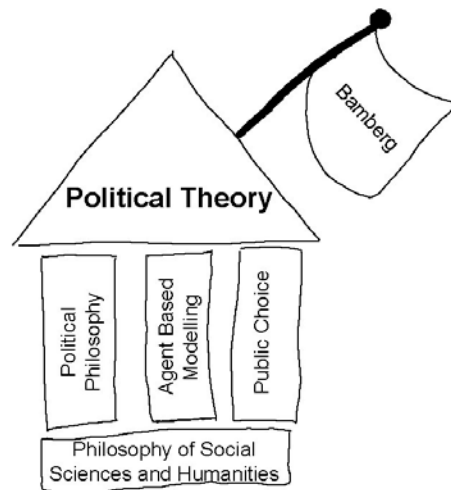




A Network-Based Explanation of Perceived Inequality

Jan Schulz
Daniel Mayerhoffer
Anna Gebhard

20 January 2021



**Bamberg Contributions
to
Contemporary Political Theory
No. 2**

Political Theory
Institute for Political Science
University of Bamberg

<https://www.uni-bamberg.de/en/poltheorie/>

A Network-Based Explanation of Perceived Inequality

Jan Schulz¹, Daniel M. Mayerhoffer², and Anna Gebhard³

¹Institute for Economics, University of Bamberg:

jan.schulz@uni-bamberg.de

²Institute for Political Science, University of Bamberg, Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) -

430621735

³MathOpt Research Group, University of Magdeburg

Abstract

Across income groups and countries, the public perception of economic inequality and many other macroeconomic variables such as inflation or unemployment rates is spectacularly wrong. These misperceptions have far-reaching consequences, as it is *perceived* inequality, not *actual* inequality informing redistributive preferences. The prevalence of this phenomenon is independent of social class and welfare regime, which suggests the existence of a common mechanism behind public perceptions. We propose a network-based explanation of perceived inequality building on recent advances in random geometric graph theory. The literature has identified several stylised facts on how individual perceptions respond to actual inequality and how these biases vary systematically along the income distribution. Our generating mechanism can replicate all of them simultaneously. It also produces social networks that exhibit salient features of real-world networks; namely, they cannot be statistically distinguished from small-world networks, testifying to the robustness of our approach. Our results, therefore, suggest that homophilic segregation is a promising candidate to explain inequality perceptions with strong implications for theories of consumption behaviour.

Keywords: Homophily, network, inequality, perception, random geometric graph.

1 Introduction

Conventional modern macroeconomics has long recognised the crucial relevance of expectations and belief-formation for aggregate dynamics (Galí 2015). In particular, beliefs about economic inequality and perceptions of social hierarchy can inform individuals in such diverse fields as consumption decisions (Duesenberry 1949; Veblen 2001; Frank et al. 2014), redistributive preferences and voting behaviour (Gimpelson & Treisman 2018; Kim et al. 2018; Choi 2019) or subjective well-being and ethical convictions (Kuhn 2019; Clark & Senik 2010). Even in the most sophisticated behavioural models, belief-formation is, however, typically either assumed to be atomistic (Gabaix 2020) or does not systematically account for the impact of individual embeddedness within heterogeneous social contexts on those beliefs, even if social interaction is explicitly modelled (Flieth & Foster 2002; Lux 2009). We propose a parsimonious network-based model for the interaction of macro-level inequality, micro-level beliefs and the mediating effects of heterogeneous social contexts. In contrast to the assumption of deductive reasoning in orthodox models, we build on the empirically well-established notion that economic agents reason *inductively* and generalise from finite samples. Recent theoretical and empirical work has demonstrated that potency of this approach in explaining phenomena in such diverse fields as human probability assessment (Sanborn & Chater 2016; Chater et al. 2020) or regional inequality (Collier & Tuckett 2020). Thus, the model is both consistent with several stylised facts about inequality perceptions and the micro-level evidence on the composition of social networks.

The relevance of individual beliefs is perhaps best exemplified by spelling out its political economy implications. Across income groups and countries, the public perception of economic inequality and many other macroeconomic variables is empirically wrong, often spectacularly so. Errors in those beliefs might be due to conceptually different problems: *uninformed* beliefs or *misinformed* beliefs (Kuklinski et al. 2000). Uninformed voters are ignorant about the actual state of affairs, while misinformed voters' beliefs are consistently deviating from it in one direction. The distinction is a crucial one. Uninformed voters' beliefs would cluster around the actual state of affairs and, with no systematic deviations, be correct in expectations. For uninformed voters, we only need one informed voter to tip elections under majority rule into the correct direction; a majority of ignorant individuals might nevertheless vote for the correct policy, known as the 'miracle of aggregation' (Page & Shapiro 1993). However, this miraculous aggregation breaks down when we consider misinformed rather than uninformed voters with beliefs that are no longer randomly distributed but consistently tend in a (false) direction (Caplan 2011). The type of error in perceptions is thus intimately linked to the efficacy of democratic systems.

For inequality perceptions, beliefs appear to be indeed the result of misinformation in this technical sense and they are consistently biased across income groups and welfare regimes. The literature has identified four particular stylised facts for any theory of perceived inequality to be evaluated against: *(i)* Irrespective of their objective status, all individuals perceive themselves to be in the middle of the social hierarchy (Kelley & Evans 1995; Evans & Kelley 2004); *(ii)* as an immediate corollary of *(i)*, poor individuals overestimate their social position, rich individuals tend to underestimate it (Knell & Stix 2020); *(iii)* poor individuals tend to perceive inequality to be higher and are closer to objective inequality on average (Osberg & Smeeding 2006; Newman et al. 2018) and *(iv)* the evolution of objective inequality is detached from the evolution of subjective inequality, that is, increases in objective inequality do not necessarily increase perceived inequality (Kenworthy & McCall 2008; Bartels 2018; Gimpelson & Treisman 2018; Hvidberg et al. 2020). The ubiquity of misperceptions across states and welfare regimes that the literature identifies calls for a common mechanism independent of differences in actual inequality or institutional framework.

In contrast to much of the behavioural literature, we refrain from ad-hoc assumptions about possible biases, e.g., assuming that individuals tend to perceive themselves in the middle of social hierarchies (cf., e.g. Knell & Stix 2020). Instead, we assume unbiased information processing capabilities for all the economic agents. Information is, however, asymmetric and agents form estimates about aggregate variables according to their local information. Notice the similarity to the canonical monetary misperceptions model here, where individuals also form (rational) expectations about aggregate inflation from local information about price movements (Lucas 1972, 1973, 1975). We show that a parsimonious process can generate sufficiently skewed information sets to replicate the aforementioned stylised facts and generate perceived inequality levels that are quantitatively in line with recent empirical evidence for a large sample of 32 OECD countries (Choi 2019). In essence, we assume that agents (correctly) observe inequality within their local social network and (correctly) form estimates about the total population from them but still generate biased perceptions due to their network contacts not being representative for the overall population. Employing a new variant of a random geometric graph network, the assumption of income homophily alone can generate substantial misperception in line with the empirical evidence. The derived network topology also corresponds to empirically observed social networks across the world and features a small-world structure. Given the ubiquity of these topological features, our homophilic process appears to be a plausible candidate to explain the equally ubiquitous inequality misperceptions.

The remainder of this paper is organised as follows: The second section situates our model within the pertinent literature and reviews the evidence on empirical network topologies

and individual belief formation. Section 3 introduces the basic model of homophilic graph formation and reviews the main mechanisms generating heterogeneity in information sets. Section 4 presents our analytic and simulative results, shows that they are consistent with the outlined stylised facts regarding network topologies as well as inequality perceptions and derives some important implications regarding heterogeneous segregation patterns across the income distribution. Section 5 concludes and discusses several promising avenues for further research, especially regarding consumption dynamics and voting behaviour.

2 Related Literature

Our model brings three different strands of literature together. Its two major sets of explananda are the empirical social network structure, mainly their homophily and their small-world character as well as the main empirical findings on belief-formation. We review those findings below and introduce the family of random geometric graphs as the third strand of literature and as a promising candidate to explain both sets of phenomena simultaneously in the model section.

2.1 Empirical Social Networks

Empirical networks exhibit ubiquitous and salient features that can serve as stylised facts to guide the validation of proposed theoretical graph formation processes. Probably the most prominent one is the small-world property, indicating that paths between nodes in real-world social networks are unexpectedly short. At the same time, those networks also feature high degrees of clustering. Small-worldiness has obvious implications for any contagion process, be it rumours, diseases or information and where contagion across the whole network happens much faster than our intuition would suggest (Watts 1999; Moore & Newman 2000; Kleinberg 2001). The empirical research has identified small-world features across many different social groups, including friendship networks in schools (Weeden & Cornwell 2020), corporate board networks (Kogut & Walker 2001; Borgatti & Foster 2003; Davis et al. 2003; Conyon & Muldoon 2006; Galaskiewicz 2007) and scientific and artistic collaboration (Watts & Strogatz 1998; Newman 2001; Uzzi & Spiro 2005). Given this ubiquity, it appears safe to say that a graph-generating process for social networks needs to simultaneously produce low average path lengths and large degrees of clustering to be consistent with this stylised fact.

We situate our model in the random graph literature, where graph formation happens according to a stochastic process and is not the result of deliberate optimisation. Random graphs have been very successful in replicating structural stylised facts about network

topologies, with the Watts-Strogatz model famously able to replicate those small-world properties (Watts & Strogatz 1998). Since the graph-generating process is, however, stochastic in nature, it does not feature explicit behavioural microfoundations. Therefore, we extend the purely stochastic notion with a behavioural ingredient, notably, that link-formation is *homophilic*.

First introduced by Lazarsfeld et al. (1954), another salient feature of empirical social networks is *homophily*, the tendency of similar individuals to connect with each other. This tendency is not only an empirical curiozum but has relevant theoretical implications, e.g., for information transmission, where homophilic segregation can severely slow the speed of learning (Golub & Jackson 2012). The similarity can come in many dimensions such as gender, ethnicity or socio-economic status (McPherson et al. 2001). We focus on the latter in the narrow sense of the empirically well-established *economic* homophily. One strand of literature focuses on friendship networks in schools and colleges and provides evidence significant homophily according to income or social class (Cohen 1979; Huckfeldt 1983; Mayer & Puller 2008; Boucher 2015; Malacarne 2017). Even one of the earliest contributions in the field, however, shows that homophily in socio-economic status is not fixed in time and varies with cultural norms and the importance of class distinctions (Cohen 1979). In light of this result, it appears unsurprising that we find considerable variation in implied degrees of country-level homophily in our model, perhaps reflecting cultural norms not in the structure but the degree of the graph formation process. Homophily in income also exists in social media friendship networks (Lewis et al. 2012, for a large sample of Facebook friends), where spatial segregation should not confound findings. Economic homophily becomes apparent in the choice of romantic partners, typically under the label of ‘homogamy’ (Kalmijn 1991; Kalmijn & Flap 2001). Finally, a very recent contribution by Cepić & Tonković (2020) for a representative sample of Croatian adults finds evidence for homophilic tie formation according to social class and income, with however considerable variability in cross-class ties, hinting at possible confounding factors we aim to capture with a parsimonious stochastic process. Importantly, Cepić & Tonković (2020) show that there is considerable variability in cross-class ties, though, which we show might be crucial for unbiased individual inference.

