Homophily and minority 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], **Mirta Galesic**[8,9]

[1]Department of Mathematics, University of North Carolina at Chapel Hill, 27599, USA
[2]Department of Computational Social Science, GESIS, Cologne, 50667, Germany [3]Institute for Web Science and Technologies, University of Koblenz-Landau, 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, Germany [8]Santa Fe Institute, Santa Fe, NM [9]Harding Center for Risk Literacy, Max Planck Institute for Human Development, Berlin, Germany

Abstract

People's perceptions about the frequency of attributes in their social networks sometimes show *false consensus*, or overestimation of the frequency of own attributes, and sometimes *false uniqueness*, or underestimation of the frequency. Here we show that both perception biases can emerge solely from the structural properties of social networks. Using a generative network model, we show analytically that perception biases depend on the level of homophily and its asymmetric nature, as well as on the size of minority group. Model predictions correspond to empirical data from a cross-cultural survey study and to numerical calculations on six real-world networks. We also show in what circumstances individuals can reduce their biases by relying on perceptions of their neighbors. This study advances our understanding of the impact of network structure on 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, ranging 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. Instead, their samples are constrained by the local structure of their personal networks,

which can bias their perception of the frequency of different attributes in the general population. For example, during the 2016 U.S. presidential election Twitter users, including journalists, who supported one candidate isolated themselves from users who supported a different candidate[7]. Members of such insular communities often tend to overestimate the frequency of their own attributes in the overall society. Such *social perception biases* have been studied in the literatures on false consensus[8–11]. However, it has also been documented that people holding a particular view sometimes tend to underestimate the frequency of that view. This apparently contradictory bias has been studied in the literature on false uniqueness[12–14], pluralistic ignorance[15,16] and majority illusion[17].

It has been observed that diverse social perception biases can be related to the structural properties of personal networks[18,19], which can strongly affect the samples of information that individuals rely on when forming their social perceptions[20,21]. However, the impact of different network properties on social perception biases has not yet been systematically explored. Most existing explanations of these biases instead rely on assumptions about motivational and cognitive processes, including feeling good when others share one's own view[22], wishful thinking[23], easier recall of the reasons for having own view[8], justifying one's undesirable behaviors by overestimating their frequency in the society[24], and rational inference of population frequencies based on own attributes[25]. These accounts can explain biases such as false consensus but not false uniqueness.

Here we show empirically, analytically, and numerically how a simple network model can explain social perception biases resembling both false consensus and false uniqueness, without further assumptions about biased motivational or cognitive processes. Empirical results from a cross-cultural survey show that structural network properties, namely, level of 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 sample biases that could occur in six empirical networks, and point to the importance of an additional structural property, asymmetric homophily. We also show when 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 related social phenomena.

Results

Defining social perception biases

We focus on perceptions of the prevalence of binary attributes (e.g., smoking, believing in a god, or donating to charity) in the overall network. We define social perception biases on the individual and the group level. Unless otherwise stated, we use $P_{\text{indv.}}$ and P_{group} to denote individual- and group-level perception bias. If necessary, we add a superscript a for the minority and b for the majority group.

On the individual level, we assume that individuals' perceptions are based on the prevalence of an attribute in their personal network (direct neighborhood). We define individual i's bias

in perception of the fraction of individuals with a given attribute in the overall network as follows:

$$P_{\text{indv.}} = \frac{1}{f_a} \frac{\sum_{j \in \Lambda_i} x_j}{k_i},$$ (1)

where k_i is the degree of individual i, that is, the size of i's personal network, Λ_i is a set of i's neighbors, x_j is the attribute of individual i's neighbor j ($x_j = 1$ for a minority attribute and 0 for a majority attribute), and f_a is the true fraction of minority nodes in the overall network.

The group-level perception bias is defined by aggregating perception biases of all individuals in a minority or a majority group:

$$P_{\text{group}} = \frac{1}{f_a} \frac{\sum_{i \in N_g} \sum_{j \in \Lambda_i} x_j}{\sum_{i \in N_g} k_i},$$ (2)

where N_g is the set of individuals in a group g, which can be a minority or a majority group. This definition is comparable to those used in studies of the friendship paradox[26,27] and enables us to derive analytic predictions at the group level using a mean-field approach.

The minimum value of the perception bias is 0 and the maximum value is $1/f_a$ (see Method). A perception-bias 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 fraction of a minority attribute in the overall network.

As an example, Fig. 1 illustrates how we define the perception bias at the individual level and group level for a high-homophily (homophilic) network and a low-homophily (heterophilic) network. Here, 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 his personal network (dashed circle). In the homophilic network (Fig. 1a), his individual-level perception bias is $(1/6)/(1/3) = 0.5$, which means that he underestimates the size of the minority group by a factor of 0.5. Consequently, he overestimates the size of his own, majority group in the overall network. In the heterophilic network (Fig. 1b), the perception bias of individual i is $(4/6)/(1/3) = 2$, implying that he overestimates the size of the minority group by a factor of 2. At the group level, the majority group (blue) perceives the size of the minority group to be $4/18$ in the homophilic network and $10/18$ in the heterophilic network. Therefore, the majority group underestimates the size of the minority group by a factor of $(4/18)/(1/3) = 0.67$ in the homophilic network and overestimates it by a factor of $(10/18)/(1/3) = 1.67$ in the heterophilic network.

Survey of social perception biases

To gain insight into the role of homophily and minority-group size in social perception biases, we conducted a survey with $N = 101$ participants from the United States and $N = 99$ participants from Germany, recruited from Amazon's Mechanical Turk, and $N = 100$ South Korean participants recruited from the survey platform Tillion Panel. For American and German participants, we asked questions about 10 attributes, listed in Table S1. The questions were taken from existing national surveys that provided objective frequencies of different attributes in the general population of each country. For Korean participants, we asked questions about seven of these attributes for which we had objective frequencies from national surveys (see Table S2).

