tau+1}\mathbb{E}_P[s(\theta_{\tau+1})|\theta_0]$ vanishes as $\tau \to \infty$, yielding the following result:

Lemma 1. *There exists a unique Nash equilibrium. The equilibrium strategy profile s satisfies*

$$s(\theta_0) = \sum_{t=0}^{\infty} \gamma^t \mathbb{E}_P [\theta_t \mid \theta_0] + \frac{\beta \mathbb{E}_F[\theta]}{(1-\gamma)(1-\gamma-\beta)}$$

for all types θ_0, and is strictly increasing and continuous. The average action is $\mathbb{E}_F[s(\theta)] = \frac{\mathbb{E}_F[\theta]}{1-\gamma-\beta}$.

Thus, any type θ's equilibrium action is a γ-discounted sum of θ's expected t-step ahead matches under the Markov process P plus a constant that depends on F, γ, and β. The equilibrium action is increasing in θ for two reasons. First, because of the direct effect that higher types prefer higher actions; second, because assortativity means that higher types are more likely to meet other high types. The latter effect is reflected by the fact that the t-step ahead expectation $\mathbb{E}_P [\theta_t \mid \theta_0]$ is (weakly) increasing in θ_0 *for all* $t \geq 1$, which is a consequence of the monotonicity of the Markov process induced by P.[21]

3.2 Assortativity and Action Dispersion

We now show that the extent of assortativity in society manifests itself in a simple way: More assortative societies are characterized precisely by more *dispersed* behavior.

For any (P, γ, β), the ***Nash action distribution*** is the cdf $H = F \circ s^{-1}$ over actions when types are drawn according to F and behave according to the Nash strategy profile s at (P, γ, β). Throughout the paper, we use the mean-preserving spread order to capture dispersion; natural analogs of our results can also be derived under other measures of dispersion (see Section 7.2). Recall that cdf H_1 is a ***mean-preserving spread*** of cdf H_2, denoted $H_1 \succsim_m H_2$, if $\int \varphi(a)\, dH_1(a) \geq \int \varphi(a)\, dH_2(a)$ for any convex function $\varphi : \mathbb{R} \to \mathbb{R}$ for which the integrals are well-defined.

The following result establishes an equivalence between the more assortative order over interaction structures and the mean-preserving spread order over Nash action distributions.

[21] In a finite network setting with incomplete information, Golub and Morris (2017) also highlight the connection between Nash equilibria of linear best response games and higher-order expectations of a related Markov process. However, their setting does not feature assortativity/monotonicity, which is the key property driving our subsequent analysis. Relatedly, in finite network settings with complete information, equilibria of linear best-response games can also be expressed as infinite sums by iterating the underlying adjacency matrix. But again, without further assumptions on the structure of this matrix, it would be difficult to conduct comparative statics of higher-order terms and we are not aware of any work exploiting this infinite sum expression.

Theorem 1. *Fix $C_1, C_2 \in \mathcal{C}$, and let $H_i^{F,\gamma,\beta}$ denote the Nash action distribution under (F, C_i, γ, β) for each $i = 1, 2$. The following are equivalent:*

1. $C_1 \succsim_{MA} C_2$
2. $H_1^{F,\gamma,\beta} \succsim_m H_2^{F,\gamma,\beta}$ *for all* (F, γ, β).

Proof. See Appendix B. □

The equivalence result in the theorem shows a tight connection between assortativity and action dispersion. Not only does greater assortativity lead to greater action dispersion (by 1. ⇒ 2.), but greater action dispersion is indeed the defining feature of more assortative societies (by 2. ⇒ 1.). Note that $H_1^{F,\gamma,\beta}$ and $H_2^{F,\gamma,\beta}$ have the same mean, since by Lemma 1 the average Nash action does not depend on the interaction structure. Moreover, since the underlying type distributions are the same, $H_1^{F,\gamma,\beta} \succsim_m H_2^{F,\gamma,\beta}$ can be expressed directly in terms of strategies as $\mathbb{E}_F[s_1^{F,\gamma,\beta}(\theta)|\theta \geq \theta^*] \geq \mathbb{E}_F[s_2^{F,\gamma,\beta}(\theta)|\theta \geq \theta^*]$ and $\mathbb{E}_F[s_1^{F,\gamma,\beta}(\theta)|\theta \leq \theta^*] \leq \mathbb{E}_F[s_2^{F,\gamma,\beta}(\theta)|\theta \leq \theta^*]$ for all θ^*. That is, more assortative societies are characterized by the fact that on average high and low types both take more extreme actions; for example, wealthier people may consume more luxury goods, or left-/right-leaning individuals may each engage in more partisan political activities.

To see the intuition, consider the effect of greater assortativity on agents' best responses against some *fixed* monotone strategy profile s. In this case, greater assortativity implies that agents with higher types are more likely to face matches with higher types and hence, since s is monotone, are more likely to face higher actions. As a result, local coordination motives ($\gamma \geq 0$) will induce higher types to choose higher actions. Likewise, greater assortativity leads lower types to best-respond to s with lower actions. Thus, the distribution of best responses against s is more dispersed. Note the importance of local coordination motives for this argument; indeed, if $\gamma = 0$, greater assortativity has no effect as best responses do not depend on the interaction structure C.

Of course, the intuition in the previous paragraph is incomplete, because it concerns only one direction of the theorem and applies only to best responses against a fixed monotone strategy profile (instead of to equilibrium behavior, i.e., the fixed point of the best response correspondence). To establish the full-fledged equivalence result in Appendix B, we exploit the Markov process representation of equilibrium strategies from Lemma 1. This allows us to reduce the problem to a comparison of t-step ahead expectations of the Markov process induced by P, where assortativity of P (i.e., monotonicity of the Markov process) plays an essential role at several steps of the proof.

More concretely, 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 across society. The key observation is the following "duality lemma:"[22]

Lemma 2 (Duality lemma)**.** *For any $C_1, C_2 \in \mathcal{C}$, the following are equivalent:*

[22] Lemma 2 itself does not appear in Appendix B, but follows from two other lemmas. The equivalence between (i) and (ii) follows from Lemma B.4, which shows that the mean-preserving spread order is the "dual order" of the more-assortative order $\succsim_{MA}$. Given this, the fact that (ii) implies (iii) follows from Lemma B.3, which establishes that $\succsim_m$ is an "isotone" order.

(i) $C_1 \succsim_{MA} C_2$

(ii) $D^1_{F,C_1} \succsim_m D^1_{F,C_2}$ *for all* F

(iii) $D^t_{F,C_1} \succsim_m D^t_{F,C_2}$ *for all* F *and* t*.*

The equivalence between (i) and (ii) establishes a duality between the more assortative order over interaction structures C_i and the mean-preserving spread order over the corresponding expected match distributions D^1_{F,C_i} under all possible type distributions F. Moreover, by (iii), this duality extends to higher-order expected match distributions. The step from (ii) to (iii) employs tools from majorization theory (Marshall, Olkin, and Arnold, 2010). Given this duality result, we complete the proof of Theorem 1 by exploiting the representation of Nash as a discounted sum of expected t-step ahead matches along with the fact that $\succsim_m$ is a linear and continuous order.

The following example illustrates Theorem 1 in Gaussian societies.

Example 2. Consider a Gaussian society P parametrized by (μ, σ^2, ρ) as in Example 1. For each type θ_0, the distribution $P(\cdot|\theta_0)$ of θ_0's 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. Thus, by Lemma 1 the Nash equilibrium takes the following linear form:

$$s(\theta_0) = \sum_{t=0}^{\infty} \gamma^t \left((1-\rho^t)\mu + \rho^t\theta_0\right) + \frac{\beta\mu}{(1-\gamma)(1-\gamma-\beta)} = \frac{\theta_0 - \mu}{1-\gamma\rho} + \frac{\mu}{1-\gamma-\beta}.$$

Hence, the equilibrium action distribution is normally distributed with mean and variance

$$\mathbb{E}[s] = \frac{\mu}{1-\gamma-\beta}, \quad \mathrm{Var}[s] = \frac{\sigma^2}{(1-\gamma\rho)^2}.$$

Illustrating Theorem 1, $\mathrm{Var}[s]$ is increasing in the correlation coefficient ρ while $\mathbb{E}[s]$ is independent of ρ. At the same time, $\mathrm{Var}[s]$ is also increasing in the local coordination motive γ and type heterogeneity σ^2. The following subsection will generalize the latter observations. □

3.3 Other Sources of Action Dispersion

We conclude the analysis of the Nash benchmark by highlighting two additional sources of action dispersion that will play a role in subsequent sections: Type heterogeneity and local coordination motives.

Proposition 1 (Effect of type heterogeneity)**.** *Let* H_i *denote the Nash action distribution under* (F_i, C, γ, β) *for* $i = 1, 2$*. Then* $F_1 \succsim_m F_2$ *implies* $H_1 \succsim_m H_2$*.*

Since local coordination motives affect the mean of the Nash action distribution (see Lemma 1), we isolate the effect of γ on action dispersion by considering mean-adjusted Nash action distributions. Given any cdf H, its ***mean-adjustment*** $\bar{H}$ is the mean-zero cdf that is obtained by subtracting the mean from all values of H; i.e., $\bar{H}(a) := H(a + \mu_H)$ for all $a \in \mathbb{R}$, where $\mu_H := \mathbb{E}_H[a]$.

Proposition 2 (Effect of local coordination motives)**.** *Let H_i denote the Nash action distribution under $(F, C, \gamma_i, \beta_i)$ for $i = 1, 2$. Then $\gamma_1 \geq \gamma_2$ implies $\bar{H}_1 \succsim_m \bar{H}_2$.*

Note that the global coordination coefficients β_i are arbitrary in Proposition 2. This is because, by Lemma 1, β only affects each type's strategy via the same constant term $\frac{\beta \mathbb{E}_F[\theta]}{(1-\gamma)(1-\gamma-\beta)}$ and hence has no impact on action dispersion.

To see the intuition for Proposition 2, consider the effect of an increase in γ on best responses against a fixed monotone strategy profile. When γ increases, higher/lower types have stronger incentives to play high/low actions, given that, because interactions are assortative, they are likely to meet other high/low types (and hence high/low actions). Paralleling the discussion following Theorem 1, this intuition for best responses is again only partial, and to obtain the full equilibrium comparison we again exploit the Markov process representation of equilibrium. The same approach also underlies the proof of Proposition 1.

Finally, note that assortativity of C is key in bringing about a strict increase in action dispersion under higher γ; indeed, under the independent interaction structure $C = C_I$, γ does not influence action dispersion. This echoes the observation following Theorem 1 that local coordination motives $\gamma > 0$ are necessary for an increase in assortativity to have an effect on action dispersion.[23]

4 Local Perception Equilibrium (LPE)

So far, we have studied Nash equilibrium behavior, which assumes that agents best-respond to correct perceptions about the underlying strategy profile s and society P. However, as motivated in the introduction, assortative interactions create scope for misperception about both s and P.

Intuitively, we can think of our coordination game as capturing the steady state of a setting where agents repeatedly interact with matches drawn according to the matching technology.[24] In such a situation, we might expect agents to possess a good understanding of the action distribution among their matches (e.g., consumption levels of luxury goods among their peers). However, as this is only a very partial snapshot of society, this information is not enough for them to correctly identify s and P. In particular, they may not know how representative their matches' behavior is of the overall population, and consequently may draw wrong inferences about the action distribution in society as a whole. Likewise, without knowing whom their matches interact with, they may not be able to separate to what extent matches' actions reflect their types (e.g., wealth) or their coordination incentives (e.g., keeping up with their peers' consumption), and hence may be incorrect about the type distribution, both among their matches and in society as a whole.

[23] More generally, Proposition E.1 in Appendix E.3 formalizes a sense in which local coordination motives and assortativity act as complements in increasing action dispersion.

[24] Note that the usual interpretation of Nash equilibrium is also based on steady states, which is used to justify the assumption that players are correct about others' strategies. We allow for potentially incorrect perceptions, since interaction structures, non-matches' actions, and all agents' types are not directly observable. Also, as in many games with imperfect observations, we assume that players do not observe their own utilities. This assumption is particularly natural in our main examples, where players' utilities may partly reflect non-material consequences, such as psychological conformity motives.

To capture this, we now introduce a solution concept that in any given society P jointly pins down true behavior s as well as each agent's perceptions of s and P, subject to the following three requirements: Each agent (i) correctly observes the action distribution among his matches; (ii) holds perceptions that can "rationalize" these observations, but need not be correct; and (iii) best-responds to these perceptions.

To formalize this idea, we first define 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,[25]

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

We now define our solution concept:

Definition 4. A ***local perception equilibrium (LPE)*** at P is a strategy profile s together with a perceived society $\hat{P}_\theta$ and perceived strategy profile $\hat{s}_\theta$ for each type θ satisfying:

1. ***Observational consistency***:
   (a) $H_\theta^{s,P} = H_\theta^{\hat{s}_\theta, \hat{P}_\theta}$
   (b) $\hat{s}_\theta(\theta) = s(\theta)$;
2. ***Nash rationalization***: For all θ', $\hat{s}_\theta(\theta') = \mathrm{BR}_{\theta'}(\hat{s}_\theta, \hat{P}_\theta)$.

Underlying any LPE is a true society P and true strategy profile s. Under Nash equilibrium, players are correct about both these objects. LPE generalizes this, assuming instead that each type θ has in mind a perceived society $\hat{P}_\theta$ and perceived strategy profile $\hat{s}_\theta$. These perceptions need not be correct, but are disciplined by two requirements.

First, observational consistency captures the idea ((i) above) that θ is correct about the distribution of his matches' actions, in the sense that (a) θ's perceived local action distribution $H_\theta^{\hat{s}_\theta, \hat{P}_\theta}$ coincides with the true local action distribution $H_\theta^{s,P}$; additionally, (b) θ is correct about his own behavior. Informally, based on the steady state interpretation above, observational consistency can be thought of as capturing the idea that agents' misperceptions are *persistent*, because they are not contradicted even under perfect knowledge of what happens in agents' local interactions.

The second bullet captures the requirement ((ii) above) that θ's perceptions "rationalize" his local observations: We require that the behavior, $\hat{s}_\theta(\theta')$, that θ attributes to any other type θ' should be a best response given θ's perceived strategy profile $\hat{s}_\theta$ and society $\hat{P}_\theta$. Note that in rationalizing others' behavior, θ adopts the simple and dogmatic worldview that all other types θ' *share* his perceptions $\hat{s}_\theta$ and $\hat{P}_\theta$. We refer to this as Nash rationalization, because it implies that $\hat{s}_\theta$ is the Nash equilibrium profile at $\hat{P}_\theta$.

[25]Here $\mathbb{1}_{\{s(\theta') \leq a\}}$ denotes the indicator random variable on the event $\{\theta' : s(\theta') \leq a\}$.

Observe that if θ had no views about other agents' perceptions, then he could believe others to employ arbitrary strategies, and this would severely limit his ability to draw inferences about the global strategy profile and society from his local action distribution. This would be the case under self-confirming equilibrium (Battigalli, 1987; Fudenberg and Levine, 1993), which we discuss in more detail in Section 7.1. On the other hand, we could relax Nash rationalization to allow θ to believe θ' to hold different perceptions than his own (and to capture more general forms of higher-order belief disagreement), as long as we impose common certainty of observational consistency and of rationality. In Section 7.1, we consider such a solution concept by adapting Esponda (2013)'s notion of "rationalizable conjectural equilibrium" to our setting, and we discuss to what extent our main results remain valid under this concept.

Finally, note that observational consistency and Nash rationalization jointly imply the requirement ((iii) above) that θ best-responds to his perceptions: Indeed, by part (b) of observational consistency, θ plays action $\hat{s}_\theta(\theta)$, and by Nash rationalization, the latter is θ's best response to his perceptions $(\hat{s}_\theta, \hat{P}_\theta)$.

5 LPE Analysis

Clearly, one special case of LPE is Nash equilibrium. We now proceed to analyze LPE in which agents are incorrect about the underlying society; i.e., $\hat{P}_\theta \neq P$ for some θ. Analogous to Section 2.2, we can decompose each perceived society $\hat{P}_\theta$ into a perceived type distribution $\hat{F}_\theta = \operatorname{marg} \hat{P}_\theta$ and perceived interaction structure $\hat{C}_\theta$ satisfying $\hat{C}_\theta(x, x') = \hat{P}_\theta\left(\hat{F}_\theta^{-1}(x), \hat{F}_\theta^{-1}(x')\right)$ for all $x, x' \in (0, 1)$.

We will structure our analysis by first asking which perceived interaction structures $\hat{C}_\theta$ can be sustained in LPE and then examining the corresponding perceived type distributions and behavior. We take a two-pronged approach to this question: Section 5.1 asks which $\hat{C}_\theta$ can be sustained in *arbitrary* environments. We find that this "robustness" requirement selects a particularly simple form of misperception, *assortativity neglect*, and analyze its implications in Sections 5.2 and 5.3. Section 5.4 specializes to the case of Gaussian societies and fully solves for *all* LPE (with linear strategies and Gaussian perceptions) that can arise in this setting.

5.1 Robust Sustainability of Assortativity Neglect

We say that type θ suffers from ***assortativity neglect*** if his perceived interaction structure $\hat{C}_\theta$ is the independent interaction structure C_I; that is, he believes everyone to interact with a representative sample of society as a whole. Assortativity neglect is broadly in line with empirical findings of "location effects" in the network cognition literature, in particular the fact that individuals exhibit a "projection bias" (Dessi, Gallo, and Goyal, 2016), viewing their own neighborhoods as representative of the global network. Additionally, assortativity neglect bears some resemblance to well-documented forms of inferential naiveté in the behavioral economics literature; notably "selection neglect" (e.g., Enke, 2017; Levy and Razin, 2017; Jehiel, 2018), where agents fail to take into account that the information they see may be subject to selection effects, and "correlation neglect" (e.g., Enke and Zimmermann, 2017; Eyster and Rabin, 2010; Ortoleva and Snowberg, 2015; Levy and Razin, 2015), where agents do not correctly account for correlation across different information sources.[26]

The following theorem provides a theoretical foundation for assortativity neglect. First, in any society and coordination game, there is a unique LPE in which all types suffer from assortativity neglect. Thus, assortativity neglect is a very robust form of misperception that can persist regardless of how assortative the actual society P is. Second, assortativity neglect is the *only* perception of the interaction structure that is robust in this way: For any interaction structure $\hat{C}$ other than the independent interaction structure and any type θ, there are environments in which θ cannot sustain perception $\hat{C}$ in any LPE. For the second statement, we impose the following very mild regularity requirement: Interaction structure $\hat{C}$ is called ***regular*** if there exists some $y \in (0,1)$ such that $|\{x \in (0,1) : \hat{C}(x|x) = y\}| = 1$.[27]

Theorem 2 (Robust sustainability of assortativity neglect)**.**

1. *For any (P, γ, β), there exists a unique LPE such that $\hat{C}_\theta = C_I$ for all θ.*
2. *For any regular $\hat{C} \neq C_I$ and any θ, there exists (P, γ, β) at which all LPE satisfy $\hat{C}_\theta \neq \hat{C}$.*

Theorem 2 is of interest for at least two reasons. First, the robustness result suggests that misperception in the form of assortativity neglect is distinguished by the fact that it can be a very stable phenomenon: Even if society changes over time (e.g., becomes more assortative) or if agents are engaged in social interactions along many separate dimensions that each might involve different type distributions and coordination motives, agents will never be forced to give up this misperception, as long as they suitably adapt their perceptions of the underlying population and behavior. Second, as we will discuss below, since assortativity neglect *uniquely* pins down the corresponding LPE, it yields sharp predictions for agents' behavior and their misperceptions about the distribution of population characteristics, paving the way for comparative statics analysis of how these are shaped by the nature of social interactions.

We prove Theorem 2 in Appendix D.2. The key observation is that assortativity neglect has the unique feature that it can explain *any* behavior: Formally, under C_I, any action distribution $\hat{H}$ can be rationalized as Nash under some appropriate type distribution $\hat{F}$, and C_I is the only interaction structure with this property.

Lemma 3. *For any $\hat{C} \in \mathcal{C}$, $\gamma > 0$ and $\beta \geq 0$, the following are equivalent:*

1. *$\hat{C} = C_I$.*

[26] Enke (2017) and Enke and Zimmermann (2017) provide experimental evidence for these phenomena. The remaining papers exogenously impose these forms of naiveté and study their implications in specific settings. By contrast, we endogenously derive assortativity neglect as the unique misperception of the interaction structure that can be robustly sustained in LPE.

[27] The 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 the possibility that this function oscillates arbitrarily, in particular requiring there to be at least one value y that is achieved exactly once. This is much weaker than the requirement that the map be monotone, which is satisfied under both Gaussian interactions and the independent interaction structure.

2. *For any action distribution $\hat{H} \in \mathcal{F}$,*[28] *there exists a type distribution $\hat{F}$ such that $\hat{H}$ is the Nash action distribution under $(\hat{F}, \hat{C}, \gamma, \beta)$.*

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

Proof. See Appendix D.1. □

The intuition for Lemma 3 is closely related to Section 3, where we saw that Nash action dispersion is increasing in assortativity (Theorem 1), but also in type heterogeneity and local coordination motives (Propositions 1 and 2). Concretely, if $\hat{H}$ is the Nash action distribution in society $(\hat{F}, \hat{C})$, then the action difference between any two quantiles $x > y$ can be decomposed into two terms—the corresponding type difference and the difference in coordination incentives, where the latter results from differences in x and y's *matches'* behavior and is greater under more assortative $\hat{C}$:

$$\underbrace{\hat{H}^{-1}(x) - \hat{H}^{-1}(y)}_{\Delta(\text{actions})} = \underbrace{\hat{F}^{-1}(x) - \hat{F}^{-1}(y)}_{\Delta(\text{types})} + \underbrace{\gamma \int \hat{H}^{-1}(z)\, d\left(\hat{C}(z|x) - \hat{C}(z|y)\right)}_{\Delta(\text{coordination incentives})}. \tag{5}$$

Under the independent interaction structure $\hat{C} = C_I$, there is no difference between x and y's coordination incentives as x and y face the same distribution of matches. As a result, we can rationalize $\hat{H}$ by attributing all action dispersion to type dispersion, and for any $\hat{H}$ there is a unique type distribution $\hat{F}$ that achieves this while also yielding the correct action mean $\mathbb{E}_{\hat{H}}[a] = \frac{\mathbb{E}_{\hat{F}}[\theta]}{1-\gamma-\beta}$. By contrast, in the presence of assortativity, coordination incentives may differ too much across agents for (5) to be satisfied: If $\hat{C} \neq C_I$, consider $x > y$ such that the distribution $\hat{C}(\cdot|x)$ of x's matches strictly stochastically dominates that of y. For any $\gamma > 0$, we can then find 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 (5) regardless of $\hat{F}$, because x and y's coordination incentives already differ by more than their actions.

