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

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

[1]Department of Mathematics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
[2]Department of Computational Social Science, GESIS, Cologne 50667, Germany
[3]Institute for Web Science and Technologies, University of Koblenz-Landau, 55118 Koblenz, Germany
[4]Asia Pacific Center for Theoretical Physics, Pohang 37673, Republic of Korea
[5]Department of Physics, Pohang University of Science and Technology, Pohang 37673, Republic of Korea
[6]Department of Computer Science, Aalto University, Espoo FI-00076, Finland
[7]HumTec Institute, RWTH Aachen University, 52062 Aachen, Germany
[8]Santa Fe Institute, Santa Fe, NM 87501, USA
[9]Complexity Science Hub Vienna, 1080 Vienna, Austria
[+]These authors contributed equally to this work.
[*]Corresponding authors. Email: fariba.karimi@gesis.org, eunfeel@email.unc.edu

Abstract

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 analytically that these biases depend on the level of homophily and its asymmetric nature, as well as on the size of the minority group. We find that our model predictions correspond well with empirical data from a cross-cultural survey and with numerical calculations on six real-world networks. We also show under what circumstances individuals can reduce their biases by relying on perceptions of their neighbors. 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.

Introduction

People's perceptions of their social worlds determine their own personal aspirations[1] and willingness to engage in different behaviors, from voting[2] and energy conservation[3] to health behavior[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 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 U.S. 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 judgments 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 that individuals rely on 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 homophily, the tendency to be connected to similar others, a fundamental structural property of many social networks[23]. The second property is that homophily can be 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 we study is the relative size of minority and majority subgroups. 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 math, 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 (e.g., 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 behaviors by overestimating their frequency in society[28]. However, these processes cannot explain the opposite effect, underestimating the frequency of own view (e.g., 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 effects 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 further assumptions about biased motivational or cognitive processes. Results from a cross-cultural survey show that homophily and minority-group size 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 model predictions are consistent with biases that could occur in six empirical networks and point to the importance of asymmetric homophily. We also show when social perception biases can be reduced by aggregating one's own perceptions with the perceptions of one's neighbors. 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 (e.g., 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 superscript 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 neighborhoods). We define individual i's social perception bias as follows:

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

where Λ_i is the set of i's neighbors, $k_i = |\Lambda_i|$ 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.

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

$$B_{\text{group}} = \frac{1}{|N_g|}\sum_{i\in N_g} B_{\text{indv},i}, \tag{2}$$

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-level and individual-level perception biases is 0 and their maximum value is $1/f_m$ (see Method). A value below 1 indicates an underestimation of the minority-group size, and 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 color 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 her personal network (enclosed in a dashed circle). In the homophilic network (Fig. 1a), per Eq. 1 her individual-level perception bias is $(1/6)/(1/3) = 0.5$, which means that she underestimates the size of the minority group by a factor of 0.5. Consequently, she overestimates the size of her 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 she overestimates the size of the minority group by a factor of 2. At the group level, per Eq. 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 Method for details). We asked questions about different attributes (e.g., donating to charity, worship attendance, and smoking) that were taken from existing national surveys in each country (Table S1). These surveys provided true frequencies of different attributes in the general population of these countries (10 attributes in Germany and the United States, 7 in South Korea; Table S2).

Participants answered three groups of questions. First, they answered questions about their own attributes (e.g., whether they smoke or not). 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 Method for details). For example, if a participant was a smoker and 70% of her social contacts were smokers, the probability of a friendship link between this participant and a smoker is 70%. We used this information to estimate the homophily parameter for each individual's personal network using Eq. 9 and Eq. 13. Homophily can vary from 0 (complete heterophily; e.g., smoker only interacts with nonsmokers) to 1 (complete homophily; e.g., smoker only interacts with smokers).

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

For each country, we analyzed 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 (Table S2). 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%).

Fig. 2 shows the survey results. Perception bias for the size of the minority group is shown separately for participants who belonged to the minority (left column) and majority (right column) groups. The value of the individual perception bias indicates how accurately 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 increment) and they are shown by fitted lines. A perception bias of 1 means perfect accuracy, values above 1 indicate overestimation of the minority-group size, and values below 1 indicate underestimation of the minority-group size. We observed clear effects of the objective fraction of the minority group in the overall population (f_m) and of homophily of personal networks (h) on perception biases. As minority-group size in the overall population decreased, its overestimation increased. 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. 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. Similar relationships between perception biases, homophily, and minority-group size were observed in all three countries.

Generative network model with tunable homophily and group size

Our survey results revealed a clear correlation between homophily and perception biases. For example, we observed that smokers (the minority in all of our survey samples) who have many friends that do not smoke (i.e., have heterophilic personal networks) tend to underestimate the population of smokers while smokers who are surrounded by other smokers (i.e., have homophilic personal networks) tend to overestimate the size of the smoker population. We also observed that the extent to which smokers (the minority) overestimate the smoker population increases with minority-group size. 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 that allowed us to create scale-free networks with tunable homophily and minority-group sizes.

In our model, nodes have a binary attribute (e.g., 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, as $h_{mm} = h_{MM} = h$ (i.e., $h_{mM} = h_{Mm} = 1 - h$). On the other hand, for the asymmetric case we need two homophily parameters, h_{mm} and h_{MM}, as they 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 larger 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 that regulates nodes' attractiveness to other nodes. In our model, nodes have an intrinsic tendency to connect to other nodes, which depends on the attractiveness between a pair of nodes rather than an individual's characteristic. We call this intrinsic characteristic homophily. Mathematically speaking, 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". One property of the BA-homophily model is that it generates networks with the scale-free degree distributions observed in many large-scale social networks[30].

Figure 3 depicts analytically derived perception biases of minority-group size among the members of minority (Fig. 3a) and majority (Fig. 3b) groups, as a function of the true fraction of the minority in the entire network (f_m) and the homophily parameter (h). The solid lines show the analytic results (see Method) and the circles are numerical results obtained from the BA-homophily model. 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 Eq. 7 and Eq. 6. Also note that in the mean-field approximation we assume that nodes with the same attribute behave similarly on average.

These analytic derivations can help us describe the functional form of the biases observed in the survey (Fig. 2). As shown in Eq. 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 Eq. 12 so that perception biases can increase (or decrease) nonlinearly with group size (see the Supplementary Materials). 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 analytic 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 some 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 when $h > 0.5$ for majority participants. A possible explanation that is in line with previous studies in social cognition 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. However, for smaller minority-group sizes the estimate cannot be lower than 0, meaning that the former errors (overestimated - true sample frequency) could be larger than the latter errors (true - underestimated sample frequency) 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.

Social perception biases in real-world networks

The BA-homophily model offers a very 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 Method for detailed descriptions of the data sets and references). The network characteristics of the data sets are presented in Table 1. These empirical networks have different structural characteristics and show different levels of homophily or heterophily with respect to one specific attribute (see the Supplementary Materials). In five of the networks this attribute is gender (female or male), while in one — the American Physical Society (APS) network — the attribute is whether a paper belongs to the field of classical statistical mechanics or quantum statistical mechanics.

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 of correlation between attributes of connected nodes (e.g., race[31]). 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 the homophily parameter 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 the Supplementary Materials).