Participants answered three groups of questions. First, they answered questions about their own attributes (e.g., smoking). 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 more 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 to a smoker is 70%. We used this information to estimate the homophily parameter (h) for each individual's personal network using Eq. 10. Homophily can vary from 0 (complete heterophily) to 1 (complete homophily).

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 national surveys to calculate their social perception bias (see Method). The perception bias was calculated as a ratio of an individual's estimate and the objective frequency from national surveys. For example, if a participant reported that she believes that 60% of the population smoke tobacco whereas the national survey suggested that 40% do so, the perception bias was 60/40 = 1.5.

For each country, we analyzed perception biases separately for attributes that in that country are objectively found to be held by a small ($f_a < 0.2$), medium ($0.2 \leq f_a < 0.4$), or large ($0.4 \leq f_a < 0.5$) fraction of the minority group. For example, the attribute "not having money for food" is held by a small minority in all three countries (Table S2). In contrast, the attribute "donating to charity" is held by a large minority in Germany and a medium minority in South Korea. In the United States, this attribute is held by a majority, and its reverse, "not donating to charity," is held by a large minority.

Fig. 2 shows the survey results for the United States and Germany (Fig. S1 for the South Korea), with a different country in each row. 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 perception bias indicates how accurately each participant (each point in the plot) perceived the size of the minority group to be in the overall population. 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_a) and of homophily of personal networks (h) on perception

biases. As minority-group size in the overall population decreased, the size of the perception bias 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 false-uniqueness tendencies for both minority and majority participants. Similar relationships between perception biases, homophily, and minority-group size were observed in all three countries.

The survey results provide evidence that homophily of personal networks and minority-group size partially explain perception biases regardless of cultural differences. They stress the necessity of considering the social structure in which an individual is embedded to understand how social perception biases are formed. In the next section we therefore explore whether these biases can be explained by a simple generative network model.

Generative network model with tunable homophily and minority-group size

To systematically study the relation between perception biases and network structure, we developed a network model that allows us to create scale-free networks with tunable homophily and minority-group sizes[28]. This network model is a variation of the Barabási–Albert preferential attachment model with the addition of a homophily parameter h (we call this model BA-homophily). The probability of an attachment of a newly arriving individual (node) w to an existing node v (ϕ_{wv}) is proportional to the node's degree (k_v) and the homophily between the two nodes (h_{wv}); thus $\phi_{wv} \propto h_{wv} k_v$.

Thus, the degree and the homophily parameter regulate the probability of connection between individuals who share the same attribute. The value of homophily ranges from 0 to 1. If homophily is low ($0 \le h < 0.5$), nodes have a tendency to connect to nodes with opposite attribute values. As the homophily parameter increases above 0.5, the probability of connection between nodes with the same attribute increases. In the case of extreme homophily ($h = 1$), nodes strictly connect to other nodes with the same attribute value and therefore two separate communities emerge. One advantage of using the BA-homophily model is that it generates networks with the scale-free degree distributions observed in many large-scale social networks. Another advantage is that it generates networks in which homophily can be symmetric or asymmetric. Given nodes with two attributes, a and b, when homophily is symmetric the tendency to connect to nodes with the same attribute is the same for both groups, $h_{aa} = h_{bb} = h$. In this case one parameter is sufficient to generate the network. In the asymmetric homophily case, two homophily parameters are needed to regulate the connectivity probabilities for each group separately, thus $h_{aa} \ne h_{bb}$.

Fig. 3 depicts analytically derived biases in perceptions of minority-group size among individuals who themselves belong to a minority (Fig. 3a) or a majority (Fig. 3b) group, as a function of the true fraction of the minority in the overall network (f_a) and the homophily of individuals' personal networks (h). The solid lines show the analytic results (see Method) and the circles are numerical results obtained from the BA-homophily model. In heterophilic networks ($0 \le h < 0.5$), perception biases resemble false uniqueness. The minority underestimates its own size, and the majority overestimates the size of the minority, the more so the smaller the minority group (smaller f_a). In homophilic networks ($0.5 < h \le 1$),

perception biases resemble false consensus. The minority overestimates its own size (the more so the smaller the minority group), and the majority underestimates the size of the minority. Slight asymmetries between biases expected for minority and majority groups (see insets in Fig. 3) are due to the disproportionate number of links for two groups, affecting the results of Eq. 7.

These analytic derivations can help us describe the functional form of the biases observed in the survey (Fig. 2). As shown in Eq. 8, the minority's perception bias (P^a_{group}) is proportional to the density of links between minorities (p_{aa}), which increases with the homophily between minority nodes, that is, h_{aa}. Similarly, the majority's perception bias (P^b_{group}) is proportional to the density of intergroup links (p_{ab}), which decreases as the homophily (h_{aa} and h_{bb}) increases. In addition, the relative sizes of minority and majority groups influence the growth rate of links for each group according to Eq. 9 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 $p_{aa} = p_{bb} = 1$, while $p_{ab} = p_{ba} = 0$, leading to the minority's group-level perception bias of $1/f_a$. 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 calculations, there are some differences that call for more detailed investigation in the future. One main difference is that in 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 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. This does not matter on average when minority size is relatively large as errors of over- and underestimation cancel out. But for small minority-group size the estimate cannot be lower than 0, meaning that the former errors (overestimated - true sample frequency) will be larger than the latter errors (true - underestimated sample frequency) and will not cancel out[20]. Hence, people's estimates of the frequency of attributes in their samples will be overestimated for small minority groups, which is what we observed in 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. The network characteristics of the data sets are presented in Table 1. The detailed descriptions of the data sets and references can be found in the Method section. These empirical networks have different structural characteristics and show different levels of homophily or heterophily with respect to one specific attribute (see 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 first assumed that homophily is symmetric in all networks. The symmetric homophily is equivalent to Newman's assortativity measure (q), which is the Pearson coefficient of correlation between pairs of linked nodes according to some attribute (e.g., race[29]). 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 ($q = -1$), $h = 0.5$ indicates no relationship between structure and attributes ($q = 0$), and $h = 1$ indicates complete homophily ($q = 1$; see the Supplementary Materials).

