Utrecht University

Master Thesis

Contemporary Social Problems

Eleni Vittoria

June 2018

Preventing the spread of fake news:

Is diversity an adequate mechanism that enhances crowd wisdom and reduces the spread of fake news in social networks?

Master thesis

Sociology: Contemporary Social Problems

Utrecht, June 29, 2018

Author: Eleni Vittoria (6173217)

Preventing the spread of fake news: *Is diversity an adequate mechanism that enhances crowd wisdom and reduces the spread of fake news in social networks?*

Thesis Supervisor: Prof. Dr. Arnout van de Rijt

Second Reader: Dr. Zoltán Lippényi

Preface

This thesis is written to fulfill the graduation requirements of the Master of Science named "Sociology: Contemporary Social Problems" at Utrecht University in the Faculty of Social and Behavioral Sciences.

Being an international student in a foreign country is not an easy thing. It takes you time to adapt to the different cultural and educational system. I could not have achieved my current level of success without a strong support group.

First of all, I would like to thank my thesis and internship supervisor Arnout van de Rijt for his excellent guidance and support during this process. Second, special thanks to my parents and my brother who are always by my side, believe in me, and they support me in every step of my life. Even if they were abroad, they were mentally next to me and they encouraged me every day to continue this challenging trip. Last but not least, I would like to thank my international friends and fellow students with who we got through difficulties together and we created great memories.

I hope you enjoy your reading.

Eleni Vittoria

Utrecht, June 29, 2018

Summary

With the rise of the internet and social media, there is a massive spread of fake news and misinformation in social networks. Individuals that lack of diversity collectively consume and spread fake news. Hence, diversity could help people escape their enclosed systems and expand their critical thinking against a false belief. Furthermore, researches have shown a crowd with independent decision-makers tend to be closer to the truth than a single expert. Thus, by combining diversity and wisdom of the crowd, the present thesis research if heterogeneous groups in which members differ along multiple demographic lines are less vulnerable to the spread of a false belief.

By analyzing the data from an experiment conducted in ELSE laboratory at Utrecht University by Frey & van de Rijt (2018), we created artificial heterogeneous and homogeneous groups in order to measure the crowd wisdom across different demographic categories. We showed that even if heterogeneous groups are enriched in qualities, skills and knowledge their correct majority was equal or slightly greater than the correct majority of homogeneous groups in the decision-making process with independent decision-makers. However, there were no statistically significant differences between homogeneous and heterogeneous groups.

Based on the findings of this study, we report that diversity does not enhance the crowd wisdom and thus it could not be used as an adequate mechanism in reducing or even preventing the spread of fake news in social networks. Nevertheless, the wisdom of the crowd proved to be a positive factor in the reduction of misinformation. Having independent decision-makers that are not under the social influence could be proved beneficial in the limitation of misinformation on social networks.

Table of Contents

Chapter 1 – Introduction

Social media and online social networking sites have become a major propagator of false facts, urban legends, fake news and generally misinformation (Kim, Tabibian, Oh, Schölkopf and Gomez-Rodriguez, 2018). While the meaning of the popular term "fake news" may vary across users of it, it may be defined as: "… information that mimics the output of the news media in form, but not in organizational process or intent—e.g., lacking editorial norms and processes to weed out the untrue in favor of the true. Fake news is thus a subgenre of the broader category of misinformation—of incorrect information about the state of the world." (Lazer et al., 2018). Hence, a piece of fake news involves a claim about the world that is incorrect but not obviously so that there is some incertitude about its veracity (Aymanns, Foerster and Georg, 2017).