Lemma 3 can be viewed as providing a microfoundation for the "fundamental attribution error," a central bias studied in social psychology (e.g., Ross, 1977) that refers to people's tendency to attribute others' behavior to intrinsic characteristics rather than to external factors such as social influence. This corresponds to the mechanism we highlighted above, whereby agents who suffer from assortativity neglect rationalize any observed behavior by attributing all action dispersion to type dispersion. Lemma 3 suggests a sense in which this is the unique view of society that can explain any behavior, thus possibly shedding some light on the prevalence of this bias.

In Appendix D.2, we use Lemma 3 to show that if type θ's perceived interaction structure is C_I, then no matter what strategy s is being played, there is a unique perceived type distribution $\hat{F}_\theta$ and perceived strategy $\hat{s}_\theta$ that allows θ to maintain observational consistency and Nash rationalization. Conversely, if $\hat{C}_\theta \neq C_I$, then in some environments θ will be unable to sustain this perception in any LPE.

[28] Recall that $\mathcal{F}$ is the set of all cdfs that are absolutely continuous, L^1 and have a connected support.

Finally, we complete the proof of Theorem 2 by showing that if all types suffer from assortativity neglect, then this uniquely pins down LPE behavior. Suppose that s is sustained in an LPE by perceptions $(\hat{s}_\theta, \hat{P}_\theta)_{\theta\in\Theta}$ with $\hat{C}_\theta = C_I$ for all θ. For any type θ, the Nash rationalization requirement implies

$$\hat{s}_\theta(\theta) = \theta + \gamma\mathbb{E}_{\hat{P}_\theta}[\hat{s}_\theta(\theta')|\theta] + \beta\mathbb{E}_{\hat{F}_\theta}[\hat{s}_\theta(\theta')].$$

But since $\hat{C}_\theta = C_I$, θ's perceived local action mean and perceived global action mean coincide, i.e., $\mathbb{E}_{\hat{P}_\theta}[\hat{s}_\theta(\theta')|\theta] = \mathbb{E}_{\hat{F}_\theta}[\hat{s}_\theta(\theta')]$. Moreover, observational consistency implies that θ is correct about both the local action action mean and his own action, i.e., $E_{\hat{P}_\theta}[\hat{s}_\theta(\theta')|\theta] = E_P[s(\theta')|\theta]$ and $\hat{s}_\theta(\theta) = s(\theta)$. Combining these observations, we have

$$s(\theta) = \theta + (\gamma + \beta)E_P[s(\theta')|\theta].$$

Thus, iterating expectations, behavior for all θ_0 must satisfy

$$s^{AN}(\theta_0) = \sum_{t=0}^{\infty}(\gamma+\beta)^t\mathbb{E}_P[\theta_t|\theta_0]. \tag{6}$$

Henceforth, for any (P, γ, β), we refer to the unique LPE in which all types suffer from assortativity neglect as the ***assortativity neglect LPE (ANLPE)***.

5.2 Implications of Assortativity Neglect

We now turn to examining the implications of assortativity neglect for behavior and perceptions about the population distribution.

Corollary 1 (Implications of assortativity neglect)**.** *For any (P, γ, β), the unique ANLPE has underlying strategy profile s^{AN} given by (6) and satisfies the following properties:*

1. ***False Consensus Effect:*** *Type θ's perceived population mean $\hat{\mu}_\theta = \mathbb{E}_{\hat{F}_\theta}[\theta']$ is increasing in θ, with*
$$\hat{\mu}_\theta = (1-\gamma-\beta)\sum_{t=0}^{\infty}(\gamma+\beta)^t\mathbb{E}_P[\theta_{t+1}|\theta_0 = \theta]. \tag{7}$$
2. ***Increased action dispersion:*** *The induced action distribution $H^{AN} = F \circ s^{AN^{-1}}$ is a mean-preserving spread of the Nash action distribution H^{NE} at (P, γ, β).*

Proof. See Appendix E.1. □

We refer to the first bullet point as a "false consensus effect," based on the eponymous finding in social psychology (Ross, Greene, and House, 1977; Marks and Miller, 1987) that people's perceptions of others' attributes tend to be positively correlated with their own attributes. This effect has been documented in a variety of settings. For example, individuals' perceptions of the median income in their country or city tend to be increasing in own income (e.g., Cruces, Perez-Truglia, and Tetaz, 2013); and in the context of political attitudes, supporters (opponents) of particular policies tend to overestimate (underestimate) its support in the population (e.g., Bauman and Geher, 2002), by as much as 67% vs. 33% in the case of abortion.

In our setting, this effect emerges in equilibrium as a consequence of assortativity neglect. Moreover, it generates and is further amplified by a second key implication of assortativity neglect, namely an increase in action dispersion relative to the Nash benchmark with correct perceptions.

To see the intuition, consider first the Nash equilibrium strategy profile s. Under assortativity neglect, all types believe their local action distribution to be perfectly representative of the global action distribution. However, since interactions are assortative and s is monotone, higher types observe higher actions. Hence, to maintain observational consistency, they must perceive a higher population mean, generating a false consensus effect. But then, higher types also perceive a higher global action mean, and hence, since $\beta \geq 0$, face stronger global coordination incentives than lower types. Thus, relative to Nash, assortativity neglect leads best responses against s to exhibit greater type sensitivity. This in turn exacerbates the false consensus effect, which feeds back into yet more dispersed behavior.

Formally, observe that by (6) and Lemma 1, behavior under ANLPE is the same as Nash equilibrium behavior in the modified game with local coordination motive $\gamma + \beta$ and global coordination motive 0. Since greater local coordination motives induce a more dispersed Nash action distribution (Proposition 2) while the average Nash action depends only on the *sum* of local and global coordination coefficients, this immediately implies that H^{AN} is a mean-preserving spread of H^{NE}. Finally, note that ANLPE and Nash strategy profiles coincide when $\beta = 0$, but the false consensus effect still obtains.[29]

5.3 Comparative Statics under Assortativity Neglect

We have seen that assortativity neglect differs from Nash in two important ways: First, behavior is more dispersed; second, perceptions about population characteristics are dispersed, as reflected for instance by the false consensus effect. Since assortativity neglect can persist in all environments, our framework lends itself to a number of comparative statics predictions.

Throughout this section, let M denote the distribution of perceived population means $\hat{\mu}_\theta$ when θ is distributed according to F; we sometimes refer to M as ***perception distribution*** for short. Note that by (7), agents' perceptions are on average correct; i.e., $\mathbb{E}_F[\hat{\mu}_\theta] = \mathbb{E}_F[\theta]$.

In the following, we show how M and the action distribution H^{AN} are shaped by the underlying environment. We first consider the effect of the true level of assortativity in society:

Theorem 3 (Effect of greater assortativity). *Fix $C_1, C_2 \in \mathcal{C}$, and let $H_i^{F,\gamma,\beta}$ and $M_i^{F,\gamma,\beta}$ denote the action and perception distributions under the ANLPE at (F, C_i, γ, β) for each $i = 1, 2$. Then the following are equivalent:*

[29] More generally, when $\beta = 0$, *all* LPE strategy profiles coincide with Nash equilibrium, while players' perceptions need not (see Lemma I.1 in Appendix I).

1. $C_1 \succsim_{MA} C_2$

2. $H_1^{F,\gamma,\beta} \succsim_m H_2^{F,\gamma,\beta}$ *for all* (F, γ, β)

3. $M_1^{F,\gamma,\beta} \succsim_m M_2^{F,\gamma,\beta}$ *for all* (F, γ, β).

Proof. See Appendix E.2. □

In Theorem 1, we saw that a defining feature of more assortative societies is more dispersed Nash behavior. Theorem 3 shows that more assortative societies also correspond to more dispersed ANLPE behavior. In addition, under assortativity neglect, greater assortativity leads to a more dispersed perception distribution M, i.e., a stronger false consensus effect. Because of the latter effect, for any given increase in assortativity, the increase in action dispersion is *more severe* under ANLPE than under Nash, as formalized in the following corollary. Thus, assortativity neglect amplifies the effect of assortatitivity on action dispersion.

Corollary 2 (Assortativity neglect amplifies effect of assortativity)**.** *Fix any* (F, γ, β) *and* $C_1 \succsim_{MA} C_2$. *Let* s_i^{AN} *and* s_i^{NE} *denote ANLPE and Nash strategy profiles under* (F, C_i, γ, β) *for* $i = 1, 2$. *Then for all types* θ^**,*

$$\mathbb{E}_F[s_1^{AN}(\theta) - s_2^{AN}(\theta)|\theta \geq \theta^*] \geq \mathbb{E}_F[s_1^{NE}(\theta) - s_2^{NE}(\theta)|\theta \geq \theta^*]. \tag{8}$$

Proof. See Appendix E.3 □

Greater action dispersion under C_1 corresponds to the fact that higher types on average take higher actions than under C_2 ; (8) implies that this effect is stronger under assortavity neglect than under Nash.

The next proposition studies the effect of local and global coordination motives. Similar to the effect of assortativity, greater coordination motives also lead to more dispersed behavior. This is again reminiscent of the result under Nash (Proposition 2); however, while Nash action dispersion is only affected by the local coordination motive γ, ANLPE action dispersion is affected by the *sum* of local and global motives $(\gamma + \beta)$. This is because under assortativity neglect, unlike under Nash, global coordination incentives vary across agents, as perceived global action means are increasing in agents' types. At the same time, as far as perceptions are concerned, greater coordination motives have the *opposite* effect of increased assortativity: Greater coordination motives *diminish* the false consensus effect, leading to a *less* dispersed perception distribution. Under mild regularity conditions, all agents' perceived population means converge to the truth as $\gamma + \beta \to 1$.

Proposition 3 (Effect of greater coordination motives)**.** *Fix* (F, C) *and consider* $\gamma_1 + \beta_1 \geq \gamma_2 + \beta_2$. *Let* H_i *and* M_i *denote the ANLPE action and perception distributions at* $(F, C, \gamma_i, \beta_i)$ *for* $i = 1, 2$. *Then:*[30]

1. $\bar{H}_1 \succsim_m \bar{H}_2$

[30] Recall that $\bar{H}_i$ denotes the mean-adjustment of H_i.

2. $M_2 \succsim_m M_1$*.*

Moreover, if $\lim_{t\to\infty} \mathbb{E}_P[\theta_t|\theta_0] \to \mathbb{E}_F[\theta]$*, then* $\hat{\mu}_{\theta_0} \to \mathbb{E}_F[\theta]$ *as* $\gamma + \beta \to 1$*.*

Proof. See Appendix E.4. □

To see the idea, recall that ANLPE behavior is given by $s(\theta) = \sum_{t=0}^{\infty}(\gamma+\beta)^t \mathbb{E}_P[\theta_t|\theta_0 = \theta]$ for all θ. Since this coincides with Nash behavior in the game with local coordination motives $\gamma + \beta$, Proposition 2 implies that higher $\gamma + \beta$ leads to a more dispersed action distribution.

One might expect this to also entail a more dispersed perception distribution. However, as far as agents' perceptions of the population mean are concerned, all that matters is the distribution of *normalized* actions $\tilde{s}(\theta) = (1-\gamma-\beta)s(\theta) = (1-\gamma-\beta)\sum_{t=0}^{\infty}(\gamma+\beta)^t \mathbb{E}_P[\theta_t|\theta_0 = \theta]$, as all agents know $\gamma + \beta$. Indeed, recall from Theorem 2 that type θ_0's perceived population mean is

$$\hat{\mu}_{\theta_0} = (1-\gamma-\beta)\sum_{t=0}^{\infty}(\gamma+\beta)^t \mathbb{E}_P[\theta_{t+1}|\theta_0] = \mathbb{E}_P[\tilde{s}(\theta_1)|\theta_0].$$

Now note that under higher $\gamma + \beta$, any type θ's normalized action $\tilde{s}(\theta)$ shifts relative weight away from θ's own type and θ's expected match $\mathbb{E}_P[\theta_1|\theta_0 = \theta]$ towards more distant t-step ahead matches $\mathbb{E}_P[\theta_t|\theta_0 = \theta]$. Since the latter are less sensitive to θ than the former, this renders $\tilde{s}(\theta)$ less sensitive to θ, which by observational consistency translates into less dispersed perceptions of the mean.

If additionally $\lim_{t\to\infty} \mathbb{E}_P[\theta_t|\theta_0] \to \mathbb{E}_F[\theta]$, so that in the limit expected t-step ahead matches perfectly approximate the average type in the population, then θ_0's perceived mean $\hat{\mu}_{\theta_0}$ will converge to the truth as $\gamma+\beta \to 1$. The requirement that $\lim_{t\to\infty} \mathbb{E}_P[\theta_t|\theta_0] \to \mathbb{E}_F[\theta]$ is a form of ergodicity. This is satisfied, for example, in any Gaussian society $P = (\mu, \sigma^2, \rho)$, as in this case $\mathbb{E}_P[\theta_t|\theta_0] = (1-\rho^t)\mu + \rho^t\theta_0 \to \mu$ by Example 2. More generally, the following lemma provides a simple sufficient condition for $\lim_{t\to\infty} \mathbb{E}_P[\theta_t|\theta_0] \to \mathbb{E}_F[\theta]$, which reflects the idea that the matching technology is not too local and the society is well connected.

Lemma 4. *Suppose P satisfies the following conditions:*

1. *There exists $\eta \in [0,1)$ and $K \in \mathbb{R}$ such that $\int |\theta_1|\, dP(\theta_1 \mid \theta_0) \leq \eta|\theta_0| + K$ for all θ_0.*

2. *On any compact interval, $I \subseteq \mathbb{R}$, $\inf_{\theta_1 \in I, \theta_0 \in I} p(\theta_1 \mid \theta_0) > 0$, where p denotes the density of P.*

Then $\lim_{t\to\infty} \mathbb{E}_P[\theta_t|\theta_0] \to \mathbb{E}_F[\theta]$ *for all* θ_0*.*

Proof. See Appendix E.5. □

Finally, in characterizing perceptions, our analysis so far has focused on perceived population *means* $\hat{\mu}_\theta$. Of course, other moments of agents' perceptions may also be of interest, and our framework lends itself to predictions about such higher moments as well. As an illustration, we briefly consider agents' perceptions of type *dispersion* in the population. When types represent income, this can be viewed as a measure of perceived income inequality, while in the case of political attitudes it could represent perceived polarization. Both are the subject of an active empirical literature, which also documents their importance in shaping individuals' preferences (e.g., for redistributive policies) and behavior (e.g., political participation).

The following proposition highlights an interesting difference between perceived population means and perceived dispersion. While we have seen that greater coordination motives lead to more accurate perceptions of the population mean, this need not be the case for perceptions of dispersion; indeed, perceived dispersion is strictly increasing in coordination motives.

Proposition 4 (Perceived population dispersion)**.** *Fix* (F, C) *and consider* $\gamma_1 + \beta_1 \geq \gamma_2 + \beta_2$. *For each* θ, *let* $\hat{F}^i_\theta$ *denote* θ*'s perceived type distribution under the ANLPE at* $(F, C, \gamma_i, \beta_i)$. *Then* $\bar{\hat{F}}^1_\theta \succsim_m \bar{\hat{F}}^2_\theta$.

Proof. See Appendix E.6. □

To see the intuition, note that greater coordination motives lead to more dispersed ANLPE behavior (Proposition 3), and this is true not only at the global level, but also for each type's local action distribution. However, while agents are correct about the parameters γ and β, under assortativity neglect the fact that their perceived interaction structure is C_I means that increases in these parameters do not affect their perceived action distribution (recall the discussion following Lemma 3). Thus, they must instead rationalize increases in local action dispersion by attributing them to increased type dispersion.

Proposition 4 leaves open the possibility that perceived type dispersion may be either smaller or greater than actual. Both possibilities are consistent with empirical work, which documents under- or overestimation depending on the context. For instance, a number of studies across different countries establish widespread underestimation of income inequality (e.g., Norton and Ariely, 2011; Engelhardt and Wagener, 2015); on the other hand, recent work on perceived political attitudes in the United States suggests that individuals tend to overestimate political polarization (e.g., Ahler, 2014; Westfall, Van Boven, Chambers, and Judd, 2015). Focusing on the Gaussian environment, Corollary 3 in the next subsection derives exact conditions under which agents under- or overestimate dispersion, showing that which of the two occurs depends on the relative strength of coordination motives and assortativity.

5.4 Linear-Gaussian LPE

So far, we have considered arbitrary societies and focused on LPE under assortativity neglect, which we showed to be the only form of misperception about the interaction structure that is sustainable in any environment.

In this section, we specialize to Gaussian societies as in Example 1, where P is jointly normally distributed and parametrized by the mean μ and variance σ^2 of the type distribution along with a correlation coefficent $\rho \in [0, 1)$. The purpose is twofold: First, this setting is particularly tractable, allowing us to fully solve for *all* LPE (with linear strategies and Gaussian perceptions) in closed form. Second, these LPE can feature *partial* assortativity neglect, where agents underestimate but do not completely neglect the assortativity in society, and we illustrate the robustness of our key findings in the previous section to this weaker form of misperception.

Our analysis focuses on the subclass of LPE where strategies are linear in types (as under Nash, cf. Example 2) and each type's perceived society is itself Gaussian.

Definition 5. A ***linear-Gaussian LPE*** at P is an LPE $(s, (\hat{s}_\theta, \hat{P}_\theta)_{\theta\in\Theta})$ such that

1. strategy profile s is linear and non-constant in types;
2. each type θ's perceived society $\hat{P}_\theta$ is Gaussian parametrized by $(\hat{\mu}_\theta, \hat{\sigma}^2_\theta, \hat{\rho}_\theta)$.

The following proposition characterizes the set of all linear-Gaussian LPE under the assumption that $\beta > 0$.[31]

Proposition 5 (Linear-Gaussian LPE)**.** *Consider a Gaussian society (μ, σ^2, ρ) and suppose $\beta > 0$. For any $\hat{\rho} \in [0, 1)$, there exists a unique linear-Gaussian LPE in which $\hat{\rho}_\theta = \hat{\rho}$ for all θ. Conversely, for any linear-Gaussian LPE, there exists $\hat{\rho} \in [0, 1)$ such that $\hat{\rho}_\theta = \hat{\rho}$ for all $\theta \neq \mu$.*

Proof. See Appendix F.1. □

While Definition 5 allows different types θ to have different perceived correlations $\hat{\rho}_\theta$, Proposition 5 shows that every linear-Gaussian LPE is in fact characterized by a *single* perceived correlation $\hat{\rho} \in [0, 1)$ that is common across all θ (except possibly the mean type).[32] Moreover, any perceived correlation $\hat{\rho} \in [0, 1)$ can be sustained, and $\hat{\rho}$ uniquely pins down the true strategy profile s, each type θ's perceived strategy profile $\hat{s}_\theta$ and perceived mean $\hat{\mu}_\theta$, and the type-independent perceived variance $\hat{\sigma}^2$; the proof of Proposition 5 exhibits closed-form expressions for all these quantities.

Note that if $\hat{\rho} = \rho$, then linear-Gaussian LPE and Nash behavior coincide and all types' perceptions are correct. Consider instead the case of ***partial assortativity neglect*** where $\hat{\rho} < \rho$, so that all agents underestimate assortativity in society but do not necessarily completely neglect it. From the closed-form expressions in the proof of Proposition 5, it is easy to see that all qualitative implications of the analysis under assortativity neglect remain valid in this case. In particular, perceived population means are subject to a false consensus effect and behavior is more dispersed than under Nash equilibrium; moreover, both effects are stronger in more assortative societies, while stronger coordination motives lead to more dispersed behavior but more accurate perceptions of the mean. Finally, more severe assortativity neglect (smaller $\hat{\rho}$) also leads to more dispersed behavior and perceptions. Figure 2 illustrates perceptions across different types under partial assortativity neglect.

The following result expands on the analysis of perceived type dispersion in Proposition 4. Recall that in any linear-Gaussian LPE, perceived population variance $\hat{\sigma}^2_\theta = \hat{\sigma}^2$ is the same for all

[31]When $\beta = 0$, perceived correlation $\hat{\rho}_\theta$ can vary across types θ in an arbitrary manner, but the corresponding values $(\hat{\mu}_\theta, \hat{\sigma}^2_\theta)$ are uniquely pinned down by the same expressions as in the proof of Proposition 5. This is because perceived correlation does not influence the actual strategy profile, which coincides with the Nash equilibrium.

[32]The perceived correlation $\hat{\rho}_\mu$ of the mean type $\theta = \mu$ can take arbitrary values in $[0, 1)$. However, μ's perceived mean and true strategy are given by $\hat{\mu}_\mu = \mu$ and $s(\mu) = \frac{\mu}{1-\gamma-\beta}$ in every linear-Gaussian LPE, and every $\hat{\rho}_\mu \in [0, 1)$ uniquely pins down μ's corresponding perceived variance $\hat{\sigma}^2_\mu$ and perceived strategy profile $\hat{s}_\mu$.