In reality, however, the tendency of groups to connect to other groups can be asymmetric. For example, it has been observed that in scientific collaborations, homophily among women is stronger than homophily among men[30]. Given the relationship between number of edges that run between nodes of the same group and homophily in Eqs. 10 and 11, we can estimate the asymmetric homophily, which differs for the minority (h_{aa}) and the majority (h_{bb}) group (see Method). As we describe 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 gain insights about the impact of homophily and minority-group size differences on possible individual- and group-level perception biases.

Fig. 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 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[31], social learning[32], and wisdom of crowds[33]. To evaluate the impact of degree on shaping perception biases, we plotted individual perception biases, $P_{indv.}$, versus individual degree in Fig. S2 in supplementary. 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. This analysis enables us to understand the impact of influential individuals on perception biases exhibited by other individuals in the network.

Reducing social perception biases

We investigated to what extent and under what structural conditions individuals can reduce their perception bias by considering the perceptions of their neighbors. We built on DeGroot's weighted belief formalization by aggregating an individual's own perception (*ego*) with the averaged perceptions of the individual's direct neighbors (*1-hop*)[34]. For simplicity, we assumed symmetric homophily in the BA-homophily model (the results for asymmetric homophily are in the Supplementary Materials).

Fig. 5 shows a comparison of the average perception bias ($\bar{P}_{indv.}$) for individuals who belong to the (a) minority and (b) majority and the weighted averages of their own perceptions and those of their 1-hop neighbors. The minority-group size is fixed to 0.2. The results show that taking into account perceptions of 1-hop neighbors improves estimates of individuals in heterophilic networks (blue triangles are closer to the gray dashed line compared to orange circles in the log scale). The improvement in the perception is the result of nodes 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.

This suggests that in homophilic networks, individuals cannot improve their perceptions by consulting their peers because they are similar to them and do not add new information that might not increase the accuracy of their estimates. However, in heterophilic networks, individuals benefit from considering their neighbors' perceptions because they are not exposed only to other like-minded individuals but to more diverse perceptions. While the overall trend is not surprising, our results reveal how the accuracy of these 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. However, many studies have documented biases in these social perceptions, including both false consensus and false uniqueness. Here

we investigated to what extent both of these biases can be explained merely by the structure of the social networks in which individuals are embedded, without any assumptions about biased cognitive or motivational processes.

Using empirical survey data, analytical investigation, and numerical simulations, we show that structural properties of personal networks strongly affect the samples people draw from the overall population. We find 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[22], our analyses establish a baseline level of biases that can occur without assuming biased information processing[20,35–38].

Our results suggest that homophily impacts the accuracy of the estimates of individuals in both minority and majority groups. In homophilic networks, minority group tends to overestimate its own size, and majority group tends to underestimate the size of the minority. In contrast, in heterophilic networks, minority group tends to underestimate its own size, and majority group overestimates the size of the minority. In other words, when homophily is high, both minority and majority groups tend to show social perception biases similar to false consensus, whereas when it is low, both group show biases similar to false uniqueness.

We further show that the relative sizes of the majority and minority groups influence perception biases. Specifically, the smaller the size of the minority, the higher the false consensus among members of the minority group and false uniqueness among members of the majority group.

To explain the underlying structural mechanisms of perception biases in social networks, we developed a generative network model with tunable homophily and minority-group size. We find that predictions from this theoretical model correspond to empirical observations well, especially when we assume asymmetric rather than symmetric homophily. In addition, 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.

Most previous explanations of perception biases in social psychology did not include precise quantitative models of social environments. Many of them did not explicitly include the structure of social networks at all. Those that did, such as selective-exposure explanations, typically involved qualitative statements that people tend to socialize with similar others[8,18]. The network model described in the present paper enabled a thorough quantitative investigation of how different levels of homophily, its asymmetric nature, and different relative sizes of majority and minority groups, affect social perception biases. This investigation did not include a quantitative specification of the cognitive processes underlying people's sampling from their personal networks. Such specifications[19,21,35,37] could be combined with the network model described here.

Besides providing a theoretical account of perception biases, our results have 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[39] to group performance[40] and social identity[41], it is crucial to consider obstacles to improving 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[39,42,43]. When people overestimate the prevalence 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. 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[44,45], 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 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 that call for future investigation on the impact of minority-group size on perceptions. In particular, experimental group studies could lead to a better understanding of how different network properties affect perceptions.

In sum, this study shows that both over-and under-estimation of the frequency of own view can be explained by different levels of homophily, asymmetric nature of homophily, and 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 how to measure and reduce perception gaps between different groups 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 allows us to create scale-free networks with tunable homophily and minority-group size[28]. 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 m 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 m guarantees the lower bound of degree and in our model is set as 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

Let us refer to the minority as a and the majority as b. $K_a(t)$ and $K_b(t)$ are 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 m existing nodes in the network. Therefore, the total degree of the growing network at time t is $K(t) = K_a(t) + K_b(t) = 2mt$. In this model, the degree growth is linear for both groups. Denoting C as the minority's degree growth factor, we have

$$K_a(t) = Cmt, \quad K_b(t) = (2 - C)mt. \tag{4}$$

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

$$p_{aa} = \frac{h_{aa}K_a(t)}{h_{aa}K_a(t) + h_{ab}K_b(t)} = \frac{h_{aa}C}{h_{aa}C + h_{ab}(2 - C)}, \tag{5}$$

where h_{aa} is the homophily between minority nodes, and h_{ab} is the homophily between a minority and a majority in a relation of $1 - h_{aa}$. The connection probability from a minority to a majority is the complement of p_{aa} as

$$p_{ab} = \frac{h_{ab}K_b(t)}{h_{aa}K_a(t) + h_{ab}K_b(t)} = \frac{h_{ab}(2 - C)}{h_{aa}C + h_{ab}(2 - C)}. \tag{6}$$