2.2 Belief Formation in Networks

The literature on belief formation itself appears to be much more scarce than the literature on the effects of perceptions and misperceptions. While very different in detail, the two currently dominant theories of public opinion formation suggest that the beliefs an individual holds are in a broad sense averages over the idiosyncratic messages she receives (Zaller 1992;

Lodge et al. 1995).¹ This literature has focused on the specific ‘averaging’ individuals use to process their information sets. We develop on the notion that information is local but assume unbiased processing with skewed information sets as implied by the well-documented homophilic social network formation on which we expand in section 3. The psychological literature on ‘social comparison theory’ (Festinger 1954) supports the notion that individual self-perceptions are much more responsive to local knowledge about small groups than to information about aggregates, e.g., knowing the population average (Buckingham & Alicke 2002; Zell & Alicke 2009; Alicke et al. 2010). Thus, belief formation about inequality appears to be indeed primarily based on local knowledge. However, informational treatments in the form of reported averages may change individual change beliefs. Providing information about the actual degree of inequality seems to exhibit a significant effect on redistributive preferences for Argentina, Sweden and the US (Cruces et al. 2013; McCall et al. 2017; Karadja et al. 2017), with however small and insignificant effects for Germany (Engelhardt & Wagener 2018). Finally, two recent studies for the whole of Europe and Denmark demonstrate that individuals indeed tend to know the income levels of their immediate friends and family rather well, with non-negligible effects on inequality and fairness perceptions as well as perceived social positions (Clark & Senik 2010; Hvidberg et al. 2020).

In a series of articles close in spirit to our approach, Chiang (2011, 2015a,b) exploits this notion and shows experimentally and computationally that individuals base their beliefs about inequality on local perceptions within referent networks and that income homophily has a potentially strong effect on those perceptions. While his approach is exploratory and does not account for the outlined stylised facts on inequality perceptions and empirical social networks, we provide a tractable model, readily calibrated with regards to those phenomena. We will discuss this model in the upcoming section.

3 Model

This section provides a content-oriented presentation; we provide a technical description following the ODD protocol upon request. The model consists of three distinct phases run in sequential order:

1. Agent initialisation and income allocation
2. Network formation
3. Gini perception and network evaluation

¹Cf. Stevenson & Duch (2013) for a summary on those views.

Each phase runs only once and phases one and two build the structure for subsequent analysis in phase three. This sequence implies that during network generation, agents adapt to others' income level. However, they do not react to others' linking behaviour or perception and, thus, the model does not feature interaction in a narrow sense. Moreover, in the model, an agent's social contacts depend on their income. We choose this direction of causality for technical reasons and because it seems empirically likely (cf. section 2). Nevertheless, our process scheduling would also be consistent with the opposite direction of causality or positive feedback effects between income and social contacts.

The model is designed that way because it focuses entirely on income perceptions given defined income distributions and network structures. Hence, both an agent's income and their social contacts remain constant for the evaluated time frame or, put differently, that the simulation outcome is a snapshot of a certain point in time.

3.1 Agent Initialisation and Income Allocation

There are 1000 agents in the model; each agent draws their income from an exponential distribution with a mean of $\lambda = 1$. Such a distribution normalises the empirical observed (pre-tax or market) income distributions in various industrialised countries for the vast majority of individuals (Drăgulescu & Yakovenko 2001; Silva & Yakovenko 2004; Tao et al. 2019). Thus, one can understand the model population as constituting a representative sample of empirical populations of these countries. The upper tail of 1 to 5 % of the income distributions empirically follows a Pareto law (Silva & Yakovenko 2004). We deliberately choose to exclude this small minority from our model, since their population size would induce another degree of freedom in our model and we want to demonstrate that segregation is indeed endogenous and not driven by differences in actual income regime. We use an identical, pre-validated exponential distribution for all Monte Carlo runs and also all levels of homophily to ensure comparability between simulation runs. Agents store their true income decile for evaluation purposes, too.

3.2 Network Formation

Each agent draws five other agents to link to. Like for real-world networks, links are therefore created *by agents*, not imposed on them. The number of five link choices is also empirically validated, as humans tend to only know the income of close friends or family (Clark & Senik 2010; Hvidberg et al. 2020), with typically only five individuals at this closest layer

of emotional connection (Zhou et al. 2005; Hamilton et al. 2007; Mac Carron et al. 2016).² The relative weight in the draws are a function of the homophily strength and the respective income levels. Thereby, agent j 's weight in agent i 's draw is denoted by w_{ij} and determined as follows:

$$w_{ij} = \frac{1}{\exp[\rho |I_j - I_i|]} \quad (1)$$

I denotes the income of an agent, and $\rho \in \mathbb{R}^+$ denotes the homophily strength in income selection, externally set, and identical for all agents. $\rho = 0$ represents a random graph, and for an increasing positive value of ρ , an agent becomes ever more likely to pick link-neighbours with incomes being closer to their own. The exponential character of the link function ensures that those others with are large income difference become unlikely picks even at low homophily strengths.

Figures 1 and 2 illustrate the linkage probabilities implied by the weighted draw based on the exponentially distributed income levels. As can be seen, the decay within the left tail is always more rapid than for the right tail, indicating differences in the ‘selectivity’ above or below a relative position. We understand ‘selectivity’ according to rank as the effect a decrease in income rank distance of one agent to another has on the linkage probability between them. Consequently, the local maxima of individual linkage probability densities exhibit a bi-modal shape with peaks at the highest and lowest rank but are also heavily skewed to the left, i.e., agents with the high incomes are most selective in their link picks. General selectivity increases with ρ . Notice also that largest income ranks are extremely selective in all scenarios, in some cases exceeding linkage probabilities of incomes close to the median by more than two orders of magnitude in linkage probabilities.

The resulting network is a member of the family of Random Geometric Graphs (Dall & Christensen 2002), which Talaga & Nowak (2020) showed to reproduce core features of many social networks efficiently. Specifically, we combine the notions of homophily (Boguná et al. 2004) with pre-setting node degrees (Newman et al. 2001; Newman 2009). However, concerning our application, we are able to simplify both approaches by pre-determination of only the global minimum degree, like in Preferential-Attachment networks, and consequently defining relative weights rather than absolute probabilities.

Links are undirected and have identical weights for evaluation purposes. Agents pick their neighbours in random sequential order. If an agent i picks agent j who had themselves

²For a recent review on the large literature on ‘Dunbar’s number’, cf. the first section of Mac Carron et al. (2016).

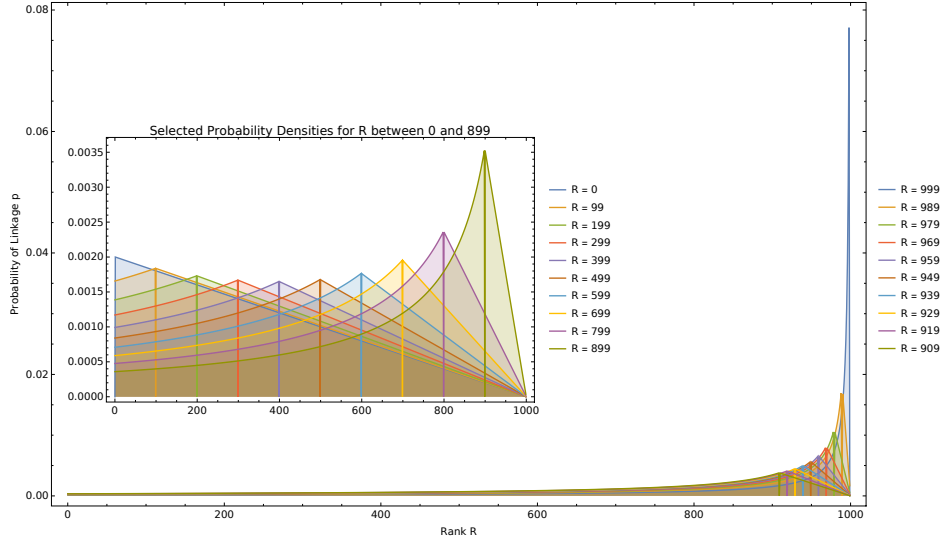


Figure 1: Theoretical PDF of Linkage Probabilities for Ranks R and $\rho = 1$.

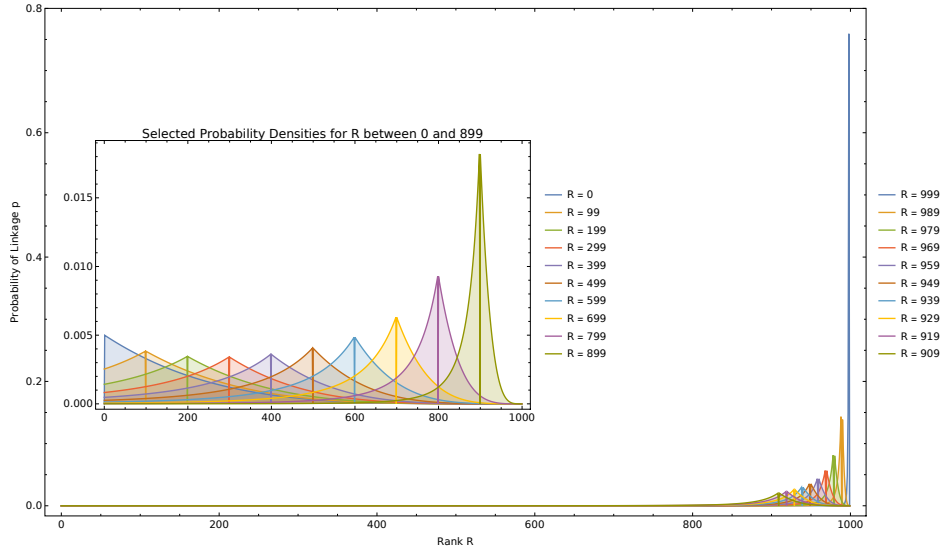


Figure 2: Theoretical PDF of Linkage Probabilities for Ranks R and $\rho = 4$.

Note: The Figures plot the Probability Density Functions (PDFs) of a node with a given income rank for linkage with another node for the whole support of income ranks. The above Figure assumes a homophily strength $\rho = 1$, whereas the below Figure assumes $\rho = 4$.

picked i before that, the already existing link between the two agents remains untouched, but i does not pick another neighbour instead of j . Consequently, each agent has at least 5 link-neighbours (i.e. social contacts) but may have more.

3.3 Gini Perception and Network Evaluation

Agents know about their own income and also their social contact’s incomes. However, they do not possess knowledge about any other agent or structural features of the whole income distribution. Thus, agents judge income inequality in the population as well as their own income position solely based on themselves and their link-neighbours. Besides the agents’ perceptions, there is a global assessment of various network parameters in order to validate the model.

Subjective inequality perceptions mirror standard Gini calculation on the level of individual personal networks: Each agent finds the mean of all income differences between themselves and each link neighbour and between any two of their link-neighbours and divides this by the mean overall income of themselves and all link-neighbours. Then, the overall perceived Gini is simply the mean of individual perceptions.

To estimate their income decile, an agent compares the number of link-neighbours having a higher income than the agent themselves to the link-neighbours having a lower income than the agent themselves.

