ARTICLES
https://doi.org/10.1038/s41562-019-0677-4

nature human behaviour

Homophily and minority-group size explain perception biases in social networks

Eun Lee[1,11*], Fariba Karimi[2,3,11*], Claudia Wagner[2,3], Hang-Hyun Jo[4,5,6], Markus Strohmaier[2,7] and Mirta Galesic[8,9,10]

People's perceptions about the size of minority groups in social networks can be biased, often showing systematic over- or underestimation. These social perception biases are often attributed to biased cognitive or motivational processes. Here we show that both over- and underestimation of the size of a minority group can emerge solely from structural properties of social networks. Using a generative network model, we show that these biases depend on the level of homophily, its asymmetric nature and on the size of the minority group. Our model predictions correspond well with empirical data from a cross-cultural survey and with numerical calculations from six real-world networks. We also identify circumstances under which individuals can reduce their biases by relying on perceptions of their neighbours. This work advances our understanding of the impact of network structure on social perception biases and offers a quantitative approach for addressing related issues in society.

People's perceptions of their social worlds determine their own personal aspirations[1] and willingness to engage in different behaviours, from voting[2] and energy conservation[3] to health behaviour[4], drinking[5] and smoking[6]. Yet, when forming these perceptions, people seldom have an opportunity to draw representative samples from the overall social network, or from the general population. Instead, their samples are constrained by the local structure of their personal networks, which can bias their perception of the relative frequency of different attributes in the general population. For example, supporters of different candidates in the 2016 US presidential election formed relatively isolated Twitter communities[7]. Such insular communities can overestimate the relative frequency of their own attributes in the overall society. This has been documented in the literature on overestimation effects including false consensus, looking-glass perception and, more generally, social projection[8–12]. In an apparent contradiction, it has also been documented that people holding a particular view sometimes underestimate the frequency of that view, as described in the literature on false uniqueness[13,14], pluralistic ignorance[15,16] and majority illusion[17]. These over- and underestimation errors, which we call social perception biases, affect people's judgements of minority- and majority-group sizes[18].

It has been observed that social perception biases can be related to the structural properties of personal networks[19,20], which can strongly affect the samples of information on which individuals rely when forming their social perceptions[21,22]. However, the impact of different network properties on social perception biases has not yet been systematically explored. Here we explore three such properties. The first is the level of homophily, or how likely the one is to be connected to similar others, which is known as a fundamental structural property of many social networks[23]. The second property is the asymmetry of homophily, or whether homophily is larger in some subgroups than in others. For example, it has been observed that in scientific collaborations, homophily among women is stronger than homophily among men[24]. The third property is the relative size of minority and majority groups in the society. Many social networks are characterized by a large majority group and a much smaller minority group. Examples are the proportions of different genders in science, technology, engineering and maths, of people with different levels of income and of people who smoke or not.

Most existing explanations of social perception biases invoke motivational and cognitive processes rather than social network structure. For example, processes that explain overestimation of the frequency of one's own attributes (for example, false consensus) include wishful thinking[25], easier recall of the reasons for having one's own view[9], rational inference of population frequencies based on one's own attributes[26], feeling good when others share one's own view[27], and justifying one's undesirable behaviours by overestimating their frequency in society[28]. However, these processes cannot explain the opposite effect, underestimating the frequency of our own view (for example, false uniqueness). Instead, this opposite bias is typically explained by a different set of cognitive or motivational processes, such as differential attention to one's own and other groups[13] and bolstering perceived self-competence[14]. Ideally, both overestimation and underestimation biases would be explained by a single mechanism[18].

Here we show empirically, analytically and numerically that a simple network model can explain both over- and underestimation in social perceptions, without assuming biased motivational or cognitive processes. Results from a cross-cultural survey show that the level of homophily and size of the minority group influence people's social perception biases. Analytical results from a generative network model with tunable homophily and minority-group size align well with the empirical findings. Numerical investigations show that

[1]Department of Mathematics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [2]Department of Computational Social Science, GESIS, Cologne, Germany. [3]Institute for Web Science and Technologies, University of Koblenz-Landau, Koblenz, Germany. [4]Asia Pacific Center for Theoretical Physics, Pohang, Republic of Korea. [5]Department of Physics, Pohang University of Science and Technology, Pohang, Republic of Korea. [6]Department of Computer Science, Aalto University, Espoo, Finland. [7]Department for Society, Technology and Human Factors & Department of Computer Science, RWTH Aachen University, Aachen, Germany. [8]Santa Fe Institute, Santa Fe, NM, USA. [9]Complexity Science Hub Vienna, Vienna, Austria. [10]Harding Center for Risk Literacy, Max Planck Institute for Human Development, Berlin, Germany. [11]These authors contributed equally: Eun Lee, Fariba Karimi.
*e-mail: eunfeel@email.unc.edu; fariba.karimi@gesis.org

model predictions are consistent with biases that could occur in six empirical networks, and point to the importance of accounting for asymmetries in homophily. We also show when social perception biases can be reduced by aggregating one's own perceptions with those of one's neighbours. We discuss the implications of these results for the understanding of the nature of human social cognition and diverse social phenomena.

Results

Defining social perception biases. We focus on individual perceptions, or estimates, of the frequency of binary attributes (for example, smoking, attending worship or donating to charity) in the overall social network. We define social perception bias as a ratio of perceived frequency and the true frequency of an attribute. We study these perception biases at the individual level (B_{indv}) and at the group level (B_{group}). Whenever necessary, we add a subscript m for the minority and M for the majority group.

At the individual level, we assume that individuals' perceptions are based on the frequency of an attribute in their personal networks (their direct neighbourhoods). We define individual i's social perception bias as follows:

$$B_{\mathrm{indv},i} = \frac{i\text{'s perception of the minority}}{\text{true fraction of the minority}} = \frac{1}{f_m} \frac{\sum_{j \in \Lambda_i} x_j}{k_i} \qquad (1)$$

where Λ_i is the set of i's neighbours, $k_i = |\Lambda_i|$ is the degree of i, x_j denotes the attribute of individual j, which has the value of 1 for a minority attribute and 0 for a majority attribute, and f_m is the true fraction of the minority in the entire network. Another measure of perception bias could be a simple difference between i's perception of the minority and the true fraction of the minority. However, this measure depends on the minority-group size, making it difficult to compare the biases for different minority-group sizes. To be able to compare the biases we need to normalize the difference with the minority-group size. That normalized difference produces essentially the same results as the ratio measure of equation (1).

The group-level perception bias is defined as the average of perception biases of all individuals in the group:

$$B_{\mathrm{group}} = \frac{1}{|N_g|} \sum_{i \in N_g} B_{\mathrm{indv},i} \qquad (2)$$

where N_g is the set of individuals in a group g, which is either a minority group or a majority group.

We focus on perception biases in estimates of the size of the minority group. The minimum value of the group- and individual-level perception biases is 0 and their maximum value is $1/f_m$ (see Methods). A value below 1 indicates an underestimation of the minority-group size while a value above 1 indicates an overestimation. If the value equals 1, a group or an individual perfectly perceives the frequency of a minority attribute in the entire network.

As an example, Fig. 1 illustrates how we define the perception bias at the individual and group levels for a high-homophily (homophilic) network and a low-homophily (heterophilic) network. The colour of an individual node depicts its group membership: orange nodes belong to the minority and blue nodes to the majority. We focus on the central individual i, who is in the majority in both networks. This individual estimates the size of the minority group on the basis of the fraction of orange nodes in their personal network (enclosed in a dashed circle). In the homophilic network (Fig. 1a), as per equation (1), their individual-level perception bias is (1/6)/(1/3) = 0.5, which means that they have underestimated the size of the minority group by a factor of 0.5. Consequently, they overestimate the size of their own majority group in the entire network. In the heterophilic network (Fig. 1b), the perception bias of individual i is (4/6)/(1/3) = 2, implying that the size of the minority group is