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

To calculate the group perception bias (P_{group}), one needs to consider the number of inter- and intragroup edges as expressed in the definition of Eq. 2. Thus, the group perception bias for the minority and majority is as follows:

$$P^a_{\text{group}} = \frac{1}{f_a} \frac{2M_{aa}}{2M_{aa} + (M_{ab} + M_{ba})}, \quad P^b_{\text{group}} = \frac{1}{f_a} \frac{M_{ab} + M_{ba}}{2M_{bb} + (M_{ab} + M_{ba})}. \tag{7}$$

Here, M represent the number of edges between different groups. For example, M_{aa} is the number of edges between minority nodes. Note that we distinguish number of edges between minority and majority M_{ab} and between majority and minority M_{ba}. These values are equivalent when homophily is symmetric but they are unequal when the homophily is asymmetric. Since each M has a relation as a product of the total number of edges and the edge probability, such as $M_{aa} = mN_a p_{aa}$, we can reduce Eq. 7 to the following equations:

$$P^a_{\text{group}} = \frac{1}{f_a} \frac{2p_{aa}}{2p_{aa} + p_{ab} + (N_b/N_a)p_{ba}}, \quad P^b_{\text{group}} = \frac{1}{f_a} \frac{(N_a/N_b)p_{ab} + p_{ba}}{2p_{bb} + (N_a/N_b)p_{ab} + p_{ba}}, \tag{8}$$

where N_a and N_b 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 $f_a = 0.5$ in extreme homophily ($h = 1.0$) with the degree growth $C = 1$ (a symmetric homophily condition), $P^a_{\text{group}} = 2$ from Eq. 8, 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_{aa}, p_{ab}, and p_{bb} (see the Supplementary Materials). Thus,

$$C = f_a(1 + p_{aa}) + f_b p_{ba}. \tag{9}$$

In empirical networks we know the edge density for the minority ($r_{aa} = M_{aa}/M$), between groups ($r_{ab} = M_{ab}/M$), and for the majority ($r_{bb} = M_{bb}/M$). Thus, the probability of connection within a group can be written as $r_{aa} = f_a p_{aa}$ and $r_{bb} = f_b p_{bb}$.

Connection probabilities are defined as follows:

$$p_{aa} = \frac{h_{aa}C}{h_{aa}C + h_{ab}(2 - C)}, \quad p_{bb} = \frac{h_{bb}(2 - C)}{h_{bb}(2 - C) + h_{ba}C}. \tag{10}$$

From Eq. 10 and the relation between r_{aa} and p_{aa} (or r_{bb} and p_{bb}), we can derive the empirical homophily by using edge density r_{aa}, r_{bb} as follows:

$$h_{aa} = \frac{r_{aa}(2 - C)}{f_a C + 2r_{aa}(1 - C)}, \quad h_{bb} = \frac{r_{bb}C}{f_b(2 - C) - 2r_{bb}(1 - C)}. \tag{11}$$

This calculations allow to estimate the homophily from the empirical networks assuming that the BA-homophily model is a valid model of a social network. The homophily can be symmetric or asymmetric in its nature. In order to understand the role of asymmetricit homophily in perception biases, we first assume all empirical networks have symmetric homophily. We approximate symmetric homophily by projecting the homophily to Newman's assortativity measure (q) (see Supplementary Materials and Fig. S4).

Empirical networks

The first network, Brazilian network, captures sexual contact between sex workers and sex buyers[46]. The network consists of 16,730 nodes and 39,044 edges. There are 10,106 sex workers and 6,624 sex buyers (minority-group size $f_a = 0.4$). 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)[47]. 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[48]. Since the group definition is arbitrary, 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)[49]. The network is composed of 6,253 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 small heterophily[49] ($q = -0.06, h = 0.48$).

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[50]. After we removed ambiguous names, the network consisted of 120,338 men and 7,330 women. Here, women belong to the minority group and represent only about 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)[51]. We used a new method that combines names and images to infer the gender of the scientists with high accuracy[50]. 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 (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 ($f_a = 0.37$). This network shows the highest homophily compared to the other empirical data sets ($h = 0.94$ and $q = 0.87$).

Survey study

For all countries (the Unite State, Germany, and South Korea), we restricted participant age to at least 18 years. Regarding gender, 61.4% of the U.S. participants, 85.9% of the German participants, and 50.0% of the Korean participants were male. The age distribution in the United States was 18–30 years: 34.6%, 31–40 years: 36.6%, 41–50 years: 9.8%, 50+ years: 19.0%; in Germany it was 18–30 years: 61.4%, 31–40 years: 26.3%, 41–50 years: 6.0%, 50+ years: 6.3%; and in South Korea it was 18–30 years: 26.0%, 31–40 years: 26.0%, 41–50 years: 24.0%, 50+ years: 24.0%.

Participants were asked questions about their own attributes, the frequency of these attributes in their personal networks, and their frequency in the overall population of their country. Question texts and objective sizes of minority and majority groups in the overall populations were taken from publicly available results of large national surveys conducted in each country. Details are provided in Table S1. U.S. and German participants were asked about 10 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 and following an approach similar to that of Coleman[52]. Each survey 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 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_{aa} or p_{bb}, and their complementaries correspond to p_{ab} or p_{ba} in the BA-homophily model. Using Eqs. 9 and 10 we can calculate the homophily h_{aa} (or h_{bb}) of each participant's personal network. In addition, we can evaluate h_{ab} and h_{ba} using the relations $h_{ab} = 1 - h_{aa}$ and $h_{ba} = 1 - h_{bb}$.

