Gender Bias and Learning in Social Networks

Xiaomin Bian*, Mir Adnan Mahmood†, Samantha Stelnicki‡

May 2024

Abstract

Our experiment identifies biases in information updating by men and women in exogenously-formed networks. We include circle and star network structures to see how the structure of networks and level of information available interact with gender. We find that both men and women in star networks exhibit homophily. In contrast, circle network participants do not exhibit any gender-based bias. Rather, including gender information in circle networks increases men's willingness to update their own information. These differences result in opposing aggregate behavior; gender information increases the rate of consensus in circle networks but decreases the rate of consensus in star networks.

Keywords: Gender, Learning, Networks, Biased Beliefs

We thank Paul J. Healy, John Rehbeck, Ian Krajbich, John Kagel, Yaron Azrieli, Sally Sadoff, Kirby Nielsen, attendees of the Decision Sciences Collaborative Forum, all members of The Ohio State University Theory and Experimental Reading Group, and the audience of the Economic Science Association 2022 North America Meeting for their helpful comments and feedback. Additionally, we are grateful to Friederike Mengel for allowing us to her avatars and Katherine Coffman for allowing us to use her trivia questions. Financial support from the Decision Sciences Collaborative Small Grant at The Ohio State University is gratefully acknowledged. This experiment was reviewed and determined exempt by the Institutional Review Board at The Ohio State University.

*xiaominbian@jhu.edu; Krieger School of Arts and Sciences, Johns Hopkins University

†mahmood.69@osu.edu; Department of Economics, The Ohio State University

‡stelnicki.4@osu.edu; Department of Economics, The Ohio State University

1 Introduction

Research shows that women and men exhibit differences in representation and actions in professional settings. Whether this difference is described by preferences for competition (Niederle and Vesterlund, 2007), risk and social preferences (Croson and Gneezy, 2009), wage earnings (Blau and Kahn, 2000), job titles (Bertrand and Hallock, 2001), preference for workplace characteristics (Wiswall and Zafar, 2018), or beliefs about which types of activities are most suitable (Coffman et al., 2021b), the conclusion is still the same: the representation and actions of men and women in workplace environments differ. Due to this understanding, many efforts have been made towards equal representation of the genders. Programs like Affirmative Action mandate that the minority gender represent a certain percentage of workers, but companies have become more focused on minority representation without legislation as well. Recently, entire countries and regions have mandated equal representation of men and women. For example, the European Union has put into law that by 2026 all companies need at least 40% of the underrepresented sex in non-director positions, and at least 33% in director positions (European Commission, 2022). These initiatives are helpful to have equal representation in roles, but may present a larger problem: in environments where there's a push for equal representation, individual beliefs about inferiority of a genders' abilities may lower aggregate welfare.

Our main motivation is thus to measure the extent of individual biases between genders and determine if initiatives aimed at producing equal representation of men and women in workplace settings are effective. Additionally, we investigate how the structure of an organization that carries out such a program may affect the effectiveness of the program.

In this paper, we design an experiment to measure gender differences in information transmission and the aggregate effects that result from this bias to investigate how individual beliefs about information stemming from men and women affect groups with different structures. Our first point of interest is how information transmission stemming from men or women differs. We measure how much weight individuals put on others' information, as well as their belief in the accuracy of others. These two measurements allow us to examine how differences in information weighting between men and women plays a role in overall welfare. We study how this bias between genders affects the efficiency of information spread across networks. This allows us to examine how changing gender composition of organizations influences productivity. Additionally, we look at how the network structure and the number of network neighbors interact with potential gender discrimination. We are interested in seeing if having a larger number of neighbors affects how others view the accuracy of one's belief. We investigate if having an increased amount of information can be used to correct gender differences in information transmission. We exogenously determine gender composition of networks, to mimic professional environments where quotas are used to determine number of men and women to investigate these questions.

In the experiment each subject first takes a trivia quiz that determines their initial beliefs about the state. The state is determined by the correct answers of this trivia quiz. Each trivia question has four questions with five possible multiple-choice answers, where the last answer is always "(e) None of the Above." The true state is determined by randomly selecting a number zero through four, which is the number of trivia questions where the correct answer is (e). Then networks of participants try to converge to the true state through belief updating. We use this updating task to identify individual biases. Each participant is simultaneously asked their belief of the true state, and their belief in the accuracy of their own guess. They then are shown their neighbors' guesses about the true state, and asked to determine the accuracy of their neighbors' guesses. They are then allowed to update their own guess after. This process continues for 20 rounds and is repeated 4 times with different trivia quizzes each time. The belief updating in networks allows us to measure how these individual beliefs are affected by the beliefs of their male and female neighbors, and how that affects overall consensus and efficiency of the networks.

The experiment consists of six treatments. These treatments vary in the network structure and gender composition. We use the star and circle network structures in order to vary the amount of information each subject has, as well as to mimic typical workplace settings. We vary gender composition in three ways: the control treatment, where gender information is unknown, the Mixed treatment, which has both men and women in a network, and the Same treatment, which are networks comprised solely of men or women.

We find that there is substantial homophily, meaning that both men and women update more on information from their same-gender neighbors, in star networks. This same-gender preference is present for both men and women in the star network, where information, i.e. number of direct network neighbors, is limited for most participants. Networks with this same-gender preference have lower levels of consensus when compared to treatments without gender and the circle network. In the circle network, where the number of network neighbors increases, we no longer see the same-gender preference. Specifically, we find that in the circle network men are more likely to update their guess of the state when gender information is available, but do not update differently on information from men and women. We find higher levels of consensus among the gendered treatments in the circle network when compared to the Baseline. These results suggest that network structure is a characteristic that drives behavior in this setting. The difference in ability of the networks to reach consensus between the network structures, as well as the individual biases present, suggest that increasing the number of network members coupled with gender information might reduce the homophily we see in limited information environments of the star networks.

The rest of the paper is as follows. Section 2 discusses related literature. Section 3 details the experimental design. Section 4 elaborates on our results. Section 5 concludes the paper.

2 Literature Review

2.1 *Networking and Social Learning*

The most closely related paper to ours in the area of networking is Mengel (2020). As formation and use of networks has become one of the leading explanations for differences in workplace outcomes of men and women, Mengel investigates how men and women differ when allowed to freely form these connections. The paper finds that difference in networking connections formed in the laboratory experiment lead to earnings and promotions that are substantially higher for men than women. Men display higher levels of homophily in network formation, and connect with more individuals, than women. We add to this by finding that there exists homophily in exogeneously-set networks, but that this is dependent on the type of network. Additionally, there is a large body of work on networking differences between men and women (Ibarra, 1993; Campbell, 1988; Aguiar et al., 2009; Gersick et al., 2000) that describe how the differences in networking form and the repercussions that persist because of such a disparity in networking ability. These motivate our research, but are slightly different as they call into question the formation of networks. Our experiment tests the validity of creating networks with the differences of men and women in mind. Our experiment is the first, to the best of our knowledge, to test how creating structured, equal opportunity networks affects welfare.

More generally, our paper is also related to the literature examining the structure of networks. Grimm and Mengel (2020) show that the degree of information transmission is dependent on network structure. Participants in their experiment pay attention to their neighbors' position in the networks and weight each neighbor accordingly. This paper has motivated our attention to network size and structure, as these attributes seemingly play a large role in the efficiency of a network. We too find that network structure matters for information transmission, and in fact can ease some of the homophily that results from different gender compositions in our setting. Chandrasekhar et al. (2020) shows that network efficiency varies greatly depending on the culture of the network participants. This experiment highlights how networks can be sensitive to the background of the the participant. Efficiency rates, as well as the transmission process, highly differ between cultures. Our experiment adds to this literature by establishing how biased information can impact the goal of the network, as well as how transmission differs between individuals of different genders.

Outside of networks, our paper is connected to the literature on social learning and information transmission. Weizsäcker (2010) provides a meta-analysis of social learning games, looking at how much people earn from listening to information other than their own. Similar to our results, this paper finds that subjects tend to listen to their own information over that of others. They find that this is optimal in regards to earnings in social learning games, whereas we find that subjects earn about the same if they update versus if they do not. Evdokimov and Garfagnini (2023) look at how cognitive ability can help or hurt social learning with disagreements. This paper is similar to ours, as the availability of characteristics of the source of information changes how one may update based on information they receive from the source. The catalyst of their paper is cognitive ability, whereas our paper shows the change in information transmission arising from gender.

2.2 *Gender Differences*

Our paper also relates more generally to the literature examining gender differences in professional settings. There's already a large literature discussing what attributes cause the gender gap in professional settings (Niederle and Vesterlund, 2007; Croson and Gneezy, 2009; Blau and Kahn, 2000; Bertrand and Hallock, 2001; Wiswall and Zafar, 2018; Coffman et al., 2021b). These papers show

how the gap manifests and the outcomes that occur. Our paper adds to this existing literature by documenting and measuring the level of bias that exists between genders. At an individual level, we can show if information is being weighted differently based on the gender of the source of the information. Coffman et al. (2021b) is most closely related to our paper. Their experiment shows how beliefs about the best line of work for both genders furthers discrimination in the workplace. They show that employers prefer to hire men for male-type tasks and women for female-type tasks. The discrimination is derived from beliefs that different groups of people will be worse at certain tasks. Thus inaccurate beliefs about the ability of each gender can, and do, add to discrimination in the workplace. Our experiment relates directly to this paper, as our task is gender-neutral, but may still generate underlying biases about which gender is better. Their results show that holding ability constant, employers prefer to hire men over women. With more information about ability available, employers hire the more qualified individual, regardless of gender. Our task is supported by this conclusion, as our task is performed equally well by men and women, but we still find differences in weighting of information from men and women. Similarly, we find this difference may be mitigated by more available information.

Our finding that gender discrimination may be lessened through addition of information relates to Coffman et al. (2021a) as well. This paper shows the same pattern for group deliberations. Mengel (2021) adds to gender bias in groups, as it shows that gender bias increases with deliberation. Similarly, Bordalo et al. (2019) discusses how stereotypes of tasks most appropriate for each gender affect self confidence, as well as the belief in ability of others. Our paper adds to this literature by determining the level of gender bias in beliefs at an individual level, as well as determining the aggregate welfare effects due to gender discrimination for a group working together.

3 Experimental Design

Our experiment's main goal is to first determine the extent of individual biases in information transmission based on gender, then evaluate the welfare effects caused by such biases at the network level. In order to do so, the experiment is comprised of three treatments that vary the gender composition of exogenously-determined networks. The first treatment is the Baseline treatment, where no gender information is revealed to participants. The other two treatments are the Mixed treatment and the Same treatment. Networks in the Mixed treatment consist of both men and women, whereas networks in the Same treatment consist of either all men or all women. In both of these treatments gender information is revealed to network members.

In all three treatments, participants first provide demographic information, including race, gender, and age. In the Baseline treatment, participants' genders are never revealed to any other participants in the experiment. When placed into a network, they are labeled as "Neighbor X" where X is 1,2,3,..., when shown to other participants. In the Mixed and Same treatments, participants are represented by an avatar that correspond to their reported gender, and stays constant throughout the experiment. These avatars are displayed next to their "Neighbor X" label. Avatars are used to minimize any potential demand effects from revealing gender, such as mitigating any existing bias by participants recognizing the experiment is about gender. The avatar set we used for the experiment comes from Mengel (2020). This set of avatars is gender-obvious and designed to remove confounds with race, age, and attractiveness. Figure 1 shows examples of the avatars.

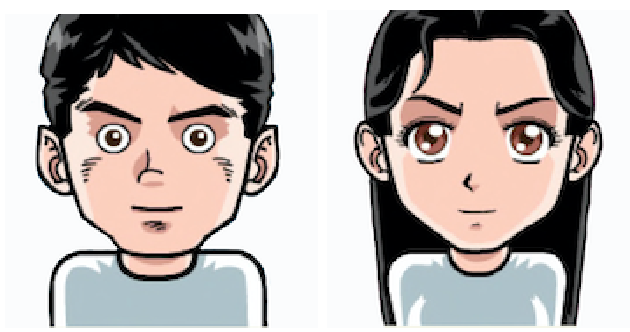

Figure 1: Examples of two of the avatars used.

The task in all treatments is first an individual multiple-choice trivia quiz, followed by 20 rounds of reporting updated beliefs about the state. The trivia quiz consists of four questions, each with five possible answers. The last answer of each multiple-choice question is "(e) None of the above." The state variable used in the updating portion of the task is the number of trivia questions whose correct answer is (e). We use the trivia quiz as our state-determining task in order for beliefs of the state variable to be determined by the ability of each participant. The trivia quiz consists of questions from Coffman (2014). Using their data, we determined which questions were answered correctly by men and women equally.[1] In total, 16 questions from Coffman (2014) are used in

[1] Gender differences were analyzed using Chi-Squared tests and all questions used in our trivia task had p-values greater than 0.3. This procedure was used so that the questions themselves are gender-neutral in performance, but that beliefs of ability may vary between genders.

the trivia quiz. All participants, regardless of treatment, see the same 16 questions in the same order throughout the experiment.[2] In the Baseline treatment, participants see and answer the trivia questions with no other information. In the Mixed and Same treatments, participants see their avatars at the top of the page. Examples of the trivia quiz from the gendered and no-gender treatments are shown in Figure 2.