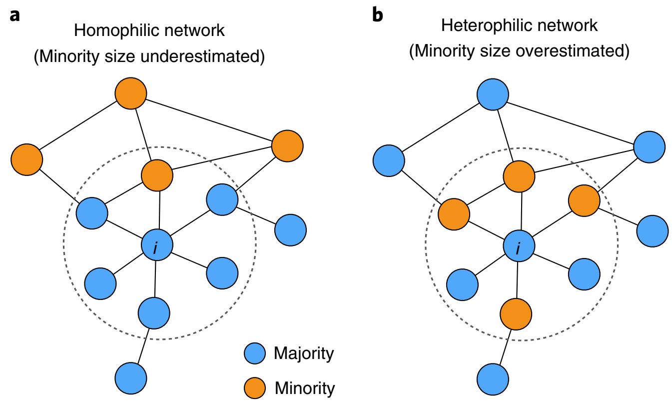

Fig. 1 | Individual- and group-level social perception bias. a,b, Individuals belong to one of two groups: the majority (blue) or the minority (orange). The minority fraction is 1/3 in both ($f_m \approx 0.33$) the (**a**) homophilic network and (**b**) heterophilic network. We studied social perception biases originating at both the individual and group level. On the individual level, individual i perceives the size of the minority group in the overall network based on their personal network, denoted by dashed circles. In the homophilic network, i perceives the size of the minority to be approximately 1/6 ≈ 16%, while in the heterophilic network, i perceives the size of the minority to be approximately 4/6 ≈ 67%. Therefore, in the homophilic network, individual i underestimates the minority-group size by a factor of 0.5 and in the heterophilic network i overestimates the minority-group size by a factor of 2 (see equation (1)). At the group level, the majority group perceives the size of the minority group to be (1/3 + 1/6 + 2/3)/8 = 7/48 ≈ 0.15 in the homophilic network, and (1/2 + 1/3 + 2/3 + 2/3 + 1 + 1)/8 = 25/48 ≈ 0.52 in the heterophilic network. Thus, the majority group underestimates the size of the minority group by a factor of 0.45 in the homophilic network (group-level perception bias $= \frac{0.15}{f_m} = \frac{0.15}{0.33} = 0.45$) and overestimates the minority-group size by a factor of 1.6 in the heterophilic network ($\frac{0.52}{0.33} = 1.6$; see equation (2)). In sum, depending on the structure of the network, individuals' and groups' perceptions about their own and other groups' sizes can be distorted.

overestimated by a factor of 2. At the group level, as per equation (2), the majority group (blue) perceives the size of the minority group to be 7/48 in the homophilic network and 25/48 in the heterophilic network. Therefore, the majority group underestimates the size of the minority group by a factor of (7/48)/(1/3) = 0.45 in the homophilic network and overestimates it by a factor of (25/48)/(1/3) = 1.6 in the heterophilic network.

Survey of social perception biases. To investigate the role of network structure in social perception biases, we conducted a survey with $n = 99$ participants from Germany, $n = 100$ from South Korea and $n = 101$ from the United States (see Methods). We asked questions about different societal attributes for which we knew the true frequencies in the general population from existing national surveys in these countries (for example, donating to charity, worship attendance and smoking; Supplementary Tables 1 and 2).

Participants answered three groups of questions. First, they answered questions about their own attributes (for example, whether they smoke). Second, they estimated the frequency of people with each attribute in their personal networks, defined as 'all adults you were in personal, face-to-face contact with at least twice this year'. We used these answers to calculate the homophily in their personal networks (see Methods). Homophily can vary from 0 (complete heterophily—for example, smoker interacts only with nonsmokers) to 1 (complete homophily—for example, smoker interacts only with smokers).

Third, participants estimated the frequency of people with a particular attribute in the general population of their country. We used these answers to calculate the participants' social perception biases as a ratio of their population estimates and the true population frequency of each attribute taken from national surveys (Supplementary Table 2). For example, if a participant estimated that 60% of the country's population smoke tobacco, whereas the national survey suggested that only 40% do so, that participant's perception bias was $60/40=1.5$ (equation (1)). We obtained group-level perception biases by averaging individual participants' perception biases for different levels of homophily (equation (2)).

For each country, we analysed individual- and group-level perception biases separately for attributes that, in that country, are objectively held by a small ($f_m<0.2$), medium ($0.2\leq f_m<0.4$) or large ($0.4\leq f_m<0.5$) minority group. For example, the attribute 'not having money for food' is held by less than 20%, or a small minority of the general population in all countries (Supplementary Table 2). The attribute 'worship attendance' is held by a small minority of the general population in Germany (10%), a mid-sized minority in South Korea (30%) and a large minority in the United States (41%; Supplementary Table 2).

Figure 2 shows individual and group perception bias for the size of the minority group, separately for participants who belonged to the minority and majority groups. Visually, patterns of perception biases differ for different sizes of the minority group in the overall population (f_m) and for different levels of homophily in personal networks (h). As minority-group size in the overall population decreased, its overestimation increased (compare red and yellow lines in Fig. 2a–f). Moreover, when homophily in personal networks was large ($h>0.5$), minority participants overestimated and majority participants underestimated the size of the minority, resembling false consensus (see right-hand side of Fig. 2a–f). In contrast, for low levels of homophily in personal networks ($h<0.5$), we observed a much smaller false consensus, or even a false uniqueness tendency, for both minority and majority participants (see left-hand sides of Fig. 2a–f). Similar visual patterns for perception biases, homophily and minority-group size could be observed in all three countries (compare the panels across the three rows in Fig. 2).

Generative network model with tunable homophily and group size. Our survey results suggest a correlation between the level of homophily and perception biases. To gain insights on how the structure of social networks (homophily and heterophily) and minority-group size influence perception biases, we developed a generative network model (see Methods) that allowed us to create heavy-tailed networks with tunable homophily and minority-group sizes.

In our model, nodes have a binary attribute (for example, smoker and nonsmoker, male and female). When the attributes are distributed unequally among the nodes, we call the smaller group the minority and the larger group the majority. Each newly added node creates links to existing nodes: the probability of an attachment of a new node w to an existing node v, denoted by ϕ_{wv}, is proportional to node v's degree (k_v) and the homophily between the two nodes (h_{wv})—that is, $\phi_{wv}\propto h_{wv}k_v$. The degree of the existing node and the homophily parameter regulate the probability of connection between nodes. Here homophily h_{wv} represents an intrinsic tendency of nodes having the same attribute to be connected, and its value ranges from 0 to 1. By assuming that all nodes having the same attribute behave similarly, we can study the model only in terms of $h_{\alpha\beta}$, with α,β being m for the minority or M for the majority. For example, h_{mm} represents the homophily between minority nodes and h_{MM} the homophily between majority nodes. We then consider two cases—that is, symmetric and asymmetric homophily. For the symmetric case, the tendency of nodes having the same attribute to be connected is the same for both groups. Thus, we need only one parameter, h, because $h_{mm}=h_{MM}=h$ (that is, $h_{mM}=h_{Mm}=1-h$). On the other hand, for the asymmetric case we need two homophily parameters, h_{mm} and h_{MM}, as these are different from each other. In the case of symmetric homophily, when $h<0.5$, nodes tend to connect to other nodes with the opposite attribute, whereas if $h>0.5$, nodes have a greater tendency to connect to nodes with the same attribute. In the case of the extremely homophilic situation, $h=1$, two separate communities of the same attributes will emerge.

The model we present here was partly inspired by the Bianconi–Barabási fitness model[29]. In that model, each node has an intrinsic fitness that is independent of other nodes and regulates nodes' attractiveness to other nodes, whereas nodes in our model have an intrinsic tendency to connect to other nodes, which is dependent on the attractiveness between a pair of nodes rather than an individual's characteristic. Considering the difference, this network model is a variation of the Barabási–Albert preferential attachment model (BA model) with the addition of a homophily parameter, h. Therefore, we call our model the BA-homophily model. .

Figure 3 depicts analytically derived perception biases of minority-group size among the members of minority and majority groups, as a function of the true fraction of the minority in the entire network and the homophily parameter. The perception biases in heterophilic networks ($0\leq h<0.5$) resemble false uniqueness. The minority underestimates its own size while the majority overestimates the size of the minority, the more so the smaller the minority group (smaller f_m). In homophilic networks ($0.5<h\leq 1$), perception biases resemble false consensus. The minority overestimates its own size (the more so the smaller the minority group), while the majority underestimates the size of the minority. Slight deviations between biases expected for minority and majority groups (see insets in Fig. 3) are due to the disproportionate number of links for the two groups, affecting the results of equations (6) and (7). Also note that, in the mean-field approximation, we assume that nodes with the same attribute behave similarly on average.