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 ($f_a < 0.2$), medium ($0.2 \leq f_a < 0.4$), and large ($0.4 \leq f_a < 0.5$). For example, small minority attributes in the U.S. are no money for food, experienced theft, and smoking, 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments:

References

1. Cialdini RB & Trost MR Social influence: Social norms, conformity and compliance In The Handbook of Social Psychology, Vol. 2, 151–192 (McGraw-Hill, Boston, MA, 1998).

2. Bond RM et al. A 61-million-person experiment in social influence and political mobilization. Nature 489, 295–298 (2012). [PubMed: 22972300]

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). [PubMed: 20813952]

5. Borsari B & Carey KB Descriptive and injunctive norms in college drinking: a meta-analytic integration. J. Stud. Alcohol 64, 331–341 (2003). [PubMed: 12817821]

6. Botvin GJ, Botvin EM, Baker E, Dusenbury L & Goldberg CJ The false consensus effect: Predicting adolescents' tobacco use from normative expectations. Psychol. Rep 70, 171–178 (1992). [PubMed: 1565717]

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. 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).

9. Mullen B et al. The false consensus effect: A meta-analysis of 115 hypothesis tests. Journal of Experimental Social Psychology 21, 262–283 (1985).

10. Krueger J & Clement RW The truly false consensus effect: An ineradicable and egocentric bias in social perception. J. Pers. Soc. Psychol 67, 596–610 (1994). [PubMed: 7965607]

11. Krueger J From social projection to social behaviour. Eur. Rev. Soc. Psychol 18, 1–35 (2007).

12. Frable DE Being and feeling unique: Statistical deviance and psychological marginality. J. Pers 61, 85–110 (1993). [PubMed: 8463952]

13. Mullen B, Dovidio JF, Johnson C & Copper C In-group-out-group differences in social projection. J. Exp. Soc. Psychol 28, 422–440 (1992).

14. Suls J & Wan CK In search of the false-uniqueness phenomenon: Fear and estimates of social consensus. Journal of Personality and Social Psychology 52, 211–217 (1987). [PubMed: 3820073]

15. Miller DT & McFarland C Pluralistic ignorance: When similarity is interpreted as dissimilarity. J. Pers. Soc. Psychol 53, 298–305 (1987).

16. Prentice D & Miller T, D. Pluralistic ignorance and alcohol use on campus: Some consequences of misperceiving the social norm. J. Pers. Soc. Psychol 64, 243–256 (1993). [PubMed: 8433272]

17. Lerman K, Yan X & Wu X-Z The" majority illusion" in social networks. PLoS One 11, e0147617 (2016). [PubMed: 26886112]

18. Sherman SJ, Presson CC, 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).

19. Galesic M, Olsson H & Rieskamp J A sampling model of social judgment. Psychol. Rev 125, 363 (2018). [PubMed: 29733664]

20. Juslin P, Winman A & Hansson P The naïve intuitive statistician: A naïve sampling model of intuitive confidence intervals. Psychol. Rev 114, 678–703 (2007). [PubMed: 17638502]

21. Pachur T, Hertwig R & Rieskamp J Intuitive judgments of social statistics: How exhaustive does sampling need to be? Journal of Experimental Social Psychology 49, 1059–1077 (2013).

22. Marks G & Miller N Ten years of research on the false-consensus effect: An empirical and theoretical review. Psychol. Bull 102, 72 (1987).

23. Miller MK, Wang G, Kulkarni SR, Poor HV & Osherson DN Citizen forecasts of the 2008 us presidential election. Politics & Policy 40, 1019–1052 (2012).

24. Suls J, Wan CK & Sanders GS False consensus and false uniqueness in estimating the prevalence of health-protective behaviors. J. Appl. Soc. Psychol 18, 66–79 (1988).

25. Dawes RM Statistical criteria for establishing a truly false consensus effect. Journal of Experimental Social Psychology 25, 1–17 (1989).

26. Eom Y-H & Jo H-H Generalized friendship paradox in complex networks. Sci. Rep 4, 4603 (2014). [PubMed: 24714092]

27. Jo H-H & Eom Y-H Generalized friendship paradox in network with tunable degree-attribute correlation. Phys. Rev. E 90, 022809 (2014).

28. Karimi F, Génois M, Wagner C, Singer P & Strohmaier M Homophily influences ranking of minorities in social networks. Sci. Rep 8 (2018).

29. Newman ME Mixing patterns in networks. Phys. Rev. E 67, 026126 (2003).

30. 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).

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

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

33. 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). [PubMed: 28607070]

34. DeGroot MH Reaching a consensus. J. Am. Stat. Assoc 69, 118–121 (1974).

35. Fiedler K Beware of samples! a cognitive-ecological sampling approach to judgment biases. Psychological review 107, 659 (2000). [PubMed: 11089402]

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

37. Le Mens G & Denrell J Rational learning and information sampling: On the "naivety" assumption in sampling explanations of judgment biases. Psychological review 118, 379 (2011). [PubMed: 21480741]

38. Denrell J & Le Mens G Information sampling, belief synchronization, and collective illusions. Management Science 63, 528–547 (2016).

39. Centola D An experimental study of homophily in the adoption of health behavior. Science 334, 1269–1272 (2011). [PubMed: 22144624]

40. Mollica KA, Gray B & Treviño LK Racial homophily and its persistence in newcomers' social networks. Organ. Sci 14, 123–136 (2003).

41. Mehra A, Kilduff M & Brass DJ At the margins: A distinctiveness approach to the social identity and social networks of underrepresented groups. Acad. Manag. J 41, 441–452 (1998).