In reality, however, the tendency of groups to connect to other groups can be asymmetric[24]. Given the relationship between the number of edges between nodes of the same group and homophily in Eqs. 9 and 13, we can estimate the asymmetric homophily, which differs for the minority (h_{mm}) and the majority (h_{MM}; see Method). 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 similar characteristics 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 about the impact of 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. 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. The overall trends shown in Fig. 4 are in agreement with the results obtained from the survey and from the synthetic networks. In heterophilic networks, the minority group is likely to underestimate its own group size and the majority group is likely to overestimate the size of the minority. Conversely, in the homophilic networks, the minority group is likely to overestimate its own size and the majority group is likely to underestimate the size of the minority.

We can compare perception biases estimated directly from empirical networks (crosses in Fig. 4) with those estimated from synthetic networks with similar homophily and minority-group size. In Fig. 4, triangles correspond to networks with symmetric homophily and squares to networks with asymmetric homophily. Although symmetric homophily traces empirically observed perception biases in most instances, it fails to capture the biases in the GitHub network, especially for the minority group. This network exhibits a higher level of asymmetric homophily compared to other networks (see Table 1). When perception biases are estimated from a synthetic model that assumes asymmetric homophily, they approximate perception biases of both the minority and the majority groups very well in all networks. 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[32], social learning[33], and wisdom of crowds[34]. To evaluate the impact of degree on shaping perception biases, we plotted individual perception biases, B_{indv}, versus individual degree in Fig. S1. The distribution of individual perception biases estimated from the BA-homophily model mostly corresponds to the empirically estimated distribution. In addition, nodes with low degrees display a higher variation in perception biases compared to nodes with high degrees. The model does not explain all the variation observed in the empirical networks. This can be due 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.

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 neighbors. We aggregated each individual's own perception of frequency of different attributes (*ego*) with the averaged perceptions of the individual's neighbors[35]. 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 the Supplementary Materials).

Figure 5 shows a comparison of the average perception bias (B_{group}) for individuals who belong to the (a) minority and (b) majority and the weighted averages of their own perception biases and those of their direct neighbors. The minority-group size is fixed to 0.2. The results show that taking into account the estimates of direct neighbors improves estimates of individuals in heterophilic networks (on the log scale, blue triangles are closer to the gray dashed line than are orange circles). The reduction in perception biases is the result of individuals being more likely to be exposed to neighbors with opposing attributes. In homophilic networks, including neighbors' perceptions does not lead to a significant improvement because individuals are exposed to neighbors with similar attributes to their own.

Our results suggest that in homophilic networks, individuals cannot improve their perception, because their peers do not add enough new information that would increase the accuracy of their estimates. However, in heterophilic networks, individuals benefit from considering their neighbors' 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 people perceive their social networks influences their personal beliefs and behaviors 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 survey data, analytical investigation, and numerical simulations, we showed that structural properties of personal networks strongly affect the samples people draw from the overall population and their resulting social perception biases. First, our survey study revealed that people's estimates of frequency of different attributes in the general population are related to to the extent of homophily in their personal networks and to the minority-group size in the general population. Second, our simple network model with tunable homophily and minority-group size provided analytical insights about the relation between social perception biases and network properties. Third, our numerical calculations demonstrated applicability of the network model to empirical social networks.

The results of these three lines of investigation 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[21,36–39]. We find that predictions from our generative network model correspond well with empirical observations as well as survey results collected in three different countries, suggesting that the model is not only theoretically interesting but also might be actually relevant for explaining human behavior.

Our results suggest that homophily impacts the accuracy of the estimates of individuals in both minority and majority groups. 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. We further show that the relative sizes of the majority and minority groups influence social perception biases. Specifically, the smaller the true minority, the more its size is overestimated by both minority and majority groups. Finally, we show that perception biases can be reduced by aggregating individuals' perceptions with those of their direct neighbors, though only in heterophilic networks. In homophilic networks, these socially informed estimates do not lead to more accurate perceptions due to similarity of the nodes to their neighbors.

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[40,41], 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 judgments about the overall population. Our study provides a quantitative elaboration of a previously only verbally postulated mechanism 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 behavior[43] to group performance[44] and social identity[45], it is crucial to consider obstacles that are facing minorities and majorities in improving their social perceptions. Perceptions of the frequency of different beliefs and behaviors 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 neighbors' perceptions. However, this can be useful only in conjunction with promoting more diversity in people's personal networks. Promoting more 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 neighborhood. 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 of minority-group size and heterogeneity of homophily at the individual level on perceptions. Finally, this investigation did not include a quantitative specification of the cognitive processes underlying people's sampling from their personal networks. Such specifications[20,22,36,38] 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.

Method

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[51]. This network model is a variation of the Barabási–Albert model with the addition of 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 depends 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: (i) the size of the minority group and (ii) 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.

Analytic derivation for group-level perception bias

In mean-field approximation, we estimate the group-level perception bias by behavior 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 in-group 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 at time t, respectively. 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

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

Similar relationships can be found for the connection probability of majority to majority and majority to minority.

Since L_{mm} and L_{MM} in Eqs. 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 Eqs. 6 and 7 to the following equations:

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

where N_m and N_M represent the number of nodes in each group. The analytic 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 Eq. 11, and it 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 the Supplementary Materials.

Measuring homophily in empirical networks

From the linear degree growth shown in Eq. S3, we can derive the relation between the degree growth C and the inter- and intralink probabilities p_{mm}, p_{mM} in Eq. 9 and Eq. 10. Thus,

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

In empirical networks we know the edge density for the minority ($r_{mm} = L_{mm}/L$) and for 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 Eq. 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:

$$h_{mm} = \frac{r_{mm}(2-C)}{f_m C + 2r_{mm}(1-C)}, \quad h_{MM} = \frac{r_{MM}C}{f_M(2-C) - 2r_{MM}(1-C)}. \tag{13}$$

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 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. German and the U.S. participants were recruited from Amazon's Mechanical Turk crowdsourcing platform, and the South Korean survey was conducted through the survey platform Tillion Panel.

Regarding gender, 85.9% of the German participants, 50.0% of the Korean participants, and 61.4% of the U.S. participants were male. The age distribution in Germany was 18–30 years: 61.4%, 31–40 years: 26.3%, 41–50 years: 6.0%, 50+ years: 6.3%; in South Korea it was 18–30 years: 26.0%, 31–40 years: 26.0%, 41–50 years: 24.0%, 50+ years: 24.0%; and in the United States it was 18–30 years: 34.6%, 31–40 years: 36.6%, 41–50 years: 9.8%, 50+ years: 19.0%.

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 Table S1. German and U.S. participants were asked about 10 attributes and Korean participants about 7 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 in our survey 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 her social circle are nonsmokers. We used this fraction to calculate the probability that any two nonsmokers in her social circle are connected. As a complementary relation of connection between attributes, we furthermore used the fraction of smokers in her 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 Eqs. 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 analyzed 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 U.S. population is smaller than 0.2 (see Table S2). 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 Eq. 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[52]. 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 the 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)[53]. 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[54]. 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)[55]. 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[55] ($q = -0.06, h = 0.47$).

The fourth network is extracted from the collaborative programming environment GitHub. The network is a snapshot of the community (extracted August 4, 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[56]. 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 & Library Project's website (DBLP)[57]. We used a new method that combines names and images to infer the gender of the scientists with high accuracy[56]. 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 coauthorships) 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 last network is a scientific citation network of the American Physical Society (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 sizes. To make the data comparable with our model, we chose two relevant subtopics, namely, classical statistical mechanics (CSM) and quantum statistical mechanics. 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 data sets ($h = 0.92$ and $q = 0.83$).