4 Results

The homophilic graph model will be evaluated against the five stylised facts outlined earlier. As we have shown in section 3, we only require the homophily strength parameter $\rho \in \mathbb{R}_0^+$, the number of links each node chooses C and the income distribution as inputs for initialisation. Since link formation is stochastic, we run the graph formation routine 100 times and report model averages, if not otherwise indicated. Most of the results are obtained with initialisation by the same set of incomes generated from an exponential distribution with location parameter $\lambda = 1$ and 1,000 observations for $C = 5$ choices of link-neighbours each agent undertakes to make results comparable for variation in ρ . The overall Gini coefficient for these 1,000 randomly generated income levels is with $G \approx 0.50701$ within 1.5 % deviation from the theoretical Gini of $G = 0.5$, indicating that the observed effects of ρ are not artefacts of initialisation. Results are also robust for different numbers of links chosen per node, as long as $C \ll N$. We also evaluated the null model for $\rho = 0$, where we did not find any significant deviations in the mean inequality perceptions and the actual overall inequality of $G = 0.5$, testifying to the robustness of our approach.³

³The results for the null model as well as for different C are available upon request.

4.1 Small-Worldiness

We use state-of-the-art methods to test for the existence of small-world features against an appropriate network null model, here an Erdős-Rényi (ER) graph with the corresponding number of nodes and mean degree first introduced by Erdős & Rényi (1960). ER graphs appear to be the correct null model for two reasons: First, they are a particular case of our model with $\rho = 0$, i.e. without homophily. Hence, allows isolating the impact of homophily and examining whether the model indeed tends to yield 'smaller worlds' for homophilic formation in the precise sense outlined below. Second, we can establish an exact one-to-one correspondence between a graph generated by our model and the ER model, as ER graphs only require the number of nodes and a linkage probability for initialisation that is fully determined by the mean degree of the correspondent network. Other prominent generating models such as Watts-Strogatz graphs have additional degrees of freedom like the 'rewiring probability' without clear correspondence to our model.

We construct three summary metrics to test our model against, as introduced by Humphries & Gurney (2008). First, Λ measures the deviation in average path lengths L , that is,

$$\Lambda_i := \frac{L_i}{L_i^{ER}}, \quad (2)$$

where L_i is the average path length of network i with L_i^{ER} as the average path length of a correspondent ER graph with equivalent number of nodes and mean degree. 'Small-worldiness' requires $\Lambda \approx 1$, as our network should not deviate too much from the random benchmark that indeed features short paths. $E[L_i^{ER}] = (\log[N] - \gamma)/(\log[k]) + 1/2$ with γ as Euler's constant, N as the number of nodes and k as the average degree can be analytically derived which we use in our calculation (Fronczak et al. 2004).

Second, we also require a high clustering coefficient which an ER graph cannot generate. The deviation in the clustering coefficients Γ is defined as

$$\Gamma_i := \frac{C_i}{C_i^{ER}}, \quad (3)$$

with C_i as the clustering coefficient of graph i and C_i^{ER} as the clustering coefficient of the corresponding ER graph. Here, again, analytical results are available which we utilise, mainly that $E[C_i^{ER}] = k/N$ with again k as the average degree and N the number of nodes (Watts 1999). Since ER graphs typically do not exhibit clustering, we require here that $\Gamma_i > 1$ for a small-world to be present.

Finally, we use a summary measure Φ introduced by Humphries & Gurney (2008). We define Φ as

$$\Phi_i := \frac{C_i}{C_i^{ER}} / \frac{L_i}{L_i^{ER}} = \frac{\Gamma_i}{\Lambda_i}. \quad (4)$$

Humphries & Gurney (2008) show that Φ_i features desirable statistical properties when confronted with the conventional Watts-Strogatz model for graph formation and shows a unique maximum between the extreme cases of a random network and an ordered lattice. This is in line with our intuition that small-worldiness results from the interaction of order (in the form of high clustering near the lattice) and randomness (in the form of the random graph featuring low average path lengths), as shown by Watts & Strogatz (1998). We require $\Phi > 1$ for small-worlds. Notice that $\Phi > 1$ is an immediate corollary of the two requirements $\Gamma > 1$ and $\Lambda \approx 1$, but $\Phi > 1$ does not imply the two individual requirements. We call the first sufficient condition 'strong small-worldiness' and $\Phi > 1$ with a violation of either $\Gamma > 1$ or $\Lambda \approx 1$ 'weak small-worldiness', where we now only require normalised clustering to increase *faster* than average path lengths.

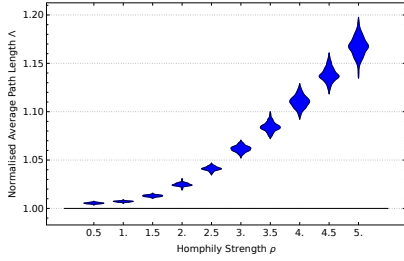


Figure 3: Violin Plots of Normalised Average Path Lengths Λ as a function of Homophily Strength ρ .

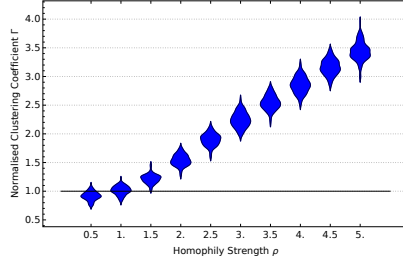


Figure 4: Violin Plots of Normalised Clustering Coefficients Γ as a function of Homophily Strength ρ .

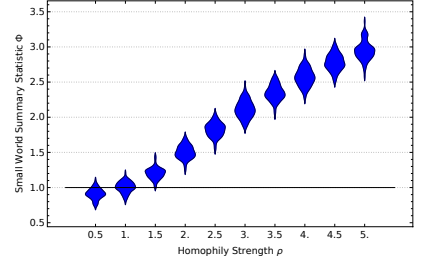


Figure 5: Violin Plots of Small-World Summary Statistic Φ as a function of Homophily Strength ρ .

Note: The Figures report violin plots for the relevant statistics for 'small-worldiness'. The average path length is significantly higher than the ER benchmark for all depicted ρ although by a small margin, indicating violation of the 'strong small-worldiness condition'. Normalised clustering coefficients are for $\rho > 1.5$ significantly higher than the ER benchmark and increase at a much faster rate than average path lengths, indicating that indeed the 'weak small-worldiness' condition is fulfilled.

We indeed find that homophily induces path lengths to grow significantly above the ER benchmark, although by rather small margins between 3 and 20%. Normalised clustering coefficients increase much more rapidly with homophily than average path lengths, demonstrating that our model can achieve relatively high clustering without simultaneously increasing path lengths in the same way. The proposed process thus violates the strong

condition but fulfills the weak condition for small-worlds and is thus broadly by the topological patterns found in real-world social networks. We note further the symmetry to the canonical Watts-Strogatz approach (Watts & Strogatz 1998). While we build on a random network with short average path lengths and interpolate to the desired high clustering through homophily, Watts and Strogatz start from an ordered state with high clustering and approach the random graph benchmark by rewiring to generate shorter average path lengths. Arguably, however, our approach starts from a plausible and empirically well-established behavioural principle in contrast to the purely stochastic process in the Watts-Strogatz world without such behavioural foundations. Besides providing empirical validation, this finding might also point to relatively rapid contagion throughout the homophilic network, be it in the form of rumours or ‘expenditure cascades’.

4.2 Perceived Social Hierarchy

For unbiased hierarchy perceptions, the reported frequency of perceived social position would coincide with the actual positions. Unbiased perceptions thus entail reported perceived positions of equal frequency, as they coincide with the actual population shares. As we show both analytically in Appendix B and by simulation, perceived social positions for homophilic graph formation are far from the equiprobable benchmark. We find a tendency of the vast majority of individuals to place themselves in the middle of the perceived hierarchy, in line with the empirical evidence. We prove that the tendency exists for all $\rho \in (0, \infty)$. Its strength is a function of ρ , though, as we show exemplarily in Figures 6 to 8. The figures plot the empirical densities of income ranks which the respective individuals perceive to hold. For $\rho = 0.5$, the tendency is relatively weak, while for $\rho = 2$ and $\rho = 4$, the densities display a distribution that notably peaks at the centre. In fact, the displayed densities indeed seem to feature all the salient features of the densities of empirical perceived social positions, as shown in Choi (2019).

Notice that this a necessary *outcome* of homophilic graph formation under very mild and general conditions, in contrast to models that take this tendency as an *assumption*. The latter strand of literature has typically taken a bounded rationality stance on the issue and argued that it is failures in information processing which explain the persistent errors in perceptions of social positioning. Our model replicates stylised fact (i) purely by virtue of the network formation process. In contrast to the literature on bounded rationality, we can hence show that stylised fact (i) is consistent with purely rational actors that form correct beliefs based on their available information, as long as homophilic graph formation constrains their information sets. Our model thus entails very different policy implications to

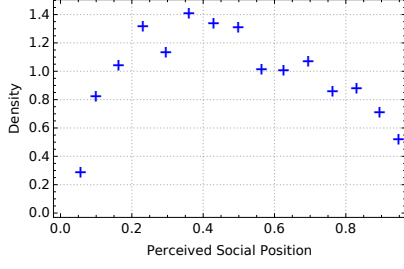


Figure 6: Empirical Density of Perceived Quantile for $\rho = 0.5$.

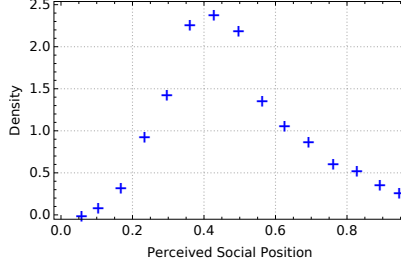


Figure 7: Empirical Density of Perceived Quantile for $\rho = 2$.

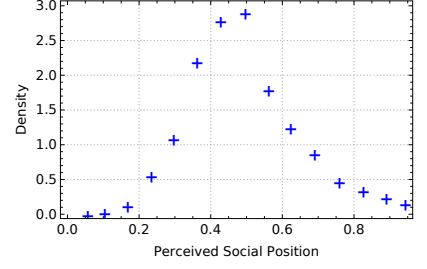


Figure 8: Empirical Density of Perceived Quantile for $\rho = 4$.

Note: The Figures report the perceived social positions for $\rho \in \{0.5; 2; 4\}$ with 15 bins each. All Figures exhibit significant deviation from the benchmark with equal densities. The tendency for individuals to place themselves in the middle of the income hierarchy is, however, only apparent for the middle and right panel, indicating that a homophily strength ρ of 0.5 might be too low to replicate the empirically observed tendency. For $\rho = 2$ and 4, the densities approximate the empirical densities rather well, though.

improve self-perceptions. Since information processing is assumed to be correct in our model, information treatments, i.e., increasing the information received from nodes with heterophile incomes, have mitigating effects on perceptions. Influencing information processing itself, as implied by the established models, is arguably a much harder task for policy.

An immediate corollary of the population perceiving themselves to earn the median income is the tendency for rather poor individuals to overestimate their position and the rich to underestimate it, as all perceive themselves to be in the middle. The simulation runs corroborate this finding, as shown in Figure 9 for $\rho = 4$, where the fit for the median perception tracks the trend in the simulations reasonably well for the vast majority of observations. As we discuss in more detail in Appendix B, there is no tendency to the median for the left and right tail of the distributions which the simulation results reflect, too. Indeed, approaching the minimum or maximum improves the accuracy of individual estimates. The intuition for this is quite simple: The poorest and the richest individual will always correctly perceive their social position, independent of $\rho \in \mathbb{R}_0^+$. The rationale for this is that the actual minimum (maximum) of the whole will always be the minimum (maximum) of any potential subset. Apart from such boundary effects, however, we indeed replicate stylised fact (ii) insofar as the poorer half of the population seems to overestimate their social position, while the richer half underestimates it. This finding is in line with the empirical evidence and suggests that the whole population tends to underestimate the degree of inequality, as we will show in the upcoming subsection.