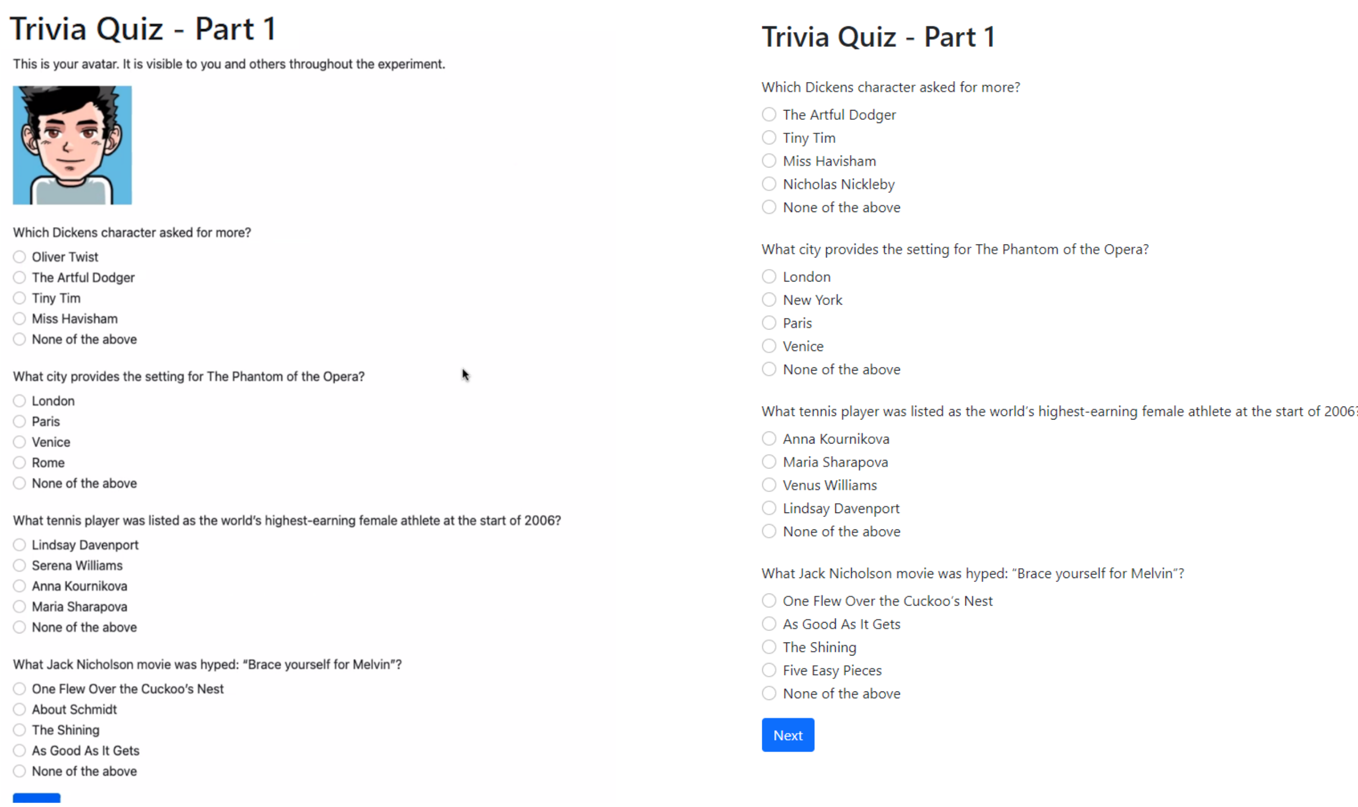

Figure 2: Example of one of the four trivia quizzes that are shown throughout the experiment. One of the gendered treatments is on the left, where the avatar shown corresponds to the individual who is taking the trivia quiz. The Baseline treatment is on the right for comparison.

After the trivia quiz, participants begin the second part of the experiment: belief updating in networks. This updating occurs through repeated guessing of the state after subjects see the information of their neighbors. Participants have complete information about the network structure and size, but only see their neighbors' genders and guesses of the state.[3] Each network consists of five

[2]The possible answers for each question are different in each session. We randomly select 4 out of 5 answers from the answer pool and add "(e) None of the above."

[3]Participants are told the number of network members and how many network neighbors each participant has. They are not shown an image of the network structure, nor told the gender composition of the network.

participants who participate in four repetitions of the trivia and updating task combination. Network positions and members are shuffled between repetitions, so that no participant has the same neighbor more than once. During the updating task, participants first guess the state by answering "For how many questions do you believe the answer was (e)?" The answers to this question will be called their *guess*.[4] Participants are then shown the guess of the state of each of their neighbors and can update their guess of the state based on this new information. The process of updating their guess is repeated 20 times.[5] The trivia task and updating task are repeated four times. In the Baseline treatment, participants see no gender information of their neighbors or themselves. In the Mixed and Same treatments, participants see their avatars and their neighbors' avatars throughout the updating process. An example of the screen each participant saw during updating is in Figure 3.

The Same and Mixed treatments are within subject. Two repetitions of the task are done in the Mixed treatment and two repetitions of the task are done in the Same treatment in a single experimental session. The order of Mixed and Same treatments is randomized at the session level.

In addition to varying the gender composition, we use two network structures in our experiment. The network structure is fixed for an experimental session. The network structures used are the circle network and the star network. These network structures are shown in Figure 4. In sessions with the circle network, each participant has two neighbors for each repetition of the task. Participants' positions and neighbors are reassigned between repetitions. In the star network, the participant in the center has four neighbors and is the only neighbor of all other participants. The participants in the center of the star are never the same participant between repetitions.

We chose these networks to be in line with previous research in the area, as well as to represent common group dynamics in real world settings. The circle network is one representative of equity, where each participant has the same amount of neighbors and same level of information from others. The star network on the other hand mimics more hierarchical professional settings, where one central member has all of the available information and all other network members report only to

[4]In the first round of updating we asked participants "What do you believe is the probability that you are correct?" in order to get baseline accuracy of subject's own guesses. We also asked "What do you believe is the probability that Neighbor X is correct?" for each neighbor, in order to get beliefs about their neighbors' accuracy. Both of these beliefs are uninformative to our results.

[5]After every five rounds, we elicit beliefs about each participant's neighbors' accuracy. These are uninformative to our results.

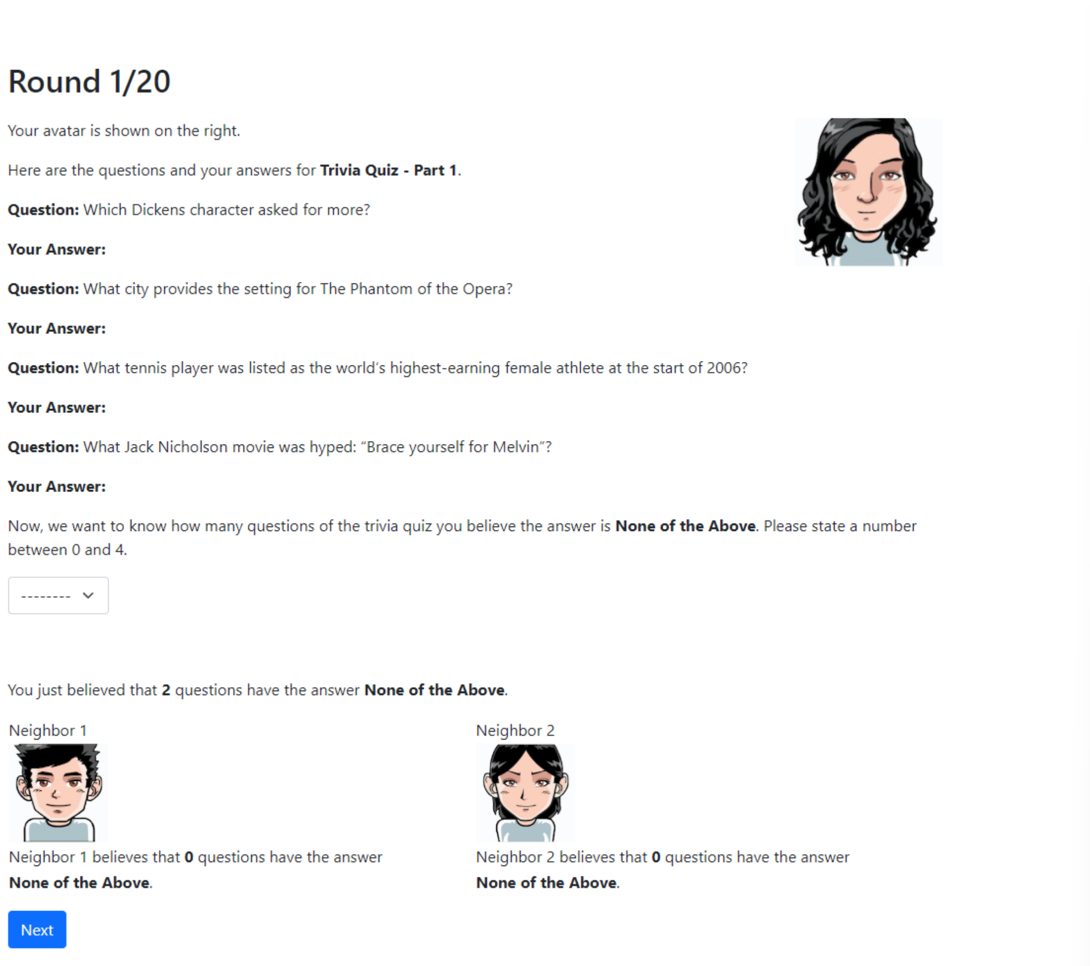

Figure 3: An example of a round of updating. The avatar in the top right corresponds to the person seeing the screen. The two avatars at the bottom are this participant's neighbors and what their guesses of the state are in that round of updating.

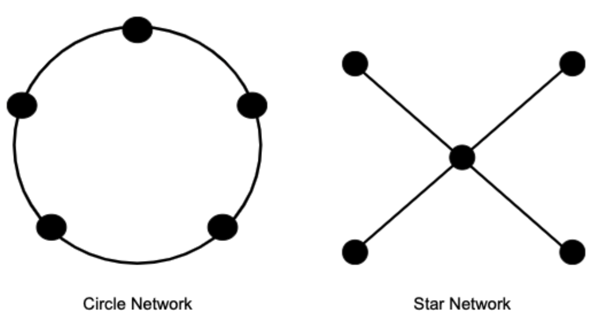

Figure 4: The network structures, circle (left) and star (right). Each participant in the circle network has two neighbors. One participant in the center of the star network has four neighbors, and the four peripheral star network participants only have the one center participant as a neighbor.

this central member. The two network structures offer the ability not only to analyze how information bias affects welfare, but the extent to which the amount of information available affects these biases as well.

Five participants are used in order for the Mixed treatment to have an uneven number of men and women in each network. This allows us to determine how the majority gender affects information transmission. In the Mixed treatment, the composition of a network is either 3 men and 2 women (Majority Men), or 3 women and 2 men (Majority Women). For a star network in the mixed treatment, a man is randomly selected as the center for the 3 men 2 women network while a woman is randomly selected as the center for the 3 women 2 men network.

All sessions were run at The Ohio State University in the Experimental Economics Laboratory. Data was collected for 64 star networks and 64 circle networks. 32 Baselines and 32 gendered treatments were run for each network structure.[6] All recruitment was done through ORSEE (Greiner, 2015). Payment was determined randomly for the session by one repetition of the task. Participants received an $8 show up fee regardless of performance. The trivia quiz was incentivized based on how many questions each participant answered correctly. $1 was given for each correctly answered question, plus $1 bonus if all four questions were answered correctly. Each participant earned $5 if they ended the updating task with the correct state.[7] One explicit belief was chosen randomly and a Multiple Price List was used to determine payment. $3, in total, was possible for belief elicitation payment. A total of $18 was possible for each participant and the experiment lasted an hour and a half. Average payment across all treatments was $12.48.

4 Results

As our main motivation of this experiment is to study difference in beliefs of ability of the genders, we must first establish whether actual ability differs in our chosen task. To do so, we looked at the average number of questions that were answered correctly by men and women across trivia quizzes. We pooled all sessions and gender treatments of the circle network and star networks. We find that there are no differences in the performance of men and women on the trivia task for any treatment

[6] The gendered treatment were within subject, so we have 16 Same networks and 16 Mixed networks.

[7] We understand this only incentivizes the last round of updating, where rounds 1 to 19 are cheap talk. This does not seem to be the case in our data, as rounds 1 to 19 do not show differences in results from that of the last round. Main results hold if we just look at the last round in isolation.

or network. Across the four repetitions of the trivia task, men on average answer 6.35 questions correctly and women answer 6.25 questions correctly in the gendered treatments. In the Baseline, men answer 6.2353 questions correctly and women answer 6.3478 questions correctly. The p-values for two-sample t-tests to compare trivia performance between men and women are shown in Table 1. We conclude that men and women on average perform equally well on these trivia tests (across treatments), which is consistent with Coffman (2014). Thus we may interpret any differences in the behavior of men and women in the updating task to be derived from other sources than true ability on individual the trivia quizzes.

	Gendered treatments	Baseline treatment	Baseline vs. Gendered Men	Baseline vs. Gendered Women
Circle Networks	0.3465	0.3659	0.4713	0.2944
Star Networks	0.8614	0.8525	0.8495	0.8641

Table 1: Two-Sample t tests for differences in trivia scores.

