Supplemental Material to
Dispersed Behavior and Perceptions in Assortative Societies

By

Mira Frick, Ryota Iijima, and Yuhta Ishii

April 2018

Revised March 2019

COWLES FOUNDATION DISCUSSION PAPER NO. 2128RS

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

http://cowles.yale.edu/

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

Mira Frick, Ryota Iijima, and Yuhta Ishii

C Omitted Proofs

C.1 Proofs for Appendix A

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

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

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, $T_C f(x) \geq T_C f(x')$ for all $x \geq x'$, so that $T_C f$ is weakly increasing. To show that $T_C f$ is absolutely continuous, note that for each $x, x' \in (0,1)$,

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

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 $\tau > \tau'$,

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

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

Proof of Lemma A.3. It is clear from the definition that $\succsim_m$ is reflexive and transitive; moreover, by Lemma A.2, $\succsim_m$ 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 A.2.

To verify that $\succsim_m$ is isotone, take any $f, g \in \mathcal{I}$ such that $f \succsim_m g$ and set $h := f - g$. Note that $\int_0^1 h(x)dx = \int_0^1 T_C h(x)dx = 0$. It suffices to show that $\int_y^1 T_C h(x)dx \geq 0$ for all $y \in (0,1)$. To see this, note that $\int_y^1 T_C h(x)dx$ is given by

$$\begin{aligned}
\int_y^1 \int_0^1 h(z)c(z|x)dzdx &= \int_0^1 \int_y^1 c(z|x)dxh(z)dz = \int_0^1 (1 - C(y|z))h(z)dz \\
&= -\int_0^1 \frac{\partial 1 - C(y|z)}{\partial z} \int_0^z h(z')dz'dz + \left[(1 - C(y|z)) \int_0^z h(z')dz'\right]_0^1 \\
&= \int_0^1 \frac{\partial C(y|z)}{\partial z} \int_0^z h(z')dz'dz \geq 0,
\end{aligned}$$

where the second equality uses $\int_y^1 c(z|x)dx = \int_y^1 c(x|z)dx = 1 - C(y|z)$, the third holds by integration by parts (using absolute continuity of c), the fourth uses $\int_0^1 h(z)dz = 0$, and the final inequality uses $\int_0^z h(z')dz' \leq 0$ (by $f \succsim_m g$) and assortativity of C. □

Proof of Lemma A.4. 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 $f \succsim_m Tf$ for all $F \in \mathcal{I}$. 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 $Tf(x) = \frac{d}{dx} \int_0^1 K(x,y)f(y)dy$ 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) $\text{essup}_y V(K(\cdot,y)) < \infty$, 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) $x_1 < x_2 \implies K(x_1,\cdot) \leq K(x_2,\cdot)$; and (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, $f \succsim_m T_C f$, proving the base case. The inductive step then follows from isotonicity of $\succsim_m$ (Lemma A.3). □

C.2 $\succsim_d$ order

This subsection records basic properties of the dispersiveness order that will be used in later proofs. Define the order $\succsim_d$ over $\mathcal{I}$: $f \succsim_d g$ if for all $x, x' \in (0,1)$ such that $x \geq x'$, we have $f(x) - f(x') \geq g(x) - g(x')$. Note that for $F, G \in \mathcal{F}$, F is more dispersive than G iff $F^{-1} \succsim_d G^{-1}$. Observe that $\succsim_d$ is *shift invariant*: for any $f, g \in \mathcal{I}$ and constant functions m, m', $f \succsim_d g$ iff $f + m \succsim_d g + m'$. We verify that $\succsim_d$ also obeys the three basic properties from Appendix A.2:

Lemma C.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

$$f_n(x) = \begin{cases} f(\frac{1}{n}) \text{ if } x \in (0, \frac{1}{n}) \\ f(x) \text{ if } x \in [\frac{1}{n}, \frac{n-1}{n}] \\ f(\frac{n-1}{n}) \text{ if } x \in (\frac{n-1}{n}, 1) \end{cases} \qquad g_n(x) = \begin{cases} g(\frac{1}{n}) \text{ if } x \in (0, \frac{1}{n}) \\ g(x) \text{ if } x \in [\frac{1}{n}, \frac{n-1}{n}] \\ g(\frac{n-1}{n}) \text{ if } x \in (\frac{n-1}{n}, 1) \end{cases} \tag{20}$$

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. □

C.3 Proof of Proposition 3

(1) ⇒ (2): Take any $F_1, F_2 \in \mathcal{F}$ such that F_1 is more dispersive than F_2. Then $F_1^{-1} \succsim_d F_2^{-1}$. First, we inductively show that for each t, $(T_C)^t F_1^{-1} \succsim_d (T_C)^t F_2^{-1}$. Indeed, supposing that the claim is true at t, isotonicity of $\succsim_d$ implies

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

as required. Next, since $\succsim_d$ is linear, we have $\sum_{t=0}^{\tau} \gamma^t (T_C)^t F_1^{-1} \succsim_d \sum_{t=0}^{\tau} \gamma^t (T_C)^t F_2^{-1}$ 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 $i = 1, 2$ and any $\gamma \in [0,1)$, continuity and linearity of $\succsim_d$ then yields

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

whence H_1^{NE} is more dispersive than H_2^{NE}.

(2) ⇒ (3): Immediate from the fact that the ANE action distribution at (P, γ, β) coincides with the Nash action distribution at $(P, \gamma + \beta, 0)$.

(3) ⇒ (1): Immediate from the fact that the ANE action distribution at $(P, 0, 0)$ coincides with F.

(1) ⇔ (4): To see that (1) implies (4), for any $x > x'$,

$$\begin{aligned}
& \left((H_1^{AN})^{-1}(x) - (H_1^{AN})^{-1}(x')\right) - \left((H_1^{NE})^{-1}(x) - (H_1^{NE})^{-1}(x')\right) \\
= \;& (1-\gamma-\beta)\sum_{t\geq 0}\left((\gamma+\beta)^t - \gamma^t\right)\left((T_C)^t F_1^{-1}(x) - (T_C)^t F_1^{-1}(x')\right) \\
\geq \;& (1-\gamma-\beta)\sum_{t\geq 0}\left((\gamma+\beta)^t - \gamma^t\right)\left((T_C)^t F_2^{-1}(x) - (T_C)^t F_2^{-1}(x')\right) \\
= \;& \left((H_2^{AN})^{-1}(x) - (H_2^{AN})^{-1}(x')\right) - \left((H_2^{NE})^{-1}(x) - (H_2^{NE})^{-1}(x')\right).
\end{aligned}$$

Here the inequality holds since by the proof of "(1) ⇒ (2)," we have $(T_C)^t F_1^{-1} \succsim_d (T_C)^t F_2^{-1}$. This establishes the first inequality in (4). The second inequality in (4) holds by the fact that (1) implies (2), as shown above.

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

C.4 Proof of Proposition 5

Fix θ, and let s_i and $(\hat{s}_\theta^i, \hat{P}_\theta^i)$ denote the actual strategy profile and θ's perceptions under the coherent ANE at (P, β_i, γ_i) for $i = 1, 2$. For any $x > y$ in $(0, 1)$

$$\begin{aligned}
(\hat{F}_\theta^i)^{-1}(x) - (\hat{F}_\theta^i)^{-1}(y) &= (1-\gamma_i-\beta_i)^{-1}\left((H_\theta^{\hat{s}_\theta^i, \hat{P}_\theta^i})^{-1}(x) - (H_\theta^{\hat{s}_\theta^i, \hat{P}_\theta^i})^{-1}(y)\right) \\
&= (1-\gamma_i-\beta_i)^{-1}\left((H_\theta^{s^i, P})^{-1}(x) - (H_\theta^{s^i, P})^{-1}(y)\right) \\
&= \sum_{t\geq 0}(\gamma_i+\beta_i)^t\left((T_C)^t F^{-1}(x) - (T_C)^t F^{-1}(y)\right),
\end{aligned}$$

where the first equality holds by coherency and the fact that $\hat{C}_\theta^i = C_I$, the second equality by observational consistency, and the final one by construction of ANE strategies. As $(\gamma_1 + \beta_1)^t \leq (\gamma_2 + \beta_2)^t$ and $(T_C)^t F^{-1}(x) - (T_C)^t F^{-1}(y) \geq 0$ for all t, it follows that $(\hat{F}_\theta^1)^{-1}(x) - (\hat{F}_\theta^1)^{-1}(y) \leq (\hat{F}_\theta^2)^{-1}(x) - (\hat{F}_\theta^2)^{-1}(y)$, so that $\hat{F}_\theta^2$ is more dispersive than $\hat{F}_\theta^1$. □

C.5 Effect of coordination incentives on behavior

Proposition C.1. *Fix any P and $\beta_1 + \gamma_1 \leq \beta_2 + \gamma_2$. Then H_1^{AN} is a mean-preserving spread of H_2^{AN} .*

Proof. Denote by h_i the inverse cdf of ANE action distribution at (P, γ_i, β_i). For each quantile $x \in (0, 1)$, we have $h_i(x) = (1 - \eta_i)\sum_{t\geq 0}\eta_i^t (T_C)^{t+1} f(x)$, where $\eta_i = \gamma_i + \beta_i$ and $f = F^{-1}$. This implies that for all $\tau \geq 0$, we have

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

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

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

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 $\eta_1 \leq \eta_2$ (so that $\frac{\sum_{t=0}^{\tau} \eta_1^t}{\sum_{t=0}^{\tau+1} \eta_1^t} \geq \frac{\sum_{t=0}^{\tau} \eta_2^t}{\sum_{t=0}^{\tau+1} \eta_2^t}$ and $\frac{\eta_1^{\tau+1}}{\sum_{t=0}^{\tau+1} \eta_1^t} \leq \frac{\eta_2^{\tau+1}}{\sum_{t=0}^{\tau+1} \eta_2^t}$) and that $(T_C)^{t+1} f \succsim_m (T_C)^{\tau+2} f$ for all $t \leq \tau + 1$ (by Lemma A.4).