These analytical derivations can help us describe the functional form of the biases observed in the survey (Fig. 2). As shown in equation (11), the minority's perception bias (B^m_{group}) is proportional to the density of links between minority nodes (p_{mm}), which increases with the homophily between minority nodes—that is, h_{mm}. Similarly, the majority's perception bias (B^M_{group}) is proportional to the density of intergroup links (p_{mM}), which decreases as the homophily (h_{mm} and h_{MM}) increases. In addition, the sizes of minority and majority groups influence the growth rate of links for each group according to equation (12) so that perception biases can increase (or decrease) nonlinearly with group size (see Supplementary Methods 1). For instance, in the extreme homophily case with $h=1$, one gets $p_{mm}=p_{MM}=1$, while $p_{mM}=p_{Mm}=0$, leading to the minority's group-level perception bias of $1/f_m$. In sum, the proposed BA-homophily model and its analytical derivations facilitate systematic understanding of how network structure affects perception biases.

While we find general agreement between the survey results and our BA-homophily model, there are certain differences that call for more detailed investigation in the future. One main difference is that in the survey results (Fig. 2) we observed perception bias >1 in some cases when Fig. 3 predicts it to be <1. Specifically, this tends to happen for small minority-group sizes, when $h<0.5$ for minority and $h>0.5$ for majority participants. A possible explanation that is in line with previous studies in social cognition[30,31] is that people do not observe and report attribute frequencies in their samples (here, their personal networks) completely accurately, but with some random noise. When minority-group size is relatively large, errors of over- and underestimation can cancel out. For smaller minority-group sizes, the estimate cannot be lower than 0 so errors of overestimation could be larger than errors of underestimation and not cancel out[21]. Hence, people's estimates of the frequency of attributes in their samples could show overestimation for small minority groups, which is what we observed in the survey results.

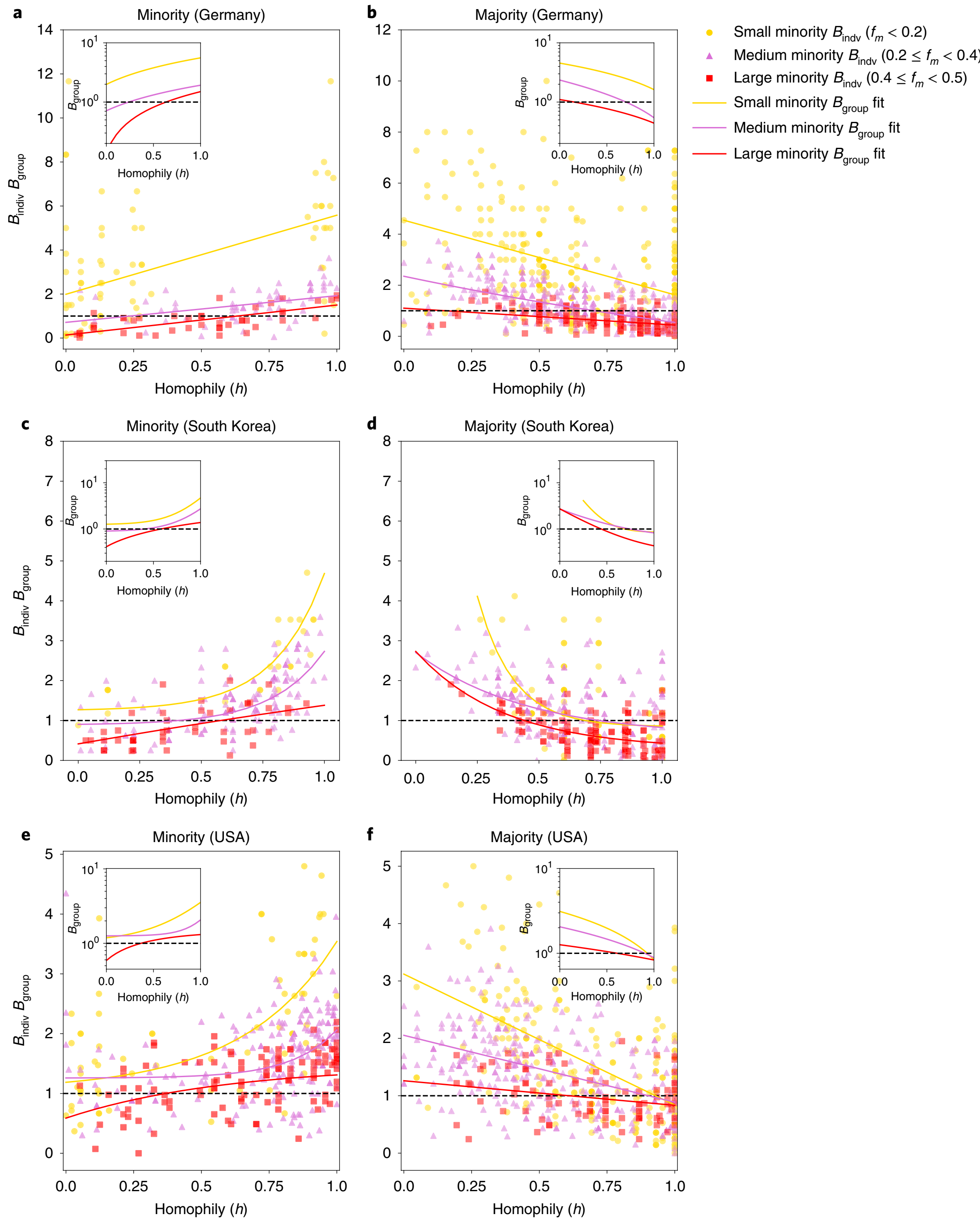

Fig. 2 | Survey results: bias in perception of minority-group size for participants whose personal networks exhibit different levels of homophily (h) and for attributes held by a small, medium or large minority group in a given country. a–f, Each row shows results from a different country: Germany (**a**,**b**, $n = 99$), South Korea (**c**,**d**, $n = 100$) and the United States (**e**,**f**, $n = 101$). Columns show perception biases of the minority (left) and the majority (right) group for each attribute. Different colours distinguish perception biases for attributes that in a given country are held by a small ($f_m < 0.2$; yellow), medium ($0.2 \le f_m < 0.4$; purple) or large ($0.4 \le f_m < 0.5$; red) minority group. The value of the individual perception bias indicates the accuracy with which each participant (each point in the plot) perceived the size of the minority group in the overall population. Group-level perception biases are calculated by averaging individual participants' perception biases for each homophily bin (0.02 increments), and they are denoted by fitted lines. A perception bias of 1 suggests perfect accuracy (horizontal line in each panel), values >1 indicate overestimation of the minority-group size and values <1 indicate underestimation of the minority-group size. The insets show fitted trends on a log scale for easier comparison of the sizes of underestimation and overestimation. These trends also approximate the results for the simple difference measure of perception biases. Homophily (h) is estimated from participants' reports about the frequency of people with each attribute in their personal networks (see Methods).

Social perception biases in real-world networks. The BA-homophily model offers a simple representation of real-world networks. To examine possible social perception biases in the real world, we studied six empirical networks with various ranges of homophily and minority-group sizes (see Methods). These networks have different structural characteristics and show different