After the trivia quiz, subjects participated in the subsequent network updating task. As a measure of how well each network performed, we report whether the network reached a consensus. A network is said to have reached consensus if all members of the network have the same guess for the state at the end of the 20 rounds of updating of our task. Figure 5 shows the percentage of networks that reached consensus across all treatments. The are 32 networks in the Baseline treatment and 16 networks for both the Same and Mixed treatment. Here we see opposite effects of gender information on consensus rates in the star and the circle networks. For the star network, access to gender information during the updating task seems to decrease the rate of consensus. The opposite is seen with the circle network, where networks with access to gender information have increased consensus rates. These result suggest that the availability of gender identifiers possibly affects the transmission of information differently depending on the network structure. For the star network, gender information may hurt the ability of the network to reach consensus, whereas in the circle networks gender information may help networks reach consensus. We will discuss consensus rates in more detail in sections 4.1.2 and 4.2.2.

This difference in consensus rates suggests that the amount of information can largely impact how gender information interacts with network structure on an aggregate level. These results suggest that gender identity can both help and harm networks. The difference we see in how gender

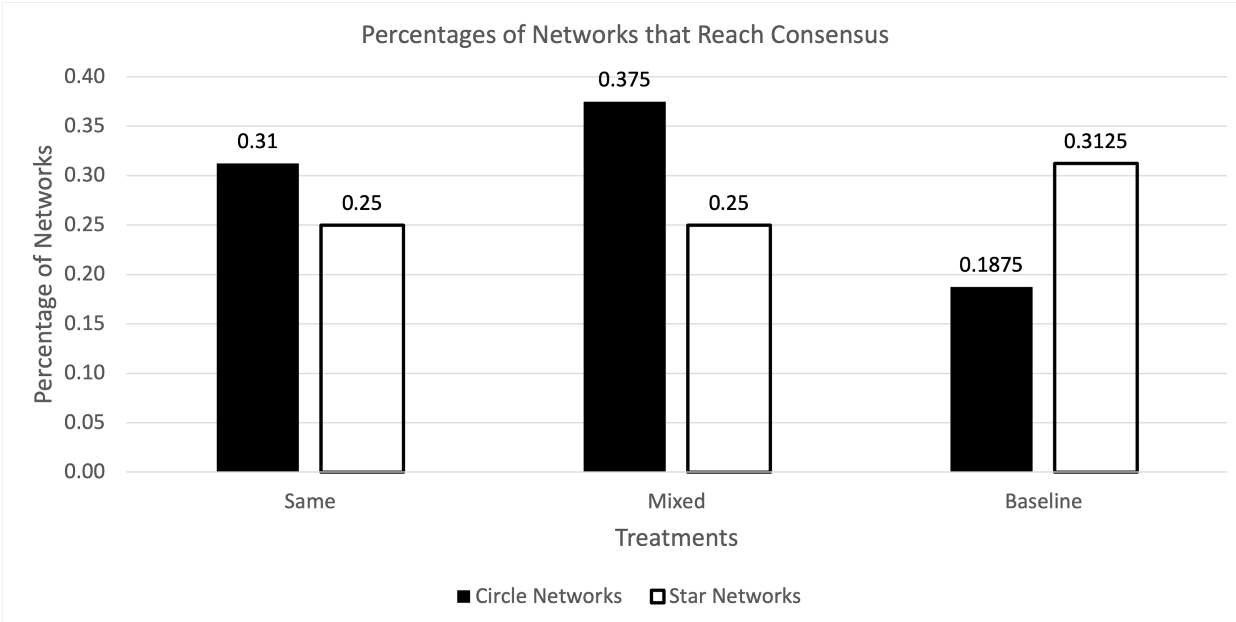

Figure 5: Percentage of networks that reached consensus for the circle and star networks.

impacts the star and circle networks caution that the type of organization structure of a work-place environment will largely impact the effectiveness of multi-gendered organizations. Since gender information affects the star networks and circle networks differently, we will report the results of these two treatments separately. All star network results are reported first then all circle network results.

Since network members are using their own guesses of the state and not being given any information from the experiment, there is no guarantee that networks have accurate information on which to update. Because of this, we may never see networks converge to accurate states regardless of how well or poorly information is moving across the network. One way to examine this is to look at the modal prior guess of the state of each network at the beginning of the updating task. If this modal prior is not the true state, networks may be updating with biased beliefs, in which we would not expect networks to converge accurately. In our networks, we use the modal response of each networks guess of the state as a measure of biased beliefs. We find that 48.4% of circle networks and 39% of star networks have unbiased beliefs. Since these numbers are less than half of the networks, it seems more likely that players have biased beliefs in our networks. Thus any results on accuracy would be affected by biased beliefs, which may never converge to the true state even with perfect information. Additionally, we find no significant differences between those who are correct in their guess of the state in the first round and last round of the updating task. The percentages of individuals with correct states in each treatment are listed in Table 2. We have pooled both men and women in the above results, but there are no significant differences between men and women and there are no significant differences between the first and last round accuracy for men or women in any treatment.[8]

	Circle		Star	
	Baseline	Gendered	Baseline	Gendered
First Round	41%	43%	29%	31%
Last Round	43%	43%	32%	36%

Table 2: Percentages of network members who have the correct state in the first and last round of updating.

4.1 Star Networks

There are three results we are most interested in for each network structure: number of network members in consensus, number of network members updating their guess of the state and how much each network member is weighting the information of their neighbors. In star networks, we find that substantial homophily, or the preference to weight information more from network neighbors of the same gender, exists. When men and women are updating, both are putting more weight on their same-gendered neighbors' guesses of the state. This homophily in belief updating is leading to Mixed treatment networks reaching consensus less often than Baseline networks. We will discuss homophily and how this affects network consensus in more detail in the next section.

4.1.1 Homophily in Star Networks

In order to discuss the aggregate welfare changes associated with gender composition of the networks, we first establish individual behavior differences between men and women in the updating task. We first look at willingness to update, which we will call switching. Here a single switch is when a network member updates their guess after seeing the guess of their neighbors. A switch does not have a direction, and is an indicator variable of 1 if a participant changed their guess of the state in a given period. Figure 6 shows the average number of switches across each treatment. A time trend version of this graph for each period of updating is available in the appendix. As there

[8]Chi-squared test p-values for differences in proportions of those correct in first and last round of updating. Circle: Baseline p-value 0.78207, Gendered p-value 0.9100; Star: Baseline p-value 0.5431, Gendered p-value 0.2862.

are many different binary comparisons we investigate for differences across treatments, we run all two-sided two-sample t tests using Bonferroni-Holm corrected p-values. The corrected p-values will be reported.

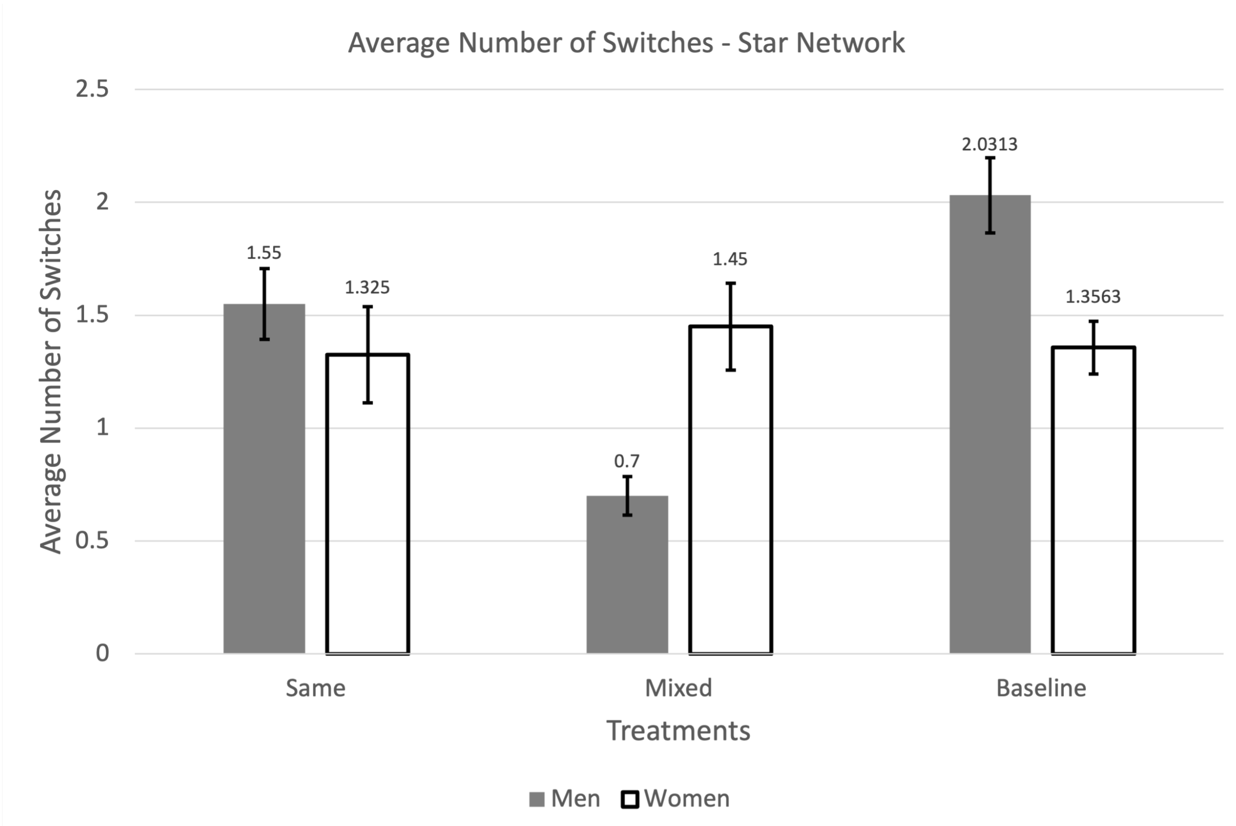

Figure 6: Average number of switches across each treatment in the star network.

Women's average updating behavior across treatments is indistinguishable, with 1.4 to 1.5 switches across the 20 periods on updating. Men's behavior is slightly more variable. Men in the Mixed treatment switch less often than men in the other treatments. This is only a significant difference when we compare to men in the Baseline treatment with a corrected p-value of 0.0097. Men's behavior in the mixed treatment is not significantly different than men in the Same treatment (p-value: 0.1564), which suggests that this change in men's willingness to update is likely due to the presence of gender information of both men and women in the network.

Since network members are shown their neighbors' guesses of the state prior to updating their own guess, we measure whether updating behavior is favoring certain network neighbors' revealed information. We do so by looking at the absolute difference of two network neighbors' guesses of the state in a given round shown in (1).

$$distance = |x_{own\, guess} - x_{neighbor's\, guess}| \tag{1}$$

If two network members have the same guess, this difference is zero. If two network neighbors have different guesses, this difference is positive. As network neighbor's guesses get farther away, this difference increases. For networks that reach consensus, we would see a distance, or difference, of zero. Network neighbors with the smallest distance are those whose information is being weighted most heavily. Figure 7 shows the average distance, or difference, of network members in each treatment based on the gender of their neighbors. A time trend version of this graph for each period of updating is available in the appendix.

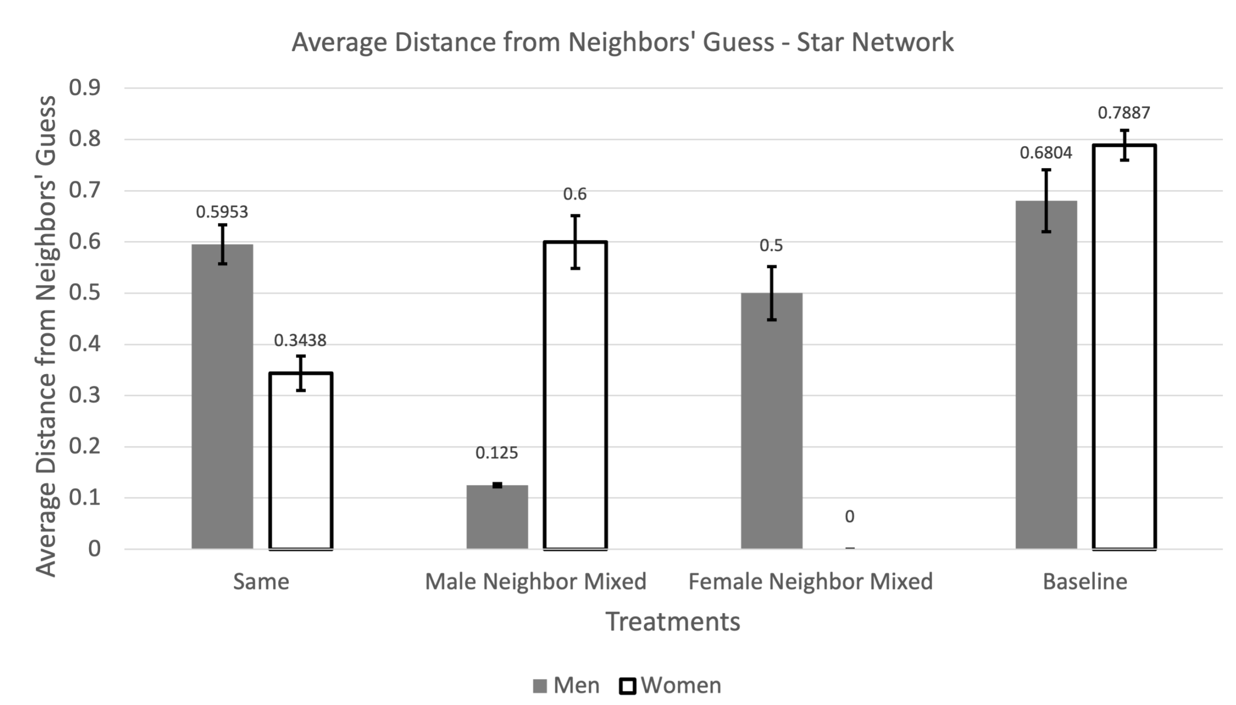

Figure 7: Average distance of each network member from their neighbor in star network.