Taking $\tau \to \infty$ in (21), continuity of $\succsim_m$ then yields $(1 - \eta_2) \sum_{t \geq 0} \eta_2^t (T_C)^{t+1} f \precsim_m (1 - \eta_1) \sum_{t \geq 0} \eta_1^t (T_C)^{t+1} f$, i.e., $h_2 \precsim_m h_1$, as claimed. □

C.6 Proof of Proposition 6

Recall that each type θ's perceived population mean under coherent ANE is given by $\hat{\mu}_\theta = \mathbb{E}_P[s^{AN}(\theta')|\theta]$. Thus, $(\hat{M}^i)^{-1} = T_C h_i^{AN}$ where h_i^{AN} denotes the inverse cdf of the ANE action distribution at (P, γ_i, β_i). Since $h_1^{AN} \succsim_m h_2^{AN}$ by Proposition C.1, isotonicity of $\succsim_m$ implies that $(\hat{M}^1)^{-1} \succsim_m (\hat{M}^2)^{-1}$. The final claim follows directly from the fact that $\hat{\mu}_\theta = \mathbb{E}_P[s^{AN}(\theta')|\theta] = (1 - \gamma - \beta) \sum_{t=0}^{\infty} (\gamma + \beta)^t \mathbb{E}_P[\theta_{t+1}|\theta_0 = \theta]$ by (7). □

C.6.1 A condition for asymptotically correct perceptions

Lemma C.2. *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. 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. □

D Details for Section 5.1

D.1 Hybrid model

We consider a setting where fraction $\alpha \in [0, 1]$ of agents of each type θ are modeled as under ANE, while the remaining fraction $1 - \alpha$ of agents of each type best-respond to correct perceptions.[60] As this section focuses only on behavior, we do not explicitly model agents' perceptions and instead employ a simplified equilibrium formulation based on the correct and misspecified best-response functions BR_θ and BR_θ^{AN}:[61]An ***α-assortativity neglect equilibrium (α-ANE)*** consists of strategy profiles s_a for assortativity neglect agents and s_c for correct agents such that for all θ, we have

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

Iterating best-responses, Appendix D.1.1 obtains the following Markov process characterizations of equilibrium strategy profiles:

$$s_a(\theta_0) = (1-\gamma-\beta)\left(\theta_0 + (\gamma+\beta)\sum_{t\geq 1}(\gamma+\alpha\beta)^{t-1}\mathbb{E}_P[\theta_t|\theta_0]\right) + \frac{(\gamma+\beta)(1-\alpha)\beta\mathbb{E}_F[\theta]}{1-\gamma-\alpha\beta} \quad (22)$$

$$s_c(\theta_0) = (1-\gamma-\beta)\left(\theta_0 + \gamma\sum_{t\geq 1}(\gamma+\alpha\beta)^{t-1}\mathbb{E}_P[\theta_t|\theta_0]\right) + \frac{1-(\gamma+\beta)\alpha}{1-\gamma-\alpha\beta}\beta\mathbb{E}_F[\theta]. \quad (23)$$

[60]We assume a continuum of copies of each type and a matching technology that depends only on types (and not on copies).

[61]Given the unique derivation of α-ANE behavior below, Lemma B.2 again yields unique perceptions for assortativity neglect agents that satisfy coherency and observational consistency.

These quasi-hyperbolic discounting expressions generalize the exponential discounting expressions for ANE and Nash: (22) reduces to (7) when $\alpha = 1$, and likewise (23) reduces to (5) when $\alpha = 0$. Letting H_a^α and H_c^α denote the induced action distributions among assortativity neglect and correct agents, this yields the following:

Proposition D.1. *Fix any (P, γ, β). For any $\alpha \in [0, 1]$, there is a unique α-ANE. Strategy profiles s_a and s_c are given by (22) and (23). Moreover, (i) $H^{AN} \succsim_m H_a^\alpha \succsim_m H_c^\alpha \succsim H^{NE}$; and (ii) H_a^α and H_c^α are increasing in α with respect to $\succsim_m$.*

Thus, behavior among correct agents is less dispersed than among assortativity neglect agents, but action dispersion among both groups is exacerbated the greater the share of assortativity neglect agents. Finally, one can show that both H_a^α and H_c^α are subject to the same comparative statics and multiplier effect as H^{AN} in Propositions 2-3.

D.1.1 Proof of Proposition D.1

Let $\mu := \mathbb{E}_F[\theta]$. Let h_a^α and h_c^α denote strategy profiles of AN and correct agents as functions of quantiles. Also let $h^\alpha := \alpha h_a^\alpha + (1-\alpha)h_c^\alpha$. In α-ANE, the best response conditions imply

$$h_a^\alpha(x) = (1-\beta-\gamma)F^{-1}(x)+(\gamma+\beta)T_C h^\alpha(x), \;\; h_c^\alpha(x) = (1-\beta-\gamma)F^{-1}(x)+\gamma T_C h^\alpha(x)+\int_0^1 h^\alpha(y)dy$$

for each $x \in (0, 1)$. Since $h^\alpha = \alpha h_a^\alpha + (1-\alpha)h_c^\alpha$, it follows that

$$h^\alpha(x) = (1-\beta-\gamma)F^{-1}(x) + (\gamma+\alpha\beta)T_C h^\alpha(x) + (1-\alpha)\beta \int_0^1 h^\alpha(y)dy$$

for each x, which implies $\int_0^1 h^\alpha(y)dy = \mu$ by integrating both sides over x. Moreover, iterating the above equation we obtain

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

where the convergence of the RHS can be shown as in the proof of Lemma 1. Note that h^α is uniquely determined for any α. By the best response conditions, we obtain

$$\begin{aligned}
h_a^\alpha(x) &= (1-\beta-\gamma)F^{-1}(x) + (\gamma+\beta)T_C h^\alpha(x) \\
&= (1-\beta-\gamma)\left(F^{-1}(x) + (\gamma+\beta)\sum_{t\geq 1}(\gamma+\alpha\beta)^{t-1}(T_C)^t F^{-1}(x)\right) + \frac{(\gamma+\beta)(1-\alpha)\beta\mu}{1-\gamma-\alpha\beta}, \\
h_c^\alpha(x) &= (1-\beta-\gamma)F^{-1}(x) + \gamma T_C h^\alpha(x) + \beta\int_0^1 h^\alpha(y)dy \\
&= (1-\beta-\gamma)\left(F^{-1}(x) + \gamma\sum_{t\geq 1}(\gamma+\alpha\beta)^{t-1}(T_C)^t F^{-1}(x)\right) + \frac{1-\alpha(\gamma+\beta)}{1-\gamma-\alpha\beta}\beta\mu
\end{aligned}$$

for each x, yielding (22)-(23). Integrating both sides of the equations over x, we obtain that $\int_0^1 h_a^\alpha(x)dx = \int_0^1 h_c^\alpha(x)dx = \mu$. Then the claim $h_a^\alpha \succsim_m h_c^\alpha$ and the comparative statics in α

can be verified directly by linearity of $\succsim_m$. Finally, observe that $(H^{AN})^{-1} = h_a^\alpha$ at $\alpha = 1$ and $(H^{NE})^{-1} = h_c^\alpha$ at $\alpha = 0$. $\square$

D.2 Weaker forms of assortativity neglect

This section shows how several of our main insights extend to weaker forms of assortativity neglect. Section D.2.1 focuses on the finding that assortativity neglect increases action dispersion. Section D.2.2 focuses on coherency.