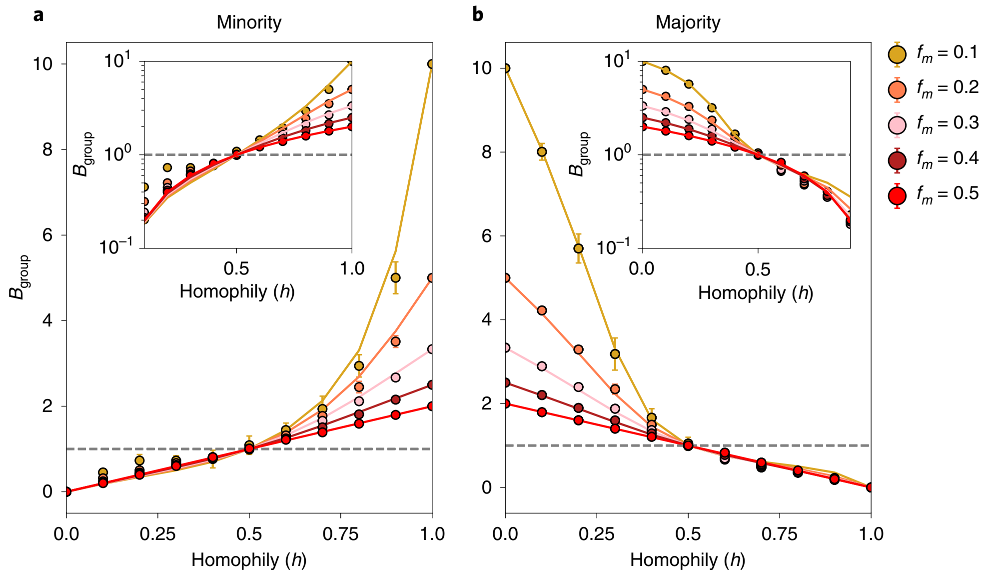

Fig. 3 | Network model results: bias in perception of minority-group size for the minority and majority groups, as a function of homophily (h) and the minority fraction (f_m) in the overall network. a,**b**, Minority (**a**) and majority (**b**) group sizes. Different colours denote networks with different minority fractions, f_m (yellow, $f_m=0.1$; orange, $f_m=0.2$; light pink, $f_m=0.3$; dark red, $f_m=0.4$; red, $f_m=0.5$). Group-level perception bias is analytically calculated based on equation (11). The dashed line in each panel indicates the point of no bias. The analytical results are displayed as solid lines and numerical results as circles. In the heterophilic networks ($0 \le h < 0.5$), the minority underestimates its own size while the majority overestimates the size of the minority, resembling false uniqueness. In homophilic networks ($0.5 < h \le 1$), the minority overestimates its own size while the majority underestimates the size of the minority, resembling false consensus. The insets show the same information on a log scale for comparison of underestimation and overestimation. The numerical estimations are averaged for 20 simulations using networks with $n=2{,}000$ nodes. Error bars are ±1 s.d.

Table 1 | Characteristics of the empirical networks.

Data	Number of nodes	Minority, n (%)	Majority (n)	Symmetric h	Asymmetric h (minority, majority)
Brazil	16,730	Sex sellers 6,624 (40%)	Sex buyers 10,106	0.0	0, 0
POK	29,341	Minority gender 12,868 (44%)	Majority gender 16,473	0.17	0.2, 0.17
USF51	6,200	Male 2,603 (42%)	Female 3,597	0.47	0.48, 0.47
GitHub	119,275	Female 6,730 (5.6%)	Male 112,545	0.53	0.69, 0.54
DBLP	280,200	Female 63,356 (22%)	Male 216,844	0.55	0.57, 0.56
APS	1,853	CMS 696 (37%)	QSM 1,157	0.92	0.9, 1.0

Each network contains nodes with binary attributes and has a minority and a majority group (see Methods for more details). The calculations of symmetric and asymmetric values of the homophily are based on the derivations described in Methods. The data can be found online at https://github.com/frbkrm/NtwPerceptionBias

levels of homophily or heterophily with respect to one specific attribute (Table 1). In five of the networks (Brazil, POK, USF51, GitHub, DBLP) this attribute is gender (female or male) while in one—the American Physical Society (APS) network[32]—the attribute is whether a paper belongs to the field of classical statistical mechanics (CSM) or quantum statistical mechanics (QSM).

To estimate homophily, we start by assuming that homophily is symmetric in all networks. The symmetric homophily has a linear relation to Newman's assortativity measure (q), which is the Pearson coefficient between attributes of connected nodes (for example, race[33]). This measure shows how assortative the network is with respect to a certain attribute. Positive assortativity means that two nodes with the same attribute are more likely to be connected compared to what would be expected from random connectivity. Negative assortativity means that two nodes with different attributes are more likely to be connected compared to what would be expected by chance. The Newman assortativity measure corresponds directly to homophily in our model when adjusted for the scale. In our model, $h=0$ means complete heterophily (negative assortativity, $q=-1$), $h=0.5$ indicates no relationship between structure and attributes (no assortativity, $q=0$) and $h=1$ indicates complete homophily (positive assortativity $q=1$; see Supplementary Fig. 1).

In reality, however, the tendency of groups to connect to other groups can be asymmetric[24], that is, different for the minority (h_{mm}) and the majority (h_{MM}). Given the relationship between the number of edges between nodes of the same group and homophily in equations (9) and (13), we can estimate the asymmetric homophily (see Methods). As we show below, it turns out that asymmetric homophily has an important impact on the predictability of perception bias in empirical networks.

We used the measured homophily and minority-group size in the empirical networks (Table 1) to generate synthetic networks with characteristics similar to those of the six empirical social networks. This enabled us to compare perception biases in empirical and synthetic networks and to gain predictive insights into the impact of

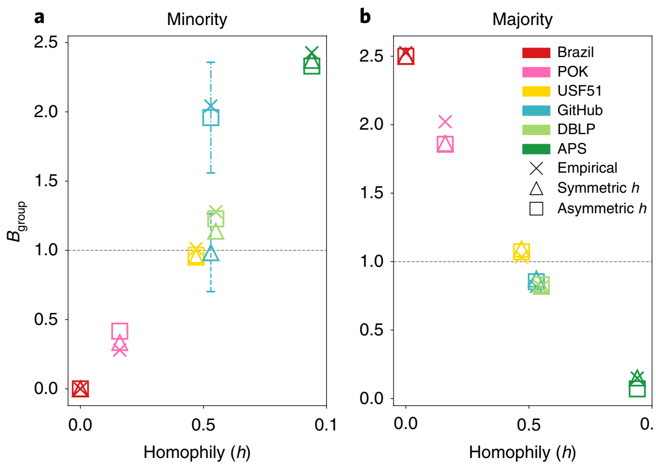

Fig. 4 | Numerical simulations: group-level social perception biases that could occur in six empirical social networks. a,b, Accuracy of estimation by (**a**) the minority group and (**b**) the majority group of the size of the minority group in real-world social networks with different levels of homophily. The symmetric homophily values of the empirical social networks are depicted on the *x* axis. Group-level perception bias of empirical networks is calculated as an average of an individual's perception bias. The horizontal line in each panel indicates the point of no bias. Homophily is measured between genders (female and male), except in the APS data, where homophily is measured between different academic fields—CSM and QSM. Empirical estimates of perception biases (crosses) can be visually compared to estimates from the BA-homophily model assuming symmetric (triangles) and asymmetric (squares) homophily. The synthetic networks were generated with $n=2{,}000$ nodes and averaged over 20 simulations. Error bars are ± 1 s.d. and are shown if they are larger than a marker size (see Supplementary Table 3 for more details).

homophily and minority-group size on possible individual- and group-level perception biases.

Figure 4 shows group-level perception biases in the empirical networks that could occur if people's perceptions were based solely on the samples of information from their personal networks. Visually these patterns are in agreement with the results obtained from the survey (Fig. 2), and from the synthetic networks (Fig. 3). Figure 4 suggests that in heterophilic networks, the minority group is expected to underestimate its own group size and the majority group is likely to overestimate the size of the minority. Conversely, in homophilic networks, the minority group is expected to overestimate its own size and the majority group is likely to underestimate the size of the minority. Because further cognitive or motivational processes could affect the final perceptions, these estimates can be taken as a baseline level of biases that could occur without any additional psychological assumptions.

We can compare perception biases estimated directly from empirical networks (crosses in Fig. 4) to those estimated from synthetic networks with similar symmetric homophily and minority-group size (triangles in Fig. 4). Although symmetric homophily traces empirically observed perception biases for most networks (as suggested by a close overlap between crosses and triangles in Fig. 4), it fails to capture the biases in the GitHub network for the minority group (see Supplementary Table 3). This network exhibits a higher level of asymmetric homophily compared to other networks (see far right column in Table 1). When perception biases are estimated from a synthetic model that assumes asymmetric homophily (squares in Fig. 4), they closely approximate perception biases of both the minority and the majority groups in all networks, including GitHub (see Supplementary Table 3). This suggests that asymmetric homophily plays an important role in shaping possible perception biases.

It is known that influential nodes in networks, usually identified by their high degree, can affect processes in networks such as opinion dynamics[34], social learning[35] and wisdom of crowds[36]. To evaluate the impact of degree on shaping perception biases, we plotted individual perception biases, B_{indv}, versus individual degrees in Supplementary Fig. 2. This figure suggests that the distribution of individual perception biases estimated from the BA-homophily model mostly corresponds to the empirically estimated distribution; however, the model does not explain all the variation observed in the empirical networks. Specifically, the model suggests that the minority members of the USF51 network should have lower perception biases than they really have, and vice versa for the majority members (see Supplementary Fig. 2). This could be due either to incomplete observations of all social contacts in real networks or to other processes that we did not consider in generating the model. However, the model can still predict the trend we observed in the empirical data, which would not be predicted assuming random connectivity among individuals (see Supplementary Fig. 3).