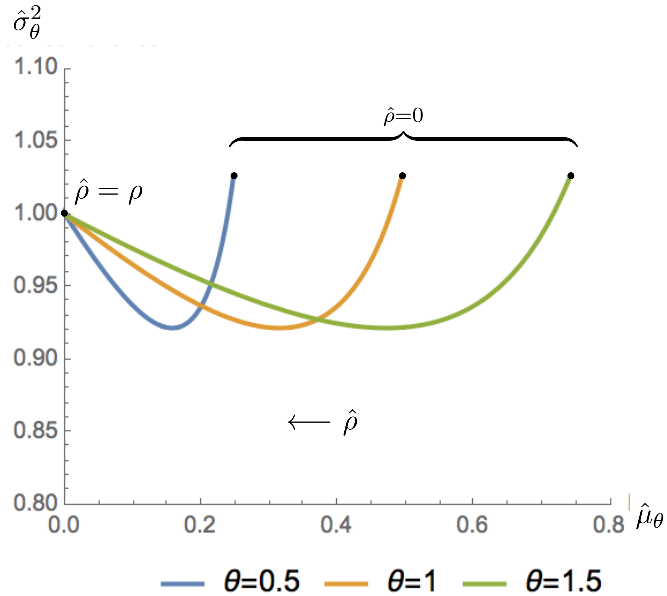

Figure 2: Perceived means and variances $(\hat{\mu}_\theta, \hat{\sigma}^2_\theta)$ under partial assortativity neglect ($\hat{\rho}$ increases westward from 0 to ρ). For $\hat{\rho} = \rho$, perceptions are equal to the true values $\mu =$ and $\sigma^2 = 1$; for $\hat{\rho} < \rho$, perceived means exhibit a false consensus effect ($\hat{\mu}_\theta$ is increasing in θ), which is more severe the smaller $\hat{\rho}$. Perceived variances are type-independent and can be higher or lower than the true variance. Parameter values: $\rho = 0.6, \gamma = 0.25, \beta = 0.1$.

$\theta \neq \mu$. Analogous to Proposition 4, under partial assortativity neglect, $\hat{\sigma}^2$ is again strictly increasing in coordination motives. At the same time, however, perceived variance can be nonmonotonic in assortativity ρ, and these two effects combine to yield overestimation of variance when coordination motives are strong relative to assortativity and underestimation in the opposite case, as illustrated in Figure 3.

Corollary 3 (Perceived variance under partial assortativity neglect)**.** *Fix (μ, σ^2) and $\hat{\rho} \in [0, 1)$. For each $\rho > \hat{\rho}$ and γ, β with $\beta > 0$, consider the commonly perceived type variance $\hat{\sigma}^2$ in the linear-Gaussian LPE with perceived correlation $\hat{\rho}$. Then*

1. *$\hat{\sigma}^2$ is strictly increasing in γ and β;*

2. *there is $\bar{\rho}(\gamma, \beta) \in [\hat{\rho}, 1)$ which is nondecreasing in γ and β such that*

$$\hat{\sigma}^2 \begin{cases} > \sigma^2 \text{ if } \rho \in (\hat{\rho}, \bar{\rho}(\gamma, \beta)) \\ < \sigma^2 \text{ if } \rho > \bar{\rho}(\gamma, \beta) \end{cases} .$$

To see the role played by assortativity, note the following two countervailing effects on local action variance: On the one hand, the greater ρ the lower the local type variance $\sigma^2(1 - \rho^2)$, which for any fixed linear strategy profile reduces local action variance. On the other hand, we have seen that greater assortativity leads to greater action dispersion, by increasing players' coordination incentives, which for any fixed local type distribution increases local action variance. Moreover, this

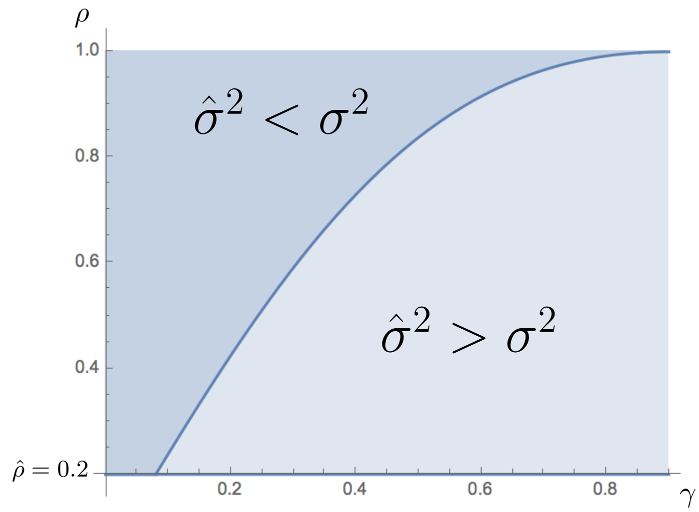

Figure 3: Over-/underestimation of population variance depending on coordination motives and assortativity. The blue curve depicts $\bar{\rho}(\gamma, \beta)$ as a function of γ for $\beta = 0.2$ and $\hat{\rho} = 0.2$. When $\rho > \bar{\rho}(\gamma, \beta)$, agents underestimate variance; when $\hat{\rho} < \rho < \bar{\rho}(\gamma, \beta)$, agents overestimate variance; for $\rho = \bar{\rho}(\gamma, \beta)$ or $\rho = \hat{\rho}$, perceived variance is correct.

latter effect is stronger the greater agents' coordination motives, and when coordination motives are strong enough relative to ρ, it can be shown to dominate the former effect. For a fixed level $\hat{\rho} < \rho$ of (partial) assortativity neglect, this means that if γ and β are large then assortativity neglect pushes agents to underestimate local action variance, which to maintain observational consistency must be counterbalanced by *overestimating* type variance; while if coordination motives are weak, the opposite is the case, and agents will underestimate σ^2, as summarized in the second bullet.

As discussed following Proposition 4, empirical work documents both over- and underestimation of type dispersion depending on the context. Corollary 3 suggests investigating the role played by coordination motives and/or assortativity. For instance, if it were the case that peer effects are stronger for political activities than for consumption decisions (or that assortativity along wealth is stronger than along political attitudes), this would be consistent with the finding that people tend to overestimate political attitude polarization but underestimate income inequality in their societies.

6 Welfare Implications

We now turn to welfare implications of our analysis. For this section only, we assume utilities of the quadratic-loss form, where type θ's payoff from action a against strategy profile s in society P is given as in (2):

$$u_P(a, \theta, s) = -\mathbb{E}_P[(a - \theta - \gamma s(\theta') - \beta \mathbb{E}_F[s(\theta')])^2 \mid \theta].$$

Note that under this specification, type θ's utility to best-responding against any strategy profile s is negatively proportional to θ's local action variance under s: For $a^* = \text{BR}_\theta(s, P)$,

$$u_P(a^*, \theta, s) = -\gamma^2 \text{Var}_P[s(\theta')|\theta]. \tag{9}$$

Intuitively, this captures the fact that variation in θ's matches' behavior creates a miscoordination cost for θ.

6.1 Welfare Implications of Misperception

We first consider the welfare implications of misperception in the form of assortativity neglect, by comparing each type's Nash equilibrium utility to that under ANLPE. In society P, θ's Nash utility is

$$u_P^{NE}(\theta) := u_P(s^{NE}(\theta), \theta, s^{NE}),$$

where s^{NE} denotes the Nash strategy profile. Under ANLPE, we distinguish between θ's objective and subjective utility. The *objective utility* $u_P^{AN}(\theta)$ is θ's payoff under the true ANLPE strategy profile and society s^{AN} and P:

$$u_P^{AN}(\theta) := u_P(s^{AN}(\theta), \theta, s^{AN}).$$

The *subjective utility* $\hat{u}_P^{AN}(\theta)$ is θ's payoff under his perceived strategy profile and society $\hat{s}_\theta$ and $\hat{P}_\theta$:

$$\hat{u}_P^{AN}(\theta) := u_{\hat{P}_\theta}(s^{AN}(\theta), \theta, \hat{s}_\theta) = u_{\hat{P}_\theta}(\hat{s}_\theta(\theta), \theta, \hat{s}_\theta),$$

where the final equality holds because $\hat{s}_\theta(\theta) = s^{AN}(\theta)$ by observational consistency.

Proposition 6 (Welfare implications of assortativity neglect)**.** *Consider any* (P, γ, β)*. Then every type* θ *is both subjectively and objectively worse off under ANLPE than under Nash; that is,*

$$u_P^{NE}(\theta) \geq \hat{u}_P^{AN}(\theta) \geq u_P^{AN}(\theta).$$

Proof. See Supplementary Appendix G.1. □

Thus, both subjectively and objectively, assortativity neglect Pareto decreases welfare relative to the correct perceptions benchmark. At the same time, under ANLPE, agents are better off under their subjective perceptions $\hat{s}_\theta$ and $\hat{P}_\theta$ than under the true strategy profile and society s^{AN} and P.

To see the idea, recall from Corollary 1 that assortativity neglect increases (global) action dispersion relative to Nash. By a similar logic, each type's *local* action variance is also higher under ANLPE than under Nash. Given (9), ANLPE thus features greater local miscoordination costs, and this effect is the same under subjective and objective ANLPE utility, as agents are correct about their local action distribution. Under objective ANLPE utility, agents are further hurt by a second, *misoptimization* effect, which arises from the fact that $s^{AN}(\theta)$ is a best response against θ's perceived strategy profile and society, but not against the true strategy profile and society.

In the context of perceptions of income distributions, a nascent empirical literature studies the possibility and implications of correcting people's misperceptions (e.g., Cruces, Perez-Truglia, and Tetaz, 2013; Perez-Truglia, 2016). Some studies find that in the short run, greater transparency may be harmful for poorer individuals because of an adverse effect on self-image (e.g., Perez-Truglia, 2016). Proposition 6 abstracts away from self-image considerations and instead highlights the fact that in the longer run corrected perceptions may change equilibrium behavior in such a way that could be beneficial for *everyone* in society.

6.2 Welfare Implications of Assortativity

A second implication of our analysis is that greater assortativity can have an ambiguous effect on welfare. Recall from the discussion following Corollary 3 that greater assortativity affects local action variance in two conflicting ways. On the one hand, it reduces local type variance, which other things equal, reduces local action variance; but on the other hand, it increases action dispersion, which tends to drive up local action variance. Moreover, the second effect dominates when coordination motives are sufficiently strong. This is true under both Nash and ANLPE, with the difference that under Nash, it is only local coordination motives that matter, while under assortativity neglect both local and global coordination motives play a role. This suggests that whether assortatitivity is beneficial or harmful depends on the strength of coordination motives. The following result illustrates this in the Gaussian setting.

Corollary 4 (Welfare implications of assortativity)**.** *Consider any Gaussian society $P = (\mu, \sigma^2, \rho)$ and $\gamma, \beta \geq 0$. Then for all types θ,*

1. *$u_P^{NE}(\theta)$ is decreasing in ρ if $\rho < \gamma$ and increasing in ρ if $\rho > \gamma$;*
2. *$\hat{u}_P^{AN}(\theta)$ is decreasing in ρ if $\rho < \gamma + \beta$ and increasing in ρ if $\rho > \gamma + \beta$.*

Proof. See Supplementary Appendix G.2. □

Corollary 4 shows that a marginal increase in assortativity is beneficial only when assortativity is already sufficiently high. In societies that exhibit a low degree of assortativity or strong coordination motives, greater assortativity may be detrimental; moreover, assortativity neglect exacerbates this problem, as it raises the threshold below which increases in ρ are harmful.

Outside the present model, there are of course many reasons why a social planner may prefer to induce a less assortative and more evenly mixed society.[33] Corollary 4 shows that even when, as in our model, coordination is the only relevant consideration, a social planner may still prefer to reduce assortativity: While in this setting welfare is maximized in perfectly assortative societies ($\rho = 1$), if the social planner faces constraints (institutional or otherwise) that preclude achieving

[33] See, e.g., Currarini, Jackson, and Pin (2009) and the references therein.

a high degree of assortativity (greater than γ or $\gamma + \beta$), then the least assortative society is in fact best.

7 Discussion

7.1 Relationship with Other Solution Concepts

In this section, we discuss the relationship between LPE and the literature on self-confirming equilibrium (SCE) (Battigalli, 1987; Fudenberg and Levine, 1993) and its variants. In general games, SCE captures possible behavior when all agents best-respond to beliefs about other agents' strategies that need not be correct, but instead need only be consistent with some limited feedback about opponents' behavior. LPE applies this idea to the specific setting in which each agent's feedback about opponents' behavior is limited to correctly observing the action distribution among his matches, where these matches are determined by some underlying society P; moreover, in addition to beliefs about strategies, LPE also treats agents' beliefs about type distributions and interaction structures as endogenous equilibrium objects.

However, in an important conceptual departure, LPE refines SCE by putting structure on how agents draw inferences from their local action distributions: While under SCE agents do not form a model of other agents' behavior and can hold *arbitrary* beliefs about opponents' strategies that are consistent with observed feedback, agents in LPE hold perceptions that rationalize others' behavior as Nash.[34] Formally, SCE applied to our setting retains the observational consistency requirement in Definition 4, but relaxes Nash rationalization to the requirement that $s(\theta) = \mathrm{BR}_\theta(\hat{s}_\theta, \hat{P}_\theta)$.[35] As a result of this relaxation, SCE has very little bite in our setting: In particular,[36] in marked contrast with our finding that under LPE assortativity neglect is the unique perception of the interaction structure that can be sustained in any environment, under SCE *any* perceived interaction structure can always be sustained; likewise, essentially any strategy profile s can always be sustained.[37] Thus,

[34]Other commonly used departures from Bayes Nash equilibrium that can be viewed as special cases of SCE include cursed equilibrium (Eyster and Rabin, 2005) and analogy-based expectation equilibrium (Jehiel, 2005; Jehiel and Koessler, 2008). These again differ from LPE in that they do not impose a counterpart of Nash rationalization; additionally, they only allow for incorrect beliefs about strategies and assume that agents are correct about the underlying type distribution. Moreover, both solution concepts *exogenously* impose specific forms of incorrect beliefs about opponents' behavior (respectively, underestimating correlation between opponents' actions and their types, and assuming that opponents' play is constant within certain "analogy classes" of states). In contrast, LPE allows for arbitrary forms of misperception about behavior and society (subject to observational consistency and Nash rationalization), and we *endogenously* derive a specific form of misperception, assortativity neglect, by showing that it is the unique perception of the interaction structure that can be sustained in LPE in any environment.

[35]Note that since $s(\theta) = \hat{s}_\theta(\theta)$ by observational consistency, this is equivalent to $\hat{s}_\theta(\theta) = \mathrm{BR}_\theta(\hat{s}_\theta, \hat{P}_\theta)$, i.e., to θ's *own* perceived behavior being a best-response to his perceptions $\hat{s}_\theta, \hat{P}_\theta$. This is strictly weaker than requiring that $\hat{s}_\theta(\theta') = \mathrm{BR}_{\theta'}(\hat{s}_\theta, \hat{P}_\theta)$ for all θ', where θ believes *all* types θ' to best-respond to $\hat{s}_\theta, \hat{P}_\theta$.

[36]Formally, consider any environment (P, γ, β) where P has full support. Then for any type θ and *any* perceived interaction structure $\hat{C}$, there exists an SCE where $\hat{C}_\theta = \hat{C}$. Likewise, if $\beta > 0$, then for *any* strictly increasing and continuous strategy profile s with $s(\Theta) = \mathbb{R}$, there exists an SCE where behavior is given by s. Details are available upon request.

[37]While we have not derived the full set of LPE strategy profiles in general environments, it is clear in specific settings that LPE is more restrictive than SCE. For example, Proposition 5 shows that the set of possible linear-Gaussian LPE strategy profiles is one-dimensional and fully parametrized by values of $\hat{\rho} \in [0, 1)$; in contrast, it is not

its lack of a model for how agents draw inferences from their observed behavior renders SCE too permissive to obtain clear predictions for possible behavior and perceptions in assortative societies.

Our Nash rationalization requirement makes LPE more similar in spirit to "rationalizable conjectural equilibrium" (RCE), a refinement of SCE due to Rubinstein and Wolinsky (1994) and Esponda (2013).[38] However, RCE is still more permissive than LPE, as it allows agents to hold potentially very complicated higher-order beliefs about other agents' perceptions: Unlike Nash rationalization, where θ believes each type θ' to share his perceptions $\hat{s}_\theta$ and $\hat{P}_\theta$ and to best-respond to these, RCE allows θ to believe that θ' might best-respond to different perceptions than his own, as long as these perceptions are in turn observationally consistent with the local action distribution that θ believes θ' to observe, where θ' may again believe his matches to hold different perceptions from his own, and so on. More generally, RCE captures the possible behavior s and perceptions about P and s that can arise when players are constrained only by observational consistency, rationality, and common certainty of these two requirements.

In Supplementary Appendix I, we formalize this relationship by defining RCE in our setting and showing that LPE is a special case thereof. Corollary 5 demonstrates that the robust sustainability of assortativity neglect generalizes to RCE and yields the same implications for behavior as under ANLPE. At the same time, unlike under LPE where assortativity neglect uniquely pins down agents' perceptions of the type distribution, under RCE it is consistent with a multitude of perceptions. Thus, LPE can be viewed as a tractable refinement of RCE that is particularly well suited to our goal of obtaining sharp predictions about agents' misperceptions and their comparative statics with respect to the environment, while also having the virtue of imposing less cognitive complexity on agents' inference process. We formalize the latter point in Supplementary Appendix I.3 by showing how certain forms of lexicographic preferences for "simpler" perceptions can refine RCE into LPE and ANLPE.

7.2 Alternative Notions of Assortativity and Dispersion

So far, we have employed the mean-preserving spread order as a measure of dispersion. In this section, we show that analogs of our results remain valid under a second, stronger notion of dispersion that is also frequently used in statistics (e.g., Shaked and Shanthikumar, 2007):

Definition 6. Given two cdfs G_1, G_2, we say that G_1 is ***more dispersive than*** G_2, denoted $G_1 \succsim_d G_2$, if for all $0 < y \leq x < 1$,

$$G_1^{-1}(x) - G_1^{-1}(y) \geq G_2^{-1}(x) - G_2^{-1}(y).$$

The more dispersive ranking requires the gap between *any* two quantiles to be greater under

difficult to see that *any* linear strategy profile can be sustained as part of an SCE with Gaussian perceptions.

[38] Fudenberg and Kamada (2015) introduce related solution concepts for extensive form games. Lipnowski and Sadler (2017) consider RCE in which the observation structure is parametrized by a network; however, in their setting the network does not affect game payoffs, and they assume that the network structure and agents' types are common knowledge.

G_1 than G_2. When G_1 and G_2 have the same mean, this is a strictly stronger requirement than the mean-preserving spread order. As a result, greater assortativity in the sense of $\succsim_{MA}$ is not in general sufficient to yield a more dispersive equilibrium action distribution (under either Nash equilibrium or ANLPE). Instead, this is ensured by the following stronger notion of comparative assortativity:

Definition 7. C_1 is ***strongly more assortative*** than C_2, denoted $C_1 \succsim_{SMA} C_2$, if for all $x, y, z \in (0,1)$ with $x \geq y$,

$$C_1(z|y) - C_1(z|x) \geq C_2(z|y) - C_2(z|x).$$

To interpret this condition, recall that assortativity of C requires the distribution of matches' quantiles to be first-order stochastically increasing in own quantile; that is,

$$C(z|y) - C(z|x) \geq 0$$

for all $x \geq y$ and z. Thus, C_1 is strongly more assortative than C_2 if the first-order stochastic increase in matches' quantile distribution is globally stronger under C_1 than C_2.

Example 3 (Mixing with independent interaction structure)**.** Any interaction structure is strongly more assortative than C_I. More generally, if $C_2 = \rho C_1 + (1-\rho)C_I$ for some $\rho \in [0,1]$ (i.e., if C_2 corresponds to drawing opponents' quantiles according to C_1 with probability ρ and uniformly with probability $1-\rho$), then we also have $C_1 \succsim_{SMA} C_2$. □

Paralleling Theorem 1, strongly more assortative societies are characterized precisely by more dispersive Nash action distributions:

Theorem 4. *Fix $C_1, C_2 \in \mathcal{C}$, and let $H_i^{F,\gamma,\beta}$ denote the Nash action distribution under (F, C_i, γ, β) for each $i = 1, 2$. The following are equivalent:*

1. $C_1 \succsim_{SMA} C_2$
2. $H_1^{F,\gamma,\beta} \succsim_d H_2^{F,\gamma,\beta}$ *for all* (F, γ, β).

Proof. See Supplementary Appendix H.1. □

Likewise, for ANLPE an analog of Theorem 3 holds, replacing $\succsim_{MA}$ and $\succsim_m$ by $\succsim_{SMA}$ and $\succsim_d$ throughout.[39]

7.3 Identification and Misestimation

In our model, the fact that agents can sustain misperceptions is related to the difficulty of identifying an underlying society (F, C) based only on observation of local action distributions. In particular, Lemma 3 establishes that any action distribution, whether it is induced by a Nash equilibrium or

[39]Moreover, as the proofs in the appendix show, greater coordination motives in fact correspond to more dispersive Nash and ANLPE behavior, so that we could have replaced $\bar{H}_1 \succsim_m \bar{H}_2$ with $H_1 \succsim_d H_2$ in Propositions 2 and 3.

not, is observationally consistent with some Nash equilibrium under C_I. We now briefly discuss implications for the related problem of an outside observer (e.g., a policy maker or an analyst) who observes the global action distribution in society and based on this estimates the corresponding interaction structure and preference distribution.