D.2.1 Increased action dispersion

Fix any society P. Since P is assortative, the distribution $C(\cdot|F(\theta))$ of θ's match quantiles is FOSD-increasing in θ. By contrast, under ANE, perceived match quantile distributions $\hat{C}_\theta(\cdot|\hat{F}_\theta(\theta))$ are the same for all agents θ, namely uniform on $[0,1]$. We now show that the finding that ANE features more dispersed behavior than Nash extends to a weaker notion of assortativity neglect, where agents' perceived match quantile distributions are not necessarily constant in θ, but are "less increasing" in θ than actual.

Formally, given any two families (G_θ), $(\hat{G}_\theta)$ of strictly increasing cdfs over $[0,1]$ such that G_θ is FOSD-increasing in θ, we say that $(\hat{G}_\theta)$ is *less increasing* in θ than (G_θ) if for all $x \in [0,1]$, $\hat{G}_\theta^{-1}(G_\theta(x))$ is weakly decreasing in θ. To understand this definition, note that since (G_θ) is FOSD-increasing in θ, $G_\theta(x)$ is decreasing in θ. If $(\hat{G}_\theta)$ is FOSD-decreasing in θ, then $\hat{G}_\theta(y)$ is increasing in θ for all y, and hence $\hat{G}_\theta^{-1}(G_\theta(x))$ is always decreasing in θ. If $(\hat{G}_\theta)$ is FOSD-increasing in θ, then $\hat{G}_\theta(y)$ is decreasing in θ, so that $\hat{G}_\theta^{-1}(G_\theta(x))$ may be increasing or decreasing in θ; if it is decreasing in θ, this captures that the "rate" at which $\hat{G}_\theta(x)$ decreases in θ does not exceed that of $G_\theta(x)$.

We say that perceptions $(\hat{P}_\theta)$ exhibit ***partial assortativity neglect (PAN)*** relative to P if the family $\left(\hat{C}_\theta(\cdot|\hat{F}_\theta(\theta))\right)$ of perceived match quantile distributions is less increasing in θ than the actual match quantile distributions $\left(C(\cdot|F(\theta))\right)$. As noted, PAN trivially holds if $\hat{C}_\theta = C_I$ for all θ, and it also holds under correct perceptions ($\hat{P}_\theta = P$ for all θ).[62] Observe that unlike assortativity neglect, PAN entails no restrictions on how any type θ perceives *other* types' match distributions. Such restrictions do not affect equilibrium behavior, but are crucial for analyzing coherency (see Section D.2.2).

For any $(\hat{P}_\theta)$ that exhibits PAN relative to P, we define a Markov process P^* over types $(\theta_0, \theta_1, ...)$ whose initial distribution is F and whose transition kernel is given by

$$P^*(\cdot|\theta) := \frac{\beta}{\beta+\gamma}\hat{C}_\theta^{-1}\left(P(\cdot|\theta)\middle|\hat{F}_\theta(\theta)\right) + \frac{\gamma}{\beta+\gamma}P(\cdot|\theta). \tag{24}$$

Note that the fact that $(\hat{P}_\theta)$ exhibits PAN implies that process P^* is monotone. We say that P^* is *bounded* if $\limsup_t \mathbb{E}_{P^*}[|s(\theta_t)|] < \infty$ for any L^1 function s, and *continuous* if $P^*(\theta'|\theta)$ admits absolutely continuous partial derivatives. We call an equilibrium *monotone* if both the true strategy profile s and each type's perceived strategy profile $\hat{s}_\theta$ are strictly increasing.

[62] In the former case $\hat{C}_\theta^{-1}(C(x|F(\theta))|\hat{F}_\theta(\theta)) = C(x|F(\theta))$ and in the latter $\hat{C}_\theta^{-1}(C(x|F(\theta))|\hat{F}_\theta(\theta)) = x$, both of which are weakly decreasing in θ.

Proposition D.2. *Fix any (P, γ, β). If $(\hat{P}_\theta)$ exhibits PAN relative to P and P^* is bounded and continuous, then there exists a unique monotone equilibrium with perceptions $(\hat{P}_\theta)$. Its strategy profile is given by*

$$s^*(\theta) = (1-\gamma-\beta)\sum_{t\geq 0}(\gamma+\beta)^t \mathbb{E}_{P^*}[\theta_t|\theta_0=\theta]. \tag{25}$$

The corresponding action distribution is more dispersive than the Nash action distribution.

Proof. See Section D.2.3. □

Expression (25) generalizes the Markov process characterizations of Nash and ANE. If $\hat{P}_\theta = P$ for each θ, then the first term in (24) is $\frac{\beta}{\beta+\gamma}F$, and (25) reduces to the Nash strategy (5). If $\hat{C}_\theta = C_I$ for each θ, then the first term in (24) is $P(\cdot|\theta)$, and (25) reduces to the ANE strategy (7). As under ANE, PAN always leads to a more dispersive action distribution than Nash.[63] Analogous to Proposition 1, this result reflects a "false consensus effect," as PAN implies that agents' perceived global action distributions are FOSD-increasing in their types.

Example 6. For any Gaussian society $P = (\mu, \sigma^2, \rho)$, Appendix E.4 constructs a family of coherent equilibria. Each equilibrium is parametrized by a correlation coefficient $\hat{\rho} \in [0, 1)$ that is commonly perceived by all agents. We show that the equilibria in this family that exhibit PAN are precisely those where $\hat{\rho} \leq \rho$. In each such equilibrium, behavior takes the form $s^*(\theta) = \frac{1-\beta-\gamma}{1-\gamma\rho-\beta\frac{\rho-\hat{\rho}}{1-\hat{\rho}}}\theta + \frac{\gamma(1-\rho)+\beta\frac{1+\hat{\rho}}{1-\rho}}{1-\gamma\rho-\beta\frac{\rho-\hat{\rho}}{1-\hat{\rho}}}\mu$. Action dispersion (in terms of both mean-preserving spread and dispersiveness) is increasing in $\hat{\rho}$ and hence greater than Nash ($\hat{\rho} = \rho$) and less than ANE ($\hat{\rho} = 0$). Moreover, each PAN equilibrium features the same comparative statics and multiplier effect for action variance as in Example 2 and the same predictions for perceived type variance (overestimation vs. underestimation) as in Example 3. ▲

D.2.2 Coherency

Propositions 4 and 7 showed that assortativity neglect is the only perception of the interaction structure that agents can coherently sustain across *all* environments. Generalizing Example 5, this section formalizes a sense in which in any *fixed* environment, it is "easier" for agents to coherently perceive lower levels of assortativity than higher levels.

We use the following strenghtening of the more-assortative order from Section 3.2. Say that 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, 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 this effect is globally stronger under C_1 than C_2. Note that any interaction structure is strongly more assortative than C_I.[64]

[63] The action distribution under (25) need not be a mean-preserving spread of Nash, as the average action may differ from Nash. This is because process P^* need not be stationary (and F need not be an invariant distribution of P^*).

[64] An analog of Proposition 2 holds by replacing $\succsim_{MA}$ with $\succsim_{SMA}$ and mean-preserving spread with the dispersive order. The proof is analogous to that of the original proposition, making use of the fact that $\succsim_{SMA}$ is the dual order of $\succsim_d$ (Lemma D.2).

The following result shows that in coherent equilibrium, perceiving less assortative interaction structures is easier than perceiving more assortative ones, in the sense that whenever θ can sustain perception $\hat{C}_1$ in a coherent equilibrium, she can also coherently sustain any perception $\hat{C}_2$ that is strongly less assortative than $\hat{C}_1$. The result is subject to one restriction: We compare perceptions that differ in the distributions of matches that θ perceives *other* agents to interact with, while holding fixed θ's perception of her *own* quantile and match quantile distribution (as was the case for θ_m's perception in Example 5).

Proposition D.3. *Fix any (P, γ, β) and θ. Suppose there exists a coherent equilibrium in which θ's perceived interaction structure is $\hat{C}_1$ and θ's perceived quantile is $\hat{x}$. Then for any $\hat{C}_2$ with $\hat{C}_1 \succsim_{SMA} \hat{C}_2$ and $\hat{C}_1(\cdot|\hat{x}) = \hat{C}_2(\cdot|\hat{x})$, there exists a coherent equilibrium in which θ's perceived interaction structure is $\hat{C}_2$ and θ's perceived quantile is $\hat{x}$.*

Proof. See Section D.2.4. □

Holding fixed θ's own perceived match distribution across the two perceptions allows us to isolate the effect that θ's perceived level of assortativity has on her ability to rationalize other agents' behavior: Indeed, in constructing the equilibrium under $\hat{C}_2$ we can ensure that θ's perceived global action distribution $\hat{H}$ is the same as in the original equilibrium under $\hat{C}_1$. Generalizing the logic behind (8), the proof of Proposition D.3 then relies on the idea that if θ can rationalize action distribution $\hat{H}$ under perception $\hat{C}_1$, she can also rationalize $\hat{H}$ under the less assortative $\hat{C}_2$ by attributing some of the action dispersion she previously attributed to differences in coordination incentives to differences in types.