As the Mixed treatment has both men and women, we specify the gender of ones neighbor. The bar represents whether a participant is a man or a woman, and then whether their neighbor is a man or a women is split into two different bars. For example, the gray bar above "Female Neighbor Mixed" means men who have neighbors that are women. This is not specified for the

Same treatment, as all of one's neighbors are their same gender.

On average, women and men in the Mixed treatment have similar guesses of the state as their same-gendered neighbors. Men are on average a distance of 0.125 from their male neighbors and women report the exact same value as their female neighbors. When we compare this difference between same-gendered neighbors and opposite-gendered neighbors we see significant differences with p-values of 0.0127 for men and 0.00004 for women. Here, the distance of both men and women from the opposite gender is much higher than that of the same gender. Both men and women are updating closer to same-gendered neighbors than opposite-gendered neighbors. Within the Mixed treatment, this result indicates that there is extensive homophily in the way that updating happens in the star network. When choosing to update guesses of the state, men and women are weighting the information of those the same to them in gender more so than those of different gender.

When we compare these results with the other treatments, we see that homophily is not present in other settings. The Baseline and Same treatments for men and women have no significant differences in their average distance from their neighbors. Men in the Same treatment and Men in the Baseline are the same distance away from their neighbors.[9] Similarly, women in the Baseline and women in the Same treatment are not significant different distances away from their neighbors.[10] Further, when we compare the weighting of information of both men and women to their opposite gender neighbors, so men with female neighbors and women with male neighbors in the Mixed treatment, most of the periods are insignificantly different from the behavior of men and women respectively in the Baseline.[11] This indicates that the homophily we see from the Mixed treatment is a true treatment affect, as the presence of men and women in a network impacts behavior significantly.

In the star networks, each non-center network member only has the center participant as their neighbor. Similarly, the center participant has all other network members as their neighbors. Here, the distance from the center to each of their neighbors and the distance from each network member to the center are the same. Because of this, Figure 7 would not change if we were to remove center participants from our distance analysis. However, we observe that the amount of updating is the

[9] t test Bonferroni-Holm corrected p-value: 1.4965

[10] t test Bonferroni-Holm corrected p-value: 1.2039

[11] t test Bonferroni-Holm corrected p-value for men: 2.1104 and women: 0.8226

same regardless of position in the network.[12] This will be discussed further in section 4.3. Thus, the homophily is not a result of changes in one's willingness to update. This is cautionary for organizations with a hierarchical structure that implement gender composition changes of employees in the highest position. From our results, a gender composition change could lead to homophily, but it could be difficult to see outwardly, as updating behavior is unchanged across treatments.

4.1.2 Does Homophily Affect Information Aggregation?

Star networks where homophily is present also have decreased ability to reach consensus in the updating task. We can measure the affect of the homophily on consensus in two ways. The first is a period-by-period time series of the size of the largest group of network members in a given network that has guessed the same state during each period of the updating task. For each period, we average the largest coalition size across all networks for each treatment. These groups can range from 0 to 5. In Figure 8, each line represents a different treatment of gender composition of the star networks. Theoretically, the star network is the most efficient network structure, and should have the fastest and highest rates of consensus (Grimm and Mengel, 2020). This is not what we find in our experiment. The number of network members in consensus does increase over time for all treatments except for the Same treatment where networks are only men. The difference in number of network members in consensus of All Men networks from the other treatments begins at period 12 and continues through the rest of the 20 periods. Although the All Men networks are decreasing in the number of network members in consensus each period, this difference is not significantly different from any other treatment.

Although fewer of the star networks with gender information reach consensus compared to the Baseline, we can see on average that the number of network members that are guessing the same state across all rounds of updating is not significantly different. We do see slight differences based on gender of the networks though. There is a divergence of male and female dominant networks, with Majority Women and All Women networks increasing to be the highest amount of consensus and All Men and Majority Men achieve lower levels of consensus, and even decreasing over time. This indicates that the gender composition of these networks does impact the aggregation of information on these networks.

[12] The similarity between center and peripheral networks members is available in Figure A3 in the appendix.

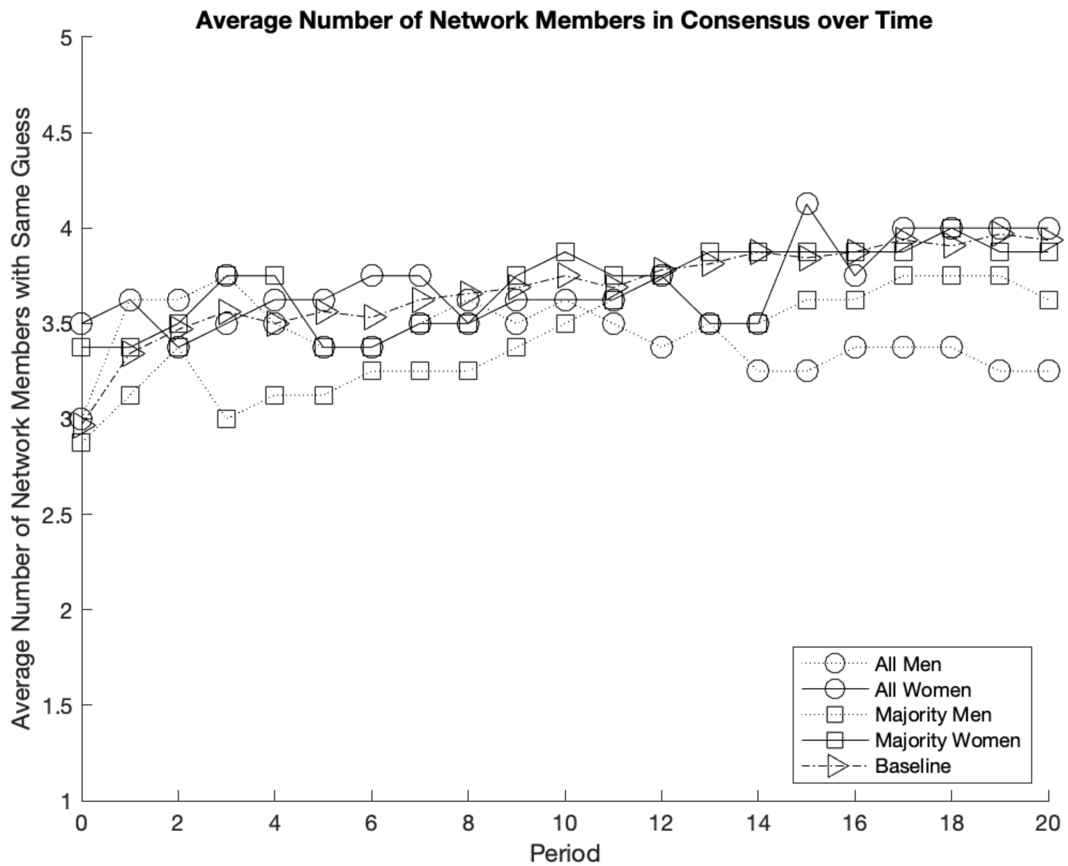

Figure 8: Time series of the maximum coalition of network members in the star network that are in consensus throughout the 20 rounds of updating. In each period, a point represents the highest number of members of a network who have stated the same guess for the state.

Additionally, we can look at the number of networks that reach consensus. In Figure 5, we find lower consensus rate in the star networks where gender information is available. We can further divide this graph into the gender composition of each treatment. Figure 9 shows that this is especially true for the Mixed treatment, where the homophily is present.

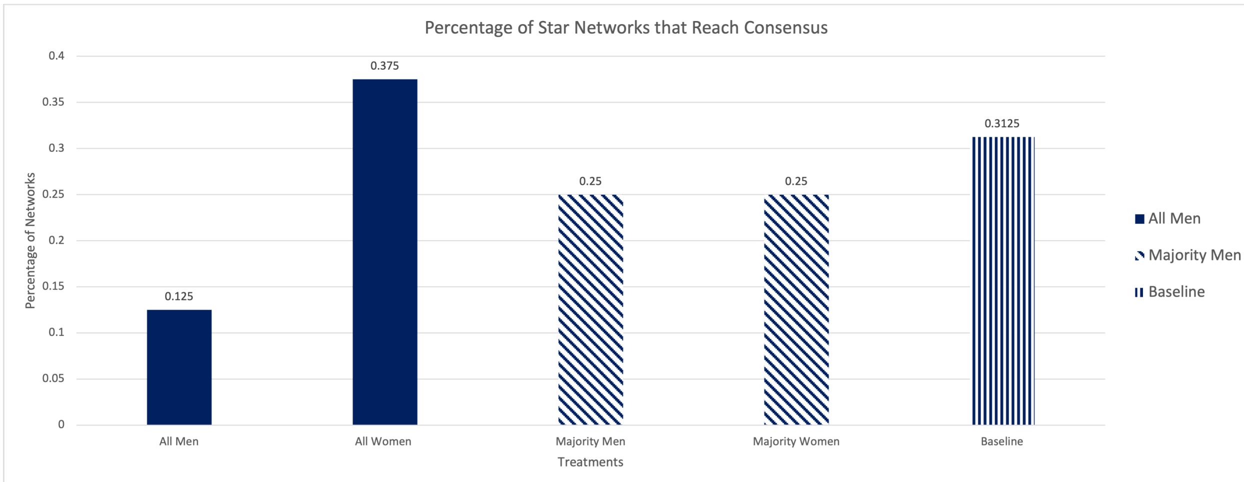

Figure 9: Percentage of star networks that reach consensus based on the gender composition of each network.

The difference between the Baseline and Mixed treatments are significant running a chi-squared test for difference in percentage of networks that reached consensus (p-value 0.0111). This difference highlights how strongly the homophily present in the Mixed gender networks affects whether the network reaches consensus. The homophily hurts Mixed treatment networks at the aggregate level, as they fail to reach consensus as often as networks without gender information. One possible implication of this is that in an organization where most workers have very limited information, anonymous decision-making processes or job rotation may be more efficient than a gender quota.

For the Same treatment, we see varying rates of consensus, with All Women networks reaching consensus at higher rates than all other treatments (p-value 0.321 for comparison to the Baseline). We see that All Women networks start with high numbers of network members in consensus, which drives this result.

These results are suggestive that changing gender composition, and highlighting these changes publicly, of hierarchical structured organizations may have unintended effects that lead to welfare reductions. Here, we see minor changes in willingness to transmit information, as updating behavior of women in all treatments is consistent, but men switch less only when in Mixed networks. This small behavior inconsistency could be cautionary foresight for the same-gender bias we see present in information weighting. When we look closer into which neighbors information transmission is favoring, we see that there's substantial homophily in the star structured networks when both men and women are present. This is cautionary for organizations with a similar structure, as this same-gender bias is leading to declines in the ability of full networks to converge to the true state. So when putting into place gender quotas, or changing the gender composition of stereotypically single-gendered organizations, there should be awareness that persistent homophily bias may lessen the cooperative nature of the organization. Ultimately, these results suggest that in organizations of this nature, that careful monitoring and adressment of such homophily should follow any gender composition shift. If policies are not put in place to acknowledge this bias, there could be declines in organizational productivity.

4.2 Circle Networks

In the circle networks, we no longer find the homophily that is present in the star networks. Rather, we now see differences in men's behavior across treatment based on whether gender information is revealed. In the Baseline treatment, men are much less willing to update their guess of the state compared to men in either the Mixed or Same treatments. Here, we find that men update their guess of the state significantly more often when gender information is shown. The increased updating of men is not coupled with a preference towards men, as in the star networks. We find that the men's increased updating equally weights information from their neighbors, regardless of gender. This increased updating is leading to higher consensus of networks with gender information when compared to the Baseline. The next section further discusses the changes in updating behavior of men across treatments and how this impacts network consensus rates.

4.2.1 Switch Rates of Men in Circle Networks

For the circle network, we see a large change in network structure. Each network member now has the same number of network neighbors and most members have more neighbors than in the star networks. This is in direct contrast from the previous star networks, where one network member has all available information and all the other network members only have one neighbor. Theoretical network predictions find that the circle network should be less efficient and converge at a slower rate than the star network (Grimm and Mengel, 2020). Instead we find that the circle networks with gender information have the highest rates of consensus across all treatments, including the

Baseline star.

Unlike the star networks, we no longer see homophily in the circle networks. Rather, we see a large difference in the behavior of men in the treatments with gender information versus the baseline. In the circle networks, we find that now gender information assists our networks in updating, as men are more likely to update their guesses of the state towards that of their neighbors when gender information is shown. We call this switching one's guess. We measure the willingness to update one's guess of the state through the switch rate. Figure 10 shows the switch rate of each treatment. A time trend version of this graph for each period of updating is available in the appendix.