To be concrete, suppose that agents are playing a Nash equilibrium.[40] The action distribution is observed by an outside observer who does not know the true society (F, C) and instead believes that $\hat{C}$ is the true interaction structure. Given this, when and how can the outside observer find some type distribution $\hat{F}$ that rationalizes her observations as Nash under $(\hat{F}, \hat{C})$? The following result summarizes the answer:[41]

Proposition 7. *Consider any $C, \hat{C} \in \mathcal{C}$ such that $C \succsim_{SMA} \hat{C}$. Then for any (F, γ, β), there exists $\hat{F}$ such that the Nash action distributions under (F, C, γ, β) and $(\hat{F}, \hat{C}, \gamma, \beta)$ coincide. Furthermore, such $\hat{F}$ is unique for each (F, γ, β) and satisfies $\hat{F} \succsim_d F$ and $\mathbb{E}_F[\theta] = \mathbb{E}_{\hat{F}}[\theta]$.*

Proof. See Supplementary Appendix H.2. □

That is, whenever the outside observer underestimates assortativity, in the sense that $C \succsim_{SMA} \hat{C}$, she can always rationalize her observations by finding an appropriate type distribution $\hat{F}$. Note that a special case of this is when the outsider observer fully neglects assortativity (i.e., $\hat{C} = C_I$), which echoes the existence result for ANLPE.

The proposition also shows that the corresponding estimate of the type distribution $\hat{F}$ must be biased in a particular direction, viz., overestimating dispersion relative to the true F. The reason is that since the outside observer underestimates assortativity in society, she misattributes the observed action dispersion to type heterogeneity.[42] At the same time, the estimated population mean under $\hat{F}$ is unbiased. (This follows from the fact that the Nash equilibrium action average is not influenced by the interaction structure, as shown by Lemma 1.) This is in contrast to agents' misperceptions under LPE, where we saw that assortativity neglect leads to a false-consensus effect.

8 Conclusion

This paper develops a framework to study the possibility and implications of misperceptions in assortative societies. Our analysis provides a theoretical foundation for assortativity neglect, which we showed to be the unique perception of the interaction structure that can persist in any environment. Assortativity neglect in turn uniquely pins down agents' behavior and perceptions about the type distribution, giving rise to two mutually reinforcing departures from the Nash benchmark: greater

[40] Since ANLPE coincides with Nash equilibrium under modified coordination incentive parameters (γ, β), an analogous exercise can be conducted for ANLPE.

[41] One could consider an alternative problem of finding $\hat{C}$ to rationalize the action observation under some fixed $\hat{F}$. We focus on the case of fixed $\hat{C}$, as this is more closely analogous to the analysis of assortativity neglect under LPE.

[42] When C is more assortative (but not strongly) than $\hat{C}$, there might not exist an $\hat{F}$ such that the Nash action distributions under (F, C, γ, β) and $(\hat{F}, \hat{C}, \gamma, \beta)$ coincide. However, if such a $\hat{F}$ exists, it satisfies $\hat{F} \succsim_m F$.

action dispersion and the well-documented "false consensus effect." Increased action dispersion adversely affects welfare through increased miscoordination costs, suggesting that correcting agents' misperceptions could be beneficial.

One virtue of our framework is that it yields testable predictions for how agents' perceptions and behavior vary with the environment. Our finding that increased assortativity manifests itself in terms of greater action dispersion is particularly relevant in light of evidence that societies are growing ever more assortative. In such a changing environment, the idea that agents might not be fully aware of the extent of assortativity appears especially plausible, and we showed that neglecting assortativity further exacerbates the effect of assortativity on action dispersion. We also highlighted the crucial role that strategic incentives play in shaping not only the magnitude, but indeed the direction of agents' misperceptions about society, as captured for instance by the finding that coordination motives determine whether agents over- or underestimate type dispersion in society.

The paper makes several methodological contributions that we believe could prove useful beyond our specific setting: Our solution concept, "local perception equilibrium," enabled us to analyze behavior and perceptions when agents only observe the distribution of actions among their matches; the definition of LPE extends readily to general games with limited feedback about opponents' behavior, where it could serve more broadly as a tractable refinement of rationalizable conjectural equilibrium. Our observation that any assortative society induces a monotone Markov process over its type space and that equilibrium strategies can be represented in terms of higher-order expectations of this process played a central role in deriving comparative statics. Finally, by comparing societies in terms of their copulas, we defined natural non-parametric notions of "more assortativity" that might find application in other areas, such as matching theory under search frictions.

It is worth highlighting that our finding that assortativity neglect is always sustainable can be extended in a variety of directions. First, this result does not rely on linear best response functions and extends to LPE in more general games that admit Nash equilibria that are strictly increasing in agents' types. A second generalization is a hybrid model where only fraction $\alpha \in (0,1)$ of agents of each type θ hold (mis)perceptions $\hat{s}_\theta$ and $\hat{P}_\theta$ that are subject to observational consistency and Nash rationalization as under LPE, while fraction $1-\alpha$ are *sophisticated*, in the sense that they know the underlying strategy profile and society as well as the fraction of LPE agents and their perceptions. In this setting, we can again show that assortativity neglect is the unique misperception of the interaction structure that the LPE agents can sustain in any environment; moreover, under assortativity neglect the action distribution among LPE agents is more dispersed than among sophisticated agents, and both are more dispersed than under Nash but less than under ANLPE.[43]

Finally, an active recent literature analyzes learning dynamics under *exogenously* given misspecified models (Heidhues, Koszegi, and Strack, 2017; Fudenberg, Romanyuk, and Strack, 2017; Esponda and Pouzo, 2017; Bohren and Hauser, 2017). The present paper differs from this literature by asking which misperceptions can be *endogenously* sustained in a static coordination game

[43]Details for both generalizations are available upon request.

environment. Nevertheless, our finding that assortativity neglect is robustly sustainable suggests model misspecification with respect to interaction structures as a relevant and important direction for this literature, and in ongoing work, we pursue this question in the context of social learning.

Appendix

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 all weakly increasing and absolutely continuous[44] 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 $F^{-1} \in \mathcal{I}$.

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. $\|T_C f\| \le \|f\|$ *for each* $f \in L^1$.
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 inductively for all $t \ge 0$ by $(T_C)^0(f) := f$ and $(T_C)^{t+1}(f) := (T_C)^t(T_C(f))$ for all f.

Proof. For the first point, note that for any $f \in L^1$,

$$\|T_C f\| = \int_0^1 |T_C f(x)| dx \le \int_0^1 \int_0^1 c(x',x)|f(x')| dx' dx = \int_0^1 |f(x')| dx' = \|f\| < \infty.$$

Thus, $T_C : L^1 \to L^1$. Moreover, since T_C is clearly linear, the above ensures that it is also continuous.

For the second point, consider $f \in \mathcal{I}$. Since C is assortative, $T_C f(x) \ge T_C f(x')$ for all $x \ge x'$, so that $T_C f$ is weakly increasing. To show that $T_C f$ is absolutely continuous, note that for each

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

$x, x' \in (0,1)$,

$$\begin{aligned} T_C f(x) &= \int_0^1 c(y,x) f(y) dy \\ &= \int_0^1 \left(\int_{x'}^{x} c_2(y,z) dz + c(y,x') \right) f(y) dy \\ &= \int_{x'}^{x} \int_0^1 c_2(y,z) f(y) dy dz + T_C f(x') \end{aligned}$$

where c_2 denotes the partial derivative of c with respect to the second argument, which exists almost everywhere by the absolute continuity assumption on c. Thus $T_C f$ is absolutely continuous with $(T_C f)'(z) = \int_0^1 c_2(y,z) f(y) dy$ for each z.

Finally, for the third point, fix any $f \in L^1$ and $\gamma \in [0,1)$. Then for any $\tau > \tau'$,

$$\| \sum_{t=0}^{\tau} \gamma^t (T_C)^t f - \sum_{t=0}^{\tau'} \gamma^t (T_C)^t f \| \le \sum_{t=\tau'+1}^{\tau} \gamma^t \| (T_C)^t f \| \le \sum_{t=\tau'+1}^{\tau} \gamma^t \|f\| \le \frac{\gamma^{\tau'+1}}{1-\gamma} \|f\|,$$

which vanishes as $\tau' \to \infty$. Thus, the sequence is Cauchy. Since the space L^1 is complete, this yields the desired result. □

A.2 Proof of Lemma 1

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

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

for each $x \in (0,1)$. Note that by construction, we have $h = F^{-1} + \gamma T_C h + \beta T_{C_I} h$, where C_I denotes the independent interaction structure. Moreover, h is strictly increasing and continuous, since $(T_C)^t F^{-1}$ is weakly increasing and continuous for each $t \ge 0$ and strictly increasing for $t = 0$.

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 a strategy. Moreover, s inherits strict monotonicity and continuity from h and F. Finally, s is a Nash equilibrium because for each type θ and $x = F(\theta)$, we have

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

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

$$\hat{h} = F^{-1} + \gamma T_C \hat{h} + \beta T_{C_I} \hat{h} \tag{10}$$

Iterating (10) yields

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

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

$$\begin{aligned}
\|\hat{h} - h\| &\leq \| \sum_{t=0}^{\tau} \gamma^t (T_C)^t \left(F^{-1} - \beta T_{C_I} \hat{h} - F^{-1} + \beta T_{C_I} h \right) \| + \gamma^{\tau+1} \|(T_C)^{\tau+1} (\hat{h} - h)\| \\
&\leq \| \sum_{t=0}^{\tau} \gamma^t (T_C)^t \left(\beta T_{C_I} (h - \hat{h}) \right) \| + \gamma^{\tau+1} \|\hat{h} - h\| \to \| \sum_{t=0}^{\infty} \gamma^t (T_C)^t \left(\beta T_{C_I} (h - \hat{h}) \right) \|
\end{aligned}$$

as $\tau \to \infty$. But integrating both sides of (10) with respect to x, we obtain $\int_0^1 \hat{h}(x)dx = T_{C_I} \hat{h} = \frac{\mu}{1-\gamma-\beta}$, and analogously $T_{C_I} h = \frac{\mu}{1-\gamma-\beta}$ from the best-response condition for h. Thus, $\|\hat{h} - h\| = 0$, whence $\hat{s} = s$. □

B Proof of Theorem 1

The proof of Theorem 1 proceeds as follows. In Section B.1, we first consider any abstract order $\succsim$ over $\mathcal{I}$ and define its dual order $\succsim^*$ over the space of interaction structures. We establish that when $\succsim$ satisfies three basic properties, then the $\succsim^*$ order over interaction structures is equivalent to Nash equilibrium behavior being ordered by $\succsim$. Section B.2 shows that the mean-preserving spread order satisfies these three properties, using tools from majorization theory. Section B.3 shows that the more-assortative order is the dual order of the mean-preserving spread order. Finally, section B.4 combines these results to yield Theorem 1.

B.1 Dual orders over $\mathcal{C}$

Consider any preorder (i.e., reflexive and transitive binary relation) $\succsim$ over $\mathcal{I}$. Define the ***dual order*** of $\succsim$ to be the preorder $\succsim^*$ over $\mathcal{C}$ given by $C_1 \succsim^* C_2$ if and only if $T_{C_1} f \succsim T_{C_2} f$ for all $f \in \mathcal{I}$.

Theorem B.1 below shows that as long as $\succsim$ satisfies three basic properties, its dual order $\succsim^*$ over interaction structures is characterized by Nash equilibrium behavior that is ordered according to $\succsim$.

Definition 8. Preorder $\succsim$ over $\mathcal{I}$ is called:

1. ***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 $\alpha_1 f + \alpha_2 h \succsim \alpha_1 g + \alpha_2 h$.
2. ***continuous*** if $f_n \succsim g_n$ for each n and $f_n \to f \in \mathcal{I}, g_n \to g \in \mathcal{I}$ imply $f \succsim g$.
3. ***isotone*** if $f \succsim g$ implies $T_C f \succsim T_C g$ for any $C \in \mathcal{C}$.

For any F, γ, β, let $H^{-1}_{F,C,\gamma,\beta}$ denote the inverse cdf of the Nash action distribution under (F, C, γ, β). By Lemma 1, we have

$$H^{-1}_{F,C,\gamma,\beta} = \sum_{t \geq 0} \gamma^t (T_C)^t F^{-1} + \frac{\beta \mathbb{E}_F[\theta]}{(1-\gamma)(1-\gamma-\beta)}. \quad (11)$$

Theorem B.1. *Suppose that $\succsim$ is a linear, continuous, and isotone preorder over $\mathcal{I}$. Then the following are equivalent:*

1. $C_1 \succsim^* C_2$

2. $H^{-1}_{F,C_1,\gamma,\beta} \succsim H^{-1}_{F,C_2,\gamma,\beta}$ *for all* (F,γ,β).

Proof. **(1) $\implies$ (2)**: Suppose that $C_1 \succsim^* C_2$ 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 $(T_{C_1})^t f \succsim (T_{C_2})^t f$ for all t. Since $C_1 \succsim^* C_2$, the claim for $t=1$ is true by definition. Suppose the claim holds for some $t \geq 1$. Then

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

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

Next, note that linearity of $\succsim$ and $C_1 \succsim^* C_2$ implies

$$\sum_{t\geq 0}^{\tau} \gamma^t (T_{C_1})^t F^{-1} \succsim \left(\gamma^\tau (T_{C_2})^\tau + \sum_{t\geq 0}^{\tau-1} \gamma^t (T_{C_1})^t \right) F^{-1} \succsim$$

$$\succsim \left(\sum_{t=\tau-1,\tau} \gamma^t (T_{C_2})^t + \sum_{t\geq 0}^{\tau-1} \gamma^t (T_{C_1})^t \right) F^{-1} \succsim \cdots \succsim \sum_{t\geq 0}^{\tau} \gamma^t (T_{C_2})^t F^{-1}$$

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

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

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

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

where $\mu = \mathbb{E}_F[\theta]$. By (11), this is equivalent to $H^{-1}_{F,C_1,\gamma,\beta} \succsim H^{-1}_{F,C_2,\gamma,\beta}$, as required.

(2) $\implies$ (1): Suppose that $H^{-1}_{F,C_1,\gamma,\beta} \succsim H^{-1}_{F,C_2,\gamma,\beta}$ for all (F,γ,β). For any $f \in \mathcal{I}$, applying this with $F = f^{-1}$ and $\beta = 0$, (11) yields that for any $\gamma \in (0,1)$,

$$H^{-1}_{F,C_1,\gamma,\beta} = \sum_{t\geq 0} \gamma^t T_{C_1} f \succsim \sum_{t\geq 0} \gamma^t T_{C_2} f = H^{-1}_{F,C_2,\gamma,\beta}.$$

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

$$T_{C_1} f + \sum_{t\geq 2} \gamma^t (T_{C_1})^t f \succsim T_{C_2} f + \sum_{t\geq 2} \gamma^t (T_{C_2})^t f. \tag{12}$$

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

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

so that, as $\gamma \to 0$, $T_{C_i} f + \sum_{t\geq 2} \gamma^t (T_{C_i})^t f \to T_{C_i} f$. Thus, by continuity of $\succsim$, (12) yields $T_{C_1} f \succsim T_{C_2} f$. As this is true for all $f \in \mathcal{I}$, we have $C_1 \succsim^* C_2$, as required. ☐

B.2 Linearity, Continuity, and Isotonicity of $\succsim_m$

Next, we show that the mean-preserving spread order $\succsim_m$, when viewed as an order over $\mathcal{I}$, is indeed linear, continuous, and isotone.

Define $\succsim_m$ 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 if F, $G \in \mathcal{F}$, then $F \succsim_m G$ if and only if $F^{-1} \succsim_m G^{-1}$. The following characterization of $\succsim_m$ is standard:

Lemma B.1. *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* $y = 0$.

Proof. See, e.g., Section 3.A.1 in Shaked and Shanthikumar (2007). □

The next two lemmas verify that $\succsim_m$ over $\mathcal{I}$ satisfies the three basic properties:

Lemma B.2. *$\succsim_m$ is a preorder over $\mathcal{I}$ that is linear and continuous.*

Proof. It is clear from the definition that $\succsim_m$ is reflexive and transitive; moreover, by part 2 of Lemma B.1 $\succsim_m$ is linear. To check that $\succsim_m$ is continuous, take sequences $f_n \to f, g_n \to g$ in $\mathcal{I}$ such that $f_n \succsim_m g_n$ for each n. For any $y \in (0,1)$, we have

$$|\int_y^1 f(x)dx - \int_y^1 f_n(x)dx| \leq \int_y^1 |f(x) - f_n(x)|dx \leq \|f - f_n\| \to 0$$

and likewise $|\int_y^1 g(x)dx - \int_y^1 g_n(x)dx| \to 0$. Since $\int_y^1 f_n(x)dx \geq \int_y^1 g_n(x)dx$ and $\int_0^1 f_n(x)dx = \int_0^1 g_n(x)dx$ for each n, this implies $\int_y^1 f(x)dx \geq \int_y^1 g(x)dx$ and $\int_0^1 f(x)dx = \int_0^1 g(x)dx$. Thus, $f \succsim_m g$ by Lemma B.1. □

Lemma B.3. *Preorder $\succsim_m$ over $\mathcal{I}$ is isotone.*

Proof. We first show that $\succsim_m$ is isotone. Take any $f, g \in \mathcal{I}$ such that $f \succsim_m g$ and any $C \in \mathcal{C}$. We assume first that f and g are bounded. To show that $T_C f \succsim_m T_C g$, consider any convex function ϕ such that $\phi \circ T_C f, \phi \circ T_C g \in L^1$. We want to show that

$$\int_0^1 \phi\left(T_C f(x)\right) dx \geq \int_0^1 \phi\left(T_C g(x)\right) dx.$$

We prove this by finite-dimensional approximation. Consider any $k \in \mathbb{N}$. Define a k-dimensional vector f^k by

$$f_i^k := k \int_{(i-1)k^{-1}}^{ik^{-1}} f(x)dx$$

for each $i \in \{1, ..., k\}$ and analogously for vector $g^k = (g_i^k)_{i=1,...,k}$. Moreover, define a symmetric $k \times k$ matrix C^k by

$$C_{ij}^k := k \int_{(i-1)k^{-1}}^{ik^{-1}} \int_{(j-1)k^{-1}}^{jk^{-1}} c(x, x')dxdx'$$

for each $i, j \in \{1, ..., k\}$. Note that C^k is doubly stochastic, i.e., $C^k_{ij} \geq 0$ and $\sum_{l=1}^k C^k_{il} = \sum_{l=1}^k C^k_{lj} = 1$ for each i, j.

Let C^k_i denote the ith row of C^k. In Section B.5, we use tools from majorization theory to prove the following claim:

Claim 1. *For any convex function ϕ, we have $\sum_{i=1}^k \phi(C^k_i \cdot f^k) \geq \sum_{i=1}^k \phi(C^k_i \cdot g^k)$.*

We can identify vectors $f^k, C^k f^k$ with simple functions in L^1 by setting

$$f^k(x) := f^k_i, \;\; (C^k f^k)(x) := C^k_i \cdot f^k$$

for each $x \in (0, 1)$ and $i \in \{1, .., k\}$ such that $x \in [(i-1)k^{-1}, ik^{-1})$; and likewise for $g^k, C^k g^k$. Moreover, we can identify matrix C^k with the interaction structure whose density c^k is given by

$$c^k(x, x') = kC_{ij}$$

for all $x, x' \in (0, 1)$ and i, j such that $x \in \big[(i-1)k^{-1}, ik^{-1}\big)$, $x' \in \big[(j-1)k^{-1}, jk^{-1}\big)$. With these identifications, we have

$$(C^k f^k)(x) = T_{C^k} f^k(x) = T_{C^k} f(x)$$

for each $x \in (0, 1)$, and likewise for $C^k g^k$. Moreover, Claim 1 yields

$$\int_0^1 \phi\left(T_{C^k} f(x)\right) dx \geq \int_0^1 \phi\left(T_{C^k} g(x)\right) dx. \tag{13}$$

Note that $T_{C^k} f(x) \to T_C f(x)$ as $k \to \infty$ for all x. Indeed, for each y, the value of $C^k(y|x)$ can be written as a weighted average of $C(y'|x')$ among pairs (x', y') with $|x' - x|, |y - y'| \leq \frac{1}{k}$. Since $C(y|x)$ is jointly continuous in (y, x) this implies that $\lim_k C^k(y|x) = C(y|x)$. Since $C(\cdot|x)$ is atomless, this yields the weak convergence of conditional distributions $C^k(\cdot|x)$ to $C(\cdot|x)$. As f is continuous and bounded, this ensures $T_{C^k} f(x) \to T_C f(x)$.

Since ϕ is continuous, this implies $\phi(T_{C^k} f(x)) \to \phi(T_C f(x))$ for each x. Hence,

$$\int_0^1 \phi\left(T_{C^k} f(x)\right) dx \to \int_0^1 \phi\left(T_C f(x)\right) dx,$$

where the bounded convergence theorem applies because $\sup_{k,x} |\phi(T_{C^k} f(x))| \leq \sup_x |\phi(f(x))| < \infty$. Analogously, we have $\int_0^1 \phi\left(T_{C^k} g(x)\right) dx \to \int_0^1 \phi\left(T_C g(x)\right) dx$. Thus, (13) implies

$$\int_0^1 \phi\left(T_C f(x)\right) dx \geq \int_0^1 \phi\left(T_C g(x)\right) dx,$$

as desired.

Next, consider arbitrary $f, g \in \mathcal{I}$ such that $f \succsim_m g$. For each $n \in \mathbb{N}$, define a bounded function $f_n \in \mathcal{I}$ by

$$f_n(x) = \begin{cases} n \int_0^{\frac{1}{n}} f(y) dy \text{ if } x \in (0, \frac{1}{n}) \\ f(x) \text{ if } x \in [\frac{1}{n}, \frac{n-1}{n}] \\ n \int_{\frac{n-1}{n}}^1 f(y) dy \text{ if } x \in (\frac{n-1}{n}, 1), \end{cases} \tag{14}$$

and analogously for g_n. Note that $f_n \succsim_m g_n$ for each n, so that by the previous part of the proof we have $T_C f_n \succsim_m T_C g_n$. Moreover, $f_n \to f$ and $g_n \to g$, so that $T_C f_n \to T_C f$ and $T_C g_n \to T_C g$ since T_C is a continuous operator. By continuity of $\succsim_m$, this yields $T_C f \succsim_m T_C g$. Hence, $\succsim_m$ is isotone. $\square$

B.3 $\succsim_{MA}$ is the dual order of $\succsim_m$

Finally, we show that the more-assortative order $\succsim_{MA}$ is in fact the dual order of $\succsim_m$.