As Nash equilibrium is coherent, Proposition D.3 immediately implies the following:

Corollary D.1. *Fix any (P, γ, β) and θ. For any $\hat{C}$ with $C \succsim_{SMA} \hat{C}$ and $C(\cdot|F(\theta)) = \hat{C}(\cdot|F(\theta))$, there exists a coherent equilibrium in which θ's perceived interaction structure is $\hat{C}$.*

For any given environment, Corollary D.1 implies that θ can coherently sustain a continuum range of perceived interaction structures $\hat{C}$, each of which is strongly less assortative than the true C and strongly more assortative than C_I. By contrast, as Example 5 illustrated, there may be limits on how much θ can overestimate assortativity.

D.2.3 Proof of Proposition D.2

Observe first that s^* given by (25) is a well-defined L^1 strategy profile. Indeed,

$$\int \left|(1-\gamma-\beta)\sum_{t\geq 0}(\gamma+\beta)^t \mathbb{E}_{P^*}[\theta_t|\theta_0=\theta]\right| dF(\theta)$$

$$\leq \int (1-\gamma-\beta)\sum_{t\geq 0}(\gamma+\beta)^t \mathbb{E}_{P^*}[|\theta_t||\theta_0=\theta] dF(\theta) = (1-\gamma-\beta)\sum_{t\geq 0}(\gamma+\beta)^t \mathbb{E}_{P^*}[|\theta_t|] < \infty$$

by the boundedness assumption on P^*. Note that PAN ensures that P^* is monotone (i.e., $P^*(\cdot|\theta)$ is FOSD increasing), which implies that s^* is strictly increasing.

We first show that strategy profile s under any monotone equilibrium with perceptions $(\hat{P}_\theta)$ must coincide with s^*. Take any type θ and her corresponding perceived strategy profile $\hat{s}_\theta$ in equilibrium. For any $\hat{\theta}$, the action $a = \hat{s}_\theta(\hat{\theta})$ must satisfy

$$P(s^{-1}(a)|\theta) = \hat{P}_\theta(\hat{\theta}|\theta) \iff a = s\left(P^{-1}(\hat{P}_\theta(\hat{\theta}|\theta)|\theta)\right)$$

by observational consistency and monotonicity. Thus, $\hat{s}_\theta(\hat{\theta}) = s\left(\phi(\hat{\theta})\right)$ where $\phi : \text{supp}\hat{F}_\theta \to \text{supp}F$ is a strictly increasing and differentiable bijection defined by $\phi(\hat{\theta}) := P^{-1}(\hat{P}_\theta(\hat{\theta}|\theta)|\theta)$. Letting $\hat{f}_\theta$ denote the density of $\hat{F}_\theta$, θ's perceived global action average (which is well-defined in equilibrium) takes the form

$$\begin{aligned}\int \hat{s}_\theta(\hat{\theta})d\hat{F}_\theta(\hat{\theta}) &= \int s(\phi(\hat{\theta}))\hat{f}_\theta(\hat{\theta})d\hat{\theta} = \int s(\theta')\hat{f}_\theta(\phi^{-1}(\theta'))\frac{d\hat{\theta}}{d\theta'}d\theta' \\ &= \int s(\theta')\frac{\hat{f}_\theta(\phi^{-1}(\theta'))}{\phi'(\phi^{-1}(\theta'))}d\theta' = \int s(\theta')d\hat{F}_\theta(\phi^{-1}(\theta')),\end{aligned}$$

based on the change of variables $\theta' = \phi^{-1}(\hat{\theta})$. Observe that $\hat{F}_\theta(\phi^{-1}(\theta')) = \hat{F}_\theta\left(\hat{P}_\theta^{-1}(P(\cdot|\theta)|\theta)\right) = \hat{C}_\theta^{-1}\left(P(\cdot|\theta)|\hat{F}_\theta(\theta)\right)$. Thus by the best-response requirement,

$$s(\theta) = (1-\gamma-\beta)\theta+\beta\int s(\theta')d\hat{F}_\theta(\phi^{-1}(\theta'))+\gamma\int s(\theta')dP(\theta'|\theta) = (1-\gamma-\beta)\theta+(\gamma+\beta)\int s(\theta')dP^*(\theta'|\theta). \tag{26}$$

Thus, any strategy profile s that arises in a monotone equilibrium with perceptions $(\hat{P}_\theta)$ must satisfy (26). To see that this implies $s = s^*$, observe that iterating (26) leads to $s(\theta) = (1-\gamma-\beta)\sum_{t\geq 0}^{\tau}(\gamma+\beta)^t\mathbb{E}_{P^*}[\theta_t|\theta_0 = \theta] + (\gamma+\beta)^{\tau+1}\mathbb{E}_{P^*}[s(\theta_{\tau+1})|\theta_0 = \theta]$ for any $\tau > 0$. Then

$$\begin{aligned}\int |s(\theta) - s^*(\theta)|dF(\theta) &= \mathbb{E}_{P^*}[|(1-\gamma-\beta)\sum_{t\geq\tau}(\gamma+\beta)^t\theta_t - (\gamma+\beta)^{\tau+1}s(\theta_{\tau+1})|] \\ &\leq (1-\gamma-\beta)\sum_{t\geq\tau}(\gamma+\beta)^t\mathbb{E}_{P^*}[|\theta_t|] + (\gamma+\beta)^{\tau+1}\mathbb{E}_{P^*}[|s(\theta_{\tau+1})|] \to 0\end{aligned}$$

as $\tau \to \infty$ as P^* is bounded and s is L^1. Hence, $s = s^*$.

The above shows that defining θ's perceived strategy profile $\hat{s}_\theta$ by $\hat{s}_\theta(\cdot) = s^*(P^{-1}(\hat{P}_\theta(\cdot|\theta)|\theta))$ is the unique way to satisfy the monotonicity, observational consistency, and best-response requirements under perceptions $(\hat{P}_\theta)$. This establishes the unique existence of a monotone equilibrium with perceptions $(\hat{P}_\theta)$.

Finally, in order to show that the equilibrium action distribution is more dispersive than the Nash action distribution, we apply the following lemma to $P_1^* = P^*$, with associated strategy profile $s_1^* = s^*$, and $P_2^* = \frac{\gamma}{\gamma+\beta}P + \frac{\beta}{\gamma+\beta}(F)^2$ (the second term denotes the product measure of F), with associated strategy profile $s_2^* = s^{NE}$. □

Lemma D.1. *Let P_i^* be monotone, bounded, and continuous and suppose that $P_1^*(\bar{\theta}|\theta') - P_1^*(\bar{\theta}|\theta'') \leq P_2^*(\bar{\theta}|\theta') - P_2^*(\bar{\theta}|\theta'')$ for each $\bar{\theta}$, and $\theta' > \theta''$. Let $s_i^*(\theta) = (1-\gamma-\beta)\sum_{t\geq 0}(\gamma+\beta)^t\mathbb{E}_{P_i^*}[\theta_t|\theta]$ for each θ. Then $s_1^*(\theta') - s_1^*(\theta'') \geq s_2^*(\theta') - s_2^*(\theta'')$ for each $\theta' > \theta''$.*

Proof. Let $P_i^{*t}(\cdot|\theta)$ denote the distribution of θ's t-step ahead matches under P_i^*. Since P_i^* is FOSD-increasing in θ, so is P_i^{*t}.

We first show inductively that for each t, $P_1^{*t}(\bar{\theta}|\theta') - P_1^{*t}(\bar{\theta}|\theta'') \leq P_2^{*t}(\bar{\theta}|\theta') - P_2^{*t}(\bar{\theta}|\theta'')$ for all $\bar{\theta}$ and $\theta' > \theta''$. The claim for $t = 1$ holds by assumption. Assume the claim holds for $t = k$. For any $\bar{\theta}$ and $\theta' > \theta''$, integration by parts (which applies by the continuity of P^*) yields

$$
\begin{aligned}
& P_1^{*k+1}(\bar{\theta}|\theta') - P_1^{*k+1}(\bar{\theta}|\theta'') = \int P_1^*(\bar{\theta}|\theta)(p_1^{*k}(\theta|\theta') - p_1^{*k}(\theta|\theta''))d\theta \\
= & \left[P_1^*(\bar{\theta}|\theta)(P_1^{*k}(\theta|\theta') - P_1^{*k}(\theta|\theta''))\right]_{-\infty}^{\infty} - \int \frac{\partial P_1^*(\bar{\theta}|\theta)}{\partial\theta}(P_1^{*k}(\theta|\theta') - P^{*k}(\theta|\theta''))d\theta \\
= & -\int \frac{\partial P_1^*(\bar{\theta}|\theta)}{\partial\theta}(P_1^{*k}(\theta|\theta') - P_1^{*k}(\theta|\theta''))d\theta \\
\leq & -\int \frac{\partial P_2^*(\bar{\theta}|\theta)}{\partial\theta}(P_2^{*k}(\theta|\theta') - P_2^{*k}(\theta|\theta''))d\theta \\
= & \left[P_2^*(\bar{\theta}|\theta)(P_2^{*k}(\theta|\theta') - P_2^{*k}(\theta|\theta''))\right]_{-\infty}^{\infty} - \int \frac{\partial P_2^*(\bar{\theta}|\theta)}{\partial\theta}(P_2^{*k}(\theta|\theta') - P_2^{*k}(\theta|\theta''))d\theta \\
= & \int P_2^*(\bar{\theta}|\theta)(p_2^{*k}(\theta|\theta') - p_2^{*k}(\theta|\theta''))d\theta \\
= & P_2^{*k+1}(\bar{\theta}|\theta') - P_2^{*k+1}(\bar{\theta}|\theta''),
\end{aligned}
$$

where p_i^{*k} denotes the conditional density of P^{*k} and the inequality holds by inductive hypothesis.