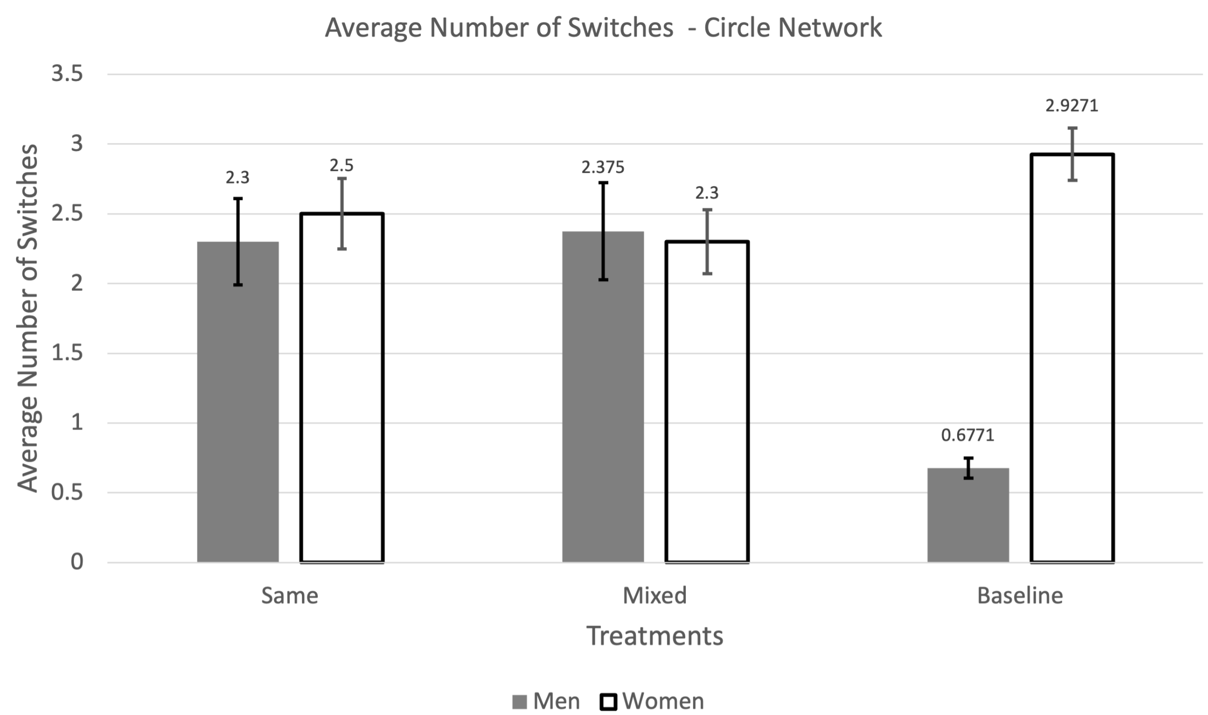

Figure 10: Average number of switches across each treatment for circle network.

Figure 10 shows the average number of switches per network member for each treatment in each round. There is a low average switch rate of only 11%, similar to the star networks. Although there are low number of switches across all treatments, we do still see differences in behavior. Most notably, men in the Baseline have the lowest number of switches on average. This low switch rate is unique to men in the Baseline treatment, as men in the Baseline treatment have lower numbers of switches than men in either treatment with gender. This difference is significant when comparing the Baseline men to Mixed treatment men (corrected p-value: 0.0421) and when comparing the Baseline men to the Same treatment men (corrected p-value: 0.0296). Men's switch rate in the Baseline is also significantly lower than that of women in the Baseline (corrected p-value: 0.0004). Additionally, there is no difference between the updating behavior of men in the Same treatment versus the Mixed treatment, which highlights the behavior difference of a low switch rate for men in the Baseline. Similarly, there are no significant differences between behavior of men and women in each treatment with gender information. Women behave indistinguishably across all treatments.

When comparing the switch rates of men in the circle network with men in the star network, we do not see the difference in Baseline and gendered treatments. In the star network, men only have lower willingness to switch in the Mixed treatment, whereas in the circle network we see that men without gender information switch at much lower rates. This highlights the interaction we see between the presence of gender information and the network structure, especially for men's switching behavior.

This suggests that the presence of gender information makes men more likely to update their own guess of the state in the circle network, whereas it does not change the behavior of women. This difference is an increase in number of switches per period, which could be contributing to the higher levels of consensus for the circle network compared to the star from Figure 5. With men more willing to update their own guesses of the state based on the information they are shown from their neighbors while women stay constant, there's more likely to be consensus among these networks.

We can further investigate whether this change in switching behavior of men is leading to bias in information transmission. Figure 11 shows the average distance of one's guess of the state from their neighbors' across each treatment. Here we find that there is no gender discrimination in this increased updating, as men do not favor any one gender over the other.

We do not see any direct gender discrimination for men or women across any treatment. Instead, we see that the high switch rates of men across gendered treatments previous discussed is leading men to be closer to their neighbors, regardless of gender. There is no significant difference

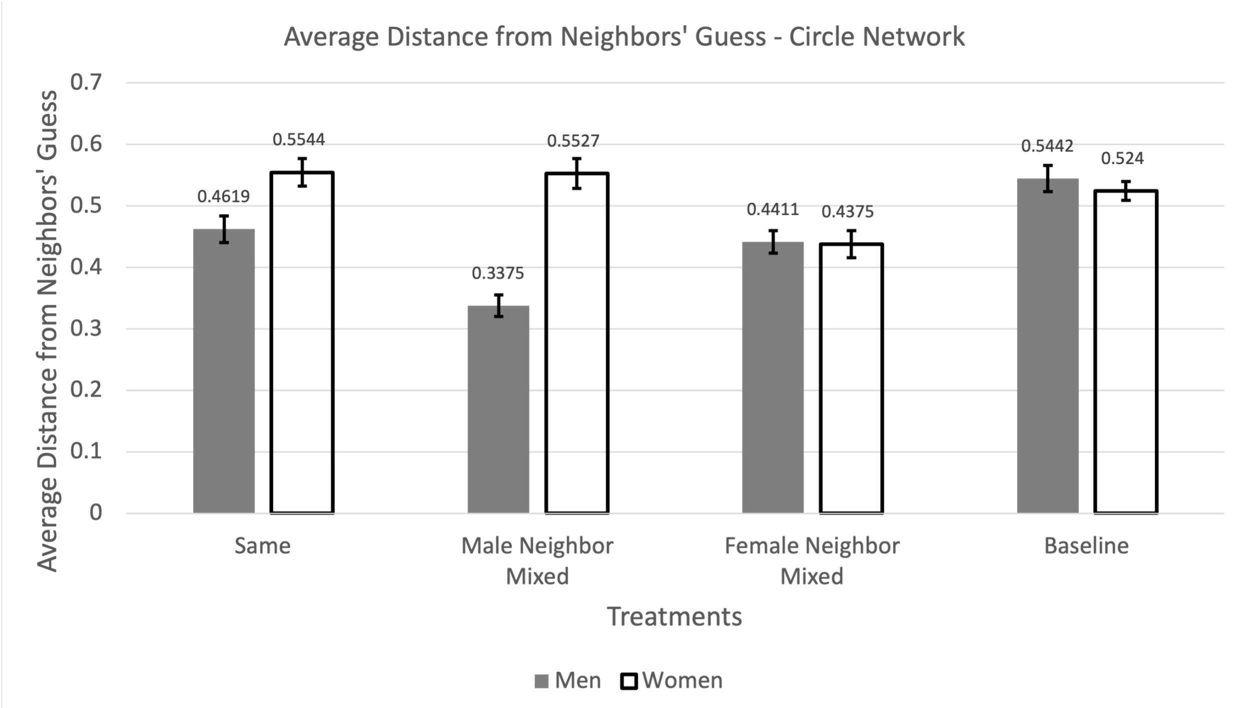

Figure 11: Average distance of each network member from their neighbors in circle networks.

for men or women across any of the treatments. So although we see increased switch rates, it is not leading to any gender bias or discrimination. Rather this is facilitating the networks with more men to reach higher levels of consensus. Women's behavior across all treatments is pretty similar and there are no significant differences. Women tend to switch less and be farther away from their neighbors than the men in these circle networks, but this is not a significant difference from any other treatment. When we see men and women in the Mixed treatment, these two behaviors work in tandem to create the highest number of networks that reach consensus across all networks and treatments. Seemingly this is because the women are staying consistent with their guesses of the state, but when gender information is revealed, the men are willing to switch their answers more often.

The difference in switching behavior of men when gender information is shown suggests that the organizational structure matters for how gender composition changes affect behavior. Here, the switching increases in the circle networks signal optimistically for organizations that have more equitable structures that gender composition changes could have a positive impact. As men are more likely to use the information of their neighbors, when implementing changes to the gender of an organization with such a structure it may be beneficial to highlight the personnel themselves, rather than the changes in composition.

4.2.2 Do Switch Rate Increases by Men Affect Information Aggregation?

We first examine the consensus rate in each period. In Figure 12, in each period we show the largest number of network members in consensus for their guesses of the state. There are no significant differences between the treatments, but there are however qualitative differences between the treatments based on gender composition. Here, we have that the treatments with All Men or Majority Men are the ones where consensus is highest amongst all treatments. The networks with All Women and Majority Women have the lowest levels of consensus for the early stages, but increase quickly towards the end of the updating period.

When comparing the circle networks to the star networks, we do not see much difference in the middle periods of updating, but the first and last periods do see differences. In the circle networks, fewer members are in consensus to begin. This is followed by a quick increase in the first couple of periods for most treatments, whereas the star networks see more network members in consensus to begin with a slower, steadier increase across all rounds. In the last periods, the Majority Men network in the circle treatment has high degrees of consensus, higher than all treatments in the star network. The All Men treatment also has high numbers of network members in consensus in the circle networks. These are the exact opposite results in the star network, where the Majority Men and All Men are the lowest two treatments, with All Men even seeing decreases in the number of network neighbors in consensus over time. Majority Women, All Women, and the Baseline are pretty consistent across these two network structures. These results suggest that the network structure impacts networks most with more men than women.

Figure 13 shows the percentage of networks that reach consensus in the last period of updating. Here, the percentage of all networks with gender information in consensus is higher than that of the Baseline, although none of significantly higher than the Baseline. The increased switching behavior of men across gendered treatments may be leading to these higher levels of consensus. In support of this conclusion, consensus rate increase is highest when there's majority men in the network. With the increased willingness to update from men and the stationarity of women, these networks

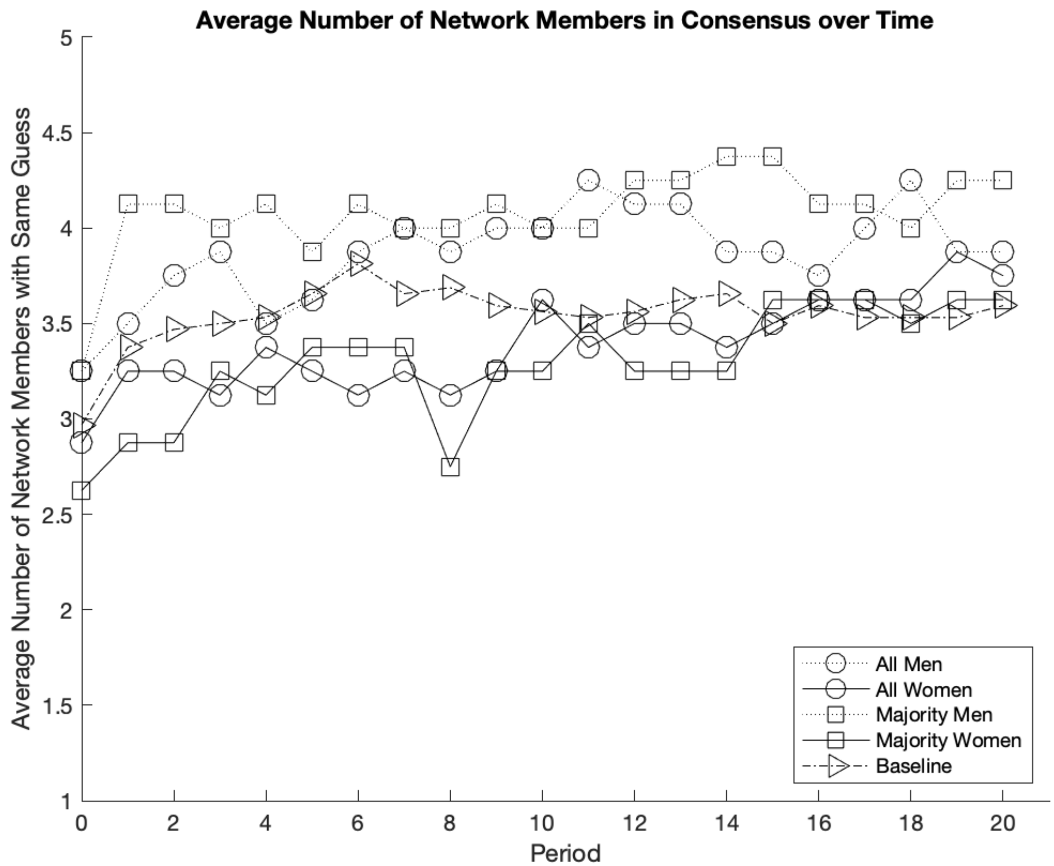

Figure 12: Time series of the maximum number of network members in circle network that are in consensus throughout the 20 rounds of updating.

are more likely to converge to the same state at the end of the updating task.