References

1. Cialdini, R. B. & Trost, M. R. Social influence: Social norms, conformity and compliance. In *The Handbook of Social Psychology*, 151–192 (McGraw-Hill, New York, 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 yse from normative expectations. *Psychol. Rep.* **70**, 171–178 (1992).
7. Journalists and Trump voters live in separate online bubbles. `https://news.vice.com/story/journalists-and-trump-voters-live-in-separate-online-bubbles-mit-analysis-shows`. 2017-12-08.
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. *Politics 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. *EPL* **54**, 436 (2001).

30. Barabási, A.-L. Scale-free networks: a decade and beyond. *Science* **325**, 412–413 (2009).

31. Newman, M. E. Mixing patterns in networks. *Phys. Rev. E* **67**, 026126 (2003).

32. Aral, S. & Walker, D. Identifying influential and susceptible members of social networks. *Science* (2012).

33. Golub, B. & Jackson, M. O. Naive learning in social networks and the wisdom of crowds. *Am. Econ. J. Microecon.* **2**, 112–49 (2010).

34. Becker, J., Brackbill, D. & Centola, D. Network dynamics of social influence in the wisdom of crowds. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E5070–E5076 (2017).

35. DeGroot, M. H. Reaching a consensus. *J. Am. Stat. Assoc.* **69**, 118–121 (1974).

36. Fiedler, K. Beware of samples! a cognitive-ecological sampling approach to judgment biases. *Psychol. Rev.* **107**, 659 (2000).

37. Gigerenzer, G., Fiedler, K. & Olsson, H. Rethinking cognitive biases as environmental consequences. In *Ecological Rationality: Intelligence in the World*, 80–110 (Oxford University Press, Oxford, 2012).

38. Le Mens, G. & Denrell, J. Rational learning and information sampling: On the "naivety" assumption in sampling explanations of judgment biases. *Psychol. Rev.* **118**, 379 (2011).

39. Denrell, J. & Le Mens, G. Information sampling, belief synchronization, and collective illusions. *Manag. Sci.* **63**, 528–547 (2016).

40. Fiedler, K. & Krueger, J. I. More than an artifact: Regression as a theoretical construct. In *Social Judgment and Decision Making*, 171–189 (Psychology Press, New York, 2012).

41. Fiedler, K. & Unkelbach, C. Regressive judgment: Implications of a universal property of the empirical world. *Curr. Dir. Psychol. Sci.* **23**, 361–367 (2014).

42. Krueger, J. On the perception of social consensus. In *Advances in Experimental Social Psychology*, 163–240 (Academic Press, Cambridge, 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. Karimi, F., Génois, M., Wagner, C., Singer, P. & Strohmaier, M. Homophily influences ranking of minorities in social networks. *Sci. Rep.* **8** (2018).

52. 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).

53. Holme, P., Edling, C. R. & Liljeros, F. Structure and time evolution of an internet dating community. *Soc. Networks* **26**, 155–174 (2004).

54. Holme, P., Liljeros, F., Edling, C. R. & Kim, B. J. Network bipartivity. *Phys. Rev. E* **68**, 056107 (2003).

55. Traud, A. L., Mucha, P. J. & Porter, M. A. Social structure of facebook networks. *Phys. A.* **391**, 4165–4180 (2012).

56. Karimi, F., Wagner, C., Lemmerich, F., Jadidi, M. & Strohmaier, M. Inferring gender from names on the web: A comparative evaluation of gender detection methods. In *Proceedings of WWWÍ6 Companion*, 53–54 (International World Wide Web Conferences Steering Committee, 2016).

57. Dblp data. http://dblp.uni-trier.de/. Accessed: 2016-09-30.

Acknowledgments: We thank Kristi Winters, Julian Kohne, Petter Holme, and Henrik Olsson for insightful conversations. **Funding:** We thank the Complex Systems Society's Bridge Fund and GESIS for funding E. Lee's research visit. H.-H.J. acknowledges financial support from the Basic Science Research Program through an NRF grant funded by the Ministry of Education (NRF-2018R1D1A1A09081919). M.G. acknowledges financial support from National Science Foundation grants no. 1745154 and 1757211, and United States Department of Agriculture, National Institute of Food and Agriculture grant no. 2018-67023-27677. **Author contributions:** All authors conceived the project, developed the argument, and wrote the paper. E.L. and F.K. conducted the analyses of synthetic and empirical networks and wrote the code. E.L. and M.G. conducted and analyzed the surveys. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The six empirical data sets and the codes for implementing the analysis can be found online at https://github.com/frbkrm/NtwPerceptionBias. The survey data can be obtained from the authors upon request.

Figures and Tables

Data	**Number of nodes**	**Minority**	**Majority**	**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	CSM 696 (37%)	QSM 1,157	0.92	0.9, 1.0

Table 1. Characteristics of the empirical networks. Each network contains nodes with binary attributes and has a minority and a majority group (see Method for more details). The calculations of symmetric and asymmetric values of the homophily are based on the derivations described in the Method and Eq.13.

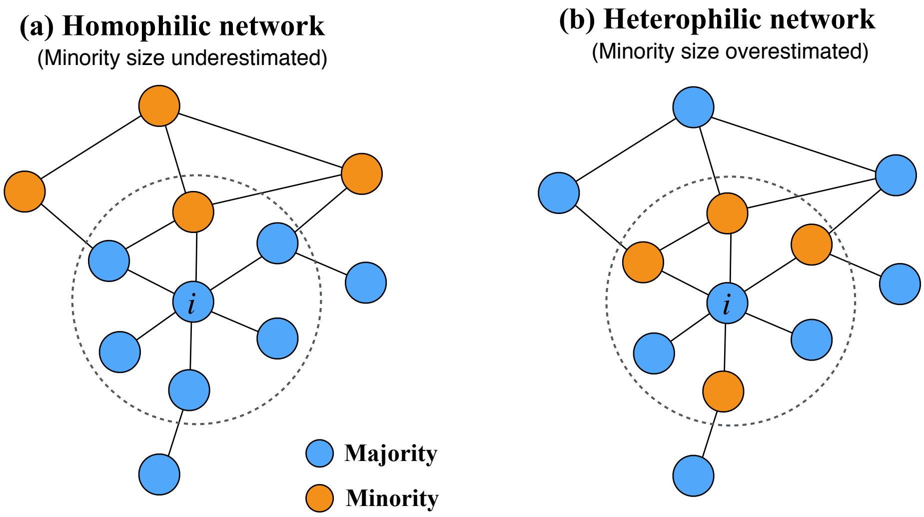

Fig. 1. Individual- and group-level social perception bias. Individuals belong to one of two groups: the majority (blue) or the minority (orange). The minority fraction is $1/3$ in both networks ($f_m \approx 0.33$). Panel (a) depicts a homophilic network and panel (b) shows a heterophilic network. We studied social perception biases originating on the individual and the group level. On the individual level, individual i perceives the size of the minority group in the overall network based on her personal network, denoted by dashed circles. In the homophilic network, i perceives the size of the minority to be approximately $1/6 \approx 16\%$, while in the heterophilic network, i perceives the size of the minority to be approximately $4/6 \approx 67\%$. 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 Eq. 1). On the group level, the majority group perceives the size of the minority group to be $(1/3+1/6+2/3)/8 = 7/48 \approx 0.15$ in the homophilic network and $(1/2+1/3+2/3+2/3+1+1)/8 = 25/48 \approx 0.52$ 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 Eq. 2). In sum, depending on the topological structure of the network, individuals' and groups' perceptions about their own and other groups' sizes can be distorted.