One explanation for the prevalence of fake news on the Internet is the lack of diversity in the relations formed on social media which traps individuals in their "filter bubbles". A filter bubble is an online personalization of internet users that isolates them in their personal preferences by not exposing them to diverse contents (Nguyen, Hui, Harper, Terveen and Konstan, 2014). The internet presents a unique mode of communication, characterized by its broad range, high speed, low costs, and the persistence of posted information (Garrett and Weeks, 2013). Through social networks people form beliefs on various topics, such as economics, politics and society, based on the information they receive from other people including not only friends, family, and coworkers also news sources and influencers[1] (Acemoglu, Ozdaglar and ParandehGheibi, 2010). Thus, whether or not a user of social media is convinced of a particular claim about the state of the world depends on his individual information about the state of the world and also on his friends' actions and aggregated opinions of anonymous people on the social networks (Aymanns, Foerster and Co-Pierre Georg, 2017).

Many studies have shown that individuals prefer to trust and to interact more often with members with the same characteristics as them (McPherson, Smith-Lovin and Cook, 2001). That results in the enclosure of individuals in their filter bubbles and echo chambers in which beliefs are intensified or fortified by communication and repetition. Hence, a diversity of voices could help people escape their enclosed systems and expand their critical thinking against a false belief (Flaxman, Goel and Rao, 2016).

This could be achieved through the "wisdom of crowds" which will consist the base of our analysis. The wisdom of the crowd is the phenomenon whereby the median judgment of *independently* deciding individuals in a large crowd tends to be close to the truth (Oliveira and Nisbett, 2017). Judgments in a diverse group could limit individual errors and lead more often to a correct majority (Larrick and Soll,

---

[1] Influencers are individuals who have the power to affect purchase decisions and opinions of others because of their authority, knowledge, position or relationship with their audience.

2006). Surowiecki (2004) who researched multiple cases of crowd wisdom at work, argues that *"under the right circumstances, groups are remarkably intelligent, and are often smarter than the smartest people in them"*.

1.1 Group homogeneity as an explanation of fake news

The present thesis focuses on explaining the spread of fake news in terms of the demographic homogeneity of groups as a reason that makes the members of the groups to jointly consume fake news. We propose that because of the "balkanization" of the Internet and its "filter bubbles", the lack of diversity in groups renders them vulnerable to the propagation of news that most people would not believe but their minority is inclined to believe. Consistent with this argument, prior theoretical work has provided formal arguments that more diverse groups achieve greater crowd wisdom. Here we test this argument using data from a recent experiment.

In addition to the literature on the role of diversity in the wisdom of crowds, there is also sizeable social-psychological literature that has argued that a key component of group performance is diversity in the composition of the group (Chen, Ren and Riedl, 2010; Hogg, 2016; Jehn, Northcraft and Neale, 1999; Williams and O'Reilly, 1998). More specifically, there are two types of groups in terms of composition: homogeneous and heterogeneous. Both homogeneity and heterogeneity refer to a multiple number of dimensions, fluctuating from age to ethnicity, from religious background to functional background, from task skills to relational skills, and from political preference to sexual preference (Knippenberg, De Dreu and Homan, 2004).

Diversity in the group characteristics can have a positive impact on the decision-making process because skills, abilities, information, viewpoints, and knowledge of the group are increased (Williams and O'Reilly, 1998). Thus, one may argue that this vast pool of resources could be a positive factor and boost group performance to the extent that it covaries with informational differences (Knippenberg, De Dreu and Homan, 2004). Hence, a heterogeneous group with independent decision makers is more enriched in different skills and knowledge and thus it has a greater probability of having a correct majority compared to a homogeneous group which has access to limited resources.

1.2 Research goals

The goal of this study is to research *if heterogeneous groups in which members differ along multiple demographic lines are less vulnerable to the spread of a false belief*. Social networks personalize contents for individuals creating the filter bubbles in which algorithms automatically recommend information that an individual is likely to like or agree with (Flaxman, Goel and Rao, 2016). This results in the isolation of individuals from other beliefs and the creation of homogeneous groups that are exposed to the same source. Thus, by researching if heterogeneous groups could achieve a greater correct majority than the homogeneous groups in the decision-making process, we could assume that heterogeneity in social media could be proved as an essential factor in reducing the proliferation of fake news.