Reducing social perception biases. To what extent and under what structural conditions can individuals reduce their perception bias? To address this question, we considered perception biases of individuals and their neighbours. We aggregated each individual's own perception of frequency of different attributes (ego) with the averaged perceptions of the individual's neighbours[37]. For simplicity, we assumed symmetric homophily in the BA-homophily model (details on DeGroot's weighted belief formalization and the results for asymmetric homophily are in Supplementary Result 1).

Figure 5 compares the bias of individual perceptions to that of individual perceptions combined with perceptions of direct neighbours. The results show that taking into account the estimates of direct neighbours improves estimates of individuals in heterophilic networks (this can be observed visually on the left-hand side of the insets in Fig. 5, where blue triangles are closer to the grey dashed line than are orange circles). The reduction in perception biases is the result of individuals being more likely to be exposed to neighbours with opposing attributes. In homophilic networks, inclusion of neighbours' perceptions does not lead to a notable improvement (this can be observed visually on the right-hand side of the insets in Fig. 5, where blue triangles and orange circles follow similar trends), because individuals are exposed to neighbours with attributes similar to their own.

Taken together, our results suggest that in homophilic networks, individuals cannot improve their perception because their peers do not add sufficient new information that would increase the accuracy of their estimates. However, in heterophilic networks, individuals benefit from considering their neighbours' more diverse perceptions. While the overall trend is not surprising, our results reveal how the accuracy of these combined estimates changes as a function of homophily.

Discussion

The way in which people perceive their social networks influences their personal beliefs and behaviours and shapes their collective dynamics. Many studies have documented biases in these social perceptions, including both overestimation and underestimation of the size of minority groups. Here we investigated to what extent these seemingly contradictory biases can be explained merely by the structure of the social networks in which individuals are embedded, without assuming biased cognitive or motivational processes.

Using a survey of 300 participants in three different countries, analytical investigations of a simple network model with tunable homophily and minority-group size, and numerical simulations on a range of real-world networks, we show that structural properties

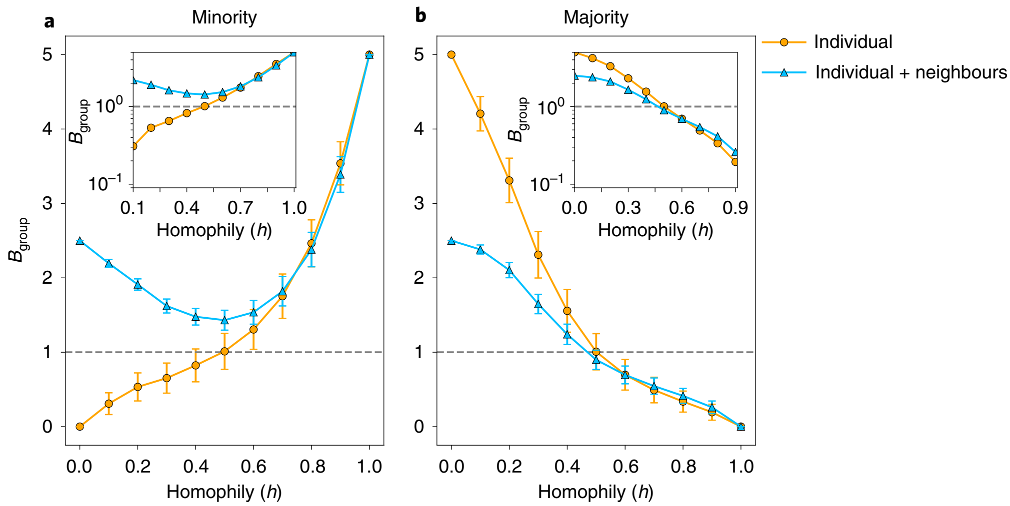

Fig. 5 | Social perception biases for individual nodes and for the weighted average of perceptions of individual nodes and their neighbours. a,b, Individual nodes and their neighbours for the minority (**a**) and majority (**b**) groups. Insets show the same results in log scale. Orange lines are calculated from equation (2) as a group-level bias. Blue lines show the perception bias of the weighted average of perceptions of individual nodes and their direct neighbours. The dashed lines indicate the point of no bias. We assume symmetric homophily, a minority fraction of 0.2 and networks with 2,000 nodes. Results, averaged over 50 simulations, show that perceptions of both minority and majority groups become slightly more accurate when taking into account their neighbours, but only in the heterophilic networks (in insets, blue triangles are closer than orange dots to the dashed line, denoting less bias). Error bars show ±1 s.d.

of personal networks strongly affect the samples people draw from the overall population. As a result, people's estimates of the frequency of different attributes in the general population are related to the level of homophily in their personal networks, asymmetries in the homophily in different groups and the minority-group size in the overall society.

The results show that biased samples alone can lead to apparently contradictory social perception biases such as false consensus and false uniqueness. While cognitive and motivational processes undoubtedly play an important role in the formation of social perceptions[27], our analyses establish a baseline level of biases that can occur without assuming biased information processing[21,38–41]. We find that predictions from our generative network model correspond well with empirical observations of perception biases in different social networks as well as self-reports from the survey study, suggesting that the model is not only theoretically interesting but also may describe actual human behaviour. Specifically, we find that when homophily is high, both minority and majority groups tend to overestimate their own size, whereas when homophily is low, both groups tend to underestimate their own size. Furthermore, the smaller the size of the minority group, the more its size is overestimated by both minority and majority groups.

Our study complements past results in the social psychology literature in several ways. It has been observed that minority groups tend to strongly overestimate, and majority groups slightly underestimate, their own frequency[18]. Our generative network model predicts when one can expect these, or different, patterns of overestimation and underestimation to be exhibited by minority and majority groups. While overestimation of small, and underestimation of large, frequencies can be expected when estimates are imperfectly correlated with the true population frequencies[30,31], our model goes further to explain how different patterns of biases occur not only because of the size of the minority group but also because of varying levels of, and asymmetries in, homophily. The fact that we found predicted relationships between homophily and perception biases in our survey data suggests that people rely on samples from their personal networks when making judgements about the overall population. Our study provides a quantitative elaboration of a mechanism, previously only verbally postulated, of selective exposure, showing that it might play an important role in the occurrence of social perception biases, over and above purely social projection effects or motivational biases[42].

Besides providing a theoretical account of perception biases, this work has practical implications for understanding real-world social phenomena. Given the importance of homophilic interactions in many aspects of social life, ranging from health-related behaviour[43] to group performance[44] and social identity[45], it is crucial to consider obstacles faced by both minorities and majorities when trying to form accurate social perceptions. Perceptions of the frequency of different beliefs and behaviours in the overall population influence people's beliefs about what is normal and shape their own aspirations[43,46,47]. When people overestimate the frequency of their own attributes in the overall population, they will be more likely to think that they are in line with social norms and, consequently, less likely to change. For example, we found that small minorities with high homophily are especially likely to overestimate their actual frequency in the overall network. If such committed minorities become resistant to change, they can eventually influence the whole network[48–50] and, when such minorities have erroneous views, the whole society could be worse off. Our results further suggest that a possible way to correct biases is to promote more communication with, and reliance on, neighbours' perceptions. However, this can be useful only in conjunction with promoting more diversity in people's personal networks. Promoting greater communication in homophilic networks does not improve perception biases.

This study is not without limitations. One strong assumption in our methodology is that one's perception is based solely on information sampled from one's personal network, or direct neighbourhood. In the real world, individuals can also rely on other sources such as news reports, polls and general education. In addition, we observe differences between the results of our survey and numerical simulations, indicating a need for future investigation of the impact on perceptions of minority-group size and heterogeneity of homophily at the individual level. Finally, this investigation did not include a quantitative specification of the cognitive processes underlying people's sampling from their personal networks. Such specifications[20,22,38,40] could be combined with the network model described here.

In sum, this study shows that both over- and underestimation of the frequency of one's own view can be explained by different levels of homophily, the asymmetric nature of homophily and the size of the minority group. Integration and quantification of the biases provide a rather comprehensive picture of the baseline level of human perception biases. We hope that this paper offers insights into measuring and reducing social perception biases, and fuels more work on understanding the impact of network structure on individual and group perceptions of our social worlds.