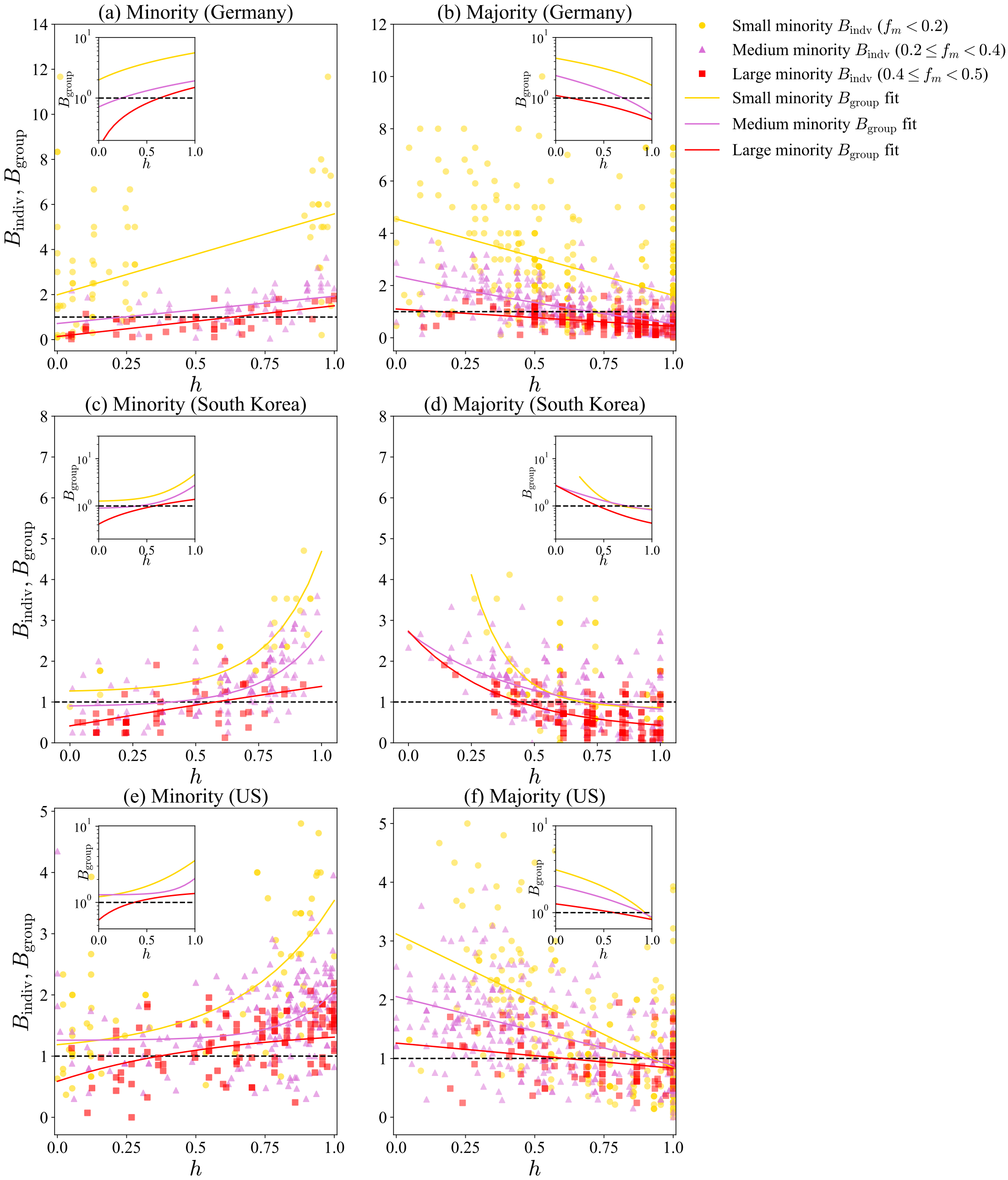

Fig. 2. Survey results: Bias in perception of minority-group size, for participants whose personal networks exhibit different levels of homophily (*h*), and for attributes held by a small, medium, or large minority group in a given country. Each row shows results from a different country: Germany (top), South Korea (middle), and United States (bottom). Columns show perception biases of the minority (left) and the majority (right) group for each attribute. Different colors distinguish perception biases for attributes that in a given country are 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. Each data point represents an individual participant's perception bias. Group-level perception bias is calculated as an average of individual's perception bias in each homophily bin (0.02 increment). The horizontal line in each panel indicates the point of no bias. The insets show fitted trends on a log scale to make the amount of underestimation and overestimation comparable. Homophily (h) is estimated from participants' reports about the minority-group fraction in their personal networks (see the Method).

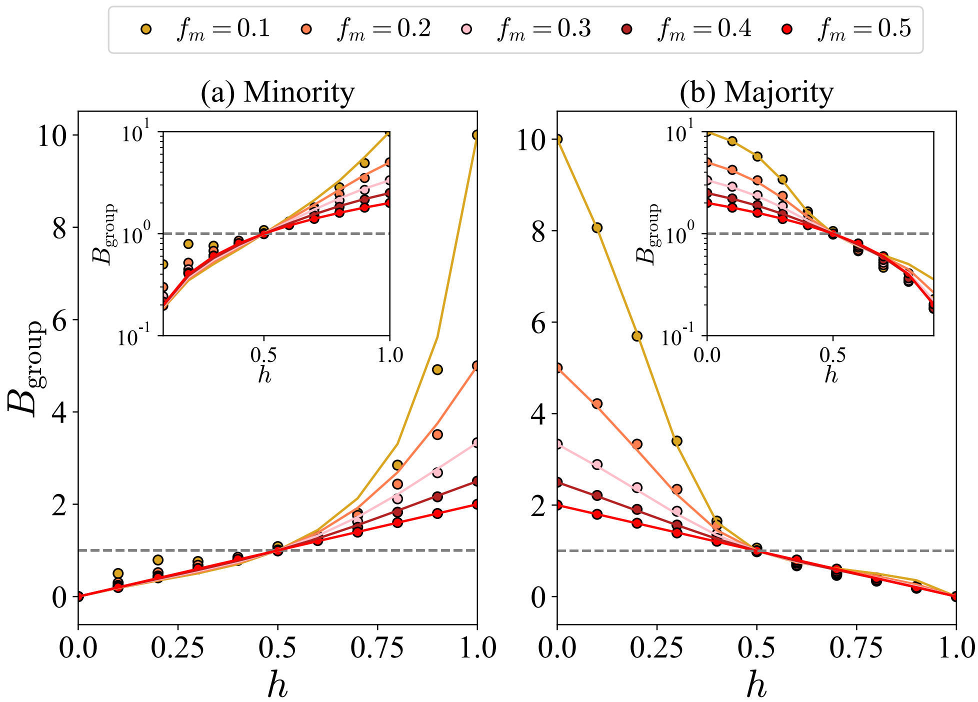

Fig. 3. Network model results: Bias in perception of minority-group size for (a) the minority group and (b) the majority group, as a function of homophily (h) and the minority fraction (f_m) in the overall network. Standard errors are not shown since they are smaller than the marker size. Different colors refer to networks with different minority fractions (f_m). Group-level perception bias is analytically calculated based on 11. The horizontal line in each panel indicates the point of no bias. The analytic results are displayed as solid lines and numerical results as circles. In the heterophilic networks ($0 \leq h < 0.5$), the minority (a) underestimates its own size, and the majority (b) overestimates the size of the minority, resembling false uniqueness. In homophilic networks ($0.5 < h \leq 1$), the minority (a) overestimates its own size and the majority (b) underestimates the size of the minority, resembling false consensus. The insets show the same information on a log scale to make the amount of underestimation and overestimation comparable. The numerical estimations are averaged for 20 runs with networks with $N = 2,000$ nodes. Standard error is not shown since it is smaller than the marker size.