We next show that for every t and $\theta' > \theta''$, $\mathbb{E}_{P_1^*}[\theta_t|\theta_0 = \theta'] - \mathbb{E}_{P_1^*}[\theta_t|\theta_0 = \theta''] \geq \mathbb{E}_{P_2^*}[\theta_t|\theta_0 = \theta'] - \mathbb{E}_{P_2^*}[\theta_t|\theta_0 = \theta'']$. Indeed, integration by parts yields

$$
\begin{aligned}
& \mathbb{E}_{P_1^*}[\theta_t|\theta_0 = \theta'] - \mathbb{E}_{P_1^*}[\theta_t|\theta_0 = \theta''] = \int \theta(p_1^{*t}(\theta|\theta') - p_1^{*t}(\theta|\theta''))d\theta \\
= & \left[\theta(P_1^{*t}(\theta|\theta') - P_1^{*t}(\theta|\theta''))\right]_{-\infty}^{\infty} - \int (P_1^{*t}(\theta|\theta') - P_1^{*t}(\theta|\theta''))d\theta \\
= & -\int (P_1^{*t}(\theta|\theta') - P_1^{*t}(\theta|\theta''))d\theta \geq -\int (P^{*t}(\theta|\theta') - P^{*t}(\theta|\theta''))d\theta \\
= & \left[\theta(P_2^{*t}(\theta|\theta') - P_2^{*t}(\theta|\theta''))\right]_{-\infty}^{\infty} - \int (P^{*t}(\theta|\theta') - P^{*t}(\theta|\theta''))d\theta \\
= & \int \theta(p_2^{*t}(\theta|\theta') - p_2^{*t}(\theta|\theta''))d\theta = \mathbb{E}_{P_2^*}[\theta_t|\theta_0 = \theta'] - \mathbb{E}_{P_2^*}[\theta_t|\theta_0 = \theta''],
\end{aligned}
$$

where the inequality uses the observation in the previous paragraph. This implies that $s_1^*(\theta') - s_1^*(\theta') \geq s_2^*(\theta') - s_2^*(\theta')$ for all $\theta' > \theta''$, as claimed. □

D.2.4 Proof of Proposition D.3

We begin with two lemmas. First, analogously to the relationship between $\succsim_{MA}$ and $\succsim_m$ (Lemma B.1), we show that $\succsim_{SMA}$ is the "dual order" of $\succsim_d$:

Lemma D.2. *Fix any $C_1, C_2 \in \mathcal{C}$. Then $C_1 \succsim_{SMA} C_2$ if and only if $T_{C_1}f \succsim_d T_{C_2}f$ for all $f \in \mathcal{I}$.*

Proof. For the "only 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 $x \in (0,1)$. Thus, 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. Hence, $T_{C_1}f \succsim_d T_{C_2}f$.

Next take an arbitrary $f \in \mathcal{I}$. Define the sequence of bounded functions (f_n) as in (20), 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 $T_{C_1}f \succsim_d T_{C_2}f$.

For the "if" part, we prove the contrapositive. Suppose 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

$$\begin{aligned} T_{C_1}f(x) - T_{C_1}f(x') = -\int f'(y)(C_1(y|x) - C_1(y|x'))dy \\ < -\int f'(y)(C_2(y|x) - C_2(y|x'))dy = T_{C_2}f(x) - T_{C_2}f(x'). \end{aligned}$$

Thus, $T_{C_1}f \succsim_d T_{C_2}f$ fails. □

Second, we show that whenever an action distribution can be rationalized as Nash under some $\hat{C}_1$, then it can also be rationalized under any strongly less assortative $\hat{C}_2$:

Lemma D.3. *Fix any (γ, β) and any $\hat{C}_1, \hat{C}_2 \in \mathcal{C}$ with $\hat{C}_1 \succsim_{SMA} \hat{C}_2$. For any $\hat{F}_1 \in \mathcal{F}$, there exists $\hat{F}_2 \in \mathcal{F}$ such that the Nash action distributions under $(\hat{F}_1, \hat{C}_1)$ and $(\hat{F}_2, \hat{C}_2)$ coincide.*

Proof. Let h denote the inverse cdf of the Nash action distribution at $(\hat{F}_1, \hat{C}_1, \gamma, \beta)$. Then for any $x > x'$ in $(0,1)$,

$$\begin{aligned} h(x) - \gamma T_{\hat{C}_2}h(x) = (1-\beta-\gamma)\hat{F}_1^{-1}(x) + \beta \textstyle\int h(y)dy + \gamma\left(T_{\hat{C}_1}h(x) - T_{\hat{C}_2}h(x)\right) \\ > (1-\beta-\gamma)\hat{F}_1^{-1}(x') + \beta \textstyle\int h(y)dy + \gamma\left(T_{\hat{C}_1}h(x') - T_{\hat{C}_2}h(x')\right) = h(x') - \gamma T_{\hat{C}_2}h(x') \end{aligned}$$

because $\hat{F}_1^{-1}$ is strictly increasing and $T_{\hat{C}_1}h \succsim_d T_{\hat{C}_2}h$ by Lemma D.2 (note that $h \in \mathcal{I}$). Thus, we can define a strictly increasing function $\hat{F}_2$ by $\hat{F}_2^{-1}(x) = \frac{h(x)-\gamma T_{\hat{C}_2}h(x)-\beta\int h(y)dy}{1-\gamma-\beta}$ for each x. Since $T_{\hat{C}_2}h$ is L^1 and absolutely continuous, so is $\hat{F}_2^{-1}$. Hence, $\hat{F}_2 \in \mathcal{F}$. Then h^{-1} is the Nash action distribution at $(\hat{F}_2, \hat{C}_2, \gamma, \beta)$ because $h(x) = (1-\gamma-\beta)\hat{F}_2^{-1}(x) + \gamma\int h(y)d\hat{C}_2(y|x) + \beta\int h(y)dy$ for each x. □

Proof of Proposition D.3. Fix any (P, γ, β) and θ. Suppose there exists a coherent equilibrium $(s, (\hat{s}_{\theta'}), (\hat{F}_{\theta'}), (\hat{C}_{\theta'}))$ with $\hat{C}_\theta = \hat{C}_1$ and $\hat{F}_\theta(\theta) = \hat{x}$, and take any $\hat{C}_2$ with $\hat{C}_1 \succsim_{SMA} \hat{C}_2$ and $\hat{C}_1(\cdot|\hat{x}) = \hat{C}_2(\cdot|\hat{x})$. We will construct a coherent equilibrium $(s', (\hat{s}'_{\theta'}), (\hat{F}_{\theta'}), (\hat{C}_{\theta'}))$ in which $\hat{C}'_\theta = \hat{C}_2$ and $\hat{F}'_\theta(\theta) = \hat{x}$. To do so, first set $s' = s$ and for all $\theta' \neq \theta$, set $\hat{s}'_{\theta'} = \hat{s}_{\theta'}$, $\hat{F}'_{\theta'} = \hat{F}_{\theta'}$, and $\hat{C}'_{\theta'} = \hat{C}_{\theta'}$. Since any type $\theta' \neq \theta$ has the same perceptions as in the original equilibrium and the true strategy profile $s' = s$ is also the same as the original equilibrium, the best-response, observational consistency, and coherency conditions continue to be satisfied for $\theta' \neq \theta$.