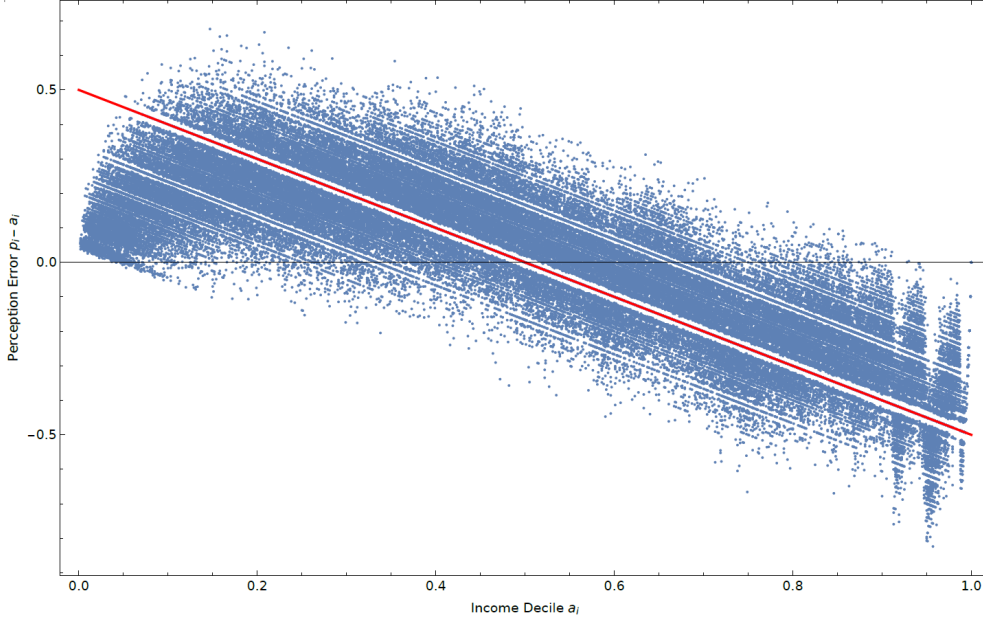


Figure 9: Errors ϵ_i show the difference between perceived position q_i and actual position a_i for all individuals i . The superposed line corresponds to $\epsilon_i = 0.5 - a_i$ or the belief for all individuals to be in a median position of the income distribution. Except for the boundary regions close to the minimum and maximum income, the theoretical fit approximates the trend in the data reasonably well. This indicates that the trend to the median is indeed present for the vast majority of the population.

4.3 Perceived Individual Inequality

We define perceived inequality as the Gini coefficient calculated over the perception set of a given individual i . In Figure 10, we plot those perceived Ginis against the income ranks of our individuals with a higher rank indicating a higher income. In line with stylised fact (iii), we find that inequality perceptions decrease almost monotonically in income rank, while all individuals underestimate the actual degree of inequality significantly. This is a corollary of our homophilic graph formation process. The Gini coefficient is conventionally defined as the ratio of (unweighted) mean differences in the incomes within the perception set of an individual to the mean income within this group. Homophilic graph formation now lets those unweighted mean differences increase linearly at most, while the mean incomes increase exponentially due to the exponential distribution by which incomes are initialised. As a result, the ratio falls almost monotonically. This result is not only plausible due to its accordance with stylised fact (iii) but might also correspond with the empirical evidence on perception formation. One of the most prominent hypotheses on perception formation from stimuli is the Weber-Fechner law (Fechner 1862) which indicates that perceived differences

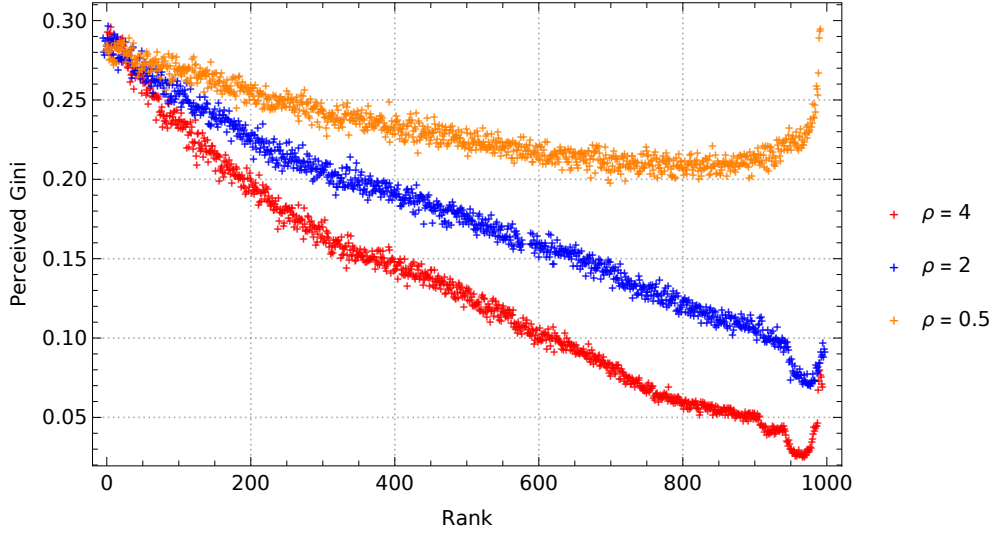


Figure 10: The Figure plots inequality perceptions against the income rank. All individuals underestimate true inequality with a Gini of 0.5. Degrees of underestimation vary, though, where bias increases almost monotonically in income rank. The intuition for this is that homophilic graph formation lets unweighted inequality (absolute income differences) increase only linearly in income rank, but the reference standard (mean income) increases exponentially.

in stimuli need to be proportional to the baseline of a given stimulus to be recognisable. The phenomenon is well-established not only for sensory stimuli (Formankiewicz & Mollon 2009; Pienkowski & Hagerman 2009) but also finds use in marketing research on price responses (Sirvanci 1993; Snell et al. 1995). In this framework, one can also understand a decreasing perceived Gini as the change in stimuli (the unweighted differences in incomes of the perception set) do not increase in the same way as the baseline of stimuli (the mean incomes of this perception set) and is thus also consistent with the psychological microevidence.

4.4 Perceived Global Inequality

For further validation, we also examine whether our graph generating process can *quantitatively* replicate empirical perception patterns. We use the mean, minimum and maximum for inequality perceptions on a national level calculated yearly for a large sample of 32 OECD countries in a 30 year time-span by Choi (2019).⁴ Over all countries, they find a minimum perceived Gini of $G_{min} = 0.1276$, a mean perceived Gini of $G_{mean} = 0.1708$ and a maximum perceived Gini of $G_{max} = 0.2534$. Analogous to their empirical results, we average over the Gini perceptions of all individuals. As we show in Figure 11, we find that our

⁴For details and descriptives of their sample, cf. Choi (2019), especially Appendix B2.

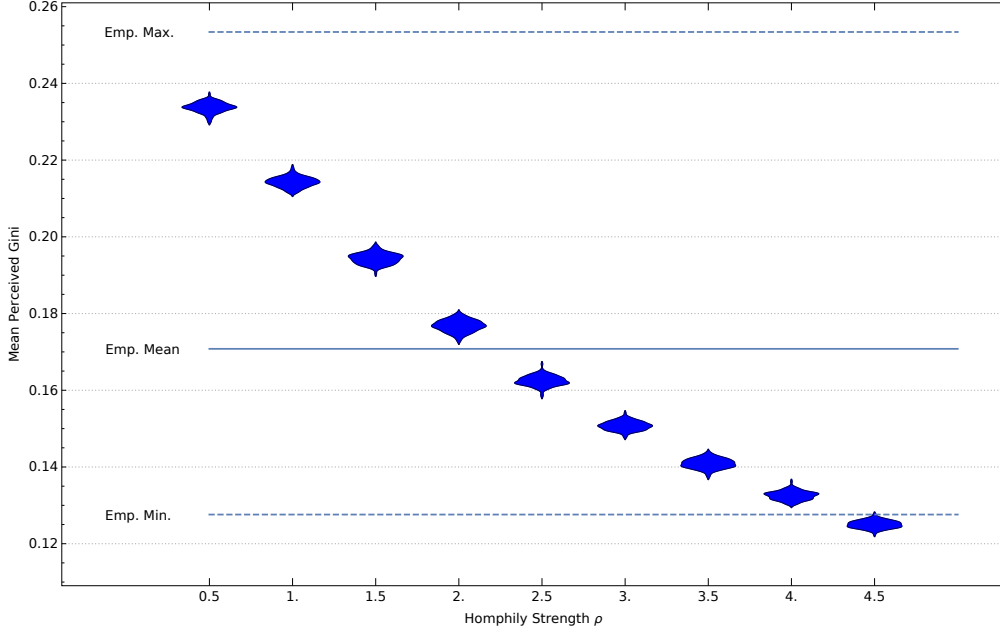


Figure 11: The figure shows the violin plots for the cross-sectional average of individual inequality perceptions per Monte Carlo run of our graph model. The dashed vertical lines correspond to the empirical sample minimum and maximum, while the bold line corresponds to the sample mean. We find that varying the homophily ρ parameter can fully quantitatively account for the variation in empirical perceptions.

process can fully account for their empirical findings and the variation between inequality perceptions by only varying the homophily parameter ρ . We also note that the sample average of national inequality perceptions implies a homophily degree $\rho \in [2.0; 2.5]$ which fits neatly into our narrative above that this is also the threshold from which onwards we observe 'small-worldiness', further testifying to the validity of our approach. Our findings imply considerable cross-country variation in homophily.

4.5 Perception Dynamics

To analyse perception dynamics, that is, the reaction of inequality perceptions to changes in actual inequality, we need to initialise the model with another distribution, as the exponential has a fixed Gini of about 0.5, irrespective of its parameter. We use the log-normal as another benchmark and vary the dispersion parameter σ to simulate changes in the Gini coefficient which is another distribution typically used to describe the skewed nature of empirical income distributions (Knell & Stix 2020). As we see, apart from implausibly low degrees of inequality, changes in actual inequality cause far less than a one-to-one change in perceived inequality. Especially for higher degrees of homophily, we find that the plot quickly reaches a plateau,

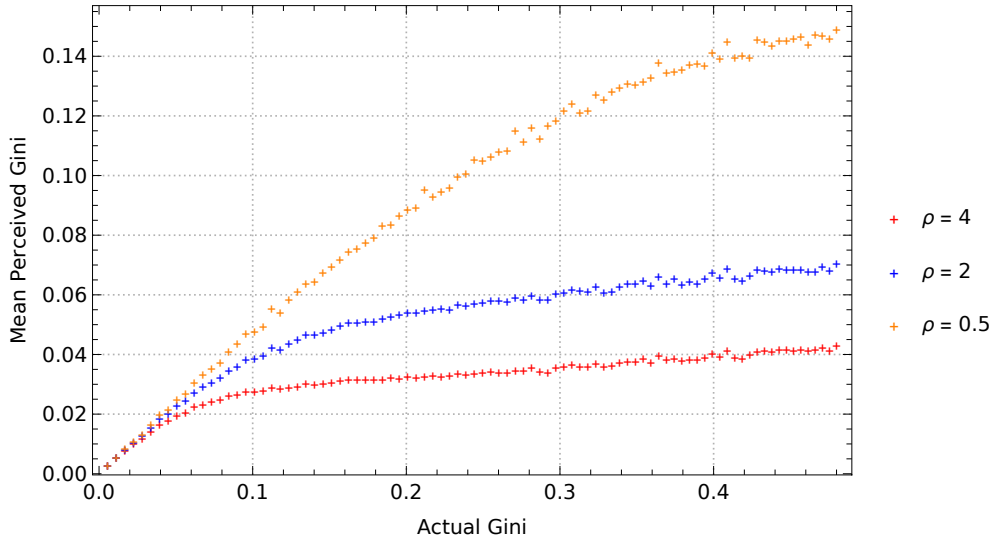


Figure 12: The figure plots varying degrees of true inequality against perceived inequality. Apart from extremely egalitarian states, increases in actual inequality are reflected in perceived inequality much less than one-to-one. For large degrees of homophily, the schedule reaches a plateau rather quickly, i.e., perceived inequality does respond extremely slow to changes in actual inequality.

where inequality perceptions are now extremely persistent invariant of actual inequality. Our model thus is consistent with stylised fact (*iv*) as our last test of validity.