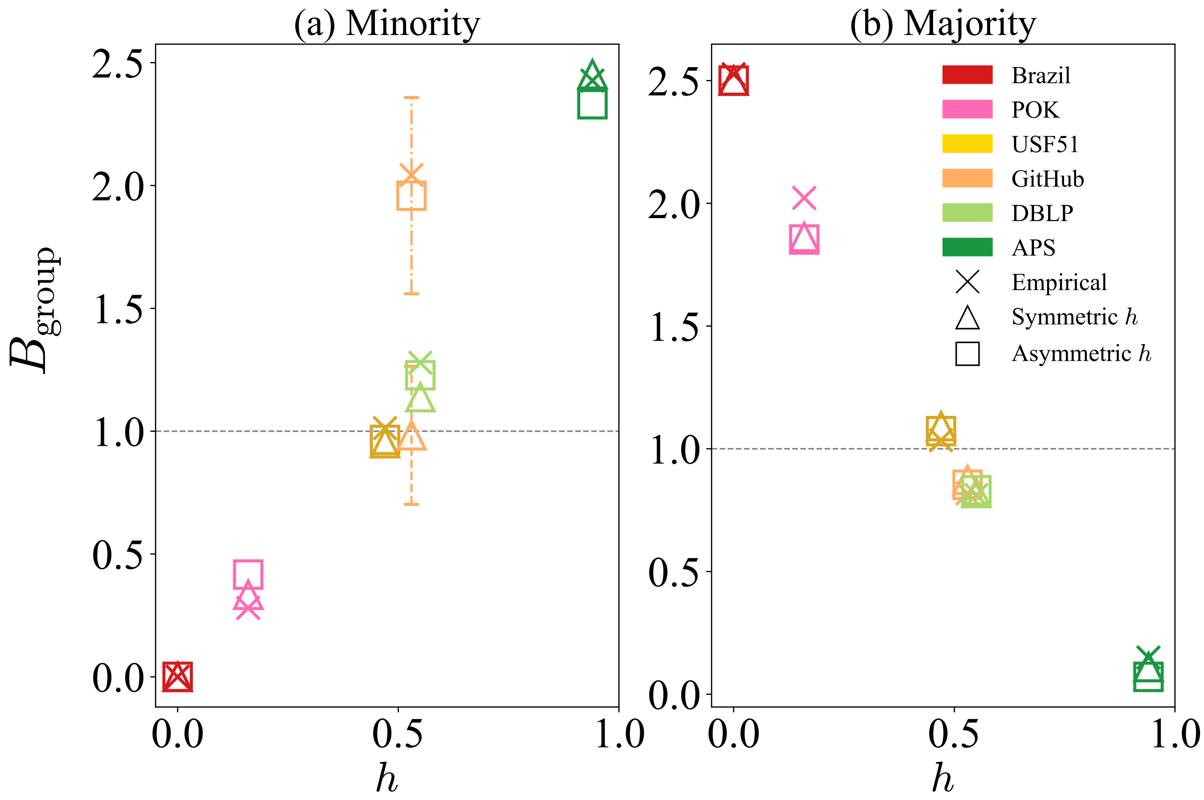

Fig. 4. Numerical simulations: Group-level social perception biases that could occur in six empirical social networks. The figure shows how accurately (a) the minority group and (b) the majority group might estimate the size of the minority group in real-world social networks with different levels of 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 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 for the American Physical Society (APS) data where homophily is measured between different academic fields: classical statistical mechanics and quantum statistical mechanics. Empirical estimates of perception biases (crosses) are compared with estimates from the BA-homophily model assuming *symmetric* (triangles) and *asymmetric* (squares) homophily. For both types of homophily, the perception bias of the minority group increases as homophily increases in a network, and that of the majority group decreases as the homophily increases. The results of the BA-homophily model with asymmetric homophily are in good agreement with the empirical estimates, highlighting the importance of considering asymmetric homophily. The results of the BA-homophily model assuming symmetric homophily predict the trend well except for networks with high asymmetric homophily. The synthetic networks were generated with $N = 2,000$ nodes and averaged over 20 simulations. Standard deviations are shown if they are larger than a marker size.

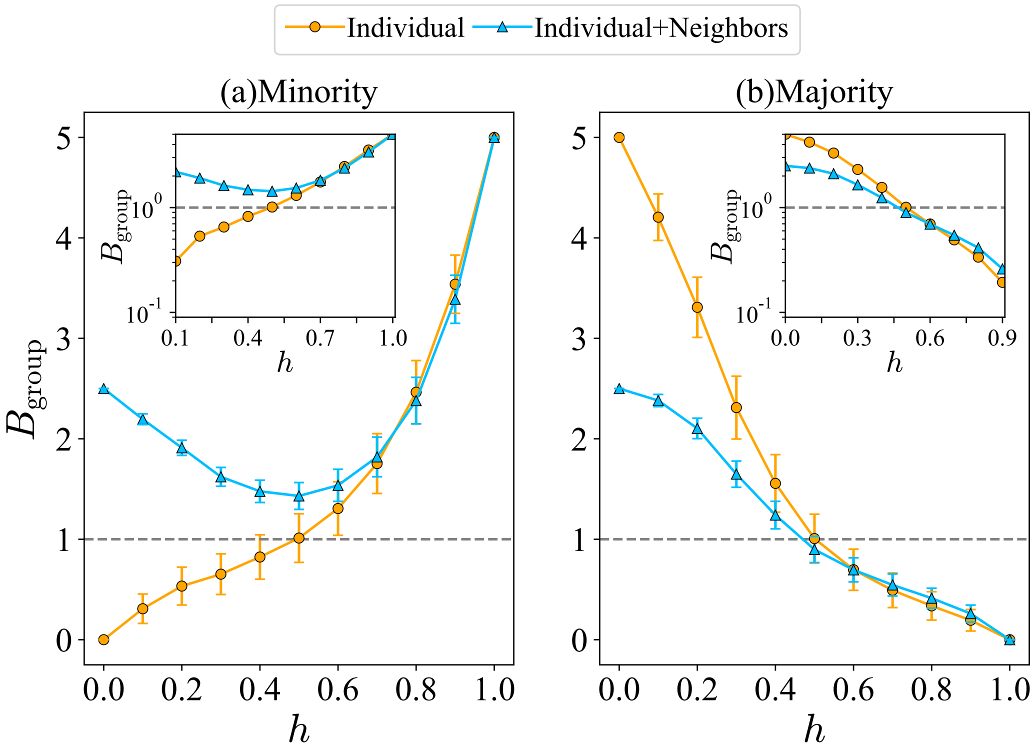

Fig. 5. Social perception biases for individual nodes and for the weighted average of perceptions of individual nodes and their neighbors. Insets show the same results in log scale. Orange lines are calculated from Eq. 2 as a group-level bias. Blue lines show the perception bias of the weighted average of perceptions of individual nodes and their direct neighbors. The horizontal line in each panel indicates the point of no bias. We assume symmetric homophily, minority fraction of 0.2, and networks with $2,000$ nodes. Results, averaged over 50 runs, show that perceptions of both minority and majority groups become slightly more accurate when taking into account their neighbors, but only in the heterophilic networks (in insets, blue triangles are closer than orange dots to the gray dashed line denoting less bias). Error bars show standard deviations.