1.3 Research questions

In order to untangle how the goals of the research will be achieved, we formed and answered a descriptive research question, an explanatory research question, and a policy question.

To begin with, the descriptive question is: *Is the majority opinion typically correct?* The answer of this is related with the wisdom of crowds in which collective opinion of a group of independent decision makers is always better than that of a single expert. Thus, we expect that if the majority pick an answer, it will probably be the right one.

The explanatory question is: *Are majority opinions in heterogeneous groups more often correct than in homogeneous groups?* The objective of this study is to research if heterogeneity of groups enhances the wisdom of the crowd. This will be achieved by analyzing the group performance of homogeneous and heterogeneous groups in terms of gender, age, ethnicity and educational background. These groups are created as synthetic sub-populations of a subject pool of experimental subjects who all participated in an earlier experiment.

Taking this approach to a more societal relevance level, the existence of false information on websites is generally unlikely to result in severe real-world consequences. However, it is conceivable that a false rumor spreading through social networks could have a crucial impact before being effectively corrected (Howell, 2013). Thus, our policy question is: *What is an adequate mechanism to prevent the spreading of fake news?* After answering the explanatory question, we will be able to see if diversity could be used as a mechanism in reducing fake news in a population.

1.4 Contribution

This thesis draws on data reported in Frey & van de Rijt (2018) who conducted an experiment that showed that social influence has a positive impact on individual level wisdom since subjects in the influence condition had higher odds of answering a trivia question correctly compared to subjects in the independent condition. However, regarding group performance/ crowd wisdom, social influence was found to have a negative effect in the study by Frey and van de Rijt, as the likelihood of the group majority to answer correctly was significantly lower in the influence condition than in the independent condition. In this thesis, we use the data from the independent condition of this experiment to answer our research questions regarding the effects of crowd diversity on the wisdom of the crowd. Because subjects in this condition are independent of one another, we can form subgroups that vary in homogeneity of composition.

The scientific relevance of this study is to make a connection between the wisdom of the crowd and the spreading of fake news in social media. Taking into account that diversity consists a critical factor for group composition, we expect to fill the gap of whether diversity on social media could be proved as an effective method to prevent fake news spreading or not. More specifically, by analyzing the performance of heterogeneous and homogeneous groups in the decision-making process, we will be able to see if it works as a preventing or enhancing mechanism in the spreading of misinformation. The combination of diversity and fake news will add a new scientific dimension, as diversity in groups was mainly researched before for the effects that it has in organizational performance, problem-solving and decision-making processes.

Chapter 2 - Theory section

2.1 Diversity

To begin with, group composition is an important determinant of group performance. More specifically, we will focus on diversity that it has been proved that has a positive impact on the process of decision-making. First of all, we should clarify what the definition of group diversity is. Group diversity refers to the fact that individuals are different from the other members of the group in terms of different attributes, and perceive themselves as such (Chen, Ren and Riedl, 2010). These attributes can vary from social characteristics such as age, gender, nationality to informational attributes such as job, education and leisure activities (Chen, Ren and Riedl, 2010). Furthermore, the differences can be in a deeper individual level such as different personalities and beliefs (Chen, Ren and Riedl, 2010).

There are three categories of diversity on groups: *informational diversity*, *social category diversity*, and *value diversity* (Jehn, Northcraft and Neale, 1999).