42. Festinger L A theory of social comparison processes. Hum. Relat 7, 117–140 (1954).

43. Suls J, Martin R & Wheeler L Social comparison: Why, with whom, and with what effect? Curr. Dir. Psychol. Sci 11, 159–163 (2002).

44. Mobilia M Does a single zealot affect an infinite group of voters? Phys. Rev. Lett 91, 028701 (2003). [PubMed: 12906515]

45. Mobilia M, Petersen A & Redner S On the role of zealotry in the voter model. J. Stat. Mech. Theory Exp 2007, P08029 (2007).

46. Rocha LE, Liljeros F & Holme P Simulated epidemics in an empirical spatiotemporal network of 50,185 sexual contacts. PLoS Comput. Biol 7, e1001109 (2011). [PubMed: 21445228]

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

48. Holme P, Liljeros F, Edling CR & Kim BJ Network bipartivity. Phys. Rev. E 68, 056107 (2003).

49. Traud AL, Mucha PJ & Porter MA Social structure of facebook networks. Phys. A 391, 4165–4180 (2012).

50. 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 the 25th International Conference Companion on World Wide Web, 53–54 (International World Wide Web Conferences Steering Committee, 2016).

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

52. Coleman J Relational analysis: The study of social organizations with survey methods. Hum. Organ 17, 28–36 (1958).

(a) Homophilic network
(False Consensus)

(b) Heterophilic network
(False Uniqueness)
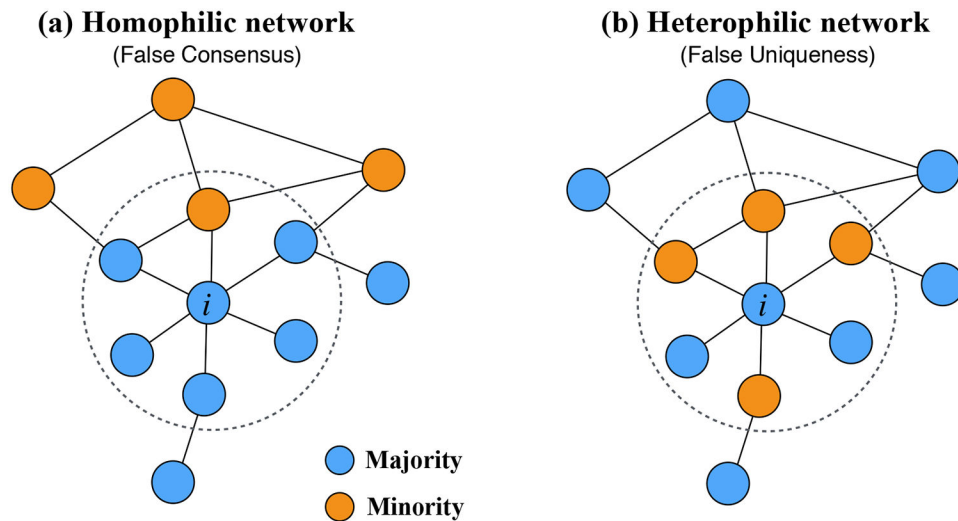
● **Majority**
● **Minority**

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 ($f_a \approx 0.33$). Panel (a) depicts a homophilic network and panel (b) shows a heterophilic network. We can study 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 his 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 size by a factor of 2. On the group level, the majority group perceives the size of the minority group to be $4/18 \approx 20\%$ in the homophilic network and $10/18 \approx 56\%$ in the heterophilic network. The majority group underestimates the size of the minority group by a factor of 0.6 in the homophilic network and overestimates the minority-group size by a factor of 1.67 in the heterophilic network. In sum, depending on the topological structure of the network, individuals' and groups' perceptions about their own and other groups' sizes can be distorted.
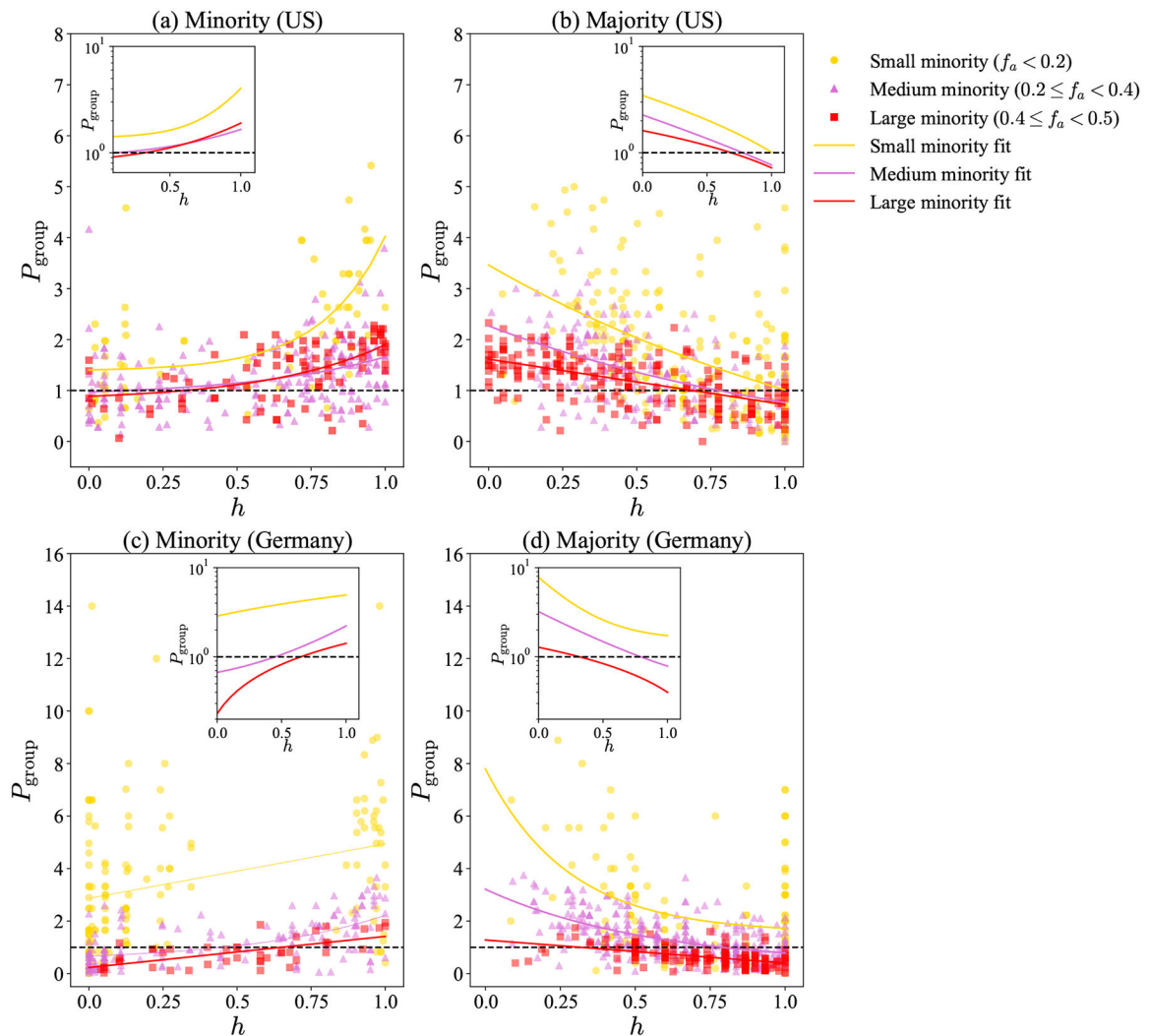
Fig. 2. Bias in perception of minority-group size, for participants whose personal networks exhibit different levels of homophily (h), and for attributes held by a small, medium, or large minority group in a given country.