Lemma B.4. *$C_1 \succsim_m^* C_2$ if and only if $C_1 \succsim_{MA} C_2$.*

Proof. By definition, $C_1 \succsim_m^* C_2$ if and only if for any $f \in \mathcal{I}$, $T_{C_1} f \succsim_m T_{C_2} f$, which by definition of $\succsim_m$ means $\int_y^1 T_{C_1} f(x)dx \geq \int_y^1 T_{C_2} f(x)dx$ for all $y \in (0,1)$ with equality if $y = 0$. But by Fubini's theorem,

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

Hence, $C_1 \succsim_m^* C_2$ is equivalent to $C_1(\cdot \mid x \geq y)$ first-order stochastically dominating $C_2(\cdot \mid x \geq y)$ for any $y \in (0,1)$, i.e., to $C_1 \succsim_{MA} C_2$. $\square$

B.4 Completing the Proof of Theorem 1

Combining Lemmas B.2-B.4 and Theorem B.1, we have that $C_1 \succsim_{MA} C_2$ if and only if $H^{-1}_{F,C_1,\gamma,\beta} \succsim_m H^{-1}_{F,C_2,\gamma,\beta}$ for all (F,γ,β), which in turn is equivalent to $H_1^{F,\gamma,\beta} \succsim_m H_2^{F,\gamma,\beta}$, as claimed. $\square$

B.5 Proof of Claim 1

Finally, we establish Claim 1 from the proof of Lemma B.3. The proof uses tools from majorization theory (Marshall, Olkin, and Arnold, 2010). Let $\mathbb{R}^n_\uparrow \subseteq \mathbb{R}^n$ denote the space of all n-dimensional vectors x that have weakly increasing coordinates (i.e., $x_1 \leq \ldots \leq x_n$). For $x, y \in \mathbb{R}^n_\uparrow$, we say that x ***majorizes*** y if

$$\sum_{j=k}^n x_j \geq \sum_{j=k}^n y_j \ \forall k = 1, \ldots n,$$

with equality for $k = 1$. The above condition is a finite-dimensional analog of the $\succsim_m$ order over $\mathcal{I}$, so that we also use $x \succsim_m y$ to denote the majorization order.

Our proof of Claim 1 will make use of the following characterization of $\succsim_m$ due to Hardy, Littlewood, and Polya. To state this, recall that a transposition over $\mathbb{R}^n$ is a permutation of the coordinates $\{1, \ldots, n\}$ that fixes all but at most two coordinates. Let $\mathcal{T}$ be the set of all matrix representations of transpositions on $\mathbb{R}^n$.[45] For any $T \in \mathcal{T}$ and $\lambda \in [0,1]$, define matrix T_λ by $T_\lambda := \lambda I + (1-\lambda)T$. That is, if $i < j$ are the unique coordinates permuted under T, then for any $x \in \mathbb{R}^n_\uparrow$, we have $(Tx)_k = x_k$ for all $k \neq i, j$ and $(T_\lambda x)_i - x_i = (1-\lambda)(x_j - x_i) = -(T_\lambda x)_j + x_j$. Thus, T_λ transfers amount $(1-\lambda)(x_j - x_i)$ from coordinate j to coordinate i. We have the following alternative characterization of $\succsim_m$:

Lemma B.5 (Hardy, Littlewood, Polya)**.** *Let $x, y \in \mathbb{R}^n_\uparrow$. The following are equivalent:*

[45] The identity matrix is an element of $\mathcal{T}$.

1. *$x \succsim_m y$*

2. *There exist finite sequences $T^1, T^2, \ldots, T^m \in \mathcal{T}$ and $\lambda_1, \ldots, \lambda_m \in [0,1]$ such that $T^k_{\lambda_k} \cdots T^1_{\lambda_1} x \in \mathbb{R}^n_\uparrow$ for all $k = 1, 2, \ldots, m$, and*
$$T^m_{\lambda_m} \cdots T^1_{\lambda_1} x = y.$$

Proof. See Lemma B.1 in Marshall, Olkin, and Arnold (2010) (p. 32). □

Returning to the proof of Claim 1, note that the construction in the proof of Lemma B.3 satisfies the following two properties:

1. $f^k, g^k \in \mathbb{R}^k_\uparrow$ and $f^k \succsim_m g^k$ for every k (since $f \succsim_m g$ and by Lemma B.1).

2. C^k is monotone for every k. That is, the sequence of row vectors, $C^k_1, C^k_2, \ldots, C^k_k$, is decreasing with respect to first-order stochastic dominance (i.e., $\sum_{l=1}^m C^k_{il} \geq \sum_{l=1}^m C^k_{jl}$ for each $i < j$ and $m = 1, ..., k$).

Given this, Claim 1 follows immediately from the following result, which we prove by invoking Lemma B.5:

Lemma B.6. *Fix any doubly-stochastic symmetric monotone matrix $C \in \mathbb{R}^{n \times n}$ with row vectors $c_1, \ldots, c_n$ and any $x, y \in \mathbb{R}^n_\uparrow$ such that $x \succsim_m y$. Then for any convex function $\phi : \mathbb{R} \to \mathbb{R}$, we have $\sum_{j=1}^n \phi(c_j \cdot x) \geq \sum_{j=1}^n \phi(c_j \cdot y)$.*

Proof. Fix any convex function $\phi : \mathbb{R} \to \mathbb{R}$.

We first show that for any $T \in \mathcal{T}$, $\lambda \in [0,1]$ and $z \in \mathbb{R}^n_\uparrow$ such that $T_\lambda z \in \mathbb{R}^n_\uparrow$, we have

$$\sum_{j=1}^n \phi(c_j \cdot z) \geq \sum_{j=1}^n \phi(c_j \cdot (T_\lambda z)).$$

To see this, let $i < k$ be the unique coordinates being permuted under T and let $\Delta := (1 - \lambda)(z_i - z_k) \geq 0$. Note that for all j, $c_j \cdot z = c_j \cdot (T_\lambda z) + \Delta \left(c^i_j - c^k_j\right)$.

Define ϕ' by

$$\phi'(u) := \sup \partial\phi(u) = \sup\{\gamma \in \mathbb{R} : \phi(v) \geq \phi(u) + \gamma(v - u)\ \forall v \in \mathbb{R}\}$$

for all $u \in \mathbb{R}$. It is well-known that ϕ' exists and is weakly increasing in u. Then

$$\sum_{j=1}^n \phi(c_j \cdot z) = \sum_{j=1}^n \phi\left(c_j \cdot (T_\lambda z) + \Delta(c^i_j - c^k_j)\right) \geq \sum_{j=1}^n \phi(c_j \cdot (T_\lambda z)) + \phi'(c_j \cdot (T_\lambda z))\Delta(c^i_j - c^k_j),$$

where $\phi'(c_1 \cdot (T_\lambda z)) \geq \phi'(c_2 \cdot (T_\lambda z)) \geq \cdots \geq \phi'(c_n \cdot (T_\lambda z))$.

Moreover, denoting the column vectors of C by $c^1, c^2, \ldots, c^n$, symmetry and monotonicity of C implies $c^1 \geq_{FOSD} c^2 \geq_{FOSD} \cdots \geq_{FOSD} c^n$. But then, we must have $\sum_{j=1}^n \phi'(c_j \cdot (T_\lambda z))\Delta(c^i_j - c^k_j) \geq 0$, whence

$$\sum_{j=1}^n \phi(c_j \cdot z) \geq \sum_{j=1}^n \phi(c_j \cdot (T_\lambda z)),$$

as claimed.

Now consider $x, y \in \mathbb{R}^n_\uparrow$ such that $x \succsim_m y$. By Lemma B.5, there exist $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$ and $T^1, \ldots T^m \in \mathcal{T}$ such that

$$T^k_{\lambda_k} \cdots T^1_{\lambda_1} x \in \mathbb{R}^n_\uparrow$$

for all $k = 1, 2, \ldots, m$ and

$$T^m_{\lambda_m} \cdots T^1_{\lambda_1} x = y.$$

The above argument shows that for any $k \leq m-1$,

$$\sum_{j=1}^n \phi\left(c_j \cdot x\right) \geq \sum_{j=1}^n \phi\left(c_j \cdot \left(T^k_{\lambda_k} \cdots T^1_{\lambda_1} x\right)\right) \geq \sum_{j=1}^n \phi\left(c_j \cdot \left(T^{k+1}_{\lambda_{k+1}} \cdots T^1_{\lambda_1} x\right)\right).$$

As a result,

$$\sum_{j=1}^n \phi\left(c_j \cdot x\right) \geq \sum_{j=1}^n \phi\left(c_j \cdot y\right),$$

as required. □

C Proofs for Section 3.3

C.1 Proof of Proposition 1

Take any F_1, F_2 such that $F_1 \succsim_m F_2$. Then $F_1^{-1} \succsim_m F_2^{-1}$. First, we inductively show that for each t, $(T_C)^t F_1^{-1} \succsim_m (T_C)^t F_2^{-1}$. Indeed, supposing that the claim is true at t, isotonicity of $\succsim_m$ implies

$$(T_C)^{t+1} F_1^{-1} = T_C (T_C)^t F_1^{-1} \succsim_m T_C (T_C)^t F_2^{-1} = (T_C)^{t+1} F_2^{-1},$$

as required. Next, since $\succsim_m$ is linear, we have

$$\sum_{t=0}^{\tau} \gamma^t (T_C)^t F^{-1} \succsim_m \sum_{t=0}^{\tau} \gamma^t (T_C)^t \hat{F}^{-1}$$

for all $\tau \in \mathbb{N}$. Since $\lim_{\tau \to \infty} \sum_{t=0}^{\tau} \gamma^t (T_C)^t F_i^{-1} = \sum_{t=0}^{\infty} \gamma^t (T_C)^t F_i^{-1}$ for each $i = 1, 2$ (Lemma A.1), continuity and linearity of $\succsim_m$ then yields

$$H_1^{-1} = \sum_{t \geq 0} \gamma^t (T_C)^t F_1^{-1} + \frac{\beta \mathbb{E}_{F_1}[\theta]}{(1-\gamma)(1-\gamma-\beta)} \succsim_m \sum_{t \geq 0} \gamma^t (T_C)^t F_2^{-1} + \frac{\beta \mathbb{E}_{F_2}[\theta]}{(1-\gamma)(1-\gamma-\beta)} = H_2^{-1},$$

whence $H_1 \succsim_m H_2$, as claimed. □

C.2 Proof of Proposition 2

Suppose that $\gamma_1 \geq \gamma_2$. We will establish the stronger result that $H_1 \succsim_d H_2$, where $\succsim_d$ denotes the dispersiveness order defined in section 7.2.

Note that $\succsim_d$ viewed as an order over $\mathcal{I}$ is linear and continuous (see Appendix H.1). Applying linearity of $\succsim_d$ together with the fact that $f \succsim_d \alpha f$ for any $f \in \mathcal{I}$ and $\alpha \in [0, 1]$, we obtain

$$\sum_{t=0}^{\tau} \gamma_1^t (T_C)^t F^{-1} \succsim_d \sum_{t=0}^{\tau} \gamma_1^t (T_C)^t (\gamma_2/\gamma_1)^t F^{-1} = \sum_{t=0}^{\tau} \gamma_2^t (T_C)^t F^{-1}$$

for any $\tau \in \mathbb{N}$.

Moreover, for any $f, g \in \mathcal{I}$ such that $f \succsim_d g$ and for any constants η_1, η_2, we have $f + \eta_1 \succsim_d g + \eta_2$. Since $\lim_{\tau\to\infty} \sum_{t=0}^{\tau} \gamma_i^t (T_C)^t F^{-1} = \sum_{t\geq 0} \gamma_i^t (T_C)^t F^{-1}$ for each $i = 1, 2$ (Lemma A.1), this observation together with continuity of $\succsim_d$ then implies that

$$H_1^{-1} = \sum_{t\geq 0} \gamma_1^t (T_C)^t F^{-1} + \frac{\beta_1 \mathbb{E}_F[\theta]}{(1-\gamma_1)(1-\gamma_1-\beta_1)} \succsim_d \sum_{t\geq 0} \gamma_2^t (T_C)^t F^{-1} + \frac{\beta_2 \mathbb{E}_F[\theta]}{(1-\gamma_2)(1-\gamma_2-\beta_2)} = H_2^{-1},$$

whence $H_1 \succsim_d H_2$, as claimed. □

D Proofs for Section 5.1

D.1 Proof of Lemma 3

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

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

Indeed, if $\hat{H} = \hat{F} \circ s^{-1}$ for the Nash equilibrium strategy profile s, then (15) follows from the best-response condition (1) via the change of variables $x = \hat{F}(\theta)$. Conversely, if (15) holds, then the corresponding strategy profile $s = \hat{H}^{-1} \circ \hat{F}$ satisfies the best-response condition (1).

(1) $\implies$ (2): Suppose $\hat{C} = C_I$ and consider any action distribution $\hat{H} \in \mathcal{F}$. Define a type distribution $\hat{F}$ by

$$\hat{F}^{-1}(x) := \hat{H}^{-1}(x) - (\gamma + \beta) \int \hat{H}^{-1}(z) dz$$

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) $\implies$ (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 under $(\hat{F}, \hat{C}, \gamma, \beta)$. 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) $x > y > y'$.[46]

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 $z \in (0,1)$.[47]

[46]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.

[47]Concretely, we can define $\hat{H}_n$ by

$$\hat{H}_n^{-1}(z) := \begin{cases} \frac{n-1}{n} \left(\frac{z}{x}\right)^n & \text{if } z \leq x \\ \frac{1}{n(1-x)}(z-1) + 1 & \text{if } z > x. \end{cases}$$

Then, under both (i) and (ii), we have that as $n \to \infty$,

$$\hat{H}_n^{-1}(y) - \hat{H}_n^{-1}(y') \to 0;$$
$$\int \hat{H}_n^{-1}(z)\, d\left(\hat{C}(z|y) - \hat{C}(z|y')\right) \to \hat{C}(x|y') - \hat{C}(x|y) > 0.$$

Hence, setting $\hat{H} = \hat{H}_n$ for some sufficiently large n, we have that

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

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

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

This is impossible since $y > y'$.

Finally, for the moreover part, note that if $\hat{H}$ is the Nash action distribution at $(\hat{F}, C_I, \gamma, \beta)$, then by (15) we have $\hat{H}^{-1}(x) = \hat{F}^{-1}(x) + (\gamma + \beta) \int \hat{H}^{-1}(z)\, dz$ for all $x \in (0,1)$, which uniquely pins down $\hat{F}$. □

D.2 Proof of Theorem 2

Part (1): Fix any (P, γ, β). We first prove the existence of an LPE such that $\hat{C}_\theta = C_I$ for all θ. Let the strategy profile s be given by (6); i.e., $s(\theta_0) = \sum_{t=0}^{\infty} (\gamma + \beta)^t \mathbb{E}_P[\theta_t|\theta_0]$ for each type θ_0. Note that s coincides with the Nash equilibrium at (P, γ', β'), where $\gamma' = \gamma + \beta$ and $\beta' = 0$. Thus, s is L^1, strictly increasing, and continuous.

For each type θ, the local action distribution under (s, P) is given by $H_\theta^{s,P}(a) = \int_0^1 \mathbb{1}_{s(\theta') \le a}\, dP(\theta'|\theta)$ for all a. Since $H_\theta^{s,P} \in \mathcal{F}$, Lemma 3 implies that there exists a type distribution $\hat{F}_\theta \in \mathcal{F}$ such that $H_\theta^{s,P}$ is the Nash action distribution at $(\hat{F}_\theta, C_I, \gamma, \beta)$. Thus, letting $\hat{s}_\theta$ denote the Nash equilibrium strategy profile at $(\hat{F}_\theta, C_I, \gamma, \beta)$ and $\hat{P}_\theta = (\hat{F}_\theta, C_I)$, Nash rationalization is satisfied by definition and $H_\theta^{s,P} = H_\theta^{\hat{s}_\theta, \hat{P}_\theta}$, proving part (a) of observational consistency. It remains to prove part (b) of observational consistency, i.e., that $s(\theta) = \hat{s}_\theta(\theta)$. To see this, note that

$$\begin{aligned} s(\theta) &= \theta + (\gamma + \beta)\mathbb{E}_P[s(\theta')|\theta] \quad \text{(by construction of } s) \\ &= \theta + (\gamma + \beta)\mathbb{E}_{\hat{P}_\theta}[\hat{s}_\theta(\theta')|\theta] \quad \text{(because } H_\theta^{s,P} = H_\theta^{\hat{s}_\theta, \hat{P}_\theta}) \\ &= \theta + \gamma \mathbb{E}_{\hat{P}_\theta}[\hat{s}_\theta(\theta')|\theta] + \beta \mathbb{E}_{\hat{P}_\theta}[\hat{s}_\theta(\theta')] \quad \text{(because } \hat{C}_\theta = C_I) \\ &= \hat{s}_\theta(\theta) \quad \text{(because } \hat{s}_\theta \text{ is Nash at } \hat{P}_\theta). \end{aligned}$$

Finally, to show uniqueness, note that by the derivation in the main text, the ANLPE strategy profile s is uniquely pinned down by (6). Given this, Nash rationalization and observational consistency uniquely pins down the perceived type distribution $\hat{F}_\theta$ for each θ, because by Lemma 3 there is a unique $\hat{F}_\theta$ such that $H_\theta^{s,P}$ is Nash at $(\hat{F}_\theta, C_I, \gamma, \beta)$. Finally, the uniqueness of the perceived strategy profile $\hat{s}_\theta$ for each θ follows from Nash rationalization and the uniqueness of the Nash equilibrium at $\hat{P}_\theta$.

Part (2): 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 LPE 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) $\Longrightarrow$ (1)" direction of Lemma 3, 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) $\Longrightarrow$ (2)" direction of Lemma 3, 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

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

We claim that in environment (F, C_I, γ, β), there exists no LPE with $\hat{C}_\theta = \hat{C}$. Indeed, suppose for a contradiction that there is an LPE 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 LPE coincides with the Nash equilibrium profile (see Lemma I.1). 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 requirement that by Nash rationalization, $\hat{G}_\theta$ 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 his 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 his own quantile to be $\hat{x}$ and the interaction structure to be $\hat{C}$, θ perceives the xth quantile among his 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

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

where the final two equalities follow from (18) and the definition of $\tilde{H}$ above. Combining this with equation (17), we have

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

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 $y > y'$. □

E Proofs for Sections 5.2 and 5.3

E.1 Proof of Corollary 1

By the proof of Theorem 2, the ANLPE strategy profile s^{AN} is given by (6), i.e., $s(\theta_0) = \sum_{t=0}^\infty (\gamma + \beta)^t \mathbb{E}_P[\theta_t|\theta_0]$ for each type θ_0.

Part 1: By Nash rationalization, each type θ_0's perceived strategy profile $\hat{s}_{\theta_0}$ is the Nash equilibrium under $(\hat{F}_{\theta_0}, C_I, \gamma, \beta)$. Hence,

$$\begin{aligned}\frac{\hat{\mu}_{\theta_0}}{1-\gamma-\beta} = \mathbb{E}_{\hat{P}_{\theta_0}}[\hat{s}_{\theta_0}(\theta)] &= \mathbb{E}_{\hat{P}_{\theta_0}}[\hat{s}_{\theta_0}(\theta_1)|\theta_0] = \mathbb{E}_P[s^{AN}(\theta_1)|\theta_0] = \\ &= \mathbb{E}_P[\theta_1 + (\gamma+\beta)\mathbb{E}_P[s^{AN}(\theta_1)|\theta_0]|\theta_0] = \cdots = \sum_{t=0}^\infty (\gamma+\beta)^t \mathbb{E}_P[\theta_{t+1}|\theta_0],\end{aligned}$$

where the first equality holds by Lemma 1, the second because $\hat{C}_\theta = C_I$, the third by observational consistency, the fourth by (6), and the convergence of the infinite sum follows from a similar argument as in the proof of Lemma 1.

Part 2: Observe that by (6) and Lemma 1, the ANLPE strategy profile s^{AN} at (P, γ, β) coincides with the Nash equilibrium strategy profile at (P, γ', β'), where $\gamma' := \gamma + \beta$ and $\beta' := 0$. Then the claim is immediate from Proposition 2 and the fact (Lemma 1) that the average action under Nash depends only on the sum of local and global coordination motives. □

E.2 Proof of Theorem 3

As noted in the proof of Corollary 1, the ANLPE action distribution at (P, γ, β) coincides with the Nash action distribution at (P, γ', β'), where $\gamma' = \gamma + \beta$ and $\beta' = 0$. Given this, the "(1) ⇒ (2)" direction is immediate from the "(1) ⇒ (2)" direction of Theorem 1. The "(2) ⇒ (1)" direction also follows from the proof of Theorem 1, by observing that in the proof of the "(2) ⇒ (1)" direction of that theorem it is without loss to assume that $\beta = 0$.