Supplementary Materials

1. Survey questions and general population data

Characteristic	Question text	Source of question text and true population frequencies		
		Germany	**S. Korea**	**U.S.**
1. Not having money for food	Have there been times in the past 12 months when you did not have enough money to buy food that you or your family needed? (a)Yes–(b)No	Gallup World Poll (2017)	Gallup World Poll (2017)	Gallup World Poll (2017)
2. Donating to charity	In the past month, have you donated money to a charity? (a)Yes–(b)No	Gallup World Poll (2018)	Gallup World Poll (2017)	Gallup World Poll (2017)
3. Experiencing theft	Within the past 12 months, have you had money or property stolen from you or another household member? (a)Yes–(b)No	Gallup World Poll (2017)	–	Gallup World Poll (2017)
4. Belief in a god	Do you believe in god or a supreme being? (a)Yes–(b)No	Ipsos (2010)	Ipsos (2010)	Ipsos (2010)
5. God and morality	Which one of these comes closer to your opinion? (a) It is not necessary to believe in God in order to be moral and have good values. (b) It is necessary to believe in God in order to be moral and have good values.	Pew (2011)	–	Pew (2011)
6. Worship attendance	Have you attended a place of worship or a religious service within the past 7 days? (a)Yes–(b)No	Pew (2010)	Pew (2013)	Pew (2012)
7. Religion importance	Is religion an important part of your daily life? (a)Yes–(b)No	Gallup World Poll (2015)	Gallup World Poll (2015)	Gallup World Poll (2015)
8. Smoking	These days, are you smoking any tobacco product at least once a day? (Tobacco smoking includes cigarettes, cigars, pipes, and any other form of smoked tobacco). (a)Yes–(b)No	World Health Organization (2010)	National Nutrition Survey (2016)	World Health Organization (2010)
9. Military force	Do you agree that it is sometimes necessary to use military force to maintain order in the world? (a)Yes–(b)No	Pew (2011)	–	Pew (2011)
10. Homosexuality acceptance	Which one of these comes closer to your opinion? (a) Homosexuality is a way of life that should be accepted by society. (b) Homosexuality is a way of life that should not be accepted by society.	Pew (2013)	Pew (2013)	Pew (2013)

Table S1. Survey questions about different attributes. Texts of questions asking about respondents' own attributes in our surveys in Germany, South Korea, and the United States, and the original sources of questions and true population data. In addition to questions about their own attributes listed above, participants answered questions about the frequency of each attribute in their personal networks and in the general population of their country: "When asked… [question text here] … what percentage of [your social contacts/of adults living in [country]] would answer [for questions 6 and 10, '(b)'; for all other questions, '(a)']?"

Attribute	Germany (%)		S. Korea (%)		U.S. (%)	
	(a)	(b)	(a)	(b)	(a)	(b)
1. Not having money for food	6	94	17	83	19	81
2. Donating to charity	44	56	40	60	61	39
3. Experiencing theft	10	89	–	–	14	86
4. Belief in a god	32	68	25	75	75	25
5. God and morality	66	33	–	–	46	53
6. Worship attendance	10	90	30	70	41	57
7. Religion importance	39	60	42	57	66	34
8. Smoking	22	78	24	76	15	85
9. Military force	50	50	–	–	75	23
10. Homosexuality acceptance	87	11	39	59	60	33

Table S2. True population frequencies for each attribute. Minority- and majority-group sizes for each attribute in the general populations of Germany, South Korea and the United States. Response options are the same as answers (a) and (b) in Table S1 except for Belief in a god, for which the general population survey offered several options: definitely believe in god(s) or supreme being(s) [categorized as (a) in our analyses] - sometimes I believe but sometimes I don't - not sure - don't believe [categorized together as (b)]. The percentages sometimes do not sum to 100 because some participants did not answer the question.

2. Distribution of individual-level perception biases as a function of degree

Figure S1 shows possible individual-level perception biases by degree in six different empirical networks. The results are compared with predictions of the BA-homophily model with similar homophily and minority size. We observe more heterogeneity in perception biases of individuals in empirical networks compared to the model, especially for Facebook data. Overall, the model predicts the empirically estimated trends well.