Rows shows results from different countries: United States (top) and Germany (bottom). Different colors distinguish perception biases for attributes that in a given country are held by a small ($f_a < 0.2$), medium ($0.2 \leq f_a < 0.4$), or large ($0.4 \leq f_a < 0.5$) minority group. Each data point represents the perception bias of a group where one individual involved for an attribute. Group level perception bias is calculated as a ratio of the perceived size of the minority group (obtained in our surveys) and the objective minority size (obtained from national surveys). Panels in the left column show perception biases of the minority, and panels in the right column show perception biases of the majority. The solid line is the curve fit of the data. 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 by mean-field assumption (see Supplementary Materials).
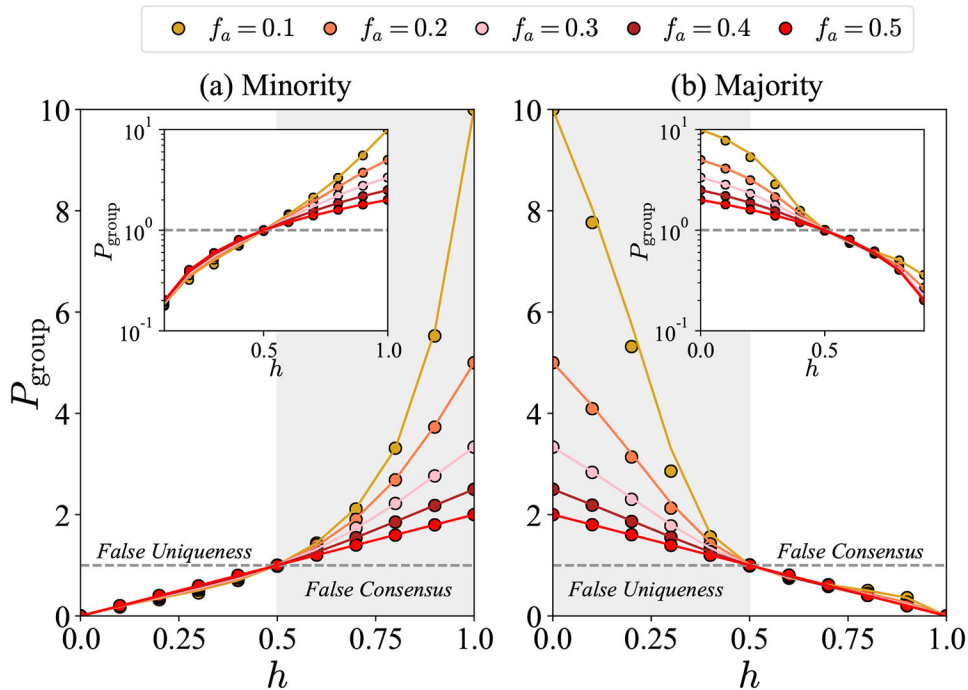
Fig. 3. Estimation of minority group size by (a) the minority group and (b) the majority group, as a function of homophily of personal networks (*h*) and the minority-fraction (*f*_a) in the overall network.