To show the equivalence of (1) and (3), recall from Corollary 1 that each type θ's perceived population mean under the ANLPE at (P, γ, β) is given by $\hat{\mu}_\theta = (1-\gamma-\beta)\sum_{t=0}^\infty (\gamma+\beta)^t \mathbb{E}_P[\theta_{t+1}|\theta_0 = \theta]$. Thus, denoting by $m_i^{F,\gamma,\beta}$ the inverse cdf of $M_i^{F,\gamma,\beta}$, we have for each quantile $x \in (0,1)$ that

$$m_i^{F,\gamma,\beta}(x) = (1-\gamma-\beta)\sum_{t\geq 0} (\gamma+\beta)^t (T_{C_i})^{t+1} F^{-1}(x).$$

Hence, if $C_1 \succsim_{MA} C_2$, then $m_1 \succsim_m m_2$ for all (F, γ, β) since $\succsim_m$ is linear, continuous, and isotone by Lemmas B.2 and B.3. Conversely, if $m_1^{F,\gamma,\beta} \succsim_m m_2^{F,\gamma,\beta}$ for all (F, γ, β), then by linearity and continuity of $\succsim_m$, it follows that $T_{C_1}F^{-1} \succsim_m T_{C_2}F^{-1}$ for any F, which implies $C_1 \succsim_{MA} C_2$ by Lemma B.4. □

E.3 Proof of Corollary 2

We instead prove the following proposition, which shows that local coordination motives and assortativity have a complementary effect on *Nash* action dispersion. Since the ANLPE strategy profile s^{AN} at (P, γ, β) coincides with the Nash equilibrium at (P, γ', β'), where $\gamma' = \gamma + \beta$ and $\beta' = 0$, this then immediately implies (8), i.e., a greater increase in action dispersion from C_2 to C_1 under ANLPE than under Nash.

Proposition E.1 (Complementarity of assortativity and local coordination motives). *Let s_{ij} denote the Nash strategy profile under $(F, C_i, \gamma_j, \beta_j)$ for $i, j = 1, 2$. If $\gamma_1 \geq \gamma_2$ and $C_1 \succsim_{MA} C_2$, then for all types θ^*,*

$$\mathbb{E}_F[s_{11}(\theta) - s_{21}(\theta) | \theta \geq \theta^*] \geq \mathbb{E}_F[s_{12}(\theta) - s_{22}(\theta) | \theta \geq \theta^*]. \tag{20}$$

Proof of Proposition E.1. By linearity and isotonicity of $\succsim_m$, we have for all t that

$$\left(\gamma_1^t (T_{C_1})^t + \gamma_2^t (T_{C_2})^t\right) F^{-1} \succsim_m \left(\gamma_2^t (T_{C_1})^t + \gamma_1^t (T_{C_2})^t\right) F^{-1},$$

since $\gamma_1^t \geq \gamma_2^t \geq 0$ and $C_1 \succsim_m^* C_2$ (by Lemma B.4). Then linearity and continuity of $\succsim_m$ also imply

$$\begin{aligned} &\sum_{t=0}^{\infty} \left(\gamma_1^t (T_{C_1})^t + \gamma_2^t (T_{C_2})^t\right) F^{-1} + \frac{\beta_1 \mathbb{E}_F[\theta]}{(1-\gamma_1)(1-\gamma_1-\beta_1)} + \frac{\beta_2 \mathbb{E}_F[\theta]}{(1-\gamma_2)(1-\gamma_2-\beta_2)} \\ \succsim_m &\sum_{t=0}^{\infty} \left(\gamma_2^t (T_{C_1})^t + \gamma_1^t (T_{C_2})^t\right) F^{-1} + \frac{\beta_1 \mathbb{E}_F[\theta]}{(1-\gamma_1)(1-\gamma_1-\beta_1)} + \frac{\beta_2 \mathbb{E}_F[\theta]}{(1-\gamma_2)(1-\gamma_2-\beta_2)}. \end{aligned}$$

By monotonicity of s_{ij}, this yields for all θ^* that

$$\mathbb{E}_F[s_{11}(\theta) + s_{22}(\theta) | \theta \geq \theta^*] \geq \mathbb{E}_F[s_{12}(\theta) + s_{21}(\theta) | \theta \geq \theta^*],$$

which is equivalent to the desired expression. □

E.4 Proof of Proposition 3

Since the ANLPE action distribution at (P, γ, β) coincides with the Nash action distribution at (P, γ', β'), with $\gamma' = \gamma + \beta$ and $\beta' = 0$, the first part is immediate from Proposition 2, indeed (by the proof of Proposition 2) under the stronger notion of dispersive ordering.

To show the second part, denoting by m_i the inverse cdf of M_i, we have for each quantile $x \in (0, 1)$ that

$$m_i(x) = (1 - \eta_i) \sum_{t \geq 0} \eta_i^t (T_C)^{t+1} f(x),$$

where $\eta_i = \gamma_i + \beta_i$ and $f = F^{-1}$.

Moreover, by an inductive argument, we have that $(T_C)^t F^{-1} \succsim_m (T_C)^{t+1} F^{-1}$ for all $t \geq 0$. Indeed, the base case $t = 0$ holds because of the following result by Ryff (1963): Call a linear operator $T : L^1 \to L^1$ an $\mathfrak{S}$-*operator* if $F^{-1} \succsim_m TF^{-1}$ for all $F \in \mathcal{F}$. The representation theorem in Ryff (1963) implies that T is an $\mathfrak{S}$-operator if there exists some measurable function $K : [0, 1]^2 \to \mathbb{R}$

such that

$$Tf(x) = \frac{d}{dx}\int\limits_0^1 K(x,y)f(y)dy$$

for all $f \in L^1$ and almost every x and the following conditions are met:

1. $K(0,y) = 0$ for all $0 \leq y \leq 1$;
2. $\operatorname{essup}_y V(K(\cdot,y)) < \infty$, where $V(\cdot)$ denotes the total variation and essup the essential supremum;
3. $\int_0^1 K(x,y)f(y)dy$ is absolutely continuous in x for all $f \in L^1$;
4. $x = \int_0^1 K(x,y)dy$;
5. $x_1 < x_2 \Longrightarrow K(x_1,\cdot) \leq K(x_2,\cdot)$;
6. $K(1,y) = 1$ for all $y \in [0,1]$.

Since $C \in \mathcal{C}$, it is easy to see that T_C satisfies these conditions with $K(x,y) := C(x \mid y)$ for all x, y, so that T_C is an $\mathfrak{S}$-operator. Thus, $F^{-1} \succsim_m T_C F^{-1}$, proving the base case. The inductive step then follows from isotonicity of $\succsim_m$.

But this implies that for all $\tau \geq 0$, we have

$$\frac{1}{\sum_{t=0}^{\tau}\eta_2^t}\sum_{t=0}^{\tau}\eta_2^t(T_C)^{t+1}f \succsim_m \frac{1}{\sum_{t=0}^{\tau}\eta_1^t}\sum_{t=0}^{\tau}\eta_1^t(T_C)^{t+1}f. \tag{21}$$

Indeed, for $\tau = 0$, there is nothing to prove. And supposing the claim holds for some $\tau \geq 0$, it follows that

$$\begin{aligned}
\frac{1}{\sum_{t=0}^{\tau+1}\eta_2^t}\sum_{t=0}^{\tau+1}\eta_2^t(T_C)^{t+1}f &= \frac{\sum_{t=0}^{\tau}\eta_2^t}{\sum_{t=0}^{\tau+1}\eta_2^t}\left(\frac{1}{\sum_{t=0}^{\tau}\eta_2^t}\sum_{t=0}^{\tau}\eta_2^t(T_C)^{t+1}f\right) + \frac{\eta_2^{\tau+1}}{\sum_{t=0}^{\tau+1}\eta_2^t}(T_C)^{\tau+2}f \\
&\succsim_m \frac{\sum_{t=0}^{\tau}\eta_2^t}{\sum_{t=0}^{\tau+1}\eta_2^t}\left(\frac{1}{\sum_{t=0}^{\tau}\eta_1^t}\sum_{t=0}^{\tau}\eta_1^t(T_C)^{t+1}f\right) + \frac{\eta_2^{\tau+1}}{\sum_{t=0}^{\tau+1}\eta_2^t}(T_C)^{\tau+2}f \\
&\succsim_m \frac{\sum_{t=0}^{\tau}\eta_1^t}{\sum_{t=0}^{\tau+1}\eta_1^t}\left(\frac{1}{\sum_{t=0}^{\tau}\eta_1^t}\sum_{t=0}^{\tau}\eta_1^t(T_C)^{t+1}f\right) + \frac{\eta_1^{\tau+1}}{\sum_{t=0}^{\tau+1}\eta_1^t}(T_C)^{\tau+2}f \\
&= \frac{1}{\sum_{t=0}^{\tau+1}\eta_1^t}\sum_{t=0}^{\tau+1}\eta_1^t(T_C)^{t+1}f,
\end{aligned}$$

as required. Here the second line holds by inductive hypothesis and the third line follows from linearity of $\succsim_m$ along with the fact that $\eta_1 \geq \eta_2$ (so that $\frac{\sum_{t=0}^{\tau}\eta_1^t}{\sum_{t=0}^{\tau+1}\eta_1^t} \leq \frac{\sum_{t=0}^{\tau}\eta_2^t}{\sum_{t=0}^{\tau+1}\eta_2^t}$ and $\frac{\eta_1^{\tau+1}}{\sum_{t=0}^{\tau+1}\eta_1^t} \geq \frac{\eta_2^{\tau+1}}{\sum_{t=0}^{\tau+1}\eta_2^t}$) and that $(T_C)^{t+1}f \succsim_m (T_C)^{\tau+2}f$ for all $t \leq \tau+1$.

Taking $\tau \to \infty$ in (21), continuity of $\succsim_m$ then yields

$$(1-\eta_2)\sum_{t\geq 0}\eta_2^t(T_C)^{t+1}f \succsim_m (1-\eta_1)\sum_{t\geq 0}\eta_1^t(T_C)^{t+1}f,$$

i.e., $m_2 \succsim_m m_1$, as claimed. □

E.5 Proof of Lemma 4

The lemma is a consequence of Harris' theorem (see Theorem 1.2 in Hairer and Mattingly (2011)).

Assumption 1 of the lemma ensures that Assumption 1 in Hairer and Mattingly (2011) is satisfied. Next, pick any $R > 2K/(1-\eta)$ and let ν be the uniform probability measure on $[-R, R]$. By the second assumption of the lemma, there exists some $\alpha > 0$ sufficiently small such that

$$\inf_{\theta_0,\theta_1 \in [-R,R]} p(\theta_1 \mid \theta_0) > \alpha \frac{1}{2R}.$$

Then for every measurable set $S \subseteq \mathbb{R}$, we have

$$\inf_{|\theta_0| \leq R} P(\theta_1 \in S \mid \theta_0) > \alpha \nu(S),$$

which ensures that Assumption 2 in Hairer and Mattingly (2011) is satisfied.

Thus, we can apply Harris' theorem to conclude that the unique invariant measure of P is given by the type distribution F and that there exist constants $C > 0$ and $\kappa \in (0, 1)$ such that for all t,

$$\sup_{\theta_0 \in \mathbb{R}} \frac{|\mathbb{E}_P[\theta_t \mid \theta_0] - \mathbb{E}_F[\theta]|}{1 + |\theta_0|} \leq C\kappa^t.$$

This implies that for all θ_0,

$$|\mathbb{E}_P[\theta_t \mid \theta_0] - \mathbb{E}_F[\theta]| \leq C\kappa^t(1 + |\theta_0|) \to 0$$

as $t \to \infty$, as required. □

E.6 Proof of Proposition 4

Fix any (P, γ, β) and let H denote the ANLPE action distribution. We show a stronger claim based on the dispersiveness order (see section 7.2) instead of mean-preserving spread. Consider any type θ and let H_θ denote θ's local action distribution under ANLPE. Recall from the proofs of Theorem 2 and Lemma 3 that θ's perceived type distribution $\hat{F}_\theta$ satisfies

$$\hat{F}_\theta^{-1}(x) := H_\theta^{-1}(x) - (\gamma + \beta) \int_0^1 H_\theta^{-1}(y)dy$$

for each x. Because the constant term $(\gamma + \beta) \int_0^1 H_\theta^{-1}(y)dy$ does not affect the dispersive ordering, it suffices to show that H_θ is monotone in $\gamma + \beta$ under the dispersive ordering. To verify this, note that for any quantiles $x > x'$, we have

$$H_\theta^{-1}(x) - H_\theta^{-1}(x') = H^{-1}(y) - H^{-1}(y')$$

where y and y' are such that $x = C(y|F(\theta))$ and $x' = C(y'|F(\theta))$. Since H becomes more dispersive as $\gamma + \beta$ increases (see the proof of Proposition 3), the desired conclusion follows. □

F Proofs for Section 5.4

F.1 Proof of Proposition 5

We prove the proposition by showing that $\left(s, \left(\hat{s}_\theta, (\hat{\mu}_\theta, \hat{\sigma}^2_\theta, \hat{\rho}_\theta)\right)_{\theta\in\mathbb{R}}\right)$ is a linear-Gaussian LPE if and only if there exists $\hat{\rho} \in [0,1)$ such that for all θ,

1. θ's action is $s(\theta) = \frac{1}{1-\gamma\rho-\beta\frac{\rho-\hat{\rho}}{1-\hat{\rho}}}(\theta-\mu) + \frac{\mu}{1-\gamma-\beta}$
2. θ's perceived society is given by

$$\hat{\mu}_\theta = \mu + \frac{(1-\beta-\gamma)(\rho-\hat{\rho})}{(1-\hat{\rho})(1-\gamma\rho-\beta\frac{\rho-\hat{\rho}}{1-\hat{\rho}})}(\theta-\mu)$$

$$\hat{\sigma}^2_\theta = \sigma^2\frac{(1-\rho^2)}{(1-\hat{\rho}^2_\theta)}\left(\frac{1-\gamma\hat{\rho}_\theta}{1-\gamma\rho-\beta\frac{\rho-\hat{\rho}_\theta}{1-\hat{\rho}_\theta}}\right)^2$$

$$\hat{\rho}_\theta \begin{cases} = \hat{\rho} \text{ if } \theta \neq \mu \\ \in [0,1) \text{ if } \theta = \mu. \end{cases}$$

3. θ's perceived strategy profile satisfies $\hat{s}_\theta(\theta') = \frac{\theta'-\hat{\mu}_\theta}{1-\gamma\hat{\rho}_\theta} + \frac{\hat{\mu}_\theta}{1-\gamma-\beta}$ for all θ'.

"Only if" direction: Suppose that $\left(s, \left(\hat{s}_\theta, (\hat{\mu}_\theta, \hat{\sigma}^2_\theta, \hat{\rho}_\theta)\right)_{\theta\in\mathbb{R}}\right)$ is a linear-Gaussian LPE. Since s is linear and non-constant in types, there exist some $x \neq 0$ and $y \in \mathbb{R}$ such that

$$s(\theta) = x(\theta-\mu) + y \text{ for all } \theta.$$

Then for every type θ, the local action distribution is distributed

$$\mathcal{N}(x\rho(\theta-\mu) + y, \sigma^2(1-\rho^2)x^2). \tag{22}$$

Moreover, by Nash rationalization, $\hat{s}_\theta$ is the Nash equilibrium profile at $(\hat{\mu}_\theta, \hat{\sigma}_\theta, \hat{\rho}_\theta)$. That is, by Example 2,

$$\hat{s}_\theta(\theta') = \frac{\theta'-\hat{\mu}_\theta}{1-\gamma\hat{\rho}_\theta} + \frac{\hat{\mu}_\theta}{1-\beta-\gamma} \text{ for all } \theta'. \tag{23}$$

Hence, θ's perceived local action distribution is distributed

$$\mathcal{N}\left(\frac{\hat{\rho}_\theta(\theta-\hat{\mu}_\theta)}{1-\gamma\hat{\rho}_\theta} + \frac{\hat{\mu}_\theta}{1-\beta-\gamma}, \hat{\sigma}^2_\theta\left(1-\hat{\rho}^2_\theta\right)\left(\frac{1}{1-\gamma\hat{\rho}_\theta}\right)^2\right). \tag{24}$$

By part (a) of observational consistency, distributions (22) and (24) must coincide for each θ. In particular, matching means for $\theta = \mu$ implies

$$y = \frac{\hat{\rho}_\mu(\mu-\hat{\mu}_\mu)}{1-\gamma\hat{\rho}_\mu} + \frac{\hat{\mu}_\mu}{1-\beta-\gamma}. \tag{25}$$

Moreover, since $y = s(\mu)$ and by part (b) of observational consistency $s(\mu) = \hat{s}_\mu(\mu)$, we also have

$$y = \frac{\mu-\hat{\mu}_\mu}{1-\gamma\hat{\rho}_\mu} + \frac{\hat{\mu}_\mu}{1-\beta-\gamma}.$$

Combining these equalities, we obtain

$$\frac{1}{1-\gamma\hat{\rho}_\mu}(1-\hat{\rho}_\mu)(\mu-\hat{\mu}_\mu)=0,$$

which implies that $\hat{\mu}_\mu=\mu$, i.e., type μ's perceived population mean is correct. Substituting this into (25) yields

$$y=\frac{\mu}{1-\beta-\gamma}.$$

Thus, for all θ, matching means for (22) and (24) yields

$$x\rho(\theta-\mu)+\frac{\mu}{1-\beta-\gamma}=\frac{\hat{\rho}_\theta(\theta-\hat{\mu}_\theta)}{1-\gamma\hat{\rho}_\theta}+\frac{\hat{\mu}_\theta}{1-\beta-\gamma}. \tag{26}$$

This implies that for all $\theta\neq\mu$, we have $\hat{\mu}_\theta\neq\theta$. Indeed, otherwise (26) implies $x\rho=\frac{1}{1-\beta-\gamma}$. But then $s(\theta)=x(\theta-\mu)+y=\frac{1}{\rho(1-\beta-\gamma)}(\theta-\mu)+\frac{\mu}{1-\beta-\gamma}$, which for $\theta\neq\mu$ is not equal to $\frac{\theta}{1-\beta-\gamma}=\hat{s}_\theta(\theta)$, violating part (b) of observational consistency.

Now observe that for all θ, we have

$$\begin{aligned}x(\theta-\mu)+y=s(\theta)=\hat{s}_\theta(\theta)&=\theta+\gamma\mathbb{E}_{\hat{P}_\theta}\left[\hat{s}_\theta(\theta')|\theta\right]+\beta\mathbb{E}_{\hat{P}_\theta}\left[\hat{s}_\theta(\theta')\right]\\&=\theta+\gamma\mathbb{E}_P\left[s(\theta')|\theta\right]+\beta\mathbb{E}_{\hat{P}_\theta}\left[\hat{s}_\theta(\theta')\right]=\theta+\gamma\left(x\rho(\theta-\mu)+y\right)+\frac{\beta}{1-\beta-\gamma}\hat{\mu}_\theta,\end{aligned}$$

where the second equality holds by part (b) of observational consistency, the third by Nash rationalization, the fourth by part (a) of observational consistency, and the final equality by (22) and (23). Substituting $y=\frac{\mu}{1-\beta-\gamma}$ and rearranging yields the linear equation

$$\hat{\mu}_\theta-\mu=k(\theta-\mu),$$

where $k:=\frac{1-\beta-\gamma}{\beta}(x(1-\gamma\rho)-1)$. Note that $k\neq 1$, since by the previous paragraph $\hat{\mu}_\theta\neq\theta$ for all $\theta\neq\mu$.

Plugging $\hat{\mu}_\theta-\mu=k(\theta-\mu)$ into (26) and rearranging yields

$$\left(x\rho-\frac{\hat{\rho}_\theta(1-k)}{1-\gamma\hat{\rho}_\theta}\right)(\theta-\mu)=\frac{k}{1-\beta-\gamma}(\theta-\mu). \tag{27}$$

Since $k\neq 1$, this implies that $\hat{\rho}_\theta$ must be constant for all $\theta\neq\mu$, say $\hat{\rho}_\theta=\hat{\rho}$ for some $\hat{\rho}\in[0,1)$. As a result, (27) yields

$$x\rho=\frac{\hat{\rho}}{1-\gamma\hat{\rho}}(1-k)+\frac{k}{1-\beta-\gamma}.$$

Substituting for k and solving for x, we obtain:

$$x=\frac{1}{1-\gamma\rho-\beta\frac{\rho-\hat{\rho}}{1-\hat{\rho}}}.$$

Combined with $y = \frac{\mu}{1-\beta-\gamma}$, this shows that behavior must follow:

$$s(\theta) = \frac{1}{1 - \gamma\rho - \beta\frac{\rho - \hat{\rho}}{1-\hat{\rho}}}(\theta - \mu) + \frac{\mu}{1 - \beta - \gamma}.$$

Finally, given x and y, matching (22) and (24) yields the desired expressions for $\hat{\mu}_\theta$ and $\hat{\sigma}^2_\theta$.

"If" direction: It is easy to verify that $\left(s, \left(\hat{s}_\theta, (\hat{\mu}_\theta, \hat{\sigma}^2_\theta, \hat{\rho}_\theta)\right)_{\theta\in\mathbb{R}}\right)$ given by points 1.–3. above indeed satisfies observational consistency and Nash rationalization, and hence constitutes a linear-Gaussian LPE. □

References

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

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.

Ballester, C., A. Calvo-Armengol, and Y. Zenou (2006): "Who's who in networks. Wanted: the key player," *Econometrica*, 74(5), 1403–1417.

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

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.

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 (2017): "Bounded Rationality And Learning: A Framework and A Robustness Result," working paper.

Bramoulle, Y., R. Kranton, and M. DíAmours (2014): "Strategic Interaction and Networks," *American Economic Review*, 104(3), 898–930.

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

Breza, E., A. G. Chandrasekhar, and A. Tahbaz-Salehi (2018): "Seeing the forest for the trees? An investigation of network knowledge," mimeo.

Calvó-Armengol, A., E. Patacchini, and Y. Zenou (2009): "Peer effects and social networks in education," *The Review of Economic Studies*, 76(4), 1239–1267.

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.

Currarini, S., M. O. Jackson, and P. Pin (2009): "An economic model of friendship: Homophily, minorities, and segregation," *Econometrica*, 77(4), 1003–1045.

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

Dekel, E., D. Fudenberg, and D. K. Levine (2004): "Learning to play Bayesian games," *Games and Economic Behavior*, 46(2), 282–303.

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

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.

Enke, B., and F. Zimmermann (2017): "Correlation Neglect in Belief Formation," 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): "Equilibrium in misspecified Markov decision processes," working paper.

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

——— (2010): "Naive herding in rich-information settings," *American economic journal: microeconomics*, 2(4), 221–43.