This behaviour occurs because homophily becomes more binding and segregation stronger when actual inequality increases. This mechanism leads *ceteris paribus* to a decrease in perceived inequality which offsets a direct impact of objective inequality on subjective perception. Compare, for illustration, the two regimes close to a completely egalitarian income distribution near $G = 0$ and relatively high degrees of inequality near $G = 0.5$. The egalitarian state is close to a random network, as homophilic segregation presupposes income differences. Small changes in actual inequality are thus not strongly reflected in segregation and almost fully impact perceived inequality, leading to a one-to-one correspondence of perceived and actual inequality in this neighbourhood. For large degrees of actual inequality and large homophily, changes in actual inequality immediately impact segregation, leading to a plateau and very persistent perceptions.

4.6 Segregation Patterns

We measure segregation as the proportion of links an individual i has in their own decile as Δ_i as one particular way to measure ‘selectivity’ without access to behavioural linkage parameters. This constitutes a straightforward but standard way to measure segregation and

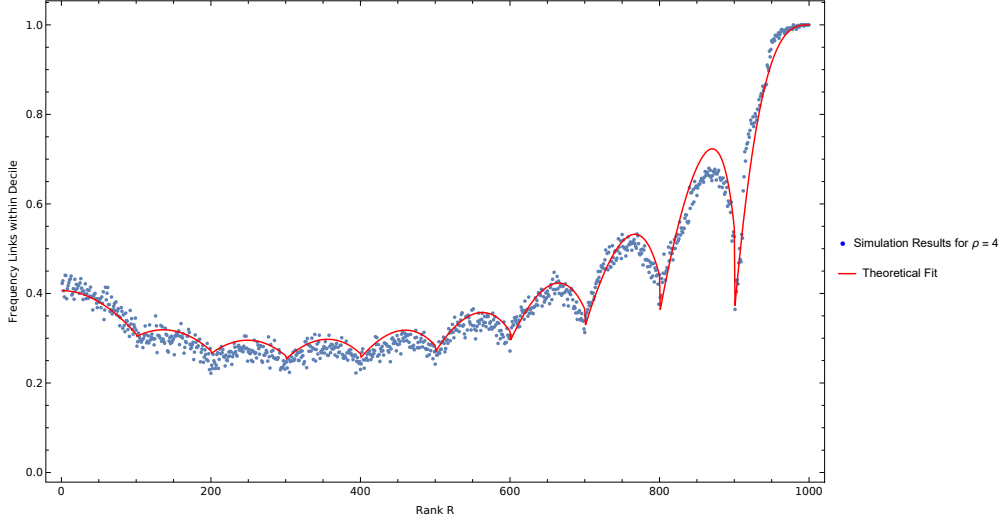


Figure 13: The figure plots our segregation measure, measured as the proportion of links of a node to nodes in the same decile, along the rank distribution. The theoretical fit is obtained for the assumption that nodes choose their neighbours themselves without other nodes choosing links incoming to them. The goodness of fit shows that this is indeed a reasonable assumption. Segregation exhibits distinct and non-trivial patterns both regarding global as well as local maxima.

is easily transformed into normalised measures of segregation like the E-I index Ξ_i which is defined as the difference between the share of between-group links and the share of within-group links (Bojanowski & Corten 2014).⁵ Figure 13 plots the simulated segregation statistics as well as a theoretical fit for $\rho = 4$. For analytical convenience, the superposed red line plots the probabilities that an individual chooses another agent to link to within their own decile as a first pick, so the total choice set consists of 999 other individuals, and does not account for the possibility that other agents already link to the agent in question, in contrast to our algorithm. As can be seen from the good fit of the predicted probabilities to the simulated shares, neither of these two effects seem to bias the approximation significantly. Appendix A details the derivation. The goodness of fit demonstrates that these incoming-links do not exhibit a significant effect on segregation patterns and tend to average out in the aggregate, showing that our analytical approximation is indeed reasonable.

Segregation exhibits two distinct patterns along the rank distribution. First, we find that segregation exhibits a skewed U-shape and increases, especially for the richest decile, which is almost completely disconnected from the other groups. In this sense, our graph formation process endogenously creates echo chambers for the richest whose information sets do not cover the poorer population at all. The echo chamber comes about due to the

⁵Both measures are simple linear transformations of each other due to $\Xi_i = 1 - 2\Delta_i$.

richest individuals being extremely selective in choosing their link-neighbours, as we have shown in the model section 3. Second, we also find a rather strong variation together with local maxima within deciles. This finding might be, however, spurious and a partial artefact of boundary effects at decile boundaries. As Appendix B shows analytically, individuals will choose those sets of link-neighbours with the highest probability that are distributed symmetrically around them in rank. Thus, individuals exactly at the decile boundary will most likely select a set with half of their neighbours across the boundary. Individuals closer to the centre of a decile, on the other hand, will by the same token choose with highest probability link-neighbours within their own decile. Arbitrarily pre-defined group boundaries can thus create within-group variability in commonly used indices like the E-I index that nevertheless exhibits desirable statistical features at an aggregate level (Bojanowski & Corten 2014). These findings extend well beyond income deciles, as variables like age group, place of living, gender, education or ethnicity are likely strongly correlated with income. Studies using E-I type indices to detect homophily in other variables might hence create spurious results if income homophily is also present. The relevance of such boundary effects has increasingly also been recognised in applied work (Hvidberg et al. 2020). Whenever the relevant dimension follows a cardinal scale, it might therefore prove more fruitful to use a rolling-window type of estimation. For rolling-windows, within-groups are defined in relation to the individual in question, such as a fixed number of income ranks or a fixed income rank interval around theirs, avoiding spurious boundary effects.

4.7 Summary of Generating Mechanisms

For non-biased individual samples, one would expect agents to both estimate global inequality correctly in aggregate and also locate their true income quantile. However, homophily triggers link selectivity and hence biased samples which in turn causes inequality perceptions based on an agent’s income level and rank (cf. Figure 14)

Selectivity in link formation depends on global income inequality. Furthermore, agents whose income is further from the global median income are more selective in their link-neighbours and so are agents with the higher income ranks, as an artefact of the cut-off in the exponential selection function at the low end of the income distribution (cf. the skewed U-shape in Figures 1 and 2). Such link-neighbour selection generates personal networks for each agent in which this agent tends to have the median income and where income rank differences are relatively small.

Moreover, the extent of relative income differences in one’s personal network now depends on the characteristics of the agent in question. Firstly, the impact of link-neighbours with

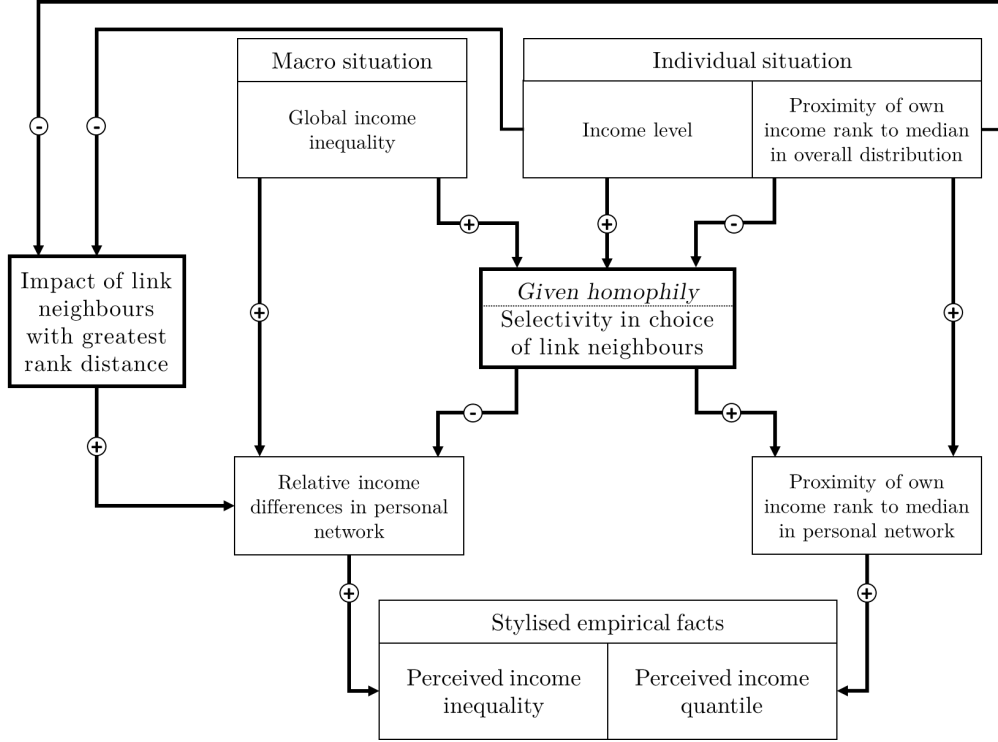


Figure 14: Causal pathway depicting the mechanisms of individual selectivity in choice of link-neighbours and subjective perception in the model given a left-skewed income distribution

great rank differences to the perceiving agent is larger if this perceiving agent and consequently the majority of link-neighbours have a low income themselves since local inequality calculations weight income differences by the local mean. Secondly, agents close to the global income distribution median *ceteris paribus* perceive lower inequality levels, for the greatest rank distances tend to be smaller in these cases.

Overall, in a homophilic linking regime, the complex interaction of actual income inequality and of individual absolute income and distribution rank cause biases in income level self-rating and inequality assessment that aggregate to biased underestimation of inequality. Due to the interplay of factors that feed into individual selectivity in choice of link-neighbours, the relation between objective income structure and individual perceptions is non-monotonic and not trivial but requires case-based assessment.

5 Discussion

Our parsimonious model demonstrates that individuals who evaluate their immediate social environment without a bias can misperceive their own rank in the overall income distribution

as well as global inequality. Homophilic formation of the immediate environment suffices to fully explain the discrepancy between actual and perceived inequality since a rising level of actual inequality causes higher selectivity in link-formation. Moreover, the further away someone’s income rank is from the global median and the higher their income, the more selective they are in their choice of link-neighbours. Hence, the less substantial the differences they encounter when assessing these link-neighbours’ incomes

Thus, public misperceptions are not necessarily driven by limitations in information processing, e.g. a behavioural tendency to place oneself near the median of social hierarchy, but by limited information sets the individuals exhibit for inductive reasoning. The seemingly subtle distinction between constraints on information processing and information sets carries important policy implications: When it is the limits of available information and not limits in cognitive ability driving misperceptions, informational treatments may be successful, as also the empirical literature suggests (Cruces et al. 2013; McCall et al. 2017; Karadja et al. 2017). Such treatments can either consist of delivering information about income inequality itself or facilitating the formation of more diverse contacts in order to overcome the segregation by income that our model finds. In other words, this means breaking up echo chambers that are caused by humans drawing confidence in their beliefs only from repeated observations while ignoring a potential lack of diversity in sources (Foster et al. 2012; Schwarz et al. 2016). Educating individual citizens about their information deficit and providing ways of overcoming it is important from a democracy theory perspective. For example, Rawls (2005, p. 224) requires “presently accepted general beliefs” as basis for arguments in the public forum. However, while one can assess the income inequality objectively without any room for disagreement if using all globally available information, citizens who work only with their individual information will agree on a belief about the Gini that underestimates its actual value. Hence, the lack of individual information access inhibits deliberation about the level inequality and its changes, e.g., in response to past policy measures.