It remains to find perceptions $\hat{s}'_\theta$ and $\hat{F}'_\theta$ with $\hat{F}'_\theta(\theta) = \hat{x}$ such that θ's best-response, observational consistency, and coherency conditions are met under $(\hat{F}'_\theta, \hat{C}_2, \hat{s}'_\theta)$. Let $\hat{H}$ denote the Nash action distribution under $(\hat{F}_\theta, \hat{C}_1)$. Since $\hat{C}_1 \succsim_{SMA} \hat{C}_2$, Lemma D.3 yields $\hat{F}'_\theta \in \mathcal{F}$ such that $\hat{H}$ is the Nash action distribution under $(\hat{F}'_\theta, \hat{C}_2)$. As $\hat{H}$ is Nash at both $(\hat{F}_\theta, \hat{C}_1)$ and $(\hat{F}'_\theta, \hat{C}_2)$, we have

$$\hat{H}^{-1}(\hat{x}) = (1-\gamma-\beta)\hat{F}_\theta^{-1}(\hat{x}) + \gamma\int \hat{H}^{-1}(y)d\hat{C}_1(y|\hat{x}) = (1-\gamma-\beta)(\hat{F}')_\theta^{-1}(\hat{x}) + \gamma\int \hat{H}^{-1}(y)d\hat{C}_2(y|\hat{x}).$$

Since $\hat{C}_1(\cdot|\hat{x}) = \hat{C}_2(\cdot|\hat{x})$, this implies that $\theta = \hat{F}_\theta^{-1}(\hat{x}) = (\hat{F}')_\theta^{-1}(\hat{x})$, so that $\hat{F}'_\theta(\theta) = \hat{x}$.

Setting $\hat{s}'_\theta(\theta') = \hat{H}^{-1}(\hat{F}'_\theta(\theta'))$ for each θ' ensures that $\hat{s}'_\theta$ is the Nash strategy profile at $(\hat{F}'_\theta, \hat{C}_2)$, so that coherency is satisfied. Moreover, $\hat{s}'_\theta(\theta) = \hat{H}^{-1}(\hat{x}) = \hat{s}_\theta(\theta) = s(\theta)$, where the final equality holds by the coherency and best-response conditions in the original equilibrium. Since $s = s'$, this implies that $s'(\theta) = \hat{s}'_\theta(\theta)$, so that $s'(\theta)$ satisfies the best-response condition by coherency of the new perceptions $(\hat{s}'_\theta, \hat{F}'_\theta, \hat{C}_2)$. Finally, to verify observational consistency for θ, it suffices to show that for each action a, $H_\theta^{s',P}(a) = \hat{C}_2(\hat{H}(a)|\hat{x})$. Since $s = s'$ and $H_\theta^{s,P}(a) = \hat{C}_1(\hat{H}(a)|\hat{x})$ by observational consistency of the original equilibrium, this is immediate from the fact that $\hat{C}_1(\cdot|\hat{x}) = \hat{C}_2(\cdot|\hat{x})$. □

D.3 Relationship with RCE

We elaborate on the relationship between coherent equilibrium and RCE discussed in Section 5.1.4. In constrast with our equilibrium notion, existing formulations of RCE do not explicitly treat players' beliefs as equilibrium objects. Thus, to make the connection transparent, we first define an extended version of RCE in our enviroment, based on the reduced-form formulation in Esponda (2013) without a full construction of an epistemic model. The state space that encodes all uncertainty relevant to agents' decisions is given by

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

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 (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 4. 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 4, 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:[65]

Definition 5. 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.

Every coherent equilibrium is an RCE, but the converse is not necessarily true. While coherent equilibrium requires each agent θ to believe that other agents follow a Nash equilibrium given his perceptions $(\hat{s}_\theta,\hat{P}_\theta)$, RCE allows for more general perceptions. However, when $\beta = 0$, RCE and coherent equilibrium 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.

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

Lemma D.4. *Every coherent equilibrium* $(s,(\hat{s}_\theta,\hat{P}_\theta)_{\theta\in\Theta})$ *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 D.3.1. □

This lemma directly implies that Proposition 4 generalizes, i.e., that assortativity neglect can be sustained as part of an RCE in any environment. The converse direction, Proposition 7 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 D.4 because in the proof of Proposition 7 we can restrict attention to $\beta=0$.

Corollary D.2 (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 (7).*

As RCE is a special case of our general equilibrium notion (Definition 2), behavior in any RCE with assortativity neglect remains uniquely pinned down by (7). At the same time, in contrast with the fact that coherent ANE also uniquely pins down all agents' perceived type distributions and perceived strategy profiles (Proposition 4), under RCE there can be many perceptions that are consistent with assortativity neglect. Thus, coherent ANE can be viewed as providing a natural way of selecting among these multiple perceptions.

D.3.1 Proof of Lemma D.4

Suppose that $(s,(\hat{s}_\theta,\hat{P}_\theta)_\theta)$ is a coherent equilibrium 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

$$\nu\{\omega':H_\theta^{\omega'}=H_\theta^{\omega}\}=1 \quad\Longrightarrow\quad \operatorname*{argmax}_{a\in A}u_\theta(a,\omega)=\operatorname*{argmax}_{a\in A}u_\theta(a,\nu) \tag{27}$$

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, (27) 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 a coherent equilibrium at P. □

D.3.2 Proof of Corollary D.2

The first part is an immediate consequence of Proposition 4 and the fact that any coherent equilibrium is also an RCE (Lemma D.4). To verify the second part, fix any regular $\hat{C} \neq C_I$ and any θ. Note that in the proof of Proposition 7, we constructed (F, C, γ, β) with $\beta = 0$ under which $\hat{C}_\theta \neq \hat{C}$ holds in any coherent equilibrium. Then the desired conclusion follows from Lemma D.4, as any RCE is a coherent equilibrium when $\beta = 0$. $\square$

E Additional Results

E.1 Non-coherent equilibria with general perceptions

We show that, without the coherency requirement, θ can sustain any perceived society as part of some equilibrium. Indeed, fix a true environment (P, γ, β) and take any type θ and perceived society $\hat{P}_\theta$. We will construct an equilibrium in which θ's perceived society is $\hat{P}_\theta$ and where all agents other than θ follow the strategy profile s^{AN} and hold perceptions as under ANE (as constructed in the proof of Proposition 1). Let H_θ^{AN} denote θ's local action distribution. Then setting $\hat{s}_\theta(\theta') := (H_\theta^{AN})^{-1}(\hat{P}_\theta(\theta'|\theta))$ for each $\theta' \in \text{supp}\hat{F}_\theta$, θ's observational consistency condition is satisfied since $(H_\theta^{AN})^{-1}(x) = \hat{s}_\theta(\hat{P}_\theta^{-1}(x|\theta)) = (H_\theta^{\hat{s}_\theta, \hat{P}_\theta})^{-1}(x)$ for each $x \in (0, 1)$. Set θ's action $s(\theta)$ to be the best response to perceptions $(\hat{s}_\theta, \hat{P}_\theta)$. Since this construction does not affect other agents' perceptions and incentives, this forms an equilibrium.

E.2 Preference for redistribution

Elaborating on Remark 4, we illustrate implications of our findings for a society's demand for redistribution with a simple example. As in Alesina and Giuliano (2011), we use the following toy version of Meltzer and Richard's (1981) classic majority voting model: Assume that voters must decide on a linear income tax $t \in [0, 1]$ to implement lump sum transfers subject to per-person wastage of wt^2. Thus, under tax t, an agent with pre-tax income θ receives $\theta + t(\mu - \theta) - wt^2$, where $\mu := \mathbb{E}_F[\theta]$ is the average pre-tax income in society. Each voter cares only about her post-tax income.

Note that θ's transfer is decreasing in her position $\theta - \mu$ relative to the mean. Thus, if agents correctly perceive their position, then whenever θ prefers tax t_1 to $t_2 < t_1$ (resp. $t_2 > t_1$), so does any type $\theta' < \theta$ (resp. $\theta' > \theta$). As a result, the median voter theorem predicts that voters implement the tax preferred by the median type θ_m, i.e., $t_C^* = \frac{\mu - \theta_m}{2w}$ under an interior solution.[66] In Example 4, $t_C^* = \frac{120-45}{2w} = \frac{75}{2w}$.

Suppose instead that each type θ evaluates taxes according to her perceived mean $\hat{\mu}_\theta$ under coherent ANE.[67] As illustrated in Example 4, perceived transfers $t(\hat{\mu}_\theta - \theta)$ need not be decreasing in θ; thus, the median income agent need not be a decisive voter. However, if we order types by $\theta \succ \theta'$ if $\hat{\mu}_\theta - \theta < \hat{\mu}_{\theta'} - \theta'$, then by the same logic as above the implemented tax is the one preferred by the $\succ$-median type $\hat{\theta}_m$; i.e., $t_{AN}^* = \frac{\hat{\mu}_{\hat{\theta}_m} - \hat{\theta}_m}{2w}$ under an interior solution. Note that in contrast to the correct perception case, the implemented tax is affected by underlying

[66] More precisely, t_C^* wins the majority vote under any sequential voting procedure over binary menus.