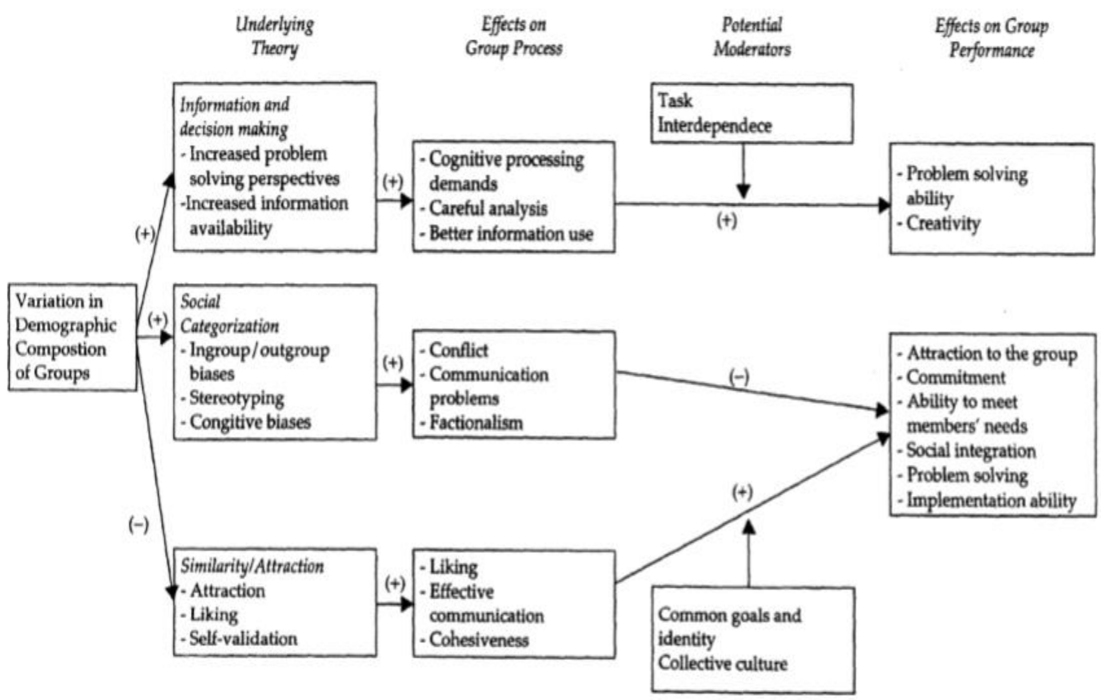
Informational diversity mainly concerns the differences in the knowledge that members of the group have (Jehn, Northcraft and Neale, 1999). More specifically, members of a diverse group tend to have different educational background, experiences and expertise which resulting in forming different opinion and perspectives on the same issue (Williams and O'Reilly, 1998). This could be proved as a positive effect in the decision-making process since they can have broader knowledge for an issue and they analyze it from many different perspectives. However, it could have negative effects as this different information, experiences and knowledge could bring disagreements and conflicts in the group members (Chatman, Polzer, Barsade and Neale, 1998), something that does not affect the independent decision-makers as there are no relations between the group members.

Social category diversity concerns the difference in characteristics such as race, gender and nationality among group members (Jehn, Northcraft and Neale, 1999; Lount and Phillips, 2007). This category is linked with social identity theory which concerns how people perceive and categorize themselves in the intergroup context. According to the founder of social identity theory, Tajfel, 1974 "the social identity of an individual conceived of as his knowledge that he belongs to certain social groups together with some emotional and value significance to him of his membership". Hence, group members even if they have differences, they share an identity of who they are, what objectives they have as group and how they should behave (Hogg, 2016). Forming a common identity as a group could work as a positive factor in the decision-making process as it enhances the group cohesion. Nevertheless, people prefer to establish connections with people that have the same characteristics with them and that means that in a diverse group this could bring conflicts and discrimination among the group members resulting less cohesion among the different individuals. Again, since our focus is on independent decision makers, we expect that diversity will only have a positive effect on the groups as it will enrich them with many different qualities.

Last but not least, is the *value diversity* in which group members have different perceptions of what are the objectives, the targets, and the missions of the group (Jehn, Northcraft and Neale, 1999). Moreover, sometimes they even have different perceptions of what are the tasks and the responsibilities of each group member (Jehn, Northcraft and Neale, 1999). On the one hand, this could be a positive thing for the group as they examine the different perspectives that emerge and end up in the best possible solution. On the other hand, again these differences could have a negative impact on groups relations as different perspectives could bring disagreements about what is right or not. Thus, for the groups with the independent decision-makers, it will be beneficial as it will give them access to multiple viewpoints.