In circle networks, we now see clear behavioral shifts as gender composition changes, where men are more willing to update their guess of the state with the presence of gender information. This suggest that for more equitably structured organizations putting emphasis on gender composition changes, such as gender quotas, leads to increased productivity at an organization level. Here we see that circle networks with gender information are more likely to converge to a single state at the end of updating. Organizations with a structure similar to circle networks may benefit from policies that highlight gender composition changes by making organization members, specifically men, more familiar with and aware of those with whom they are working. Having opportunities for members of these organizations to meet face to face with other team members can lead to an increase efficiency of the teamwork, as willingness to work together may be increased due to

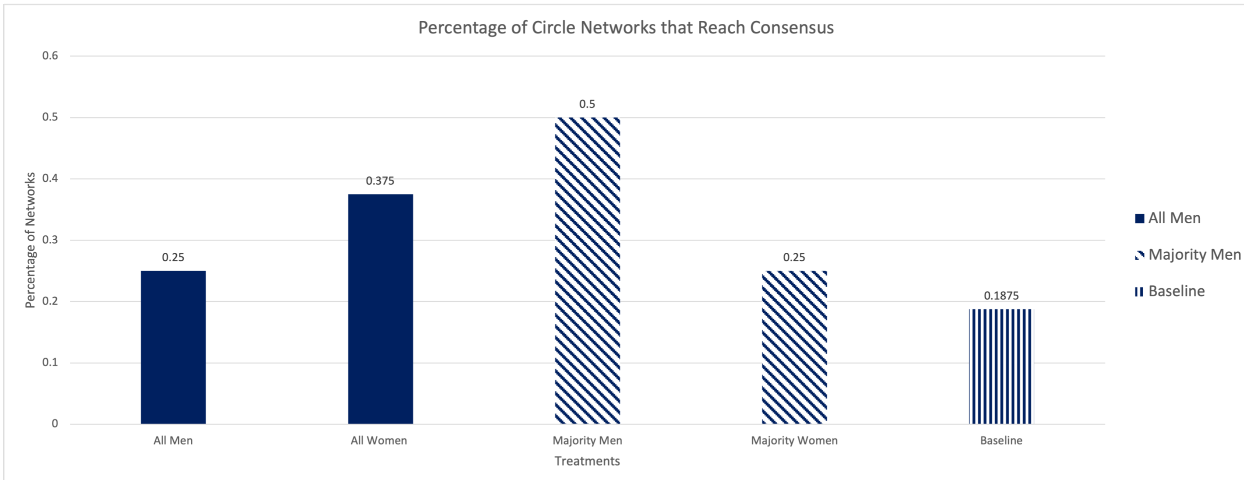

Figure 13: Percentage of circle networks that reach consensus based on the gender composition of each network.

heightened familiarity.

This behavior is starkly different than that of the star network. The star network saw large levels of homophily, whereas now we no longer see any gender favoring another in any treatment. This indicates that the network structure differences between the circle network and star network are leading to differences in how men and women view their network neighbors. Seemingly, the circle network, with more information for most network members and equal information for all network members, is leading to less gender discrimination than the star network, where network members largely have to only trust the center participant. This difference could signal that the number of network neighbors and amount of information a network member has could largely impact how likely gender discrimination is to occur.

5 Discussion

Our results suggest that the network structure greatly impacts the type of interaction we observe between men and women in the updating task. This individual behavior difference we see leads to aggregate differences, where gender can help or harm a network depending on the network structure. In the star network, where most members have limited information both men and women tend to weight the information of their own gender more than that of others. Due to the limited information one has, it could be plausible that we see individuals gravitating towards other network members who share common characteristics as them, which here is only gender.

When we change the network structure to the circle network, where there is equal information and more network neighbors, we see the homophily present in the star network vanishes completely. Men and women weight their network neighbors' information now almost perfectly evenly, without favoring same-gendered individuals. Instead, we see that the gender information actually increases the likelihood of men to put weight on the information of their neighbors. We find that in both the Same treatment and the Mixed treatment, men are updating based on their neighbors' guesses of the state more than the men in the Baseline. Women's behavior is consistent across all settings. This consistent behavior from women and the increased switching of men leads to higher rates of consensus for treatments where gender is available, compared to the Baseline.

These results are suggestive that the network structure, and more specifically, the number of network neighbors one has, could be integral to lessening gender preference for one's own gender. The differences between the star network and circle network seem indicative that the amount of information one has available interacts with gender compositions differently to determine how information is transmitted across the network. This is similar to previous research on gender discrimination, where gender differences disappear when information is increased. Here, we liken the change in network structure and increase in network neighbors as we move from star to circle networks to an increase in available information for participants. This increase in information seems to correct the homophily bias we find when information and network neighbors are limited. This suggests that when there's not enough information, men and women are more likely to put weight on information that comes from a network neighbor in which they are similar. Theoretically, the star networks should be more efficient but we find that this limited information and homophily that is resultant actually leads to lower consensus rates than the circle networks. This interaction of network structure with gender leads to welfare loss due to thereotical predictions. This increase in network neighbors and the network structures role in changing gender discrimination could be an interesting area for further research as the prevalence of gender quotas increases.

When considering gender composition changes, organizations should be managing these changes differently based on their structure. When organizations have a hierarchical structure, making these changes less publicized may help aid in the transition. Additionally, these organizations should keep watch not on the outward behavior of men and women, but rather if their willingness to be attentive and listen to their team members changes when leadership gender changes are issued. If the underlying homophily is not managed, this can lead to decreased aggregate welfare for the organization. In contrast, when organization are making gender composition changes in more equitable spaces, the highlighting of the individuals that are moving into such roles may lead to a more efficient workplace environment. The aggregate effects of making such transitions public may lead to increases in male team members willingness to cooperate.

References

Aguiar, F., Brañas-Garza, P., Cobo-Reyes, R., Jimenez, N., and Miller, L. M. (2009). Are women expected to be more generous? *Experimental Economics*, 12:93–98.

Bertrand, M. and Hallock, K. F. (2001). The gender gap in top corporate jobs. *ILR Review*, 55(1):3–21.

Blau, F. D. and Kahn, L. M. (2000). Gender differences in pay. *Journal of Economic perspectives*, 14(4):75–100.

Bordalo, P., Coffman, K., Gennaioli, N., and Shleifer, A. (2019). Beliefs about gender. *American Economic Review*, 109(3):739–73.

Campbell, K. E. (1988). Gender differences in job-related networks. *Work and occupations*, 15(2):179–200.

Chandrasekhar, A. G., Larreguy, H., and Xandri, J. P. (2020). Testing models of social learning on networks: Evidence from two experiments. *Econometrica*, 88(1):1–32.

Coffman, K., Flikkema, C. B., and Shurchkov, O. (2021a). Gender stereotypes in deliberation and team decisions. *Games and Economic Behavior*, 129:329–349.

Coffman, K. B. (2014). Evidence on self-stereotyping and the contribution of ideas. *The Quarterly Journal of Economics*, 129(4):1625–1660.

Coffman, K. B., Exley, C. L., and Niederle, M. (2021b). The role of beliefs in driving gender discrimination. *Management Science*, 67(6):3551–3569.

Croson, R. and Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic literature*, 47(2):448–474.

European Commission (2022). Gender equality: The eu is breaking the glass ceiling thanks to new gender balance targets on company boards.

Evdokimov, P. and Garfagnini, U. (2023). Cognitive Ability and Perceived Disagreement in Learning. Rationality and Competition Discussion Paper Series 381, CRC TRR 190 Rationality and Competition.

Gersick, C. J., Dutton, J. E., and Bartunek, J. M. (2000). Learning from academia: The importance of relationships in professional life. *Academy of Management Journal*, 43(6):1026–1044.

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with orsee. *Journal of the Economic Science Association*, 1(1):114–125.

Grimm, V. and Mengel, F. (2020). Experiments on belief formation in networks. *Journal of the European Economic Association*, 18(1):49–82.

Ibarra, H. (1993). Personal networks of women and minorities in management: A conceptual framework. *Academy of management Review*, 18(1):56–87.

Mengel, F. (2020). Gender differences in networking. *The Economic Journal*, 130(630):1842–1873.

Mengel, F. (2021). Gender bias in opinion aggregation. *International Economic Review*, 62(3):1055–1080.

Niederle, M. and Vesterlund, L. (2007). Do women shy away from competition? do men compete too much? *The Quarterly Journal of Economics*, 122(3):1067–1101.

Weizsäcker, G. (2010). Do we follow others when we should? a simple test of rational expectations. *American Economic Review*, 100(5):2340–60.

Wiswall, M. and Zafar, B. (2018). Preference for the workplace, investment in human capital, and gender. *The Quarterly Journal of Economics*, 133(1):457–507.

A Supplemental Material

All graphs in the following sections are time series that correspond to the average graphs shown in the main text. Figures A1-A7 are time series graphs for star networks and Figures A8-A10 are time series graphs for circle networks. Figure A1 is the time series that corresponds to Figure 8 just with error bars included.

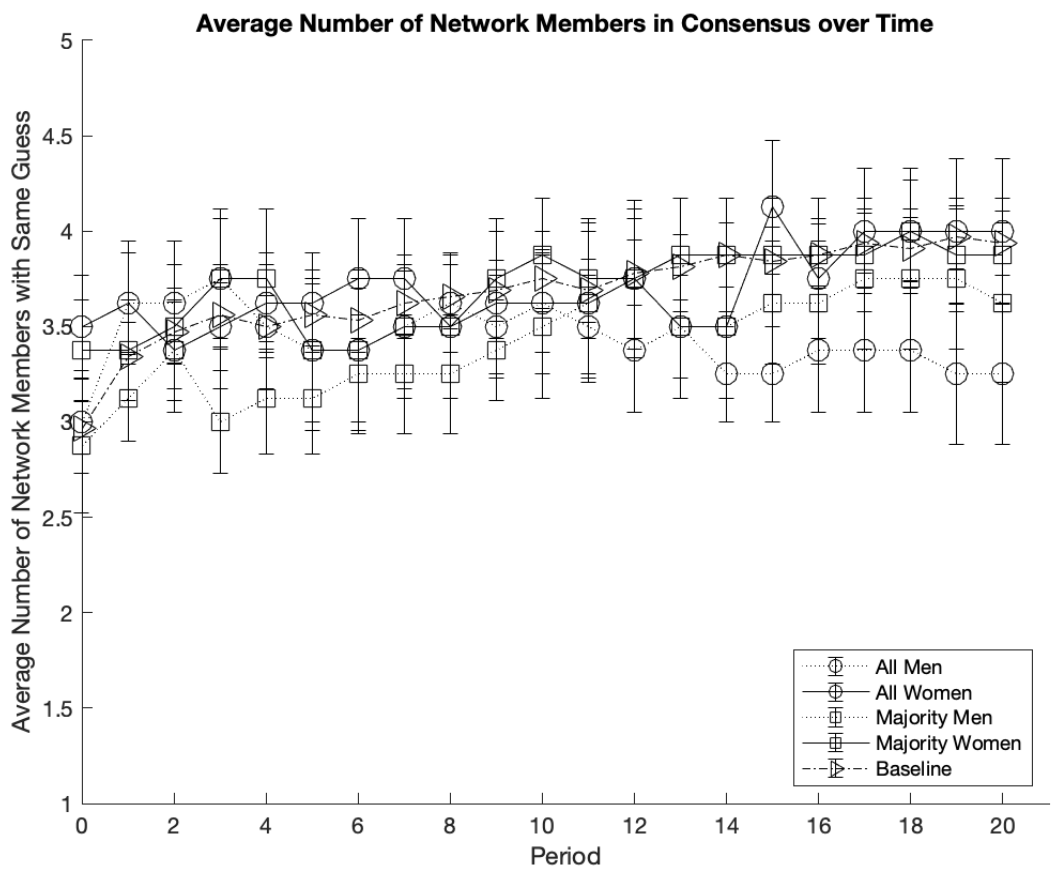

Figure A1: Time series with error bars for number of network members in consensus for star networks.

Figure A2 is the time series graph of average switches in each period of updating for star networks. This is the time series graph that corresponds to Figure 6 in the main text. Each line corresponds to one of the bars in the main text figure.