[67] The interpretation is that prior to the vote, agents engage in assortative interactions (e.g., day-to-day consumption activities) and use the perceptions derived from these interactions to evaluate taxes.

coordination motives and assortativity. In Example 4, t^*_{AN} is increasing in $\gamma+\beta$ and decreasing in ρ, with $t^*_{AN}\to 0$ as $\gamma+\beta\to 0$ and $\rho\to 1$, while $t^*_{AN}\to t^*_C$ as $\gamma+\beta\to 1$.[68]

Finally, beyond this simple example, we note that under more general redistribution schemes or more general preferences over post-tax outcomes, demand for redistribution naturally depends on features of agents' perceived income distributions $\hat{F}_\theta$ other than their perceived position relative to the mean.[69] E.g., it can be shown that if agents' preferences over post-tax outcomes are determined by their (perceived) ANE payoffs in the underlying coordination game (under utility (2)), then for linear tax schemes θ's marginal utility to redistribution is proportional to her perceived type variance $\mathrm{Var}(\hat{F}_\theta)$.

E.3 Welfare implications of assortativity neglect

We focus on the utility function (2), which captures quadratic losses due to local and global miscoordination. In any environment (P,γ,β), type θ's Nash equilibrium utility is given by $U^{NE}_\theta := U_P(\theta, s^{NE}(\theta), s^{NE})$. To contrast this with θ's utility under ANE, we distinguish between her *objective* ANE utility $U^{AN}_\theta := U_P(\theta, s^{AN}(\theta), s^{AN})$, which evaluates θ's payoff according to the true society P and true ANE strategy profile s^{AN}, and her *subjective* utility $\hat{U}^{AN}_\theta := U_{\hat{P}_\theta}(\theta, s^{AN}(\theta), \hat{s}^{AN}_\theta)$, which evaluates θ's payoff according to her perceived society $\hat{P}_\theta$ and perceived strategy profile $\hat{s}_\theta$ under the unique coherent ANE.[70] We assume that these utilities are well-defined (i.e., finite).

Proposition E.1. *Fix any (P,γ,β) such that $C\neq C_I$ and $\beta>0$. Then:*

1. *$U^{NE}_\theta > U^{AN}_\theta$ for all θ.*
2. *If $\gamma=0$, then for any θ, $\hat{U}^{AN}_\theta > U^{NE}_\theta$ if and only if $\mathrm{Var}(\hat{F}_\theta) < \mathrm{Var}(F)$.*

The first part of the proposition shows that objectively, assortativity neglect always Pareto-decreases welfare relative to Nash. This reflects two channels of welfare loss: increased miscoordination costs due to greater (local and global) action dispersion under ANE (Proposition 1); and misoptimization arising from the fact that θ's payoff is evaluated according to the true (P, s^{AN}), but she best-responds to her perceptions.

By contrast, the second part shows that subjectively, assortativity neglect can increase welfare. To see why, note first that trivially, subjective ANE utility does not entail any misoptimization. More importantly, subjective miscoordination costs under ANE may be lower than under Nash: Indeed, while θ is correct about her local action distribution (which is more dispersed than under Nash), she may misperceive the global action distribution to be less dispersive

[68] Specifically, t^*_{AN} is determined as follows: If $\hat{\mu}_L>115$, then $\hat{\theta}_m=\theta_m=45$ and $t^*_{AN}=\frac{\hat{\mu}_L-45}{2w}$. If $\hat{\mu}_L\le 115$, then there are two $\succ$-median types, $\hat{\theta}^L_m=\hat{\mu}_L-x$ and $\hat{\theta}^H_m=\hat{\mu}_H-x$, and we have $t^*_{AN}=\frac{x}{2w}$, where x is given by $\frac{2}{3}\frac{\hat{\mu}_L-x}{60}+\frac{1}{3}\frac{\hat{\mu}_H-x-60}{480}=\frac{1}{2}$. Since we saw in the main text that $\hat{\mu}_L$ is increasing in $\gamma+\beta$ and decreasing in ρ, and $\frac{2}{3}\hat{\mu}_L+\frac{1}{3}\hat{\mu}_H=120$, it follows that x is increasing in $\gamma+\beta$ and decreasing in ρ. Moreover, as $\gamma+\beta\to 0$ and $\rho\to 1$, we have $\hat{\mu}_L\to 30$ and $\hat{\mu}_H\to 300$, so that $x\to 0$.

[69] See Alesina and Giuliano (2011) for a survey of more general models of demand for redistribution without misperceptions.

[70] The use of coherent ANE perceptions to define subjective utility is not important: θ's subjective utility is the same under *any* ANE perceptions $\hat{P}_\theta$ and $\hat{s}_\theta$, as it depends only on θ's perceived local and global action distributions, which in any ANE must equal her true local action distribution $H^{s^{AN},P}_\theta$. However, the characterization of subjective welfare in terms of $\mathrm{Var}(\hat{F}_\theta)$ in Proposition E.1 relies on coherency.

than Nash. To isolate the latter effect, the proposition focuses on the case $\gamma = 0$, so that local action dispersion is not payoff-relevant.[71] In this case, we show that perceived global action dispersion is proportional to perceived type heterogeneity, and as a result θ is subjectively better off under ANE precisely when she perceives society to be more homogenous than actual (as we saw can occur in Section 4.2).

E.3.1 Proof of Proposition E.1

We begin with two preliminary observations. First we show that local and global action variance is higher under ANE than Nash. 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 ANE 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 H^{AN} is more dispersive than H^{NE} (Proposition 1). Thus, H_θ^{AN} is more dispersive than H_θ^{NE}. This implies that $\mathrm{Var}(H^{AN}) \geq \mathrm{Var}(H^{NE})$ and $\mathrm{Var}(H_\theta^{AN}) \geq \mathrm{Var}(H_\theta^{NE})$ for each θ (e.g., Shaked and Shanthikumar, 2007, p. 155). Given that $C \neq C_I$ and $\beta > 0$, H^{AN} is strictly more dispersive than H^{NE}, which implies $\mathrm{Var}(H^{AN}) > \mathrm{Var}(H^{NE})$.

Second, we can express equilibrium utilities as follows:

$$\begin{aligned}U_\theta^{NE} &= -\gamma^2 \mathrm{Var}(H_\theta^{NE}) - \beta^2 \mathrm{Var}(H^{NE}), \\ \hat{U}_\theta^{AN} &= -(\gamma^2 + \beta^2)\mathrm{Var}(H_\theta^{AN}), \\ U_\theta^{AN} &= -\int_0^1 \int_0^1 \left(s^{AN}(\theta) - (1-\gamma-\beta)\theta - \gamma (H_\theta^{AN})^{-1}(x) - \beta (H_\theta^{AN})^{-1}(x)\right)^2 dx dx' \\ &\leq -\min_a \int_0^1 \int_0^1 \left(a - (1-\gamma-\beta)\theta - \gamma (H_\theta^{AN})^{-1}(x) - \beta (H_\theta^{AN})^{-1}(x)\right)^2 dx dx' \\ &= -\gamma^2 \mathrm{Var}(H_\theta^{AN}) - \beta^2 \mathrm{Var}(H^{AN}).\end{aligned}$$

First part: The inequality $U_\theta^{NE} > U_\theta^{AN}$ for each θ follows from the above expressions for utilities because $\mathrm{Var}(H^{AN}) > \mathrm{Var}(H^{NE})$, $\mathrm{Var}(H_\theta^{AN}) \geq \mathrm{Var}(H_\theta^{NE})$, and $\beta > 0$.
Second part: Under $\gamma = 0$, the Nash equilibrium utility satisfies $U_\theta^{NE} = -\beta^2(1-\beta)^2 \mathrm{Var}(F)$ because $s^{NE}(\theta') = (1-\beta)\theta' + \beta\hat{\mu}$ for each θ'. Under coherent ANE, the subjective utilitysatisfies $\hat{U}_\theta^{AN} = -\beta^2(1-\beta)^2 \mathrm{Var}(\hat{F}_\theta)$ because $\hat{s}_\theta(\theta') = (1-\beta)\theta' + \beta\hat{\mu}_\theta$ for each θ'. Thus, since $\beta > 0$, $U_\theta^{AN} > U_\theta^{NE}$ iff $\mathrm{Var}(\hat{F}_\theta) < \mathrm{Var}(F)$. □

[71] If $\gamma > 0$, then $\mathrm{Var}(\hat{F}_\theta) < \mathrm{Var}(F)$ continues to imply that θ's perceived global action distribution $\hat{H}_\theta$ is less dispersed than H^{NE}: Indeed, $\mathrm{Var}(\hat{H}_\theta) = (1-\gamma-\beta)^2 \mathrm{Var}(\hat{F}_\theta)$, as $\hat{s}_\theta(\theta') = (1-\gamma-\beta)\theta' + (\gamma+\beta)\hat{\mu}_\theta$ for all θ' by coherency, whereas $\mathrm{Var}(H^{NE}) \geq (1-\gamma-\beta)^2 \mathrm{Var}(F)$ by (5). However, in terms of welfare, the fact that *local* action dispersion is higher than Nash can outweigh this effect when γ is large.