Methods

BA-homophily model. To gain insight into how network structure affects perception biases, we developed a network model that allowed us to create scale-free networks with tunable homophily and minority-group size[32]. This network model is a variation of the Barabási–Albert model, with the addition of the homophily parameter h. In this model, the probability that a newly introduced node w connects to an existing node v is denoted by ϕ_{wv}, and it is proportional to the product of the degree of node v, k_v, and the homophily between w and v as follows:

$$\phi_{wv} = \frac{h_{wv}k_v}{\sum_{v\in\{G\},v\neq w} h_{wv}k_v} \tag{3}$$

Here, h_{wv} is the probability of connection between nodes v and w. This is an intrinsic value that is dependent on the group membership of v and w. $\{G\}$ is a set of nodes in a graph G.

Before constructing the network, we specify two initial conditions: (1) the size of the minority group and (2) the homophily parameter that regulates the probability of a connection between minority and minority individuals, majority and majority individuals, minority and majority individuals, and majority and minority individuals. Each arrival node continues the link formation process until it finds λ nodes to connect to. If it fails to do so—for example, in an extreme homophily condition—the node remains in the network as an isolated node. The parameter λ guarantees the lower bound of degree and in our model is set to 2. Although this parameter is fixed for each node, the stochasticity of the model ensures the heterogeneity of the degree distribution.

Analytical derivation of group-level perception bias. In mean-field approximation, we estimate the group-level perception bias by behaviour of an average node in the group. In the case of the minority, let us denote the average number of links to other nodes of group m for an average node in group m as $\bar{l}_{mm}$. One can show that

$$\bar{l}_{mm} = \frac{2L_{mm}}{N_m} \tag{4}$$

where L_{mm} is total number of links between the minority nodes, and N_m is the number of nodes in the minority group m. The average degree of a node in group m is the sum of all degrees that nodes in group m have divided by the group size:

$$\bar{k}_m = \frac{K_m}{N_m} = \frac{2L_{mm} + L_{mM} + L_{Mm}}{N_m} \tag{5}$$

where K_m is the total number of degrees of the group. Thus the average perception of a minority node (about the frequency of the minority group) is proportional to the average number of links from a minority to minority $\bar{l}_{mm}$ divided by the average degree of a minority:

$$B^m_{\text{group}} = \frac{1}{f_m}\frac{\bar{l}_{mm}}{\bar{k}_m} = \frac{1}{f_m}\frac{2L_{mm}}{2L_{mm} + (L_{mM} + L_{Mm})} \tag{6}$$

Similarly, for group M,

$$B^M_{\text{group}} = \frac{1}{f_m}\frac{\bar{l}_{Mm} + \bar{l}_{mM}}{\bar{k}_M} = \frac{1}{f_m}\frac{L_{mM} + L_{Mm}}{2L_{MM} + (L_{mM} + L_{Mm})} \tag{7}$$

Here, L_{mm} is the number of edges between minority nodes and L_{MM} is the number of edges between majority nodes. Note that we distinguish the number of edges between the minority and majority L_{mM} and between the majority and minority L_{Mm}. These values are equivalent when homophily is symmetric, but they are unequal when homophily is asymmetric.

One can calculate the probability of inter- and intragroup links based on the growth mechanism of the model. Let us consider $K_m(t)$ and $K_M(t)$ as the total number of degrees for each group of the minority and the majority, respectively, at time t. At each time step, one node arrives and connects with λ existing nodes in the network. Therefore, the total degree of the growing network at time t is $K(t) = K_m(t) + K_M(t) = 2\lambda t$. In this model, the degree growth is linear for both groups. Denoting C as the minority's degree growth factor, we have

$$K_m(t) = C\lambda t, \quad K_M(t) = (2 - C)\lambda t \tag{8}$$

The probability of a connection between two minority nodes is the product of their degree and homophily:

$$p_{mm} = \frac{h_{mm}K_m(t)}{h_{mm}K_m(t) + h_{mM}K_M(t)} = \frac{h_{mm}C}{h_{mm}C + h_{mM}(2 - C)} \tag{9}$$

where h_{mm} is the homophily between minority nodes and $h_{mM} = 1 - h_{mm}$ is the tendency of minority nodes to be connected to majority nodes, or heterophily. The connection probability from a minority to a majority p_{mM} is the complement of p_{mm} as

$$p_{mM} = \frac{h_{mM}K_M(t)}{h_{mm}K_m(t) + h_{mM}K_M(t)} = \frac{h_{mM}(2 - C)}{h_{mm}C + h_{mM}(2 - C)} \tag{10}$$

Similar relationships can be found for the connection probability of majority to majority and majority to minority. Since L_{mm} and L_{MM} in equations (6) and (7) have a relation as a product of the total number of edges and the link probability, such as $L_{mm} = \lambda N_m p_{mm}$, we can reduce these two equations to the following equations:

$$\begin{aligned} B^m_{\text{group}} &= \frac{1}{f_m}\frac{2p_{mm}}{2p_{mm} + p_{mM} + (N_M/N_m)p_{Mm}} \\ B^M_{\text{group}} &= \frac{1}{f_m}\frac{(N_m/N_M)p_{mM} + p_{Mm}}{2p_{MM} + (N_m/N_M)p_{mM} + p_{Mm}} \end{aligned} \tag{11}$$

where N_m and N_M represent the number of nodes in each group. The analytical derivations are intuitive and well explained by the numerical results (solid lines in Fig. 3). For example, when $f_m = 0.5$ in extreme homophily ($h = 1.0$) with the degree growth $C = 1$ (a symmetric homophily condition), $B^m_{\text{group}} = 2$ from equation (11) matches well with the numerical result in Fig. 3a. Note that the growth parameter C is a polynomial function, and its relation to homophily is shown in Supplementary Method 1.

Measuring homophily in empirical networks. From the linear degree growth shown in Supplementary Equation (3), we can derive the relation between the degree growth C and the inter- and intralink probabilities p_{mm}, p_{mM} in equations (9) and (10). Thus,

$$C = f_m(1 + p_{mm}) + f_M p_{Mm} \tag{12}$$

In empirical networks we know the edge density for both the minority ($r_{mm} = L_{mm}/L$) and the majority ($r_{MM} = L_{MM}/L$) where L is the total number of links. Thus, the probability of connection within a group can be written as $r_{mm} = f_m p_{mm}$ and $r_{MM} = f_M p_{MM}$. From equation (9) and the relation between r_{mm} and p_{mm} (or r_{MM} and p_{MM}), we can derive the empirical homophily by using edge density, r_{mm}, r_{MM}, as follows:

$$\begin{aligned} h_{mm} &= \frac{r_{mm}(2-C)}{f_m C + 2r_{mm}(1-C)} \\ h_{MM} &= \frac{r_{MM}C}{f_M(2-C) - 2r_{MM}(1-C)} \end{aligned} \tag{13}$$

These calculations allow us to estimate the homophily from the empirical networks assuming that the BA-homophily model is a valid model of a social network. The homophily, by definition, can be either symmetric ($h_{mm} = h_{MM}$) or asymmetric ($h_{mm} \neq h_{MM}$).

Survey study. We conducted a survey on $n = 99$ participants from Germany, $n = 100$ from South Korea and $n = 101$ from the United States, from March to May 2018. The German and US participants were recruited from Amazon's Mechanical Turk crowdsourcing platform, and the South Korean survey was conducted through the survey platform Tillion Panel (see Supplementary Fig. 4 for demographic details of the samples). These sample sizes were considered to be sufficient based on the results of a previous publication using this method, where the trends of interest were detected with samples that were half this size[20]. The research was approved by the Federalwide Assurance Signatory Official of the Santa Fe Institute.

Participants were asked questions about their own attributes, the frequency of these attributes in their personal networks and their frequency in the general population of their country. Question texts and objective sizes of minority and majority groups in the general populations were taken from publicly available results of large national surveys conducted in each country. Details are provided in Supplementary Table 1. German and US participants were asked about ten attributes, and Korean participants about seven of those attributes for which we could find objective population data.

We estimated the homophily of participants' personal networks on the basis of their reports of the size of minority and majority groups in their social circles. Each participant reported the fraction of his or her personal network (or social circle) who have a specific attribute. For example, a participant who does not smoke might have estimated that 80% of their social circle is nonsmokers. We used this fraction to calculate the probability that any two nonsmokers in their social circle are connected. As a complementary relation of connection between attributes, we furthermore used the fraction of smokers in their social circle (20%) to calculate the probability that any nonsmoker and smoker are connected. These probabilities are equivalent to p_{mm} or p_{MM} in the BA-homophily model. Using equations (9) and (12) we can calculate the homophily h_{mm} (or h_{MM}) of each participant's personal network. In addition, we can evaluate h_{mM} and h_{Mm} using the relations $h_{mM} = 1 - h_{mm}$ and $h_{Mm} = 1 - h_{MM}$.