Our “model produces quantitative agreement with empirical macrostructures, as established through on-board statistical estimation routines” and also “quantitative agreement with empirical microstructures, as determined from cross-sectional and longitudinal analysis of the agent population” (Barde & Van Der Hoog 2017). Following the suggestion by Fagiolo et al. (2019), we calibrate the model using empirical micro-data and compare the simulation output to empirical macro findings, too. For calibration, we use an exponential income distribution that characterises industrialised countries. Furthermore, we build agents’ linking behaviour and perception formation on behavioural rules found in lab experiments. This calibration should guarantee resemblance (Mäki 2009) between our model and the real world. Simulation outputs of a societal structure close to a small-world one with self-segregation of

highest-income agents and severe underestimation of the income Gini across income levels mirror the corresponding empirical findings. However, we develop a specific parallel reality (Sugden 2009) that features generating mechanisms for empirical findings in our reality, and hence our results present a candidate explanation for the stylised empirical facts (Epstein 1999). Consequently, there may be different, more adequate, parallel realities featuring either these or even better mechanisms, despite to the best of our knowledge there being no existing models that fulfil these characteristics.

Given the validity of our model, for the first time, we are able to infer the composition of these reference groups from readily available observational data on perceived inequality to inform both empirical investigations as well as more comprehensive model-building in other regards. Directly investigating perception networks might provide a possible remedy for the problem that identification of interaction effects within survey data in the form of time series is hindered by sample sizes that are typically one or two orders of magnitude too low to distinguish noise from true interaction (Alfarano & Milaković 2012). Our main empirical prediction is that the homogeneity of social groups, the fraction of links to agents within the own income decile, roughly follows a U-shaped pattern with a massive decrease in diversity for the richest and (a quantitatively much more modest one) for the poorest agents. This finding might point to an endogenously emerging ‘elite discourse’ with almost no transmission of information to the poorer 90% of the population.

In terms of theory, our empirically validated random geometric graphs might provide an ideal microfoundation for theories of consumption as dependent on the relative income position and for which shocks that affect local income compositions lead to ‘expenditure cascades’ (Duesenberry 1949; Veblen 2001; Frank et al. 2014). This new mechanism might shed light on the disputed link between economic inequality and growth. We will address these questions in further research.

Finally, our model presents a way of generating random geometric graphs defining both the distribution of the feature that determines linking probability between any pair of nodes and a minimum degree value for each node. Put differently, we apply a Barabási-Albert (1999) Preferential-Attachment type procedure, which is intuitive for social scientists, to features other than degree and get network graphs that can be analysed using readily available methods from random geometric graph theory. To the best of our knowledge, there is no such way yet.

Our cross-country analysis shows that there exists considerable variability in implied homophily levels. There are several possible candidates to explain this variability, such as cultural norms, diversity in media and political representation or spatial segregation. Regarding spatial segregation patterns, Thorstein Veblen made the farsighted observation as

early as 1899 that urbanisation should increase diversity in social contacts since cities are the place “where the human contact of the individual is widest and the mobility of the population is greatest” (Veblen 2001, p. 66). This is also what the rather scarce existing evidence for Vietnam and Central and Eastern Europe suggests (Mahajan et al. 2014; Binelli & Loveless 2016). Spatial and perception network segregation might thus overlap and interact (Newman et al. 2018), where policy affecting the allocation of land can also exhibit unintentional effects on perceptions. We leave detailed analyses on these determinants for further research. Thus, perceived inequality should *ceteris paribus* be higher in urban areas resulting from the higher average income diversity per perception network, a testable hypothesis and thus a possible avenue for further research in spatial economics.

It is possible to apply generating procedure to features other than income that exhibit different distribution patterns. This approach will hopefully inspire future studies of expectation formation, e.g., regarding inflation or business sentiment. In these fields, identification of the relevant perception networks might be a crucial step to bring macroeconomic theory currently mostly building on atomistic rational expectations and empirical studies, that find little support for those types of expectations, closer together (Pesaran & Weale 2006). Hence, we also provide a toolkit for analysing the impact of homophily regarding any specified feature on network generation (and potential interactions on the resulting network) given a particular distribution of this feature. In sum, we anticipate our theoretical framework to generate numerous avenues for further studies both theoretically, regarding graph- and expectation formation, as well as empirically, pertaining to the determinants of homophily and possible policy measures, to information transmission and to the effects of inequality on aggregate consumption.

Appendix A Linkage Probabilities for Homophilic Networks

Heuristic Derivation. Linkage Probabilities.

Consider an arbitrary node K_i indexed $i \in 0, \dots, N - 1$ as their rank R increasing income that is part of a graph G with N nodes characterised by adjacency matrix A . Let I_i denote their income, where $f_\lambda(\cdot)$ defines the PDF of a exponential probability distribution defined over the real half-line $(0, \infty)$ with parameter $\lambda > 0$ and $F_\lambda(\cdot)$ the corresponding CDF. The quantile function for any population share p and with parameter $\lambda > 0$ for an exponential distribution is given as

$$\phi_\lambda(p) = \frac{-\log[1-p]}{\lambda}. \quad (5)$$

We assume without loss of generality that $\lambda = 1$ for normalisation. Calculated ρ values therefore need to be scaled up by the inverse of the mean income, λ^{-1} for empirical application.

The quantile of a node with income I_i can be approximated by their rank R , such that $p \approx R/N$ as a discrete approximation of the continuous probability density which holds for large N . We want to derive the probability that a node i with rank $R \in \mathbb{N}_0^+$ connects to a node j with a distance of d to node i . Expressing the weights as defined in the model section now in the form of quantiles, we get

$$w_{ij} = \exp[-\rho |\phi_{\lambda=1}(R/N) - \phi_{\lambda=1}((R+d)/N)|] \quad (6)$$

$$= \exp[-\rho |\log(1 - (R/N)) - \log(1 - (R+d)/N)|]. \quad (7)$$

Assume first that $d > 0$, that is, j is richer than i . Simplifying the weights yields for $d > 0$:

$$w_{ij} = \exp[-\rho (\log(1 - (R/N)) - \log(1 - (R+d)/N))] \quad (8)$$

$$= \exp[\log\left(\frac{N-R}{N}\right)^{-\rho} - \log\left(\frac{N-R-d}{N}\right)^{-\rho}] \quad (9)$$

$$= \left(\frac{N-R}{N-R-d}\right)^{-\rho} \quad (10)$$

$$= \left(\frac{N-R-d}{N-R}\right)^{\rho} \quad (11)$$

Analogously, we get for $d < 0$

$$w_{ij} = \exp[-\rho (\log(1 - ((R+d)/N)) - \log(1 - (R/N)))] \quad (12)$$

$$= \exp[\log\left(\frac{N-R-d}{N}\right)^{-\rho} - \log\left(\frac{N-R}{N}\right)^{-\rho}] \quad (13)$$

$$= \left(\frac{N-R-d}{N-R}\right)^{-\rho} \quad (14)$$

$$= \left(\frac{N-R}{N-R-d}\right)^{\rho} \quad (15)$$

To translate w_{ij} into probabilities, we need to normalise by all weights. Note that this is still a (close) approximation of the probabilities of link-formation of a given node i . Nodes

draw their C link-neighbours from the set of all neighbours. This implies that draws are not independent, as we assume here. Since $C \ll N$, however, the effect is marginal. The approximation for the probability below, however, seems to perform quite well which we verify in our subsection on segregation. The probability p that i chooses j as a link-partner can therefore be approximated as

$$p_{ij}(N, R, d) \approx \begin{cases} \left(\left(\frac{N-R}{N-R-d} \right)^\rho / \left(\sum_{\tilde{d}=-R+1}^{-1} \left(\frac{N-R}{N-R-\tilde{d}} \right)^\rho + \sum_{\tilde{d}=1}^{N-R-1} \left(\frac{N-R-\tilde{d}}{N-R} \right)^\rho \right) & \text{for } d < 0, gh \\ \left(\left(\frac{N-R-d}{N-R} \right)^\rho / \left(\sum_{\tilde{d}=-R+1}^{-1} \left(\frac{N-R}{N-R-\tilde{d}} \right)^\rho + \sum_{\tilde{d}=1}^{N-R-1} \left(\frac{N-R}{N-R} \right)^\rho \right) & \text{for } d > 0. \end{cases}$$

Notice that the function behaves as expected and is monotonically decreasing in $|d| \in \mathbb{N}^+$. The strength of selection also increases monotonically in the homophily parameter ρ . For $\rho = 0$, we recover the equiprobable case without any decay. The precise functional form of the decay for $\rho \in \mathbb{R}^+$ is, however, far from trivial and changes along the rank distribution. The right tail of the correspondent density is always a power transformation of a linear function, whereas the left tail for any given R is a power transformation of function with hyperbolic decay. In this sense, all nodes are more 'selective' regarding individuals that are poorer than regarding the richer part of the population. To see this, compare the decay for the minimum and the maximum of the distribution for $\rho = 1$ as a special case. For $R = 0$, $p_{ij} \propto 1 - (|d|/N)$ with linear decay in $|d|$, as there exists only a right tail, while for $R = N - 1$, $p_{ij} \propto 1/(1 + |d|)$ which decays extremely fast in $|d|$ by a power function, as there exists only a left tail here. In this sense, the richest individual is far more 'selective' in choosing their (poorer) link-neighbours than the poorest individual choosing their (richer) ones.

The theoretical expected segregation index we compare against our simulation results can be straightforwardly computed from those probabilities. Let δ_i be the set of nodes that are in the same group as node i such as an income quantile. The probability to connect with a link-neighbour \tilde{p}_i can then again be approximated as

$$\tilde{p}_i(N, R, d) \approx \sum_{j \in \delta_i} p_{ij}. \quad (16)$$

Appendix B Perceived Quantiles in Perception Networks

Proof Sketch. Pure Homophily implies a Tendency to the Median in Perceived Quantiles.

Consider an arbitrary node K_i indexed $i \in 0, \dots, N - 1$ in a graph G characterised by adjacency matrix A . Let I_i denote their income, where $f_\lambda(\cdot)$ defines the PDF of a exponential probability distribution defined over the real half-line $(0, \infty)$ with parameter $\lambda > 0$ and $F_\lambda(\cdot)$ the corresponding CDF. Let M be the number of links of node K_i with M even. This leaves us with $\binom{N-1}{M} = S$ possible permutations of link-neighbours. Assume further for $F_\lambda(I_i)$ that it is between $1/2 \cdot M/N$ and $1 - (1/2 \cdot M/N)$, such that

$$\frac{1}{2} \frac{M}{N} < F(I_i) < 1 - \frac{1}{2} \frac{M}{N}. \quad (17)$$

Let now θ_{ij} be an arbitrary realisation of a permitted set of incomes of nodes to which K_i linked, indexed by j out of the set of permitted sets Θ_i with $\Theta_i = \{\theta_{i1}, \dots, \theta_{iS}\}$ and size S . Assume further that all incomes in θ_{ij} are distinct. If link formation is independent of I_i as the sole characteristic differentiating K_i from all other nodes, all sets θ_{ij} of the same size M are equally likely with probability $1/S$ by extension, since K_i connects to any other node with equal probability. This would be the case for both standard preferential attachment models as well as ER random graphs.