E.4 Coherent equilibria in Gaussian societies (Example 6)

Fix any Gaussian society $P = (\mu, \sigma^2, \rho)$ and β, γ with $\beta > 0$. For each perceived correlation coefficient $\hat{\rho} \in [0, 1)$, we construct a coherent equilibrium in which for all θ,

1. θ's action is $s^*(\theta) = \frac{1-\beta-\gamma}{1-\gamma\rho-\beta\frac{\rho-\hat{\rho}}{1-\hat{\rho}}}\theta + \frac{\gamma(1-\rho)+\beta\frac{1+\hat{\rho}}{1-\rho}}{1-\gamma\rho-\beta\frac{\rho-\hat{\rho}}{1-\hat{\rho}}}\mu$.

2. θ's perceived society is Gaussian with $\hat{P}_\theta = (\hat{\mu}_\theta, \hat{\sigma}^2, \hat{\rho})$ where

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

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

3. θ's perceived strategy profile satisfies $\hat{s}_\theta(\theta') = \frac{1-\beta-\gamma}{1-\gamma\hat{\rho}}\theta + \frac{\beta+\gamma(1-\hat{\rho})}{1-\gamma\hat{\rho}}\hat{\mu}_\theta$ for all θ'.[72]

Moreover, we show that the equilibrium perceptions $(\hat{\mu}_\theta, \hat{\sigma}^2, \hat{\rho})$ exhibit PAN (Appendix D.2.1) if and only if $\hat{\rho} \leq \rho$.

Fix any $\hat{\rho} \in [0, 1)$. To verify that the above is a coherent equilibrium, let $x := \frac{1-\beta-\gamma}{1-\gamma\rho-\beta\frac{\rho-\hat{\rho}}{1-\hat{\rho}}}$ and $\hat{x} := \frac{1-\beta-\gamma}{1-\gamma\hat{\rho}}$, so that $s^*(\theta) = x\theta + (1-x)\mu$ and $\hat{s}_\theta(\theta') = \hat{x}\theta' + (1-\hat{x})\hat{\mu}_\theta$ for each θ, θ'. Since $P(\cdot|\theta)$ is distributed $\mathcal{N}(\rho\theta + (1-\rho)\mu, (1-\rho^2)\sigma^2)$, θ's actual local action distribution $H_\theta^{s^*,P}$ is distributed $\mathcal{N}(x\rho\theta + (1-x\rho)\mu, x^2(1-\rho^2)\sigma^2)$. Since $\hat{P}_\theta(\cdot|\theta)$ is distributed $\mathcal{N}(\hat{\rho}\theta + (1-\hat{\rho})\hat{\mu}_\theta, (1-\hat{\rho}^2)\hat{\sigma}^2)$, θ's perceived local action distribution $H_\theta^{\hat{s}_\theta,\hat{P}_\theta}$ is distributed $\mathcal{N}(\hat{x}\hat{\rho}\theta + (1-\hat{x}\hat{\rho})\hat{\mu}, \hat{x}^2(1-\hat{\rho}^2)\hat{\sigma}^2)$. Thus, observational consistency can be verified by noticing that the above construction ensures that the mean and variance of $H_\theta^{s^*,P}$ and $H_\theta^{\hat{s}_\theta,\hat{P}_\theta}$. To verify coherency, note that by construction θ's perceived strategy $\hat{s}_\theta$ is the Nash equilibrium strategy at $(\hat{\mu}_\theta, \hat{\sigma}^2, \hat{\rho})$ (see Example 2). Finally, to verify the best-response requirement, note that $s(\theta) = \hat{s}_\theta(\theta)$.

Next, we show that PAN holds if and only if $\hat{\rho} \leq \rho$. If $\hat{\rho} > \rho$, then $\text{Var}[s^*] = \left(\frac{1-\beta-\gamma}{1-\gamma\rho-\beta\frac{\rho-\hat{\rho}}{1-\hat{\rho}}}\right)^2\sigma^2 >$ $\text{Var}[s^{NE}] = \left(\frac{1-\beta-\gamma}{1-\gamma\rho}\right)^2\sigma^2$, which by Proposition D.2 implies that PAN is not satisfied. If $\hat{\rho} \leq \rho$, consider any θ and θ'. We want to show that $x' := \hat{C}^{-1}(P(\theta'|\theta)|\hat{F}_\theta(\theta))$ is decreasing in θ, where $C = C_\rho$ and $\hat{C} = C_{\hat{\rho}}$.[73] Since $P(\theta'|\theta) = \hat{C}(x'|\hat{F}_\theta(\theta))$, it suffices by the implicit function theorem to show that

$$\frac{\partial}{\partial\theta}P(\theta'|\theta) - \frac{\partial}{\partial\theta}\hat{C}(x'|\hat{F}_\theta(\theta)) \leq 0.$$

Since $P(\theta'|\theta) = \Phi(\frac{\theta'-\rho\theta-(1-\rho)\mu}{\sqrt{1-\rho^2}\sigma})$, the first term satisfies $\frac{\partial}{\partial\theta}P(\theta'|\theta) = -\phi(\frac{\theta'-\rho\theta-(1-\rho)\mu}{\sqrt{1-\rho^2}\sigma})\frac{\rho}{\sqrt{1-\rho^2}\sigma}$, where Φ and ϕ denote cdf and density of the standard normal distribution. To evaluate the second term, observe that we can write $\hat{C}(x'|\hat{F}_\theta(\theta)) = \hat{P}_\theta(\hat{\theta}'|\theta) = \Phi(\frac{\hat{\theta}'-\hat{\rho}\theta-(1-\hat{\rho})\hat{\mu}_\theta}{\sqrt{1-\hat{\rho}^2}\hat{\sigma}})$ for $\hat{\theta}' :=$

---

[72] More strongly, it can be shown that any coherent equilibrium in which true strategies are linear and perceptions $\hat{P}_\theta = (\hat{\mu}_\theta, \hat{\sigma}_\theta^2, \hat{\rho}_\theta)$ are Gaussian must satisfy 1–3 for some $\hat{\rho} \in [0, 1)$.

[73] Note that this is equivalent to PAN since $P(\theta'|\theta) = C(F(\theta')|F(\theta))$ for all θ, θ'.

$\hat{F}_\theta^{-1}(x')$. Also note that $\frac{\hat{\theta}'-\hat{\rho}\theta-(1-\hat{\rho})\hat{\mu}_\theta}{\sqrt{1-\hat{\rho}^2}\hat{\sigma}} = \frac{\theta'-\rho\theta-(1-\rho)\mu}{\sqrt{1-\rho^2}\sigma}$, as $P(\theta'|\theta) = \hat{C}(x'|\hat{F}_\theta(\theta))$. This yields $\frac{\partial}{\partial\theta}\hat{C}(x'|\hat{F}_\theta(\theta)) = -\phi(\frac{\hat{\theta}'-\hat{\rho}\theta-(1-\hat{\rho})\hat{\mu}_\theta}{\sqrt{1-\hat{\rho}^2}\hat{\sigma}})\frac{\hat{\rho}+(1-\hat{\rho})\hat{\mu}'}{\sqrt{1-\hat{\rho}^2}\hat{\sigma}}$, where $\hat{\mu}' := \frac{\partial\hat{\mu}_\theta}{\partial\theta}$ Thus, the proof is complete if

$$\frac{\rho}{\sqrt{1-\rho^2}\sigma} \geq \frac{\hat{\rho}+(1-\hat{\rho})\hat{\mu}'}{\sqrt{1-\hat{\rho}^2}\hat{\sigma}}. \tag{28}$$

To establish (28), recall that by observational consistency the means of $H_\theta^{s^*,P}$ and $H_\theta^{\hat{s}_\theta,\hat{P}_\theta}$ match, i.e., $x\rho\theta + (1-x\rho)\mu = \hat{x}\hat{\rho}\theta + (1-\hat{x}\hat{\rho})\hat{\mu}_\theta$. Differentiating with respect to θ yields $x\rho = \hat{x}\hat{\rho} + (1-\hat{x}\hat{\rho})\hat{\mu}'$. Thus,

$$\frac{\rho}{\hat{\rho}+(1-\hat{\rho})\hat{\mu}'} - \frac{\sqrt{1-\rho^2}\sigma}{\sqrt{1-\hat{\rho}^2}\hat{\sigma}} = \frac{\rho}{\hat{\rho}+(1-\hat{\rho})\hat{\mu}'} - \frac{\hat{x}}{x} = \frac{\hat{x}}{x}\left(\frac{x\rho}{\hat{x}\hat{\rho}+\hat{x}(1-\hat{\rho})\hat{\mu}'} - 1\right) \geq 0,$$

where the first equality holds by construction of $\hat{\sigma}$. This implies (28), as required.