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.

——— (2006): "Superstition and rational learning," *The American economic review*, 96(3), 630–651.

Fudenberg, D., G. Romanyuk, and P. Strack (2017): "Active learning with a misspecified prior," *Theoretical Economics*, 12(3), 1155–1189.

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.

Ghiglino, C., and S. Goyal (2010): "Keeping up with the neighbors: social interaction in a market economy," *Journal of the European Economic Association*, 8(1), 90–119.

Golub, B., and M. Jackson (2012): "How Homophily Affects the Speed of Learning and Best Response Dynamics," *Quarterly Journal of Economics*, 127(2), 1287–1338.

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

Goyal, S. (2012): *Connections: an introduction to the economics of networks*. Princeton University Press.

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.

Heidhues, P., B. Koszegi, and P. Strack (2017): "Unrealistic expectations and misguided learning," working paper.

Hopkins, E., and T. Kornienko (2004): "Running to Keep in the Same Place: Consumer Choice as a Game of Status," *American Economic Review*, pp. 1085–1107.

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.

Immorlica, N., R. Kranton, M. Manea, and G. Stoddard (2017): "Social Status in Networks," *American Economic Journal: Microeconomics*, 9(1), 1–30.

Jackson, M. O. (2010): *Social and economic networks*. Princeton university press.

——— (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*, forthcoming.

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.

Kilduff, M., C. Crossland, W. Tsai, and D. Krackhardt (2008): "Organizational network perceptions versus reality: A small world after all?," *Organizational Behavior and Human Decision Processes*, 107(1), 15–28.

Krackhardt, D. (1987): "Cognitive social structures," *Social networks*, 9(2), 109–134.

Krackhardt, D., and M. Kilduff (1999): "Whether close or far: Social distance effects on perceived balance in friendship networks.," *Journal of personality and social psychology*, 76(5), 770.

Kumbasar, E., A. K. Rommey, and W. H. Batchelder (1994): "Systematic biases in social perception," *American journal of sociology*, 100(2), 477–505.

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 (2017): "Peer-Confirming Equilibrium," working paper.

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.

Marshall, A. W., I. Olkin, and B. C. Arnold (2010): *Inequalities: theory of majorization and its applications*, vol. 143. Springer.

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.

——— (2005): "Heterogeneity and Uniqueness in Interaction," *The Economy As an Evolving Complex System, III: Current Perspectives and Future Directions*, p. 207.

Müller, A., and D. Stoyan (2002): *Comparison methods for stochastic models and risks*, vol. 389. Wiley.

Norton, M. I., and D. Ariely (2011): "Building a better America–One wealth quintile at a time," *Perspectives on Psychological Science*, 6(1), 9–12.

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

Perez-Truglia, R. (2016): "The effects of income transparency on well-being: evidence from a natural experiment," mimeo.

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

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. (1998): *Modeling bounded rationality.* MIT press.

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.

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

Vega-Redondo, F. (2007): *Complex social networks*, no. 44. Cambridge University Press.

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.

Supplementary Appendix to "Dispersed Behavior and Perceptions in Assortative Societies"

Mira Frick, Ryota Iijima, and Yuhta Ishii

G Proofs for Section 6

G.1 Proof of Proposition 6

Fix any type θ and let $H_\theta^{NE} = H_\theta^{s^{NE},P}$ and $H_\theta^{AN} = H_\theta^{s^{AN},P}$ denote θ's local action distribution under Nash and ANLPE at (P,γ,β). For any quantiles $x > x'$, we have

$$\begin{aligned}(H_\theta^{AN})^{-1}(x) - (H_\theta^{AN})^{-1}(x') &= (H^{AN})^{-1}(y) - (H^{AN})^{-1}(y') \\ &\geq (H^{NE})^{-1}(y) - (H^{NE})^{-1}(y') \\ &= (H_\theta^{NE})^{-1}(x) - (H_\theta^{NE})^{-1}(x')\end{aligned}$$

where y, y' are such that $x = C(y|F(\theta))$ and $x' = C(y'|F(\theta))$, and the inequality holds since by the proof of Corollary 1, the action distribution under ANLPE is more dispersive than under Nash (in the sense of $\succsim_d$ defined in Section H.1). Thus, $H_\theta^{AN} \succsim_d H_\theta^{NE}$, which implies that H_θ^{AN} has higher variance than H_θ^{NE} (e.g., Shaked and Shanthikumar, 2007, p. 155).

We can write payoffs as

$$u_P^{NE}(\theta) = -\int_0^1 \gamma^2 \left(\bar{a} - (H_\theta^{NE})^{-1}(x)\right)^2 dx, \;\; \hat{u}_P^{AN}(\theta) = -\int_0^1 \gamma^2 \left(\bar{a}' - (H_\theta^{AN})^{-1}(x)\right)^2 dx,$$

$$u_P^{AN}(\theta) = -\int_0^1 \left(\gamma\bar{a}' - \gamma(H_\theta^{AN})^{-1}(x) + \frac{\beta(\hat{\mu}_\theta - \mu)}{1-\gamma-\beta}\right)^2 dx,$$

where $\bar{a} = \int_0^1 (H_\theta^{NE})^{-1}(x)dx$, $\bar{a}' = \int_0^1 (H_\theta^{AN})^{-1}(x)dx$, and $\mu = \mathbb{E}_F[\theta]$. Then we have $u_P^{NE}(\theta) \geq \hat{u}_P^{AN}(\theta)$, since the variance of H_θ^{AN} is higher than that of H_θ^{NE}. Moreover, $\hat{u}_P^{AN}(\theta) \geq u_P^{AN}(\theta)$ since the quadratic loss $\int_0^1 \left(\gamma\bar{a}' - \gamma(H_\theta^{AN})^{-1}(x) + c\right)^2 dx$ is minimized when the constant c equals 0. □

G.2 Proof of Corollary 4

Recall that for any s, P, and θ, quadratic-loss utility satsfies $u_P(\mathrm{BR}_\theta(s,P),\theta,s) = -\gamma^2\mathrm{Var}_P[s(\theta')|\theta]$. Thus, by Example 2,

$$u_P^{NE}(\theta) = -\gamma^2\mathrm{Var}_P[s^{NE}(\theta')|\theta] = -\frac{\gamma^2\sigma^2(1-\rho^2)}{(1-\gamma\rho)^2}.$$

Moreover, evaluating the expressions derived in the proof of Proposition 5 at $\hat{\rho} = 0$,

$$\hat{u}_P^{AN}(\theta) = -\gamma^2\mathrm{Var}_{\hat{P}_\theta}[\hat{s}_\theta(\theta')|\theta] = -\gamma^2\hat{\sigma}^2 = -\frac{\gamma^2\sigma^2(1-\rho^2)}{(1-(\gamma+\beta)\rho)^2}.$$

This immediately yields the desired conclusion. $\square$

H Proofs for Section 7

H.1 Proof of Theorem 4

First, we define the dispersive order $\succsim_d$ over $\mathcal{I}$:

Definition 9. $f \succsim_d g$ if for all $x, x' \in (0,1)$ such that $x \geq x'$, we have

$$f(x) - f(x') \geq g(x) - g(x').$$

We verify that $\succsim_d$ obeys the three basic properties from Theorem B.1:

Lemma H.1. *$\succsim_d$ is a preorder that is linear, continuous, and isotone.*

Proof. It is clear from the definition that $\succsim_d$ is reflexive, transitive, and linear. To check that it is continuous, take sequences $f_n \to f$ and $g_n \to g$ in $\mathcal{I}$ such that $f_n \succsim_d g_n$ for each n. By standard results (e.g., Theorem 13.6 in Aliprantis and Border (2006)), we can find subsequences $(f_{n_k})_{k\in\mathbb{N}}, (g_{n_k})_{k\in\mathbb{N}}$ such that $f_{n_k}(x) \to f(x)$ and $g_{n_k}(x) \to g(x)$ for almost all $x \in (0,1)$. This implies $f(x) - f(x') \geq g(x) - g(x')$ for almost all $x \geq x'$, which ensures $f \succsim_d g$ since f and g are continuous.

To show that $\succsim_d$ is isotone, first consider any bounded $f, g \in \mathcal{I}$ such that $f \succsim_d g$. Since f and g are absolutely continuous, there exist integrable functions $f', g' : (0,1) \to \mathbb{R}$ such that $f(x) = f(0) + \int_0^x f'(y)\, dy$ and $g(x) = g(0) + \int_0^x g'(y)\, dy$ for all $x \in (0,1)$. Then, for any $x \geq x'$ and $C \in \mathcal{C}$, integration by parts yields

$$\begin{aligned}
T_C f(x) - T_C f(x') &= \int_0^1 f(y)(c(y|x) - c(y|x'))dy \\
&= -\int_0^1 f'(y)(C(y|x) - C(y|x'))dy + [f(y)(C(y|x) - C(y|x'))]_0^1 \\
&= -\int_0^1 f'(y)(C(y|x) - C(y|x'))dy \\
&\geq -\int_0^1 g'(y)(C(y|x) - C(y|x'))dy \\
&= -\int_0^1 g'(y)(C(y|x) - C(y|x'))dy + [g(y)(C(y|x) - C(y|x'))]_0^1 \\
&= \int_0^1 g(y)(c(y|x) - c(y|x'))dy = T_C g(x) - T_C g(x').
\end{aligned}$$

Here, the inequality holds because the fact that $f \succsim_d g$ and $f, g \in \mathcal{I}$ implies $f'(y) \geq g'(y) \geq 0$ for almost all $y \in (0,1)$.

Next, consider arbitrary $f, g \in \mathcal{I}$ such that $f \succsim_d g$. By defining bounded functions

$$f_n(x) = \begin{cases} \frac{1}{n} \text{ if } x \in (0, \frac{1}{n}) \\ f(x) \text{ if } x \in [\frac{1}{n}, \frac{n-1}{n}] \\ \frac{n-1}{n} \text{ if } x \in (\frac{n-1}{n}, 1) \end{cases} \quad g_n(x) = \begin{cases} \frac{1}{n} \text{ if } x \in (0, \frac{1}{n}) \\ g(x) \text{ if } x \in [\frac{1}{n}, \frac{n-1}{n}] \\ \frac{n-1}{n} \text{ if } x \in (\frac{n-1}{n}, 1) \end{cases} \tag{28}$$

for each $n \in \mathbb{N}$, we obtain $f_n \succsim_d g_n$ for each n and $f_n \to f, g_n \to g$. For any $C \in \mathcal{C}$, since T_C is a continuous operator, this implies $T_C f_n \to T_C f$ and $T_C g_n \to T_C g$. Thus $T_C f \succsim_d T_C g$ by continuity of $\succsim_d$, because we already know that $T_C f_n \succsim_d T_C g_n$ from the previous part of the proof. □

Next we show that the strongly more-assortative order $\succsim_{SMA}$ coincides with the dual order of $\succsim_d$:

Lemma H.2. *$C_1 \succsim_d^* C_2$ if and only if $C_1 \succsim_{SMA} C_2$.*

Proof. For the "if" part, suppose that $C_1 \succsim_{SMA} C_2$. First consider any bounded $f \in \mathcal{I}$. Then there exists an integrable function $f' : (0,1) \to \mathbb{R}$ that is nonnegative almost everywhere such that $f(x) = f(0) + \int_0^x f'(y)dy$ for all $x \in (0,1)$. Then, for any $x \geq x'$, integration by parts yields

$$\begin{aligned} T_{C_1}f(x) - T_{C_1}f(x') &= \int_0^1 f(y)(c_1(y|x) - c_1(y|x'))dy \\ &= -\int_0^1 f'(y)(C_1(y|x) - C_1(y|x'))dy + [f(y)(C_1(y|x) - C_1(y|x'))]_0^1 \\ &= -\int_0^1 f'(y)(C_1(y|x) - C_1(y|x'))dy \\ &\geq -\int_0^1 f'(y)(C_2(y|x) - C_2(y|x'))dy \\ &= -\int_0^1 f'(y)(C_2(y|x) - C_2(y|x'))dy + [f(y)(C_2(y|x) - C_2(y|x'))]_0^1 \\ &= \int_0^1 f(y)(c_2(y|x) - c_2(y|x'))dy = T_{C_2}f(x) - T_{C_2}f(x'), \end{aligned}$$

where the inequality holds because $f'(y) \geq 0$ for almost all y. Thus $T_{C_1}f \succsim_d T_{C_2}f$.

Next take an arbitrary $f \in \mathcal{I}$. Define the sequence of bounded functions (f_n) as in (28), so that $f_n \to f$. By the previous observation, we have $T_{C_1}f_n \succsim_d T_{C_2}f_n$ for each n. Since $T_{C_1}f_n \to T_{C_1}f$ and $T_{C_2}f_n \to T_{C_2}f$ by continuity of T_{C_1} and T_{C_2}, continuity of $\succsim_d$ then yields $T_{C_1}f \succsim_d T_{C_2}f$.

Finally, for the "only if" part, suppose that $C_1 \succsim_d^* C_2$. Suppose for a contradiction that C_1 is not strongly more assortative than C_2. That is, there exist y and $x > x'$ such that

$$C_2(y|x) - C_2(y|x') < C_1(y|x) - C_1(y|x') \leq 0.$$

Since C_1 and C_2 admit densities, the above inequality holds throughout some interval $(y_1, y_2) \ni y$. Define $f \in \mathcal{I}$ by $f(z) = \int_0^z f'(y')dy'$ for all z, where f' is an integrable function given by $f'(y') = 1$ for $y' \in (y_1, y_2)$ and $f'(y') = 0$ for all $y' \notin (y_1, y_2)$. Using the same integration by parts argument as above, we obtain

$$T_{C_1}f(x) - T_{C_1}f(x') = -\int f'(y)(C_1(y|x) - C_1(y|x'))dy$$
$$< -\int f'(y)(C_2(y|x) - C_2(y|x'))dy = T_{C_2}f(x) - T_{C_2}f(x'),$$

contradicting $C_1 \succsim_d^* C_2$. □

Given Lemmas H.1–H.2, Theorem 4 then follows immediately from Theorem B.1. □

H.2 Proof of Proposition 7

We begin with the following lemma:

Lemma H.3. *Fix any (F, C, γ, β) and $\hat{C} \in \mathcal{C}$. There exists $\hat{F} \in \mathcal{F}$ such that $H_{F,C,\gamma,\beta} = H_{\hat{F},\hat{C},\gamma,\beta}$ if and only if*

$$H^{-1}_{F,C,\gamma,\beta}(x) - \gamma \int_0^1 \hat{c}(y,x) H^{-1}_{F,C,\gamma,\beta}(y) dy$$

is strictly increasing in x.

Proof. Let $h(x) := H^{-1}_{F,C,\gamma,\beta}(x)$ for each x. Suppose that $H_{F,C,\gamma,\beta} = H_{\hat{F},\hat{C},\gamma,\beta}$ for some $\hat{F} \in \mathcal{F}$. Then $h(x) = H^{-1}_{\hat{F},\hat{C},\gamma,\beta}(x) = \hat{F}^{-1}(x) + \gamma \int_0^1 \hat{c}(y,x)h(y)dy + \beta \int h(y)dy$ for each x. Therefore, $h(x) - \gamma \int_0^1 \hat{c}(y,x)h(y)dy = \hat{F}^{-1}(x) + \beta \int h(y)dy$ is strictly increasing in x. Conversely, suppose that $h(x) - \gamma \int_0^1 \hat{c}(y,x)h(y)dy$ is strictly increasing in x. Then define a strictly increasing function $\hat{F}$ by $\hat{F}^{-1} = h - \gamma T_{\hat{C}} h - \beta \int h(y)dy$, or

$$\hat{F}^{-1}(x) = h(x) - \gamma \int \hat{c}(y,x)h(y)dy - \beta \int h(y)dy$$

for each x. Since $T_{\hat{C}} h$ is L^1 and absolutely continuous, so is $\hat{F}^{-1}$. Thus $\hat{F} \in \mathcal{F}$. Then h^{-1} is the Nash equilibrium strategy under $(\hat{F}, \hat{C}, \gamma, \beta)$ because $h(x) = \hat{F}^{-1}(x) + \gamma \int \hat{c}(y,x)h(y)dy + \beta \int h(y)dy$ for each x. Therefore, $H_{\hat{F},\hat{C},\gamma,\beta} = H_{F,C,\gamma,\beta}$. □

To prove the first claim in the proposition, suppose that $C \succsim_{SMA} \hat{C}$ and take any (F, γ, β). Let $h(x) := H^{-1}_{F,C,\gamma,\beta}(x)$ for each x. Then for any $x > x'$ in $(0, 1)$,

$$\begin{aligned} h(x) - \gamma T_{\hat{C}} h(x) &= F^{-1}(x) + \beta \int h(y)dy + \gamma \left(T_C h(x) - T_{\hat{C}} h(x)\right) \\ &> F^{-1}(x') + \beta \int h(y)dy + \gamma \left(T_C h(x') - T_{\hat{C}} h(x')\right) = h(x') - \gamma T_{\hat{C}} h(x') \end{aligned}$$

because F^{-1} is strictly increasing and $T_C h \succsim_d T_{\hat{C}} h$ (note that $h \in \mathcal{I}$). Thus, Lemma H.3 implies the desired claim.

The claim that there cannot be multiple $\hat{F}, \hat{F}'$ such that $H_{F,C,\gamma,\beta} = H_{\hat{F},\hat{C},\gamma,\beta} = H_{\hat{F}',\hat{C},\gamma,\beta}$ follows from the fact that $\hat{F}^{-1}(x) = h(x) - \gamma \int \hat{c}(y,x)h(y)dy - \beta \int h(y)dy = \hat{F}'^{-1}(x)$ holds for each x at such $\hat{F}, \hat{F}'$. The equality $\int \theta dF = (1 - \beta - \gamma) \int h(x)dx = \int \theta d\hat{F}$ follows by Lemma 1.

Finally, note that

$$\hat{F}^{-1} = h - \gamma T_{\hat{C}} h - \beta \int h(y)dy = F^{-1} + \gamma \left(T_C h - T_{\hat{C}} h\right).$$

Thus, for each $x > x'$ in $(0, 1)$,

$$\hat{F}(x) - \hat{F}(x') = F^{-1}(x) - F^{-1}(x') + \gamma \left(T_C f h(x) - T_{\hat{C}} h(x) - T_C f h(x') + T_{\hat{C}} h(x')\right) \geq F^{-1}(x) - F^{-1}(x')$$

because $T_C h \succsim_d T_{\hat{C}} h$. This shows that $\hat{F} \succsim_d F$. □

I Relationship with RCE

This appendix elaborates on the relationship between LPE and RCE discussed in Section 7.1. In constrast with LPE, existing formulations of RCE do not explicitly treat players' beliefs as equilibrium objects. Thus, to make the connection with LPE transparent, we first define an extended version of RCE below, based on the reduced-form formulation in Esponda (2013) without a full construction of an epistemic model. To define RCE in our environment, first define the state space that encodes all uncertainty relevant to agents' decisions:

$$\Omega = \{(s, P) : P = (F, C), \int |s(\theta)| dF < \infty\}.$$

Elements $\omega \in \Omega$ are called states. We write $\omega = (s_\omega, P_\omega)$, where P_ω is the society and s_ω the strategy profile corresponding to state ω. Let $u_\theta(a, \omega)$ denote the utility θ receives by choosing a in society P_ω against strategy profile s_ω; we consider utilities that give rise to best-responses of the linear form in (1), for example the quadratic-loss utility specification in (2). For any probability measure $\nu \in \Delta(\Omega)$, let $u_\theta(a, \nu) := \int u_\theta(a, \omega) d\nu(\omega)$. Let $H_\theta^\omega := H_\theta^{s_\omega, P_\omega}$ denote the local action distribution observed by type θ at state ω, as defined in (4).

Given some society P, RCE captures the possible behavior and perceptions about P that may arise when players are only constrained by rationality, observational consistency, and common certainty of these two requirements. As a first step, the following definition formalizes the possible beliefs that players may hold about Ω under these constraints:

Definition 10. A ***rational perception system*** consists of $\hat{\Omega} \subseteq \Omega$ and $\nu := \{\nu_{\omega,\theta}\}_{\omega \in \hat{\Omega}, \theta \in \Theta} \subseteq \Delta(\Omega)$ such that for all $\omega \in \hat{\Omega}$ and θ,

1. rationality: $s_\omega(\theta) \in \arg\max_{a \in A} u_\theta(a, \nu_{\omega,\theta})$
2. observational consistency: $\nu_{\omega,\theta}(\{\omega' : H_\theta^{\omega'} = H_\theta^\omega \text{ and } s_{\omega'}(\theta) = s_\omega(\theta)\}) = 1$
3. belief-closedness: $\nu_{\omega,\theta}(\hat{\Omega}) = 1$.

A rational perception system is a collection of type-dependent beliefs about Ω. Each belief $\nu_{\omega,\theta}$ is indexed by a state $\omega \in \hat{\Omega}$, capturing some hypothetical true behavior s_ω and true society P_ω. In each state ω, (1) ensures that each type θ best-responds to his belief $\nu_{\omega,\theta}$, while (2) requires that this belief be correct about the true local action distribution H_θ^ω and about his own strategy $s_\omega(\theta)$. Finally, (3) ensures that at each ω, there is common certainty of (1) and (2).

Suppose that the true society is P. Using Definition 10, it is now straightforward to express which behavior and beliefs about Ω can jointly arise under rationality, observational consistency, and common certainty thereof: Consider any strategy profile s together with a perceived strategy profile $\hat{s}_\theta$ and perceived society $\hat{P}_\theta$ for each type. We say that $(s, (\hat{s}_\theta, \hat{P}_\theta)_{\theta \in \Theta})$ is ***rationalized*** by the rational perception system $(\hat{\Omega}, \nu)$ at P if $(s, P) \in \hat{\Omega}$ and $\nu_{s,P,\theta}\left(\{(\hat{s}_\theta, \hat{P}_\theta)\}\right) = 1$ for all $\theta \in \Theta$. A rationalizable conjectural equilibrium is any tuple $(s, (\hat{s}_\theta, \hat{P}_\theta)_{\theta \in \Theta})$ that can be rationalized by some rational perception system:[48]

Definition 11. A ***rationalizable conjectural equilibrium (RCE)*** at P is a strategy profile s together with perceptions $(\hat{s}_\theta, \hat{P}_\theta)_{\theta \in \Theta}$ that are rationalized by some rational perception system $(\hat{\Omega}, \nu)$ at P.