In our model, the probability p_{ik} that $A_{ik} = 1$ depends negatively on the absolute distance $|I_i - I_k|$, such that $\partial p_{ik} / \partial |I_i - I_k| < 0$. By linearity, the probability p_{ij} of node i to have θ_{ij} as their chosen set of incomes to which she is linked decreases in the sum of absolute differences, that is, $\partial p_{ij} / \partial \sum_{I_k \in \theta_{ij}} |I_k - I_i| < 0$. It follows, that p_{ij} as a local probability of a set of a given length being chosen by homophilic preferential attachment is maximised for a minimisation of $\sum_{I_k \in \theta_{ij}} |I_k - I_i|$. Since the benchmark without homophily is equal probability of $1/S$ for all sets of a given size M , this condition also maximises the global probability that this set is chosen for a given size M . Formally, the minimisation problem chooses a set or sets θ_{ij} such that

$$\arg \min_{\theta_{ij}} \sum_{I_k \in \theta_{ij}} |I_k - I_i|. \quad (18)$$

It remains to be shown that this minimisation leads to the choice of a set θ_{ij} for which I_i is the median value. The median requires the same number of values above or below I_i in θ_{ij} . With M links for node K_i of income rank R and M even, this requires $M/2$ values above and below I_i . For I_i as the median being minimising for the absolute distances, this requires i) that there exists no node with rank $R + M/2 + 1$ such that their income distance to K_i is less than the income distance from node K_i to the node ranked $R - M/2$. If i) is

violated, the node with rank $R + M/2 + 1$ is part of the distance-minimising set and thus, I_i is not the median of θ_{ij} . The symmetrical condition ii) requires that there is no node with rank $R - M/2 - 1$ such that its distance to K_i is less than the distance of K_i to the node with rank $R + M/2$. In terms of a quantile function, we require

$$\phi\left(\frac{R + M/2 + 1}{N}\right) - \phi\left(\frac{R}{N}\right) > \phi\left(\frac{R}{N}\right) - \phi\left(\frac{R - M/2}{N}\right) \quad (19)$$

and

$$\phi\left(\frac{R}{N}\right) - \phi\left(\frac{R - M/2 - 1}{N}\right) > \phi\left(\frac{R + M/2}{N}\right) - \phi\left(\frac{R}{N}\right). \quad (20)$$

Rearranging yields

$$\phi\left(\frac{R + M/2 + 1}{N}\right) + \phi\left(\frac{R - M/2}{N}\right) > 2\phi\left(\frac{R}{N}\right) > \phi\left(\frac{R + M/2}{N}\right) + \phi\left(\frac{R - M/2 - 1}{N}\right). \quad (21)$$

Expressing the left-hand side of inequalities for a generic distribution in (21) for an continuous exponential such that $R/N \approx p$ and substituting the quantile function, we require

$$\frac{-\log[1 - (R + M/2 + 1)/N]}{\lambda} + \frac{-\log[1 - (R - M/2)/N]}{\lambda} > 2\frac{-\log[1 - (R/N)]}{\lambda} \quad (22)$$

The condition $R/N \approx p$ presupposes N to be sufficiently large for the discrete realisations of the sample to approximate the quantiles of the continuous exponential distribution. We find this condition fulfilled for several numerical experiments. It is easy to see that the left-hand side condition in (21) is fulfilled for a quantile function whose first derivative is monotonically increasing which is the case for $d\phi_\lambda(p)/dp = 1/((1 - p)\lambda)$ for $p \in [0, 1)$ and $\lambda > 0$. We can also show this by manipulation of (22) as

$$\log[(1 - (R + M/2 + 1)/N) \cdot (1 - (R - M/2)/N)] < \log[(1 - (R/N))^2] \quad (23)$$

$$(1 - (R + M/2 + 1)/N) \cdot (1 - (R - M/2)/N) < (1 - (R/N))^2 \quad (24)$$

which implies

$$1 - \frac{(R + M/2 + 1)}{N} - \frac{R - M/2}{N} + \frac{(R + M/2 + 1)(R - M/2)}{N^2} - 1 + 2\frac{R}{N} - \frac{R^2}{N^2} < 0 \quad (25)$$

$$-\frac{1}{N} + \frac{R - M/2 - (M/2)^2}{N^2} < 0 \quad (26)$$

$$\frac{R - M/2 - (M/2)^2 - N}{N^2} < 0. \quad (27)$$

Since $R \leq N$ per definition, condition (27) is trivially fulfilled. Notice that this implies for an exponential initial distribution, I_i cannot be below the median in the most likely set. The right hand-side of inequalities (21) is a bit more demanding. Stating the condition in terms of the quantile function for an exponential, we get

$$2\frac{-\log[1 - (R/N)]}{\lambda} > \frac{-\log[1 - (R - M/2 - 1)/N]}{\lambda} + \frac{-\log[1 - (R + M/2)/N]}{\lambda}. \quad (28)$$

Simplifying yields

$$(1 - (R/N))^2 < (1 - (R - M/2 - 1)/N) \cdot (1 - (R + M/2 - 1)/N) \quad (29)$$

$$(1 - (R/N))^2 < \frac{(N - R + M/2 + 1)(N - R - M/2)}{N^2} \quad (30)$$

$$(N - R)^2 < N^2 - RN - NM/2 - RN + R^2 + RM/2 + NM/2 - RM/2 - M^2/4 + N - R - M/2 \quad (31)$$

$$0 < N - R - M/2 - M^2/4 \quad (32)$$

$$\frac{R}{N} < 1 - \frac{1}{2}\frac{M}{N} - \frac{M^2}{4N}. \quad (33)$$

For our discrete sample, $F_\lambda(I_i) \approx R/N$ which reveals that condition (33) is only a slightly more demanding condition than boundary condition (17) that guarantees the possibility of I_i being a median in the first place and only differs by $M^2/4N$. Since we typically assume $M \ll N$, this term vanishes. Indeed, for a realistic baseline scenario with $N = 1,000$ and $M = 5$, the condition is fulfilled for the poorest 99% of the population and thus for the vast majority. Together with the lower boundary condition (17), the tendency to place themselves in the middle should exist for about 98% of the population and thus the vast majority. Minimising absolute deviations for an exponential income distribution and $M \ll N$

thus entails choosing sets that let I_i be the median of $\theta_{ij} \cup I_i$ for almost all I_i . While the strength of this mechanism will of course be dependent on $\partial p_{ij} / \partial \sum_{I_k \in \theta_{ij}} |I_k - I_i|$, the median is the most likely outcome for any homophile network as the perceived quantile for the vast majority of nodes.

References

- Alfarano, S. & Milaković, M. (2012). Identification of interaction effects in survey expectations: a cautionary note. *Studies in Nonlinear Dynamics & Econometrics*, 16(4)
- Alicke, M. D., Zell, E. & Bloom, D. L. (2010). Mere categorization and the frog-pond effect. *Psychological Science*, 21(2), 174–177
- Barabási, A.-L. & Albert, R. (1999). Emergence of scaling in random networks. *science*, 286(5439), 509–512
- Barde, S. & Van Der Hoog, S. (2017). An empirical validation protocol for large-scale agent-based models. *Bielefeld Working Papers in Economics and Management*
- Bartels, L. M. (2018). *Unequal democracy: The political economy of the new gilded age*. Princeton University Press
- Binelli, C. & Loveless, M. (2016). The urban–rural divide: Perceptions of income and social inequality in Central and Eastern Europe. *Economics of Transition*, 24(2), 211–231
- Boguná, M., Pastor-Satorras, R., Díaz-Guilera, A. & Arenas, A. (2004). Models of social networks based on social distance attachment. *Physical Review E*, 70(5), 056122
- Bojanowski, M. & Corten, R. (2014). Measuring segregation in social networks. *Social Networks*, 39(1), 14–32
- Borgatti, S. P. & Foster, P. C. (2003). The network paradigm in organizational research: A review and typology. *Journal of Management*, 29(6), 991–1013
- Boucher, V. (2015). Structural homophily. *International Economic Review*, 56(1), 235–264
- Buckingham, J. T. & Alicke, M. D. (2002). The influence of individual versus aggregate social comparison and the presence of others on self-evaluations. *Journal of Personality and Social Psychology*, 83(5), 1117

- Caplan, B. (2011). *The Myth of the Rational Voter: Why Democracies Choose Bad Policies-New Edition*. Princeton University Press
- Cepić, D. & Tonković, Ž. (2020). How social ties transcend class boundaries? Network variability as tool for exploring occupational homophily. *Social Networks*, 62, 33–42
- Chater, N., Zhu, J.-Q., Spicer, J., Sundh, J., León-Villagrà, P. & Sanborn, A. (2020). Probabilistic biases meet the bayesian brain. *Current Directions in Psychological Science*, 29(5), 506–512. doi:10.1177/0963721420954801
URL <http://dx.doi.org/10.1177/0963721420954801>
- Chiang, Y.-S. (2011). Judgment of distributional inequality in networks. *Social Networks*, 33(4), 342–349
- Chiang, Y.-S. (2015a). Good samaritans in networks: An experiment on how networks influence egalitarian sharing and the evolution of inequality. *PloS One*, 10(6), e0128777
- Chiang, Y.-S. (2015b). Inequality measures perform differently in global and local assessments: An exploratory computational experiment. *Physica A: Statistical Mechanics and its Applications*, 437, 1–11
- Choi, G. (2019). Revisiting the redistribution hypothesis with perceived inequality and redistributive preferences. *European Journal of Political Economy*, 58, 220–244
- Clark, A. E. & Senik, C. (2010). Who compares to whom? The anatomy of income comparisons in europe. *Economic Journal*, 120(544), 573–594
- Cohen, J. (1979). Socio-economic status and high-school friendship choice: Elmtown’s youth revisited. *Social Networks*, 2(1), 65–74
- Collier, P. & Tuckett, D. (2020). Narratives as a coordinating device for reversing regional disequilibrium. *Rebuilding Macroeconomics Working Paper Series*, 22
- Conyon, M. J. & Muldoon, M. R. (2006). The small world of corporate boards. *Journal of Business Finance & Accounting*, 33(9-10), 1321–1343
- Cruces, G., Perez-Truglia, R. & Tetaz, M. (2013). Biased perceptions of income distribution and preferences for redistribution: Evidence from a survey experiment. *Journal of Public Economics*, 98, 100–112
- Dall, J. & Christensen, M. (2002). Random geometric graphs. *Physical Review E*, 66(1), 016121