There are many theories that explain the effects of diversity in group performance. In figure 1, we can see an overall of the three most popular theories (*Informational and decision making*, *Social categorization*, *Similarity/Attraction*) and how each theory is linked with the group processes and group performance. It should be mentioned that in this study we will only focus in the diversity across independent decision-makers and hence we will not take into account the group relations and the influence that members of the group exert each other.

Figure 1. Three main perspectives of diversity



Source*: Figure 1*. An integrated model of demographic group process and performance. Adapted from "Demography and diversity in organizations: A review of 40 years of research". By K. Williams and A. C. O'Reilly, 1998.

In this study, we will focus on the information and *decision-making perspective* because the other two perspectives mainly involve the communication and the awareness of group membership. Since in our analysis we will make artificial groups of independent decision-makers, these mechanisms are not applicable to our study. According to *decision-making perspective,* heterogeneous groups are more enriched in qualities, skills and knowledge than homogeneous group and hence they have more resources for the decision-making process (Williams and O'Reilly, 1998). Furthermore, they are able to take into account all the distinct information related to the task at hand and make better decisions (Chen, Ren and Riedl, 2010). Thus, diversity consists a source of creativity and generation of correct answers and effective solutions (Williams and O'Reilly, 1998).

2.1.1 Gender diversity

There is evidence that mixed gender groups perform better than groups that are composed by only males or females. Females have different skills than men and that gives to gender diverse groups a broader set of skills, abilities and experiences due to the fact that they acquire different gender-specific learning experiences and different social backgrounds (Song et al., 2015; Zenger and Lawrence, 1989). For example, females tend to make different decisions than men (Dufwenberg and Muren, 2006) and thus that could be important in the decision making as mixed-gender groups will take into account multiple perspectives. Hoffman and Maier (1961) conducted a study in which they constructed sixteen homogeneous groups and twenty-five heterogeneous groups in terms of gender that interacted weekly in order to make discussions, problem-solving, and role-playing. They showed that heterogeneous groups produced a better quality of solutions compared to homogeneous (Hoffman and Maier, 1961). Besides, Song et al. (2015) in order to study how gender composition affects productivity in anonymous teams, they conducted an experiment in which female and male students had to cooperate in weekly tasks during a semester-long course project. They showed that female participation on a team has a positive effect on both quantity and quality of the group performance (Song et al., 2015). Thus, our first hypothesis is:

H1: Gender-heterogeneous groups outperform gender-homogeneous groups.

It should be mentioned that by outperform we mean that we expect a higher likelihood of a correct majority in an independent vote.

2.1.2 Age diversity

Another dimension of diversity is in terms of age. On the one hand, researches have shown that some decision-making skills maybe decrease as someone is growing older such as the of use non-compensatory choice strategies which limit the odds of arriving in the right option (Bruine de Bruin, Parker and Fischhoff, 2012). In addition, important factors for the decision-making process that are decreasing over time are primary abilities such as declarative memory and inductive reasoning (Marsiske and Willis, 1995) and hence that reduce the performance of older people. On the other hand, older people are better at abstaining from irrelevant options on choices and have more experience that could be proved important in identifying the best option (Bruine de Bruin, Parker and Fischhoff, 2012). Furthermore, older people in the decision-making process are less likely to take a risk and give an incorrect answer than young people due to the fear of perceived as incompetent if they fail to answer correct (Dror, Katona and Mungur, 1998). Hence, we expect that groups with both young and old people will consist a stronger force than homogeneous groups what will only include either young or old people. Consequently, the following hypothesis is developed:

H2: Age-heterogeneous groups outperform age-homogeneous groups.