To study the effect of minority-group size, we analysed results separately for attributes for which minority-group size in a particular country was small ($f_m < 0.2$), medium ($0.2 \leq f_m < 0.4$) and large ($0.4 \leq f_m < 0.5$). For example, small-minority attributes in the United States are experienced theft, smoking and not having enough food, because the objective frequency of these attributes in the general US population is <0.2 (see Supplementary Table 2). We measured participants' individual perception bias by dividing their estimate of minority-group size in the general population by the objective minority-group size obtained from national surveys, according to equation (1).

Empirical networks. We investigate six empirical networks. The first network is a Brazilian network that captures sexual contact between sex workers and sex buyers[51]. The network consists of 16,730 nodes and 39,044 edges. There are 10,106 sex buyers and 6,624 sex sellers (minority-group size $f_m = 0.4$). In this network, no edges among members of the same group exist, resulting in Newman's assortativity ($q = -1$) and, consequently, the network is purely heterophilic ($h = 0$).

The second network is an online Swedish dating network from PussOKram.com (POK)[52]. This network contains 29,341 nodes with strong heterophily ($h = 0.17$, $q = -0.65$). Given the high bipartivity of the network, we are able to infer the group of nodes using the max-cut greedy algorithm. The results are in good agreement with the bipartivity reported in the literature[53]. We label the nodes based on their relative group size as minority gender and majority gender. Here, the fraction of the minority in the network is 0.44.

The third network is a Facebook network of a university in the United States (USF51)[54]. We removed the nodes without links. As a result, the network is composed of 6,200 nodes and includes information about individuals' gender. In this network male students are in the minority, occupying 42% of the network, and the network exhibits a weak heterophily[54] ($q = 0.06$, $h = 0.47$).

The fourth network is extracted from the collaborative programming environment GitHub. The network is a snapshot of the community (extracted 4 August 2015) that includes information about the first name and family name of the programmers. We used the first name and family name to infer the gender of the programmers[55]. After we removed ambiguous names and nodes having no links, the network consisted of 112,545 men and 6,730 women. Here, women belong to the minority group and represent only about 5.6% of the population. The network displays a moderately symmetric gender homophily of 0.53 ($q = 0.07$).

The fifth network depicts scientific collaborations in computer science and is extracted from Digital Bibliography and Library Project's website (DBLP)[56]. We used a method that combines names and images to infer the gender of the scientists with high accuracy[55]. We used a 4-year snapshot for the network. After we filtered out ambiguous names, the resulting network included 280,200 scientists and 750,601 edges (paper co-authorships), with 63,356 female scientists and 216,844 male scientists. This network shows a moderate level of symmetric homophily ($h = 0.55$ and $q = 0.1$).

The final network is a scientific citation network of the APS. Citation networks depict the extent of attention to communities in different scientific fields. We used the Physics and Astronomy Classification Scheme (PACS) identifier to select papers on the same topics. Here, we chose statistical physics, thermodynamics and nonlinear dynamical systems subfields (PACS = 05). Within a specific subfield, there are many subtopics that form communities of various size. To make the data comparable with our model we chose two relevant subtopics, namely CSM and QSM. The resulting network consists of 1,853 scientific papers and 3,627 citation links. Among nodes, 696 are in the minority and 1,157 in the majority. Here, the minority group in these two subtopics is CSM ($f_m = 0.37$). This network shows the highest homophily compared to the other empirical datasets ($h = 0.92$ and $q = 0.83$).

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The three empirical data (DBLP, GitHub, APS) can be found online at https://github.com/frbkrm/NtwPerceptionBias. The network data for Brazil can be found in the data description of the study[51] published in *PLoS Computational Biology* in 2011. POK can be found from the corresponding authors of the study[52] published in *Social Networks* in 2004, and USF51 can be found from the corresponding author of the study[54] in *Physica A*, 2011. The survey data can be obtained from the authors upon request.

Code availability

The Python scripts used for the generative model and empirical network analyses are available online at https://github.com/frbkrm/NtwPerceptionBias. Additional information about codes is available from the corresponding authors upon request.

Received: 5 November 2018; Accepted: 3 July 2019;
Published online: 12 August 2019

References

1. Cialdini, R. B. & Trost, M. R. Social influence: social norms, conformity and compliance. in *The Handbook of Social Psychology* (eds. Gilbert, D. T., Fiske, S. T., & Lindzey, G.) 151–192 (McGraw-Hill, 1998).
2. Bond, R. M. et al. A 61-million-person experiment in social influence and political mobilization. *Nature* **489**, 295–298 (2012).
3. Allcott, H. Social norms and energy conservation. *J. Public Econ.* **95**, 1082–1095 (2011).
4. Centola, D. The spread of behavior in an online social network experiment. *Science* **329**, 1194–1197 (2010).
5. Borsari, B. & Carey, K. B. Descriptive and injunctive norms in college drinking: a meta-analytic integration. *J. Stud. Alcohol* **64**, 331–341 (2003).
6. Botvin, G. J., Botvin, E. M., Baker, E., Dusenbury, L. & Goldberg, C. J. The false consensus effect: predicting adolescents' tobacco use from normative expectations. *Psychol. Rep.* **70**, 171–178 (1992).
7. Thompson, A. Journalists and Trump voters live in separate online bubbles. *VICE News* https://news.vice.com/en_us/article/d3xamx/journalists-and-trump-voters-live-in-separate-online-bubbles-mit-analysis-shows (8 December 2017).
8. Fields, J. M. & Schuman, H. Public beliefs about the beliefs of the public. *Public Opin. Q.* **40**, 427–448 (1976).
9. Ross, L., Greene, D. & House, P. The 'false consensus effect': an egocentric bias in social perception and attribution processes. *J. Exp. Soc. Psychol.* **13**, 279–301 (1977).
10. Mullen, B. et al. The false consensus effect: a meta-analysis of 115 hypothesis tests. *J. Exp. Soc. Psychol.* **21**, 262–283 (1985).
11. Krueger, J. & Clement, R. W. The truly false consensus effect: an ineradicable and egocentric bias in social perception. *J. Pers. Soc. Psychol.* **67**, 596–610 (1994).
12. Krueger, J. From social projection to social behaviour. *Eur. Rev. Soc. Psychol.* **18**, 1–35 (2007).
13. Mullen, B., Dovidio, J. F., Johnson, C. & Copper, C. In-group-out-group differences in social projection. *J. Exp. Soc. Psychol.* **28**, 422–440 (1992).
14. Suls, J. & Wan, C. K. In search of the false-uniqueness phenomenon: fear and estimates of social consensus. *J. Pers. Soc. Psychol.* **52**, 211–217 (1987).
15. Miller, D. T. & McFarland, C. Pluralistic ignorance: when similarity is interpreted as dissimilarity. *J. Pers. Soc. Psychol.* **53**, 298–305 (1987).
16. Prentice, D. & Miller, D. T. Pluralistic ignorance and alcohol use on campus: some consequences of misperceiving the social norm. *J. Pers. Soc. Psychol.* **64**, 243–256 (1993).
17. Lerman, K., Yan, X. & Wu, X.-Z. The 'majority illusion' in social networks. *PLoS One* **11**, e0147617 (2016).
18. Krueger, J. & Clement, R. W. Estimates of social consensus by majorities and minorities: the case for social projection. *Pers. Soc. Psychol. Rev.* **1**, 299–313 (1997).
19. Sherman, S. J., Presson, C. C., Chassin, L., Corty, E. & Olshavsky, R. The false consensus effect in estimates of smoking prevalence: underlying mechanisms. *Pers. Soc. Psychol. Bull.* **9**, 197–207 (1983).
20. Galesic, M., Olsson, H. & Rieskamp, J. A sampling model of social judgment. *Psychol. Rev.* **125**, 363 (2018).
21. Juslin, P., Winman, A. & Hansson, P. The naïve intuitive statistician: a naïve sampling model of intuitive confidence intervals. *Psychol. Rev.* **114**, 678–703 (2007).
22. Pachur, T., Hertwig, R. & Rieskamp, J. Intuitive judgments of social statistics: how exhaustive does sampling need to be? *J. Exp. Soc. Psychol.* **49**, 1059–1077 (2013).
23. McPherson, M., Smith-Lovin, L. & Cook, J. M. Birds of a feather: homophily in social networks. *Annu. Rev. Sociol.* **27**, 415–444 (2001).
24. Jadidi, M., Karimi, F., Lietz, H. & Wagner, C. Gender disparities in science? Dropout, productivity, collaborations and success of male and female computer scientists. *Adv. Complex Syst.* **21**, 1750011 (2018).
25. Miller, M. K., Wang, G., Kulkarni, S. R., Poor, H. V. & Osherson, D. N. Citizen forecasts of the 2008 U.S. presidential election. *Polit. Policy* **40**, 1019–1052 (2012).
26. Dawes, R. M. Statistical criteria for establishing a truly false consensus effect. *J. Exp. Soc. Psychol.* **25**, 1–17 (1989).
27. Marks, G. & Miller, N. Ten years of research on the false-consensus effect: an empirical and theoretical review. *Psychol. Bull.* **102**, 72 (1987).
28. Suls, J., Wan, C. K. & Sanders, G. S. False consensus and false uniqueness in estimating the prevalence of health-protective behaviors. *J. Appl. Soc. Psychol.* **18**, 66–79 (1988).