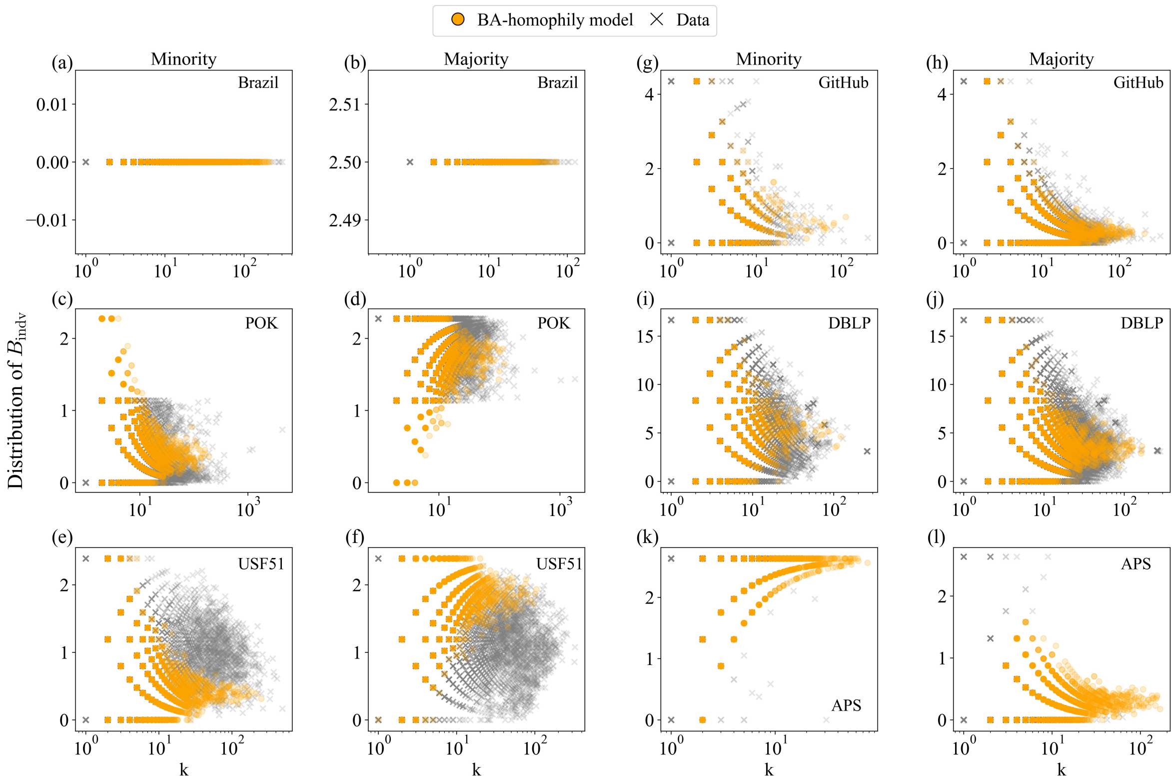

Fig. S1. Distribution of possible individual-level perception biases (B_{indv}) as a function of individual's degree (k). Each row represents one empirical network. The left two columns show the results for the heterophilic empirical networks [Brazilian sexual contact network (Brazil), Swedish online dating network (POK), Facebook network of an university (USF51)] and the right two columns for the homophilic empirical networks [GitHub developers' network (GitHub), DBLP developers' network (DBLP), American Physical Society network (APS)]. The gray crosses represent the possible perception bias of each individual estimated from the empirical network, and orange circles show the perception bias predicted by the BA-homophily model. The simulation results are aggregated over 50 iterations and the network size is $N = 2,000$. The x axis is shown in log scale.

3. Growth rate (C) in BA-homophily model

On the basis of the derivation provided by Karimi et al.[51], we can derive the exact degree dynamics of the BA-homophily model. Let us assume $K_m(t)$ and $K_M(t)$ as the sum of the degrees of each group m and M, respectively. With the number of links a node has, the total number of links at each time step will be $K(t) = K_m(t) + K_M(t) = 2\lambda t$. One can also describe the evolution of each group's degree as

$$\begin{cases} K_m(t+\Delta t) = K_m(t) + \lambda\Big(f_m\big(1 + \dfrac{h_{mm}K_m(t)}{h_{mm}K_m(t) + h_{mM}K_M(t)}\big) + f_M \dfrac{h_{Mm}K_m(t)}{h_{MM}K_M(t) + h_{Mm}K_m(t)}\Big)\Delta t, \\ K_M(t+\Delta t) = K_M(t) + \lambda\Big(f_M\big(1 + \dfrac{h_{MM}K_M(t)}{h_{MM}K_M(t) + h_{Mm}K_m(t)}\big) + f_m \dfrac{h_{mM}K_M(t)}{h_{mm}K_m(t) + h_{mM}K_M(t)}\Big)\Delta t. \end{cases} \tag{S1}$$

Here, one can separate the amount of increase of the degree for each group with the limit $\Delta t \to 0$,

$$\begin{cases} \dfrac{dK_m}{dt} = \lambda\Big(f_m\big(1 + \dfrac{h_{mm}K_m(t)}{h_{mm}K_m(t) + h_{mM}K_M(t)}\big) + f_M \dfrac{h_{Mm}K_m(t)}{h_{MM}K_M(t) + h_{Mm}K_m(t)}\Big), \\ \dfrac{dK_M}{dt} = \lambda\Big(f_M\big(1 + \dfrac{h_{MM}K_M(t)}{h_{MM}K_M(t) + h_{Mm}K_m(t)}\big) + f_m \dfrac{h_{mM}K_M(t)}{h_{mm}K_m(t) + h_{mM}K_M(t)}\Big). \end{cases} \tag{S2}$$

We can assume that $K_m(t)$ [$K_M(t)$] grows as a linear function of time. Given this assumption, we can write that $K_m(t) = C\lambda t$ [$K_M(t) = (2-C)\lambda t$] based on $K(t) = 2\lambda t$.

$$\begin{cases} \dfrac{dK_m}{dt} = C\lambda = \lambda\Big(f_m\big(1 + \dfrac{h_{mm}Ct}{h_{mm}Ct + h_{mM}(2-C)t}\big) + f_M \dfrac{h_{Mm}Ct}{h_{MM}(2-C)t + h_{Mm}Ct}\Big), \\ \dfrac{dK_M}{dt} = (2-C)\lambda = \lambda\Big(f_M\big(1 + \dfrac{h_{MM}(2-C)t}{h_{MM}(2-C)t + h_{Mm}Ct}\big) + f_m \dfrac{h_{mM}(2-C)t}{h_{mm}Ct + h_{mM}(2-C)t}\Big). \end{cases} \tag{S3}$$

Then, we can derive the relation of C with group sizes and edge density in a group (Eq. 12) from Eq. S2 by using p_{mm} (Eq. 9) and p_{mM} (Eq. 10).

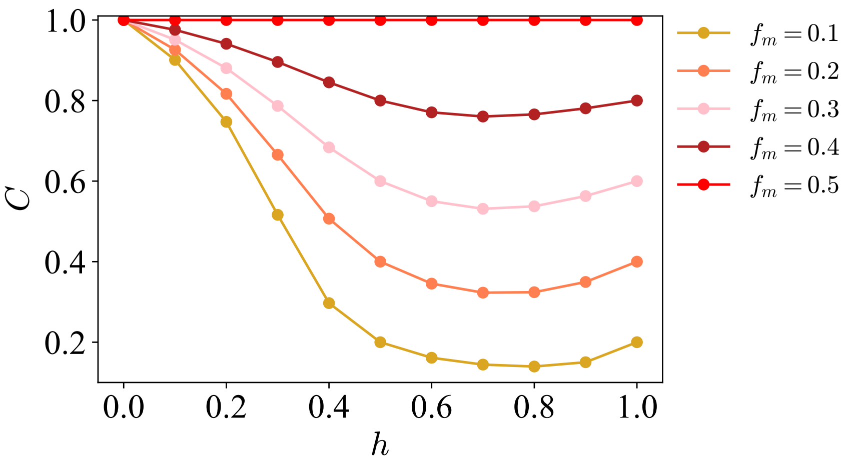

Fig. S2. The relation between the minority's degree growth rate (C) and the symmetric homophily (h). As the minority group's size f_m decreases, the degree growth rate of the minority drastically decreases with increasing symmetric homophily h ($h_{mm} = h_{MM}$). C is a function of h and f_m as described in Eq. 12.

4. Relationship between symmetric homophily (h) and Newman's assortativity (q)

The symmetric homophily is equivalent to Newman's assortativity measure (q)[31]. The latter corresponds directly to the homophily parameter in the BA-homophily model after adjusting for scale in a relation $q = 2h - 1$. In the BA-homophily model, $h = 0$ means complete heterophily ($q = -1$), $h = 0.5$ indicates no relationship between network structure and attributes ($q = 0$), and $h = 1$ indicates complete homophily ($q = 1$). The relationship deviates slightly from the linear function for small minority-group sizes but is independent of minority sizes when $h \approx 0.5$.