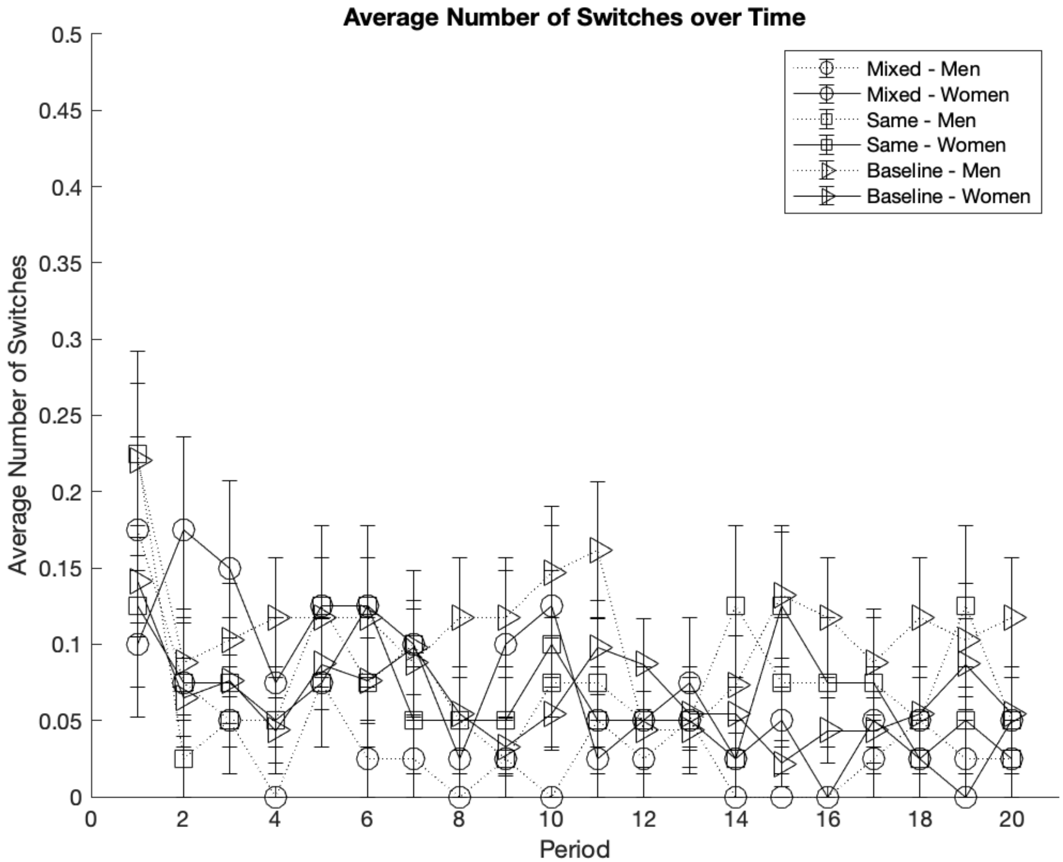

Figure A2: Time series with error bars for average number of switches in each period for star networks.

Figure A3 does not correspond to any graph in the main text, but rather shows the time series data for differences between the center and peripheral network members in the star network. Figure A3 shows the center compared to peripheral network members across each of the three treatments. Figure A4 shows the difference of center and peripheral members according to their gender for the Baseline treatment. A5 shows the difference of center and peripheral members according to their gender for the Same treatment. And A6 shows the difference of center and peripheral members according to their gender for the Mixed treatment.

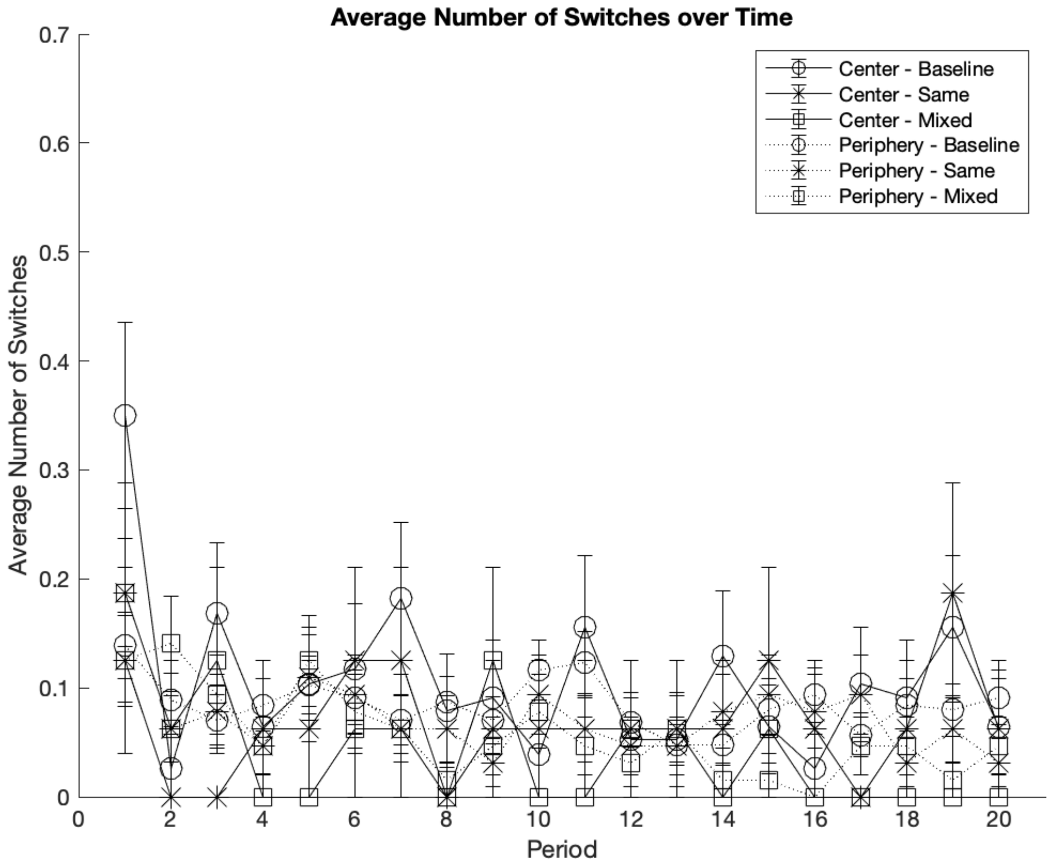

Figure A3: Time series with error bars for average number of switches in each period for star networks, center participants vs. periphery participants

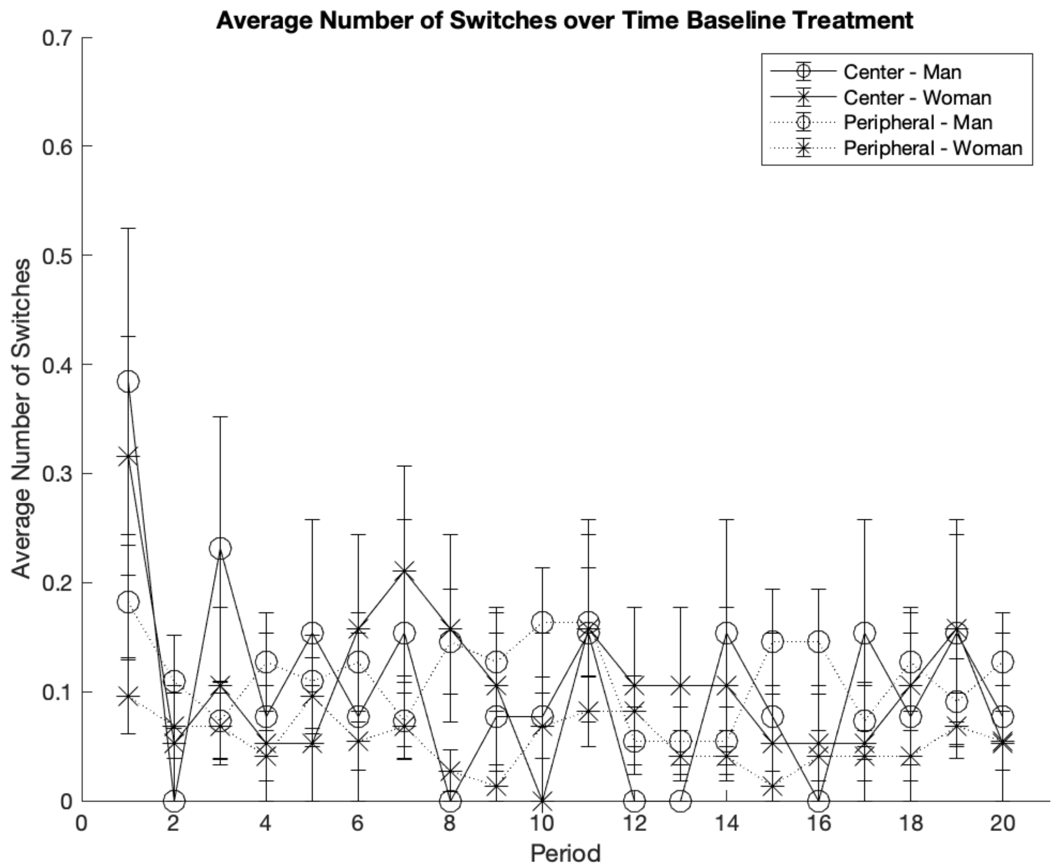

Figure A4: Time series with error bars for average number of switches in each period for star networks, center participants vs. periphery participants Baseline treatment

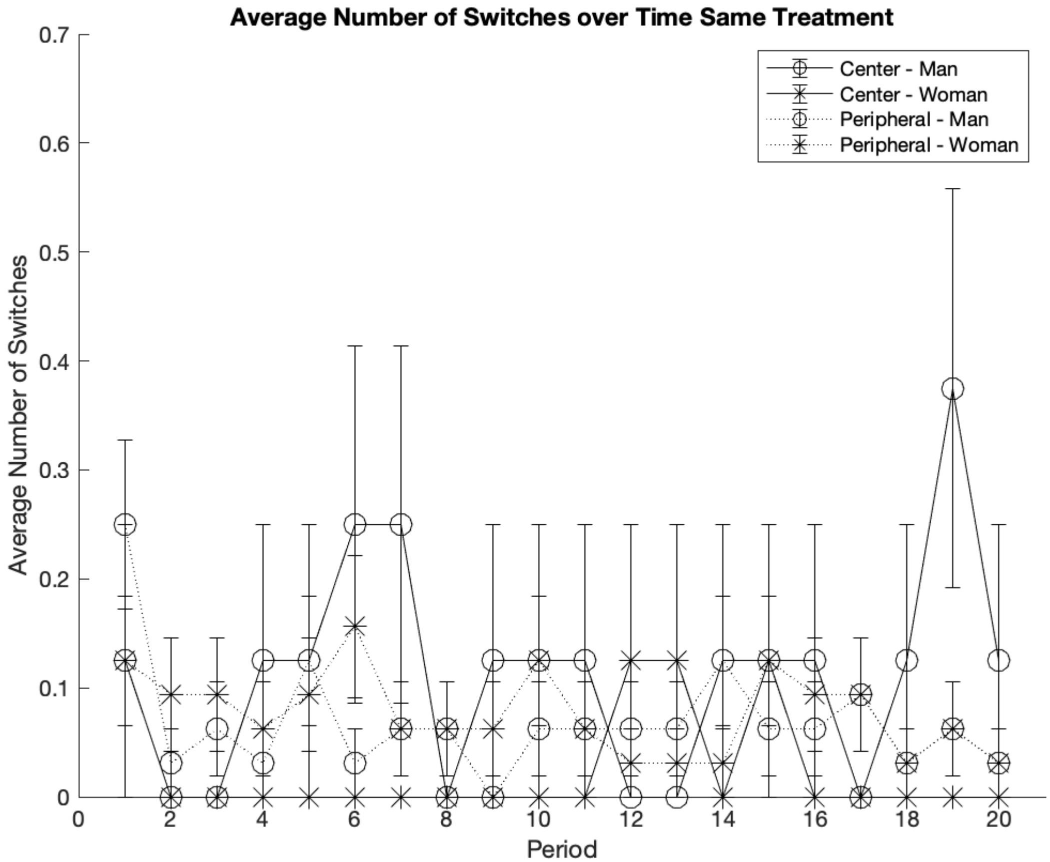

Figure A5: Time series with error bars for average number of switches in each period for star networks, center participants vs. periphery participants Same treatment

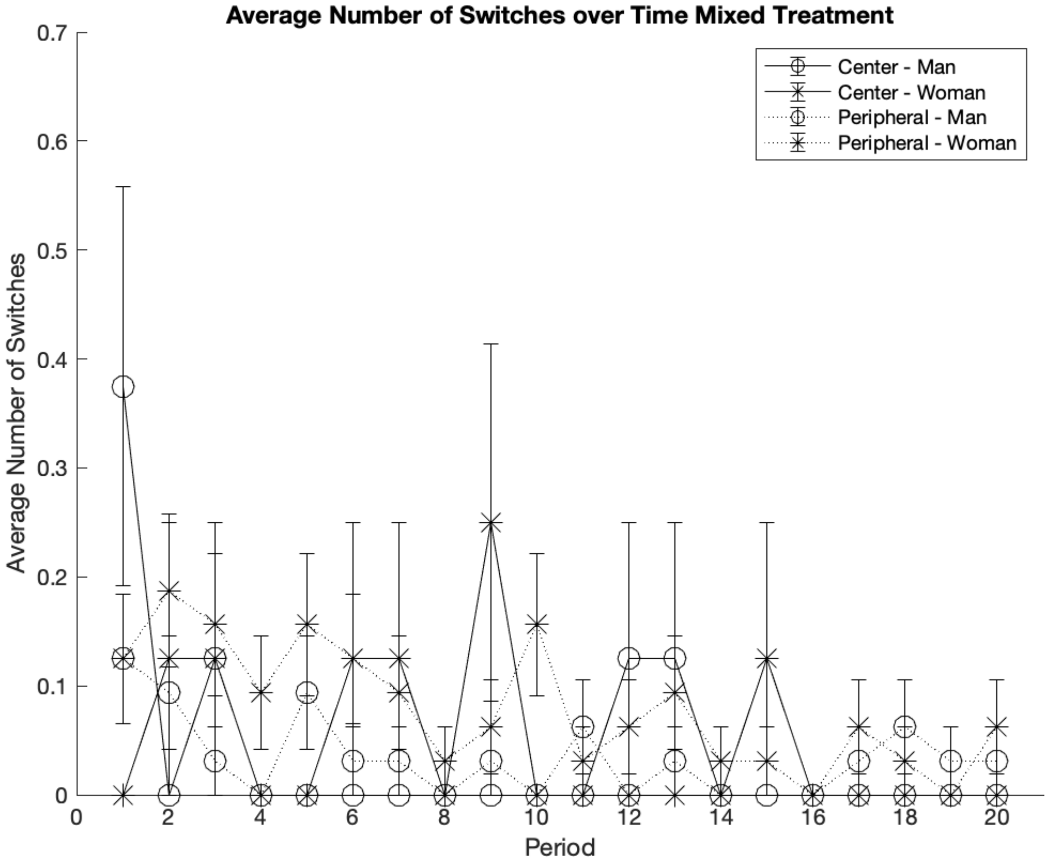

Figure A6: Time series with error bars for average number of switches in each period for star networks, center participants vs. periphery participants Mixed Treatment

Figure A7 is the time series graph that corresponds to Figure 7 in the main text. This time series shows the average distance of a participant's guess of the state from that of their neighbors across treatment for star networks. The labels for the Mixed treatment are slightly different, but each line corresponds to a bar on the main text graph.