[48] For simplicity, we focus here on the case in which players assign probability 1 to a particular state. This makes the connection with LPE more transparent.

Every LPE is an RCE, but the converse is not necessarily true. While in LPE each agent believes that other agents follow a Nash equilibrium, RCE allows for more general perceptions. However, when $\beta = 0$, RCE and LPE coincide and the underlying true behavior is in fact Nash. The key observation is that absent global coordination motives, if θ best responds to a belief that generates the same local action distribution as some true strategy profile and interaction structure, then this is enough to ensure that θ best responds to the truth. Nevertheless, even when $\beta = 0$, there is scope for misperception (i.e., we can have $(\hat{s}_\theta, \hat{P}_\theta) \neq (s, P)$) as long as the conjectured Nash and true Nash are observationally equivalent for each type.

Lemma I.1. *Every LPE $(s, (\hat{s}_\theta, \hat{P}_\theta)_{\theta \in \Theta})$ at P is an RCE at P. If $\beta = 0$, the converse is true and s is the Nash equilibrium profile at P.*

Proof. See Appendix I.1. □

This lemma directly implies that the first claim in Theorem 2 generalizes, i.e., that assortativity neglect can be sustained as part of an RCE in any environment. The second part of Theorem 2 also extends, i.e., assortativity neglect remains the only form of misperception about the interaction structure that can be sustained in RCE in any environment. The latter follows from Lemma I.1 because in the proof of Theorem 2 we can restrict attention to $\beta = 0$.

Corollary 5 (Robust sustainability of assortativity neglect under RCE)**.**

1. *For any (P, γ, β), there exists an RCE such that $\hat{C}_\theta = C_I$ for all θ.*
2. *For any regular $\hat{C} \neq C_I$ and any θ, there exists (P, γ, β) at which all RCE satisfy $\hat{C}_\theta \neq \hat{C}$.*

Moreover, in any RCE such that $\hat{C}_\theta = C_I$ for all θ, the true strategy profile s^{AN} is given by (6).

It is important to note that actual behavior under assortativity neglect remains uniquely pinned down under RCE: Specifically, as under ANLPE, this behavior s^{AN} is given by (6).[49] Thus, all implications of assortativity neglect for behavior that we derived in Section 5.1 remain valid. At the same time, in contrast with Theorem 2, where assortativity neglect also uniquely pinned down all agents' perceived type distributions and perceived strategy profiles, under RCE there are many perceptions which could be consistent with assortativity neglect. Thus, LPE can be viewed as providing one natural way of selecting among these multiple perceptions.

In Appendix I.3, we also show that LPE can be obtained as a refinement of RCE by introducing *lexicographic preferences over perceptions.* The idea is that players prefer to adopt "simple" perceptions as long as they are consistent with their local observations and do not decrease their original game payoffs. Under such preferences, Lemma 3 above, which establishes that *any* local action distribution can be rationalized as Nash equilibrium under C_I, can be used to show that any RCE with complicated higher-order beliefs unravels, eventually leading to LPE. If we additionally assume that the independent interaction structure is simpler than any other perception, then this lexicographic preference in fact yields ANLPE as the unique prediction.

[49] To see this, observe that the derivation of s^{AN} in Section 5.1 only relied on the fact that each type θ (a) best-responds to his perceptions ($s(\theta) = BR_\theta(\hat{s}_\theta, \hat{P}_\theta)$) and (b) is correct about the local action mean ($\mathbb{E}_{\hat{P}_\theta}[\hat{s}_\theta(\theta')|\theta] = E_P[s(\theta')|\theta]$), both of which remain true under RCE. The key is that under assortativity neglect, (b) also pins down the perceived global action mean, as when $\hat{C}_\theta = C_I$ the perceived global and local action mean coincide. Together with (a), this uniquely determines behavior. In fact, the same argument also establishes that assortativity neglect uniquely pins down behavior under the even more permissive solution concept of self-confirming equilibrium.

I.1 Proof of Lemma I.1

Suppose that $(s, (\hat{s}_\theta, \hat{P}_\theta)_\theta)$ is an LPE at P. We construct a rational perception system $(\hat{\Omega}, (\nu_{\omega,\theta})_{\omega\in\hat{\Omega},\theta\in\Theta})$ as follows. Let $\omega^* := (s, P)$ and $\hat{\omega}_\theta := (\hat{s}_\theta, \hat{P}_\theta)$ for all θ, set $\hat{\Omega} := \{\omega^*\} \cup \{\hat{\omega}_\theta : \theta \in \Theta\}$, and set $\nu_{\omega^*,\theta}(\{\hat{\omega}_\theta\}) = \nu_{\hat{\omega}_\theta,\theta}(\{\hat{\omega}_\theta\}) = \nu_{\hat{\omega}_\theta,\theta'}(\{\hat{\omega}_\theta\}) = 1$ for all $\theta \neq \theta'$.

Clearly belief-closedness holds. Since $H_\theta^{\hat{\omega}_\theta} = H_\theta^{\omega^*}$ and $\hat{s}_\theta(\theta) = s(\theta)$, observational consistency holds for θ at $\hat{\omega}^*$, and it trivially holds for θ at $\hat{\omega}_\theta$ and $\hat{\omega}_{\theta'}$. Finally, since $s(\theta) = \hat{s}_\theta(\theta)$ and $\hat{s}_\theta(\theta') = \mathrm{BR}_{\theta'}(\hat{s}_\theta, \hat{P}_\theta)$ for each θ and θ', rationality holds at all $\omega \in \hat{\Omega}$ and θ. Thus, $(\hat{\Omega}, (\nu_{\omega,\theta})_{\omega\in\hat{\Omega},\theta\in\Theta})$ is a rational perception system such that $(s, P) \in \hat{\Omega}$ and $\nu_{s,P,\theta}(\{(\hat{s}_\theta, \hat{P}_\theta)\}) = 1$ for all θ, whence $(s, (\hat{s}_\theta, \hat{P}_\theta)_\theta)$ is an RCE at P.

Suppose next that $\beta = 0$. Note that

$$\nu\{\omega' : H_\theta^{\omega'} = H_\theta^{\omega}\} = 1 \quad \Longrightarrow \quad \operatorname*{argmax}_{a\in A} u_\theta(a, \omega) = \operatorname*{argmax}_{a\in A} u_\theta(a, \nu) \tag{29}$$

for any θ, ω, and $\nu \in \Delta(\Omega)$.

If $(s, (\hat{s}_\theta, \hat{P}_\theta)_\theta)$ is an RCE at P, then there exists a rational perception system $(\hat{\Omega}, (\nu_{\omega,\theta})_{\omega\in\hat{\Omega},\theta\in\Theta})$ such that (i) $(s, P) \in \hat{\Omega}$ and (ii) $\nu_{s,P,\theta}(\{(\hat{s}_\theta, \hat{P}_\theta)\}) = 1$ for all θ. Fix any θ and let $\omega^* := (s, P)$. By rationality at ω^*, we have $s(\theta) \in \operatorname{argmax}_{a\in A} u_\theta(a, \nu_{\omega^*,\theta})$, and by observational consistency at ω^*, we have $\nu\{\omega' : H_\theta^{\omega'} = H_\theta^{\omega^*}\} = 1$. Hence, (29) implies $s(\theta) \in \operatorname{argmax}_{a\in A} u_\theta(a, \omega^*)$. Thus, s is Nash at P.

Moreover, by (ii) and observational consistency at ω^* , there exists $\hat{\omega}_\theta := (\hat{s}_\theta, \hat{P}_\theta) \in \hat{\Omega}$ with $H_\theta^{\hat{\omega}_\theta} = H_\theta^{\omega^*}$ and $\hat{s}_\theta(\theta) = s(\theta)$. Finally, applying rationality and observational consistency at $\hat{\omega}_\theta$, implies that $\hat{s}_\theta$ is Nash at $\hat{P}_\theta$ as in the previous paragraph. Therefore $(s, (\hat{s}_\theta, \hat{P}_\theta)_\theta)$ is an LPE at P. □

I.2 Proof of Corollary 5

The first part is an immediate consequence of the first part of Theorem 2 and the fact that any RCE is also an LPE (Lemma I.1).

To verify the second part, fix any regular $\hat{C} \neq C_I$ and any θ. Note that in the proof of the second part of Theorem 2, we constructed (F, C, γ, β) with $\beta = 0$ under which $\hat{C}_\theta \neq \hat{C}$ holds at any LPE. Then the desired conclusion follows from Lemma I.1, as any RCE is LPE when $\beta = 0$. □

I.3 Lexicographic Preference over Perceptions

Below we show that LPE can be seen as a refinement of RCE in which each agent lexicographically prefers to adopt perceptions without higher-order belief disagreement, as long as such perceptions do not reduce his utility. To formalize this idea, we say that θ is ***dogmatic*** at ω under $(\hat{\Omega}, \nu)$ if

$$\operatorname{supp} \nu_{\omega,\theta} = \operatorname{supp} \nu_{\omega',\theta'}$$

for all $\omega' \in \operatorname{supp} \nu_{\omega,\theta}$ and θ'. That is, agent θ believes that all other agents θ' believe in the same set of states as he does.

Definition 12. A rational perception system $(\hat{\Omega}, \nu)$ displays ***lexicographic preference for dogmaticity*** at ω if the following is true: For any θ such that there exists a rational perception system $(\hat{\Omega}^*, \nu^*)$ and $\omega^* \in \hat{\Omega}^*$ satisfying

1. $\nu^*_{\omega^*,\theta}(\{\omega' : H_\theta^{\omega'} = H_\theta^{\omega}\}) = 1$

2. $u_\theta(s_{\omega^*}(\theta), \nu^*_{\omega^*,\theta}) \geq u_\theta(s_\omega(\theta), \nu_{\omega,\theta})$

3. θ is dogmatic at ω^* under $(\hat{\Omega}^*, \nu^*)$,

we have that θ is dogmatic at ω under $(\hat{\Omega}, \nu)$.

The condition requires that there cannot be any agent θ who is not dogmatic at ω, but who could weakly increase his utility by finding some other observationally consistent perception system that *does* allow him to hold a dogmatic belief. This captures the idea that agents lexicographically prefer to hold perceptions without higher-order belief disagreement. Such a preference might be justified on the basis of aversion to complexity, since it is presumably "simpler" not to have higher-order belief disagreement with other agents.[50]

Proposition I.1. *Consider any strategy profile s that is strictly increasing and continuous. Then $(s, \hat{s}, \hat{P})$ is an LPE at P if and only if $(s, \hat{s}, \hat{P})$ is rationalized by some rational perception system $(\hat{\Omega}, \nu)$ at P that displays lexicographic preference for dogmaticity at (s, P).*

The key to the "if" direction is Lemma 3 above, which shows that any action distribution can be rationalized as Nash under C_I and some suitable $\hat{F}$. Given this, in any RCE, each agent can find perceptions that are dogmatic, observationally consistent, and rationalize others' behavior as Nash. Given lexicographic preference for dogmaticity, this ensures that all agents are dogmatic, yielding an LPE.

Next, we show that LPE can be further refined to ANLPE by additionally imposing a lexicographic preference for perceptions that feature independent interactions.

Definition 13. A rational perception system $(\hat{\Omega}, \nu)$ displays ***lexicographic preference for independence*** at ω if, for any θ such that there exists a rational perception system $(\hat{\Omega}^*, \nu^*)$ and $\omega^* \in \hat{\Omega}^*$ satisfying

1. $\nu^*_{\omega^*,\theta}(\{\omega' : H^{\omega'}_\theta = H^\omega_\theta\}) = 1$

2. $u_\theta(s_{\omega^*}(\theta), \nu^*_{\omega^*,\theta}) \geq u_\theta(s_\omega(\theta), \nu_{\omega,\theta})$

3. $\nu^*_{\omega^*,\theta}(\{\omega' : C_{\omega'} = C_I\}) = 1$,

we have that $\nu_{\omega,\theta}(\{\omega' : C_{\omega'} = C_I\}) = 1$.

The condition requires that there cannot be any agent θ whose perceived interaction structure at ω is not C_I, but who could weakly increase his utility by finding some other observationally consistent perception system that does allow him to believe in the independent interaction structure. This represents the idea that agents lexicographically prefer to believe that the interaction structure is independent. As before, such a preference might be justified on the basis of aversion to complexity.

Proposition I.2. *Consider any strategy profile s that is strictly increasing and continuous. Then $(s, \hat{s}, \hat{P})$ is an LPE at P if and only if $(s, \hat{s}, \hat{P})$ is rationalized by some rational perception system $(\hat{\Omega}, \nu)$ at P that displays lexicographic preference for dogmaticity and independence at (s, P).*

The proof is similar to the previous result, and the key to the "if" direction is again Lemma 3. We now show that in any LPE, each agent can find perceptions that are dogmatic, observationally consistent, and rationalize others' behavior as Nash under the independent interaction structure. Given lexicographic preference for independence, this reduces LPE to ANLPE. Note that if we only require lexicographic preference for independence (but not necessarily for dogmaticity), then s has to be the ANLPE strategy profile while $(\hat{s}, \hat{P})$ need not coincide with the ANLPE perceptions.

[50]This approach is somewhat similar in spirit to the idea of introducing preferences over the complexity of strategies in games (Rubinstein, 1998).

I.3.1 Proof of Proposition I.1

"If" direction: Suppose that $(s, \hat{s}, \hat{P})$ is rationalized by a rational perception system $(\hat{\Omega}, \nu)$ at P that displays lexicographic preference for dogmaticity at $(s, P) =: \omega$.

First, we show that any θ is dogmatic at ω under $(\hat{\Omega}, \nu)$. For this purpose, define

$$\hat{\Omega}^* := \{(s, (F, C_I)) : F \in \mathcal{F}, s \text{ is Nash at } (F, C_I)\}.$$

For each $\omega' \in \hat{\Omega}^*$ and θ', define $\nu^*_{\omega',\theta'} \in \Delta(\Omega)$ by $\nu^*_{\omega',\theta'}(\{\omega'\}) = 1$. To verify that $(\hat{\Omega}^*, \nu^*)$ is a rational perception system, first note that belief-closedness is satisfied by construction. For each $\omega' = (s, (F, C_I)) \in \hat{\Omega}^*$ and θ', observational consistency is satisfied at $\nu^*_{\omega',\theta'}$ because the belief is correct, i.e., $\nu^*_{\omega',\theta'}(\{\omega'\}) = 1$. The rationality condition is also satisfied because the belief is correct and s is Nash at (F, C_I).

By the assumption on s, cdf H^ω_θ is strictly increasing and continuous. By Lemma 3, there is $\omega^* \in \hat{\Omega}^*$ such that $H^{\omega^*}_\theta = H^\omega_\theta$. This, combined with the rationality condition, implies

$$u_\theta(\nu^*_{\omega^*,\theta}) = -\gamma \int a^2 dH^{\omega^*}_\theta(a) = -\gamma \int a^2 dH^\omega_\theta(a) = u_\theta(\nu_{\omega,\theta}).$$

Note that θ is dogmatic at ω^* under $(\hat{\Omega}^*, \nu^*)$ by construction. Therefore, by lexicographic preference for dogmaticity, it follows that θ is also dogmatic at ω under $(\hat{\Omega}, \nu)$.

Next, we show for each θ that $\hat{s}_\theta$ is Nash at $\hat{P}_\theta$. To see this, consider $\bar{\Omega} := \text{supp}\nu_{\omega,\theta}$ and the restriction of ν to $\bar{\Omega}$, denoted by $\bar{\nu}$. Then $(\bar{\Omega}, \bar{\nu})$ is a rational perception system in which there is common certainty that society is given by $\hat{P}_\theta$. Since $(\hat{s}_\theta, P) \in \bar{\Omega}$, $\hat{s}_\theta$ is an RCE strategy profile in this game, and hence is rationalizable under common certainty of $\hat{P}_\theta$. Thus, $\hat{s}_\theta$ must coincide with the unique Nash equilibrium at $\hat{P}_\theta$.

Finally, since $(s, \hat{s}, \hat{P})$ is rationalized by $(\hat{\Omega}, \nu)$ at ω, the above observations imply that the observational consistency conditions $H^\omega_\theta = H^{\hat{s}_\theta, \hat{P}_\theta}_\theta$ and $s(\theta) = \hat{s}_\theta(\theta)$ hold for each θ. Thus, $(s, \hat{s}, \hat{P})$ is an LPE at P.

"Only if" direction: Suppose that $(s, \hat{s}, \hat{P})$ is an LPE at P.

Define $\hat{\Omega} := \{(s, P)\} \cup \{(\hat{s}_\theta, \hat{P}_\theta) : \theta \in \Theta\}$. Take any $\omega \in \hat{\Omega}$. If $\omega = (\hat{s}_\theta, \hat{P}_\theta)$ for some θ, then for each θ' we construct $\nu_{\omega,\theta'} \in \Delta(\Omega)$ by setting $\nu_{\omega,\theta'}(\{\omega\}) = 1$. If instead $\omega = (s, P)$, then for each θ we construct $\nu_{\omega,\theta} \in \Delta(\Omega)$ by setting $\nu_{\omega,\theta}(\{(\hat{s}_\theta, \hat{P}_\theta)\}) = 1$.

By construction $(\hat{\Omega}, \nu)$ satisfies belief-closedness. For each ω with $\omega = (\hat{s}_\theta, \hat{P}_\theta)$ for some θ, the rationality condition is satisfied since $\hat{s}_\theta$ is Nash at $\hat{P}_\theta$, and the observational consistency condition is satisfied because $\nu_{\omega,\theta'}(\{\omega\}) = 1$ for each θ'. If $\omega \neq (\hat{s}_\theta, \hat{P}_\theta)$ for all θ, which implies $\omega = (s, P)$, the fact that $(s, \hat{s}, \hat{P})$ is LPE at P implies the rationality and observational consistency condition. Therefore, $(\hat{\Omega}, \nu)$ is a rational perception system.

To see that $(\hat{\Omega}, \nu)$ displays lexicographic preference for dogmaticity at (s, P), take any θ. Then $\text{supp}\nu_{(s,P),\theta} = \{(\hat{s}_\theta, \hat{P}_\theta)\}$. Moreover $\nu_{(\hat{s}_\theta, \hat{P}_\theta),\theta'}(\{(\hat{s}_\theta, \hat{P}_\theta)\}) = 1$ for any θ' by construction. Thus θ is dogmatic at (s, P). □

I.3.2 Proof of Proposition I.2

"If" direction: Suppose that $(s, \hat{s}, \hat{P})$ is rationalized by a rational perception system $(\hat{\Omega}, \nu)$ at P that displays lexicographic preference for dogmaticity and independence at (s, P).

We show that $\nu_{\omega,\theta}(\{\omega' : C_{\omega'} = C_I\}) = 1$ for any θ. For this purpose, as in the proof Proposition I.1, we define

$$\hat{\Omega}^* := \{(s, (F, C_I)) : F \in \mathcal{F}, s \text{ is Nash at } (F, C_I)\}$$

For each $\omega' \in \hat{\Omega}^*$ and θ', define $\nu^*_{\omega',\theta'} \in \Delta(\Omega)$ by $\nu^*_{\omega',\theta'}(\{\omega'\}) = 1$. $(\hat{\Omega}^*, \nu^*)$ was verified to be a rational perception system in the proof of Proposition I.1.

By the assumption on s, cdf H^{ω}_{θ} is strictly increasing and continuous. By Lemma 3, there is $\omega^* \in \hat{\Omega}^*$ such that $H^{\omega^*}_{\theta} = H^{\omega}_{\theta}$. This, combined with the rationality condition, implies $u_\theta(\nu^*_{\omega^*,\theta}) = u_\theta(\nu_{\omega,\theta})$. Note also that $\nu^*_{\omega^*,\theta}(\{\omega' : C_{\omega'} = C_I\}) = 1$. Therefore, by the lexicographic preference condition, it follows that $\nu_{\omega,\theta}(\{\omega' : C_{\omega'} = C_I\}) = 1$. Thus, $\hat{C}_\theta = C_I$.

Note that by Proposition I.1, $(s, \hat{s}, \hat{P})$ is an LPE at P. Since $\hat{C}_\theta = C_I$ for all θ, it is in fact the ANLPE at P.

"Only if" direction: Suppose that $(s, \hat{s}, \hat{P})$ is the ANLPE at P.

Define $\hat{\Omega} := \{(s, P)\} \cup \Big(\cup_\theta (\hat{s}_\theta, \hat{P}_\theta)\Big)$. Take any $\omega \in \hat{\Omega}$. If $\omega = (\hat{s}_\theta, \hat{P}_\theta)$ for some θ, then for each θ' we construct $\nu_{\omega,\theta'} \in \Delta(\Omega)$ by setting $\nu_{\omega,\theta'}(\{\omega\}) = 1$. If not, which means $\omega = (s, P)$, then for each θ we construct $\nu_{\omega,\theta} \in \Delta(\Omega)$ by setting $\nu_{\omega,\theta}(\{(\hat{s}_\theta, \hat{P}_\theta)\}) = 1$.

As shown in the proof of Proposition I.1, $(\hat{\Omega}, \nu)$ is a rational perception system that displays lexicographic preference for dogmaticity at (s, P). To see that it also displays lexicographic preference for independence at (s, P), note that each θ satisfies $\nu_{(s,P),\theta}(\{(\hat{s}_\theta, (\hat{F}_\theta, C_I))\}) = 1$ for some $\hat{F}_\theta$ since $(s, \hat{s}, \hat{P})$ is the ANLPE at P. □