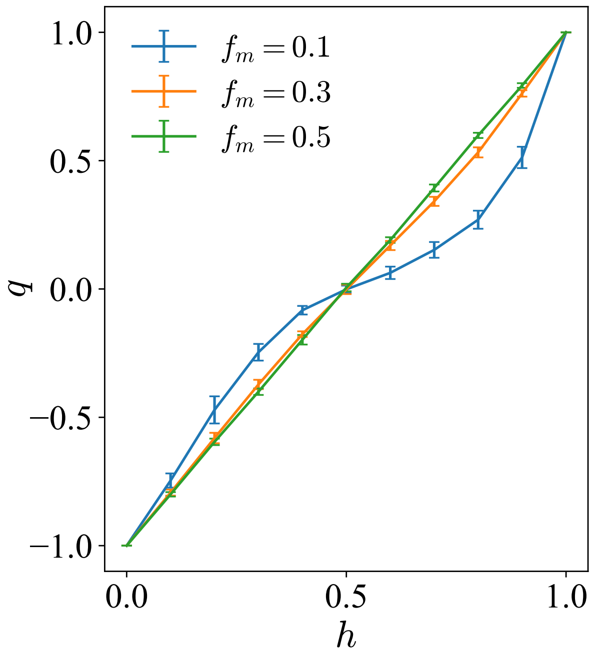

Fig. S3. Relationship between Newman's assortativity (q) and symmetric homophily (h) in the BA-homophily model, for different sizes of minority group (f_m). Newman's assortativity is proportional to h scaled as $2h - 1$.

5. Group-level perception bias aggregated with those of their neighbors, assuming asymmetric homophily

Here, we investigate to what extent and under what structural conditions individuals can reduce their perception bias for the prevalence of an attribute in a population by asking their friends about their perceptions and integrating those perceptions with their own when homophily is asymmetric. We build on DeGroot's weighted belief formalization by aggregating an individual's perception (*ego*) with the averaged perceptions of the individual's direct neighbors[35]. In this formalization, individual i with a perception bias of $B_{\text{indv},i}$ asks each neighbor j about her perception bias of that attribute frequency $B_{\text{indv},j}$. The final perception bias of individual i is the weighted average of her and her neighbors' bias: $\hat{B}_{\text{indv},i} = \frac{B_{\text{indv},i} + (\sum_{j\in\Lambda_i} B_{\text{indv},j})/k_i}{2}$, where Λ_i is the set of i's neighbors and $k_i = |\Lambda_i|$.

In the main text we assumed symmetric homophily, while here we assume asymmetric homophily in the BA-homophily model. Figure S4 shows a comparison of the group perception (B_{group}) of individuals who are in (a) the minority and (b) the majority, with the bias of their perceptions aggregated with those of their neighbors. The minority-group size is fixed to 0.1 in the top row, 0.3 in the middle row, and 0.5 in the bottom row. The homophily for the minority group h_{mm} is fixed to $0.1, 0.5, 0.9$ (depicted by lines of different colors), while homophily for the majority group h_{MM} ranges from 0 to 1 along the horizontal axis.

Results for asymmetric homophily (Fig. S4) are generally similar to those for symmetric homophily (Fig. 5): accounting for the opinion of neighbors can decrease perception bias when networks are heterogeneous. For the majority, aggregating their own perceptions with those of their neighbors leads to a robust improvement in perception accuracy, as described in Fig. S4b, d, and f. For the minority, accounting for 1-hop neighbors also helps decrease the bias, though less than for the majority. However, when minority group is small and homophily is highly asymmetric ($h_{mm} = 0.5$ and $h_{MM} < 0.5$), accounting for neighbors can indeed increase the bias see Fig. S4a).

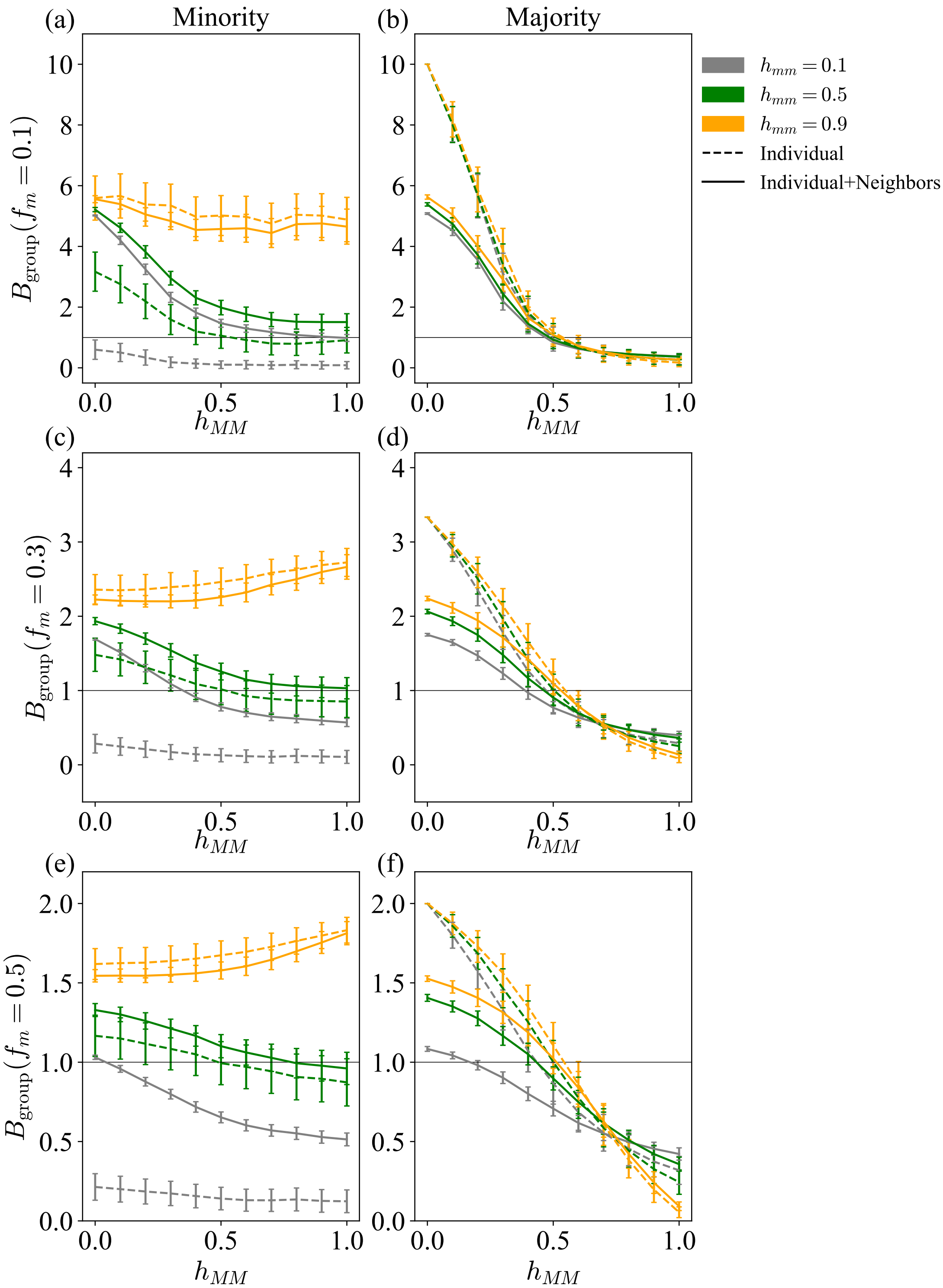

Fig. S4. Estimates of possible group-level perception biases (dashed lines) compared to possible biases of the weighted average of perceptions of individuals and their neighbors (solid lines), assuming asymmetric homophily. The minority-group size is fixed at 0.1 in the top row, 0.3 in the middle row, and 0.5 in the bottom row. The homophily for the minority h_{mm} (depicted by lines of different colors) is fixed, while homophily for the majority h_{MM} ranges from 0 to 1 along the horizontal axis.