- Davis, G. F., Yoo, M. & Baker, W. E. (2003). The small world of the american corporate elite, 1982–2001. *Strategic Organization*, 1(3), 301–326
- Drăgulescu, A. & Yakovenko, V. M. (2001). Evidence for the exponential distribution of income in the USA. *The European Physical Journal B-Condensed Matter and Complex Systems*, 20(4), 585–589
- Duesenberry, J. S. (1949). *Income, saving, and the theory of consumer behavior*. Harvard University Press
- Engelhardt, C. & Wagener, A. (2018). What do Germans think and know about income inequality? A survey experiment. *Socio-Economic Review*, 16(4), 743–767
- Epstein, J. M. (1999). Agent-based computational models and generative social science. *Complexity*, 4(5), 41–60
- Erdős, P. & Rényi, A. (1960). On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5(1), 17–60
- Evans, M. D. & Kelley, J. (2004). Subjective social location: Data from 21 nations. *International Journal of Public Opinion Research*, 16(1), 3–38
- Fagiolo, G., Guerini, M., Lamperti, F., Moneta, A. & Roventini, A. (2019). Validation of agent-based models in economics and finance. In C. Beisbart & N. J. Saam (Eds.), *Computer Simulation Validation: Fundamental Concepts, Methodological Frameworks, and Philosophical Perspectives*, (pp. 763–787). Cham: Springer International Publishing
- Fechner, G. T. (1862). *Elemente der Psychophysik*. Breitkopf und Härtel, Leipzig
- Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7(2), 117–140
- Flieth, B. & Foster, J. (2002). Interactive expectations. *Journal of Evolutionary Economics*, 12(4), 375–395
- Formankiewicz, M. A. & Mollon, J. (2009). The psychophysics of detecting binocular discrepancies of luminance. *Vision research*, 49(15), 1929–1938
- Foster, J. L., Huthwaite, T., Yesberg, J. A., Garry, M. & Loftus, E. F. (2012). Repetition, not number of sources, increases both susceptibility to misinformation and confidence in the accuracy of eyewitnesses. *Acta psychologica*, 139(2), 320–326

- Frank, R. H., Levine, A. S. & Dijk, O. (2014). Expenditure cascades. *Review of Behavioral Economics*, 1(1–2), 55–73
- Fronczak, A., Fronczak, P. & Hołyst, J. A. (2004). Average path length in random networks. *Physical Review E*, 70(5), 056110
- Gabaix, X. (2020). A Behavioral New Keynesian Model. *American Economic Review*, 110(8), 2271–2327
- Galaskiewicz, J. (2007). Has a network theory of organizational behaviour lived up to its promises? *Management and Organization Review*, 3(1), 1–18
- Galí, J. (2015). *Monetary policy, inflation, and the business cycle: an introduction to the new Keynesian framework and its applications*. Princeton University Press
- Gimpelson, V. & Treisman, D. (2018). Misperceiving inequality. *Economics and Politics*, 30(1), 27–54
- Golub, B. & Jackson, M. O. (2012). How homophily affects the speed of learning and best-response dynamics. *The Quarterly Journal of Economics*, 127(3), 1287–1338
- Hamilton, M. J., Milne, B. T., Walker, R. S., Burger, O. & Brown, J. H. (2007). The complex structure of hunter–gatherer social networks. *Proceedings of the Royal Society B: Biological Sciences*, 274(1622), 2195–2203
- Huckfeldt, R. R. (1983). Social contexts, social networks, and urban neighborhoods: Environmental constraints on friendship choice. *American Journal of Sociology*, 89(3), 651–669
- Humphries, M. D. & Gurney, K. (2008). Network ‘small-world-ness’: a quantitative method for determining canonical network equivalence. *PloS One*, 3(4), e0002051
- Hvidberg, K. B., Kreiner, C. & Stantcheva, S. (2020). Social position and fairness views. Working Paper 28099, National Bureau of Economic Research
- Kalmijn, M. (1991). Status homogamy in the United States. *American Journal of Sociology*, 97(2), 496–523
- Kalmijn, M. & Flap, H. (2001). Assortative meeting and mating: Unintended consequences of organized settings for partner choices. *Social Forces*, 79(4), 1289–1312
- Karadja, M., Mollerstrom, J. & Seim, D. (2017). Richer (and holier) than thou? The effect of relative income improvements on demand for redistribution. *Review of Economics and Statistics*, 99(2), 201–212

- Kelley, J. & Evans, M. (1995). Class and Class Conflict in Six Western Nations. *American Sociological Review*, 60(2), 157–178
- Kenworthy, L. & McCall, L. (2008). Inequality, public opinion and redistribution. *Socio-Economic Review*, 6(1), 35–68
- Kim, H., Huh, S., Choi, S. & Lee, Y. (2018). Perceptions of inequality and attitudes towards redistribution in four East Asian welfare states. *International Journal of Social Welfare*, 27(1), 28–39
- Kleinberg, J. (2001). Small-world phenomena and the dynamics of information. *Advances in neural information processing systems*, 14, 431–438
- Knell, M. & Stix, H. (2020). Perceptions of inequality. *European Journal of Political Economy*, 65, 101927
URL <https://doi.org/10.1016/j.ejpoleco.2020.101927>
- Kogut, B. & Walker, G. (2001). The small world of germany and the durability of national networks. *American Sociological Review*, 66(3), 317–335
- Kuhn, A. (2019). The subversive nature of inequality: Subjective inequality perceptions and attitudes to social inequality. *European Journal of Political Economy*, 59, 331–344
- Kuklinski, J. H., Quirk, P. J., Jerit, J., Schwieder, D. & Rich, R. F. (2000). Misinformation and the currency of democratic citizenship. *Journal of Politics*, 62(3), 790–816
- Lazarsfeld, P. F., Merton, R. K. et al. (1954). Friendship as a social process: A substantive and methodological analysis. *Freedom and control in modern society*, 18(1), 18–66
- Lewis, K., Gonzalez, M. & Kaufman, J. (2012). Social selection and peer influence in an online social network. *Proceedings of the National Academy of Sciences*, 109(1), 68–72
- Lodge, M., Steenbergen, M. R. & Brau, S. (1995). The responsive voter: Campaign information and the dynamics of candidate evaluation. *American Political Science Review*, 89(2), 309–326
- Lucas, R. E. (1972). Expectations and the neutrality of money. *Journal of Economic Theory*, 4(2), 103–124
- Lucas, R. E. (1973). Some international evidence on output-inflation tradeoffs. *American Economic Review*, 63(3), 326–334

- Lucas, R. E. (1975). An equilibrium model of the business cycle. *Journal of Political Economy*, 83(6), 1113–1144
- Lux, T. (2009). Rational forecasts or social opinion dynamics? Identification of interaction effects in a business climate survey. *Journal of Economic Behavior & Organization*, 72(2), 638–655
- Mac Carron, P., Kaski, K. & Dunbar, R. (2016). Calling dunbar’s numbers. *Social Networks*, 47, 151 – 155
- Mahajan, S., Dinh, V. T., Rab, H., Trieu, V. Q., Nguyen, L. V., Kwakwa, V. & Verghis, M. (2014). Taking stock: an update on Vietnam’s recent economic developments. Tech. rep., Washington, D.C.: World Bank Group
- Mäki, U. (2009). MISSing the world. Models as isolations and credible surrogate systems. *Erkenntnis*, 70, 29–43
- Malacarne, T. (2017). Rich friends, poor friends: Inter-socioeconomic status friendships in secondary school. *Socius*, 3, 2378023117736994
- Mayer, A. & Puller, S. L. (2008). The old boy (and girl) network: Social network formation on university campuses. *Journal of Public Economics*, 92(1-2), 329–347
- McCall, L., Burk, D., Laperrière, M. & Richeson, J. A. (2017). Exposure to rising inequality shapes americans’ opportunity beliefs and policy support. *Proceedings of the National Academy of Sciences*, 114(36), 9593–9598
- McPherson, M., Smith-Lovin, L. & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27(1), 415–444
- Moore, C. & Newman, M. E. (2000). Epidemics and percolation in small-world networks. *Physical Review E*, 61(5), 5678
- Newman, B. J., Shah, S. & Lauterbach, E. (2018). Who sees an hourglass? Assessing citizens’ perception of local economic inequality. *Research and Politics*, 5(3)
- Newman, M. E. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, 98(2), 404–409
- Newman, M. E. (2009). Random graphs with clustering. *Physical Review Letters*, 103(5), 058701

- Newman, M. E., Strogatz, S. H. & Watts, D. J. (2001). Random graphs with arbitrary degree distributions and their applications. *Physical Review E*, 64(2), 026118
- Osberg, L. & Smeeding, T. (2006). "Fair" inequality? Attitudes toward pay differentials: The United States in comparative perspective. *American Sociological Review*, 71(3), 450–473
- Page, B. I. & Shapiro, R. Y. (1993). The rational public and democracy. In *Reconsidering the Democratic Public*. Pennsylvania State University Press
- Pesaran, M. H. & Weale, M. (2006). Survey expectations. *Handbook of Economic Forecasting*, 1, 715–776
- Pienkowski, M. & Hagerman, B. (2009). Auditory intensity discrimination as a function of level-rove and tone duration in normal-hearing and impaired subjects: The "mid-level hump" revisited. *Hearing research*, 253(1-2), 107–115
- Rawls, J. (2005). *Political liberalism*. Columbia University Press
- Sanborn, A. N. & Chater, N. (2016). Bayesian brains without probabilities. *Trends in Cognitive Sciences*, 20(12), 883–893
- Schwarz, N., Newman, E. & Leach, W. (2016). Making the truth stick & the myths fade: Lessons from cognitive psychology. *Behavioral Science & Policy*, 2(1), 85–95
- Silva, A. C. & Yakovenko, V. M. (2004). Temporal evolution of the "thermal" and "superthermal" income classes in the usa during 1983–2001. *EPL (Europhysics Letters)*, 69(2), 304
- Sirvanci, M. B. (1993). An empirical study of price thresholds and price sensitivity. *Journal of Applied Business Research (JABR)*, 9(2), 43–49
- Snell, J., Gibbs, B. J. & Varey, C. (1995). Intuitive hedonics: Consumer beliefs about the dynamics of liking. *Journal of Consumer Psychology*, 4(1), 33–60
- Stevenson, R. T. & Duch, R. (2013). The meaning and use of subjective perceptions in studies of economic voting. *Electoral Studies*, 32(2), 305–320
- Sugden, R. (2009). Credible worlds, capacities and mechanisms. *Erkenntnis*, 70(1), 3–27

- Talaga, S. & Nowak, A. (2020). Homophily as a process generating social networks: Insights from social distance attachment model. *Journal of Artificial Societies and Social Simulation*, 23(2), 6
URL <http://jasss.soc.surrey.ac.uk/23/2/6.html>
- Tao, Y., Wu, X., Zhou, T., Yan, W., Huang, Y., Yu, H., Mondal, B. & Yakovenko, V. M. (2019). Exponential structure of income inequality: evidence from 67 countries. *Journal of Economic Interaction and Coordination*, 14(2), 345–376
- Uzzi, B. & Spiro, J. (2005). Collaboration and creativity: The small world problem. *American Journal of Sociology*, 111(2), 447–504
- Veblen, T. (2001). *The Theory of the Leisure Class*. New York: The Modern Library
- Watts, D. J. (1999). Networks, dynamics, and the small-world phenomenon. *American Journal of Sociology*, 105(2), 493–527
- Watts, D. J. & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, 393(6684), 440–442
- Weeden, K. A. & Cornwell, B. (2020). The small-world network of college classes: implications for epidemic spread on a university campus. *Sociological Science*, 7, 222–241
- Zaller, J. R. (1992). *The nature and origins of mass opinion*. Cambridge University Press
- Zell, E. & Alicke, M. D. (2009). Contextual neglect, self-evaluation, and the frog-pond effect. *Journal of Personality and Social Psychology*, 97(3), 467
- Zhou, W.-X., Sornette, D., Hill, R. A. & Dunbar, R. I. (2005). Discrete hierarchical organization of social group sizes. *Proceedings of the Royal Society B: Biological Sciences*, 272(1561), 439–444