29. Bianconi, G. & Barabási, A.-L. Competition and multiscaling in evolving networks. *Europhys. Lett.* **54**, 436 (2001).
30. Fiedler, K. & Krueger, J. I. More than an artifact: regression as a theoretical construct. in *Social Judgment and Decision Making* 171–189 (Psychology Press, 2012).
31. Fiedler, K. & Unkelbach, C. Regressive judgment: implications of a universal property of the empirical world. *Curr. Dir. Psychol. Sci.* **23**, 361–367 (2014).
32. Karimi, F., Génois, M., Wagner, C., Singer, P. & Strohmaier, M. Homophily influences ranking of minorities in social networks. *Sci. Rep.* **8**, 11077 (2018).
33. Newman, M. E. Mixing patterns in networks. *Phys. Rev. E* **67**, 026126 (2003).
34. Aral, S. & Walker, D. Identifying influential and susceptible members of social networks. *Science* **337**, 337–41 (2012).
35. Golub, B. & Jackson, M. O. Naive learning in social networks and the wisdom of crowds. *Am. Econ. J. Microecon.* **2**, 112–149 (2010).
36. Becker, J., Brackbill, D. & Centola, D. Network dynamics of social influence in the wisdom of crowds. *Proc. Natl Acad. Sci. USA* **114**, E5070–E5076 (2017).
37. DeGroot, M. H. Reaching a consensus. *J. Am. Stat. Assoc.* **69**, 118–121 (1974).
38. Fiedler, K. Beware of samples! A cognitive-ecological sampling approach to judgment biases. *Psychol. Rev.* **107**, 659 (2000).
39. Gigerenzer, G., Fiedler, K. & Olsson, H. Rethinking cognitive biases as environmental consequences. in *Ecological Rationality: Intelligence in the World* 80–110 (Oxford Univ. Press, 2012).
40. Le Mens, G. & Denrell, J. Rational learning and information sampling: on the 'naivety' assumption in sampling explanations of judgment biases. *Psychol. Rev.* **118**, 379–392 (2011).
41. Denrell, J. & Le Mens, G. Information sampling, belief synchronization, and collective illusions. *Manag. Sci.* **63**, 528–547 (2016).
42. Krueger, J. On the perception of social consensus. in *Advances in Experimental Social Psychology* 163–240 (Academic Press, 1998).
43. Centola, D. An experimental study of homophily in the adoption of health behavior. *Science* **334**, 1269–1272 (2011).
44. Mollica, K. A., Gray, B. & Treviño, L. K. Racial homophily and its persistence in newcomers' social networks. *Organ. Sci.* **14**, 123–136 (2003).
45. Mehra, A., Kilduff, M. & Brass, D. J. At the margins: a distinctiveness approach to the social identity and social networks of underrepresented groups. *Acad. Manag. J.* **41**, 441–452 (1998).
46. Festinger, L. A theory of social comparison processes. *Hum. Relat.* **7**, 117–140 (1954).
47. Suls, J., Martin, R. & Wheeler, L. Social comparison: why, with whom, and with what effect? *Curr. Dir. Psychol. Sci.* **11**, 159–163 (2002).
48. Mobilia, M. Does a single zealot affect an infinite group of voters? *Phys. Rev. Lett.* **91**, 028701 (2003).
49. Mobilia, M., Petersen, A. & Redner, S. On the role of zealotry in the voter model. *J. Stat. Mech. Theory Exp.* **2007**, P08029 (2007).
50. Centola, D., Becker, J., Brackbill, D. & Baronchelli, A. Experimental evidence for tipping points in social convention. *Science* **360**, 1116–1119 (2018).
51. Rocha, L. E., Liljeros, F. & Holme, P. Simulated epidemics in an empirical spatiotemporal network of 50,185 sexual contacts. *PLoS Comput. Biol.* **7**, e1001109 (2011).
52. Holme, P., Edling, C. R. & Liljeros, F. Structure and time evolution of an internet dating community. *Soc. Netw.* **26**, 155–174 (2004).
53. Holme, P., Liljeros, F., Edling, C. R. & Kim, B. J. Network bipartivity. *Phys. Rev. E* **68**, 056107 (2003).
54. Traud, A. L., Mucha, P. J. & Porter, M. A. Social structure of facebook networks. *Physica A* **391**, 4165–4180 (2012).
55. Karimi, F., Wagner, C., Lemmerich, F., Jadidi, M. & Strohmaier, M. Interring gender from names on the web: a comparative evaluation of gender detection. in *Proceedings of WWW '16 Companion* 53–54 (International World Wide Web Conferences Steering Committee, 2016).
56. Collaboration networks from DataBase systems and Logic Programming (DBLP). http://dblp.uni-trier.de/ (accessed 30 September 2016).

Acknowledgements

We thank K. Winters, J. Kohne, P. Holme and H. Olsson for insightful conversations. We thank the Complex Systems Society's Bridge Fund and GESIS for funding E.L.'s research visit. E.L. was financially supported by a MURI grant to E. B. Falk from the Army Research Office (No. W911NF-18-1-0244), with additional financial support from grants to P. J. Mucha from the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health (No. R01HD075712) and the James S. McDonnell Foundation (grant No. 220020315). H.-H.J. acknowledges financial support from the Basic Science Research Program through an NRF grant funded by the Ministry of Education (No. NRF-2018R1D1A1A09081919). M.G. acknowledges financial support from National Science Foundation grants Nos. 1745154 and 1757211, and United States Department of Agriculture, National Institute of Food and Agriculture grant No. 2018-67023-27677. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

E.L. and F.K. conducted the analyses of synthetic and empirical networks and wrote the code. E.L. and M.G. conducted and analysed the surveys. E.L., F.K., C.W., H.-H.J., M.S. and M.G. conceived the project, developed the argument and wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41562-019-0677-4.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to E.L. or F.K.

Peer review information: Primary Handling Editor: Mary Elizabeth Sutherland.

nature research

Corresponding author(s): Eun Lee

Last updated by author(s): Jun 14, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	We used the web survey platform Unipark (www.unipark.de) to collect Germand and the U.S. survey data.
Data analysis	We used Python for all data analyses.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The six empirical data sets can be found online at https://github.com/frbkrm/NtwPerceptionBias. The survey data can be obtained from the authors upon request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Web-survey study in three countries.
Research sample	N=99 participants from Germany, N=100 from South Korea, and N=101 from the United States; adults of both genders. Participants in Germany and the United States were sampled from Amazon Mechanical Turk, and in South Korea from the survey platform Tillion Panel.
Sampling strategy	These were convenient samples.). These sample sizes were considered to be sufficient based on the results of a previous publication using this method, where the trends of interest were detected with samples that were half this size (Galesic, Olsson, & Rieskamp, 2018).
Data collection	Web-survey questionnaire.
Timing	From March to May 2018.
Data exclusions	No data were excluded.
Non-participation	None of the participants dropped out. The sample was convenient so there is no data about non-response.
Randomization	There was no randomization into experimental groups.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study	
☒	☐	Antibodies
☒	☐	Eukaryotic cell lines
☒	☐	Palaeontology
☒	☐	Animals and other organisms
☐	☒	Human research participants
☒	☐	Clinical data

Methods

n/a	Involved in the study	
☒	☐	ChIP-seq
☒	☐	Flow cytometry
☒	☐	MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	See above.
Recruitment	See above.
Ethics oversight	The research was approved by the Federalwide Assurance Signatory Official of the Santa Fe Institute.

Note that full information on the approval of the study protocol must also be provided in the manuscript.