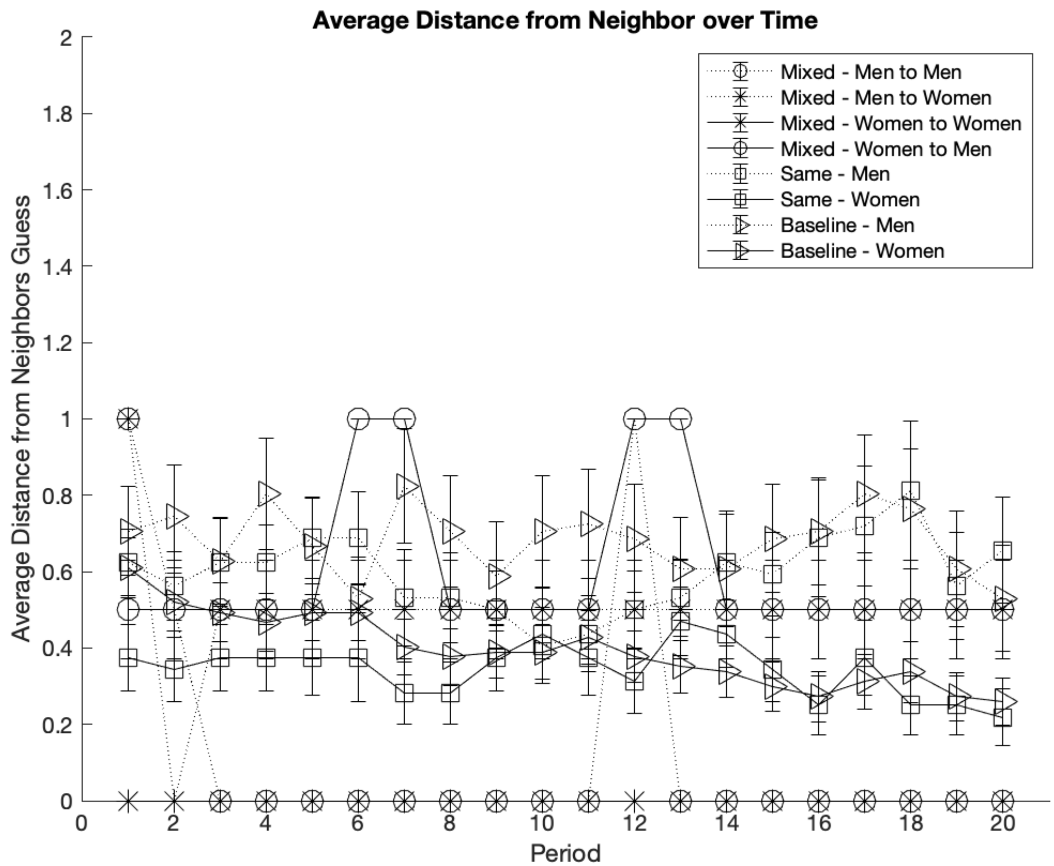

Figure A7: Time series with error bars for average distance of network members from their neighbors for star networks.

Figure A8 is the time series that corresponds to Figure 12 just with error bars included.

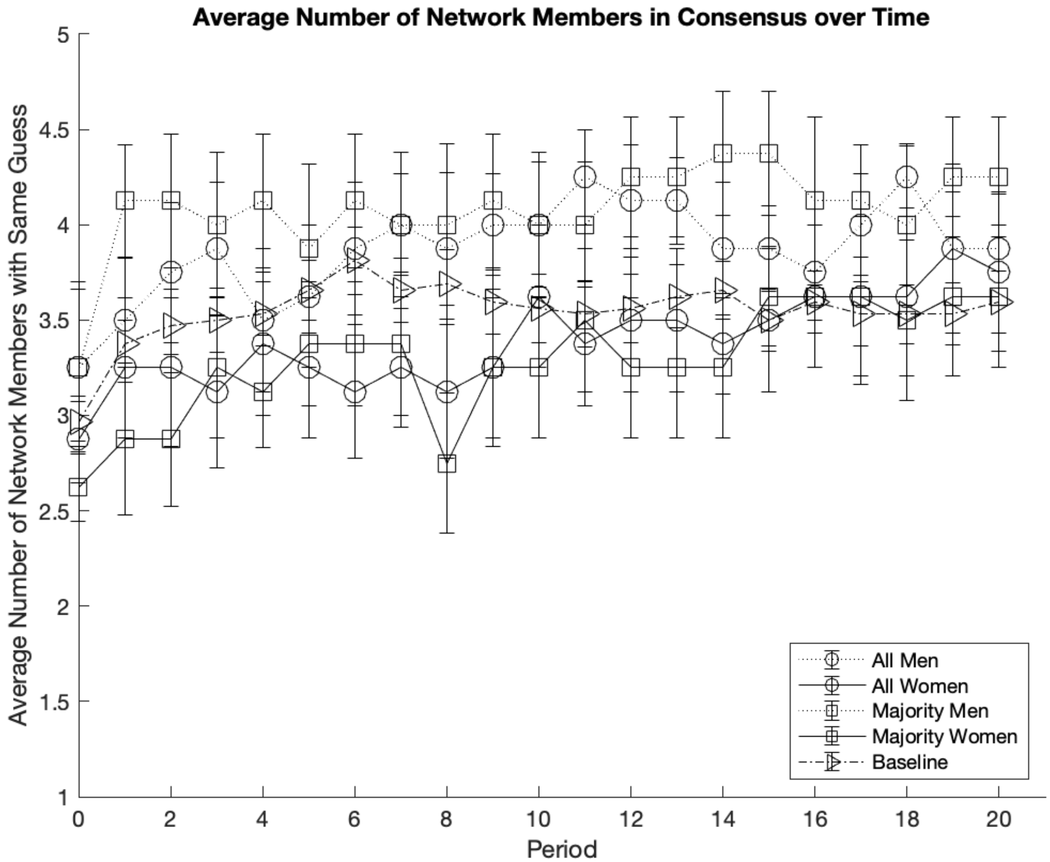

Figure A8: Time series with error bars for number of network members in consensus for circle networks.

Figure A9 is the time series graph of average switches in each period of updating for circle networks. This is the time series graph that corresponds to Figure 10 in the main text. Each line corresponds to one of the bars in the main text figure.

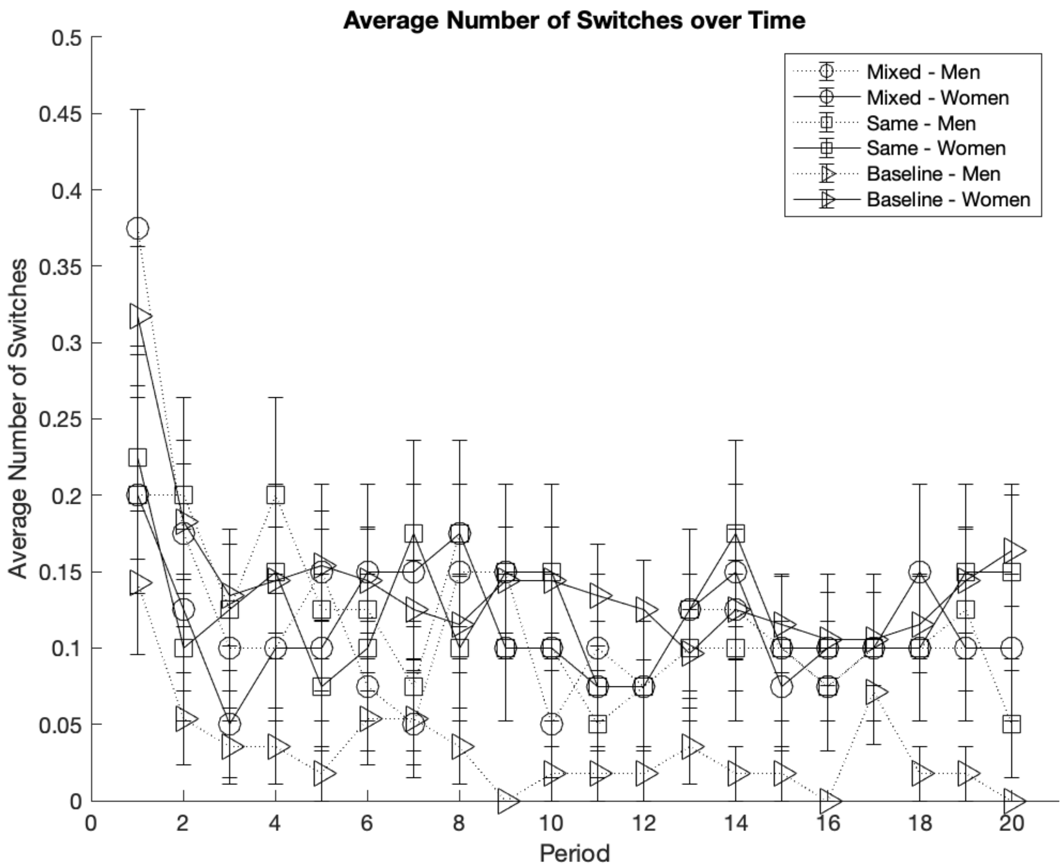

Figure A9: Time series with error bars for average number of switches in each period for circle networks.

Figure A10 is the time series graph that corresponds to Figure 11 in the main text. This time series shows the average distance of a participant's guess of the state from that of their neighbors across treatment for circle networks. The labels for the Mixed treatment are slightly different, but each line corresponds to a bar on the main text graph.

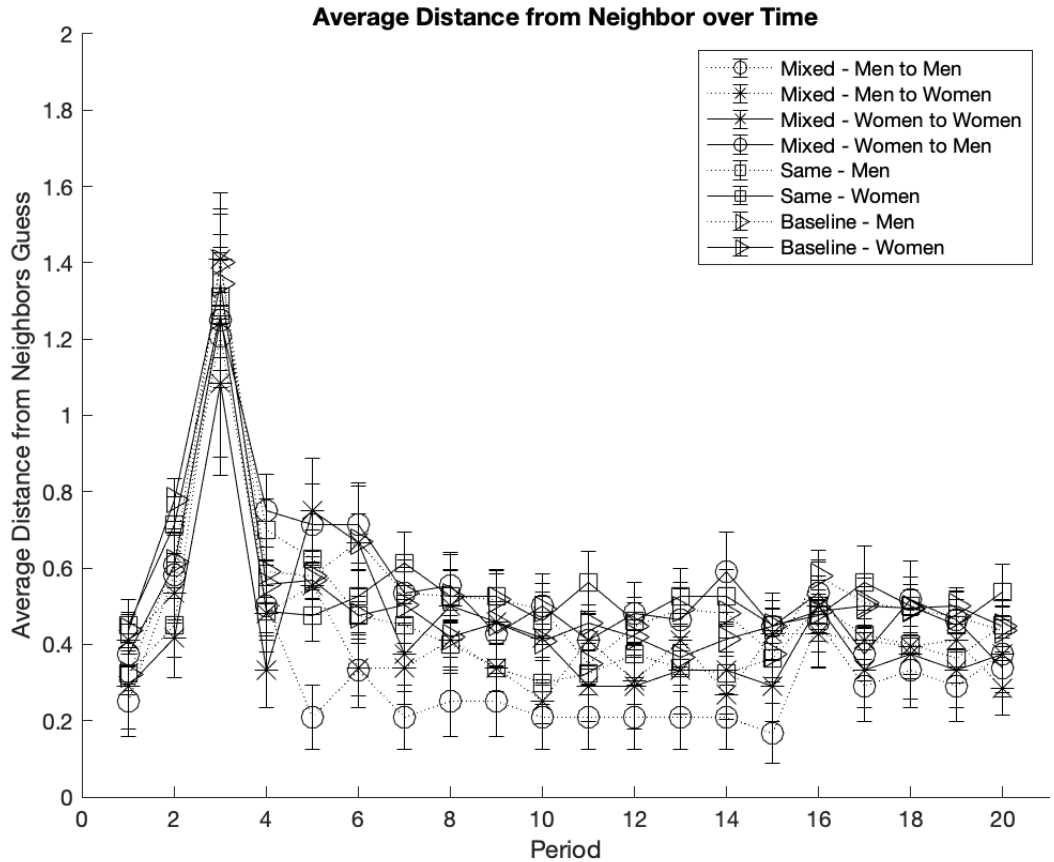

Figure A10: Time series with error bars for average distance of network members from their neighbors for star networks.