2.1.3 Ethnic diversity

Studies have shown that nationality differences have positive effects in group collaboration as they extend the perspectives and the alternatives that should be taken into consideration (Williams and O'Reilly, 1998). Furthermore, researches have shown that nationality is correlated with IQ differences (Marks, 2010). More specifically, Marks (2010) showed that "*when the average literacy of a population is relatively high, so is the population's mean IQ"* and hence in ethnically diverse groups there is a range of IQ scores that improve the overall performance of the group. In addition, Watson, Kumar and Michaelsen (1993) composed a seventeen culturally homogeneous and nineteen culturally diverse groups in order to examine how the groups perform overtime on a series of complex problem-solving tasks. They found that heterogeneous groups were more capable of identifying problem perspectives and generating alternative solutions (Watson, Kumar and Michaelsen, 1993). Another study with similar results was conducted by McLeod, Lobel and Cox (1996) who made eighteen ethnically diverse groups and sixteen ethnically homogeneous groups in order to investigate the differences between these two types of groups in brainstorming. They found that diverse groups had a greater performance than homogeneous groups in tasks that require creativity and knowledge of different cultures (McLeod, Lobel and Cox, 1996). Based on these we draw the following hypothesis:

H3: Nationality-heterogeneous groups outperform nationality-homogeneous groups.

2.1.4 Educational background diversity

In addition, as we mentioned above informational diversity which is linked with the differences in knowledge and educational background that individuals have in a diverse group, provides more information and enhances the ability of the group to perform better and more efficiently in problem-solving and decision-making processes. Furthermore, the need to find common ground among the different perspectives enhances group discussions and leads to a higher quality and more original decisions (Bantel and Jackson, 1989). Bantel and Jackson (1989) researched the relationship between the social composition of top management teams and innovation adoption and found that diversity of expertise improves the ability to solve complex problems. Thus, our next hypothesis is:

H4: Groups with heterogeneous educational backgrounds outperform groups with homogeneous educational backgrounds.

2.2 Condorcet's jury theorem

Nicolas de Condorcet who was a French philosopher, mathematician and great supporter of democracy argued that a group of individuals who have to make a decision between two choices if they take into account some rules, it would be likely that they make the correct choice (Berend and Paroush, 1998). Thus, Condorcet Jury Theorem shows that, under specific conditions, a majority of a group is always better at choosing the best choice of two alternatives than any single individual (Kanazawa, 1998). These specific conditions are: i) There are two alternative choices ii) All of the individuals have a common preference in one of the alternatives under the information they have iii) The individual decisions are independent of one another iv) Each individual makes the right decision with the probability p .0.5 v) The group decision consists the majority.

2.2.1 Comparing homogeneous and heterogeneous decision-making

A natural question that might arise is if heterogeneous groups could outperform their homogeneous counterparts in the decision-making process. The issue was addressed by Boland (1989) and further extended by Kanazawa (1998). The analysis of *Theorem 1* of Kanazawa, 1998 is exactly the answer to the aforementioned question. By using Hoeffding (1956) and Boland (1989) work he demonstrates that, for any given level of average competence ($\bar{p}$), heterogeneous groups are more capable of arriving at the right decision than homogeneous groups.

2.2.2 Mathematical details

According to Kanazawa, 1998:

"Let $X_1, \ldots, X_n$ denote independent Bernoulli random variables with probability of success $p_1, \ldots, p_n$. Independent Bernoulli random variables are analogous to a heterogeneous group where individuals make the right decision with p_i.

Let $X \equiv X_1 + \ldots + X_n$ and $\bar{p} \equiv (p_1 + \ldots + p_n/n)$.

Let Y denote a binomial $(n, \overline{p})$ random variable. A binomial random variable is analogous to a homogeneous group where individuals make the right decision with $p = \bar{p}$.

Hoeffding (1956) (Theorem 5) shows that

$Pr(a \ \leq Y \ \leq b) \leq Pr(a \ \leq X \ \leq b)$,

$if\ 0 \ \leq a \leq n\bar{p} \ \leq b \leq n$.