Different colors refer to networks with different minority fractions (f_a). 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.
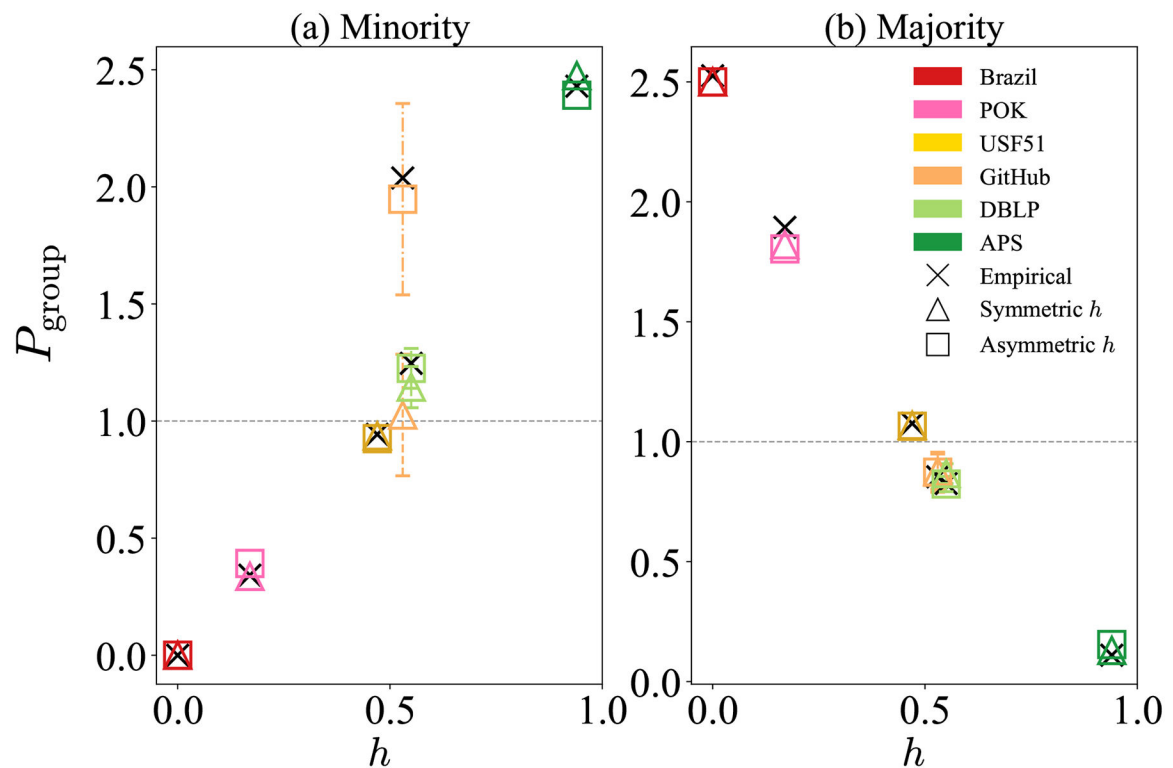
Fig. 4. Group-level social perception biases that could occur in six empirical social networks. The figure shows how accurately (a) the minority group and (b) the majority group might estimate the size of the minority group in real-world social networks with different levels of homophily. The symmetric homophily values of the empirical social networks are depicted on the *x* axis. 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 excellent 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.
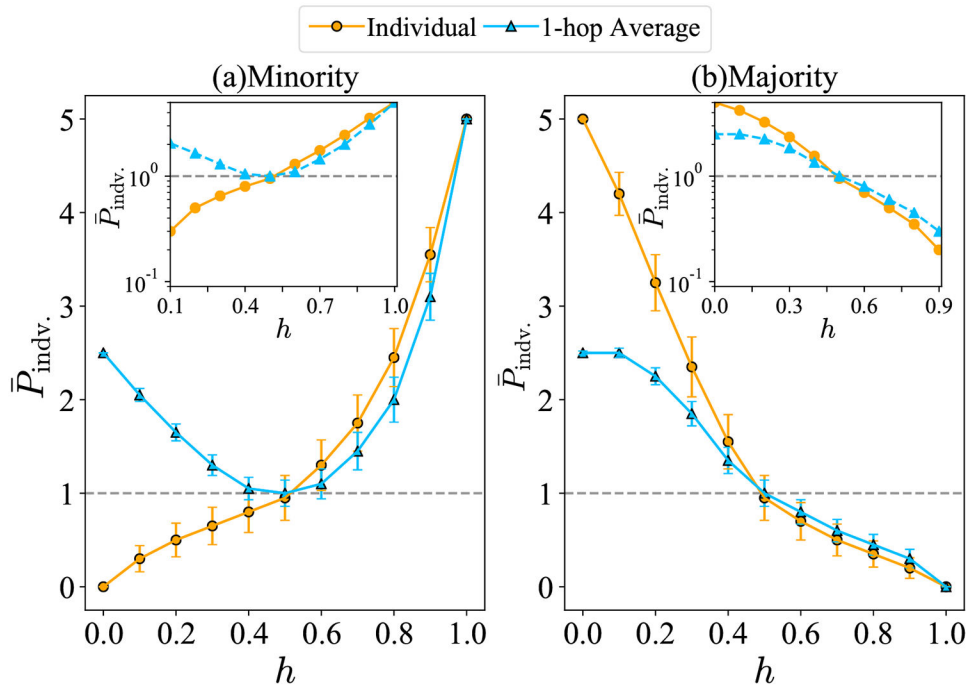
Fig. 5. Social perception biases for individual nodes and for the weighted average of perceptions of individual nodes and their 1-hop neighbors.

Insets show the same results in log scale. Orange lines are calculated from Eq. 1 and averaged over all nodes in the group. Blue lines show the perception bias of the weighted average of perceptions of individual nodes and their direct neighbors (1-hop). 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).

Table 1.

Characteristics of the empirical networks.

Shown are sizes (number of nodes) of the minority and majority groups regarding a gender attribute (for Brazil, POK, USF51, GitHub, DBLP) and an academic field (for APS), as well as the values of symmetric and asymmetric homophily. Values are rounded to the second decimal point.

Data	Number of nodes	Minority	Majority	Symmetric h	Asymmetric h (minority, majority)
Brazil	16,730	Sex workers 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	6,253	Male 2,626 (42%)	Female 3,627	0.47	0.48, 0.47
GitHub	127,668	Female 7,330 (5.7%)	Male 120,338	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 695 (37%)	QSM 1,158	0.94	0.88, 1.0