In other words, $Pr(a \leq X \leq b)$ attains its minimal value when $p_1 = p_2 = \ldots = p_n$ (when the independent Bernoulli random variables reduce to a binomial). If we let $a = (n + ½)$ and $b = n$ in, then

$$Pr\left(\frac{n+1}{2} + \leq Y \leq n\right) \leq Pr\left(\frac{n+1}{2} \leq X \leq n\right),$$

if $\frac{n+1}{2} \leq Y \leq n\bar{p} \leq n$,

or $\frac{n+1}{2n} \leq \bar{p} \leq 1$,

or $\frac{1}{2} + \frac{1}{2n} \leq \bar{p} \leq 1$

Thus, for any given $\bar{p}$ where $(\ ½\ ) + (½n) \leq \bar{p} \leq 1$, heterogeneous groups are more likely to make the right decision than homogeneous groups.

Theorem 1. (The superiority of heterogeneous to homogeneous groups).

If $p * \geq (½\ ) + (½n)$, then $P_{N_{HET}} > P_{N_{HOM}}$ where

$P_{N_{HET}} = P_N$ for a heterogeneous group with $p = p *$, and

$P_{N_{HOM}} \ = P_N$ for a homogeneous group with $p = p *$."

The interesting point in this theorem is that it shows that a pluralistic group does not need to consist of experts in order to arrive in the right decision. Furthermore, Kanazawa,1998 showed that it only needs a small fraction above ½ in order to raise the probability of success of heterogeneous groups significantly compared to the mean of decisions of homogeneous groups. Thus, that gives us a strong formal basis for our analysis that heterogeneous groups will outperform homogeneous.

Chapter 3 – Methods

3.1 Dataset

The dataset that will be used in this thesis is from an experiment conducted in ELSE laboratory at Utrecht University (Frey & van de Rijt, 2018). In this experiment groups of 10 to 14 participants had to answer 30 different questions from five categories: visual, art, equations, history, and geometry. Each question had two possible answers of which only one was correct and subjects had twenty seconds to pick an answer. Thirty questions were put in cyclic order in such a way that all of the participants of experimental groups could answer the questions at the same time and that different subjects of the same group could start at a different position in the cycle. There were eight experimental sessions with each session involving twenty-one to twenty-seven subjects, split randomly into two groups with approximately the same size. The difference between these groups was that the one was assigned to the independent condition and the other to the social influence condition. In the influence condition, the participants could see the number of prior group members who had chosen each of the possible answers. This social information was not available for the independent condition. It is important to mention that there was no access to the internet and the computer stations were isolated from each other. Thus, subjects could not seek the right answer online and they were not able to communicate with one another.

We will reanalyze the experimental data to answer a different question than the original study asked. Whereas Frey & van de Rijt (2018) wanted to test whether social influence would undermine the wisdom of the crowd and boost the spread of a false belief, we ask whether the heterogeneity of groups enhances the wisdom of the crowd.

3.2 Analytical Strategy

The analytical strategy consists of many steps. First of all, as mentioned above the given dataset had an independent condition and a social influence condition. For the purpose of this study we only kept the independent condition as we wanted to analyze diversity under independent decision-makers. Moreover, for each independent variable (gender-heterogeneity, age-heterogeneity, ethnic-heterogeneity and educational background-heterogeneity) we made the following two categories: for *gender* 1) males 2) females, for *age* 1) young 2) old, for *ethnicity* 1) Dutch 2) non-Dutch and for *educational background* 1) sociologists 2) non-sociologists. From these categories, we later draw participants in order to construct our artificial groups. In the next step, we checked the individual-level relationship between gender, age, ethnicity and educational background, and wisdom of the crowd and for the actual group-level relationship between gender, age, ethnicity and educational background composition and wisdom of the crowd.

In the last and most important step, we constructed artificial groups by sampling 1,000 times a homogeneous group, and 1,000 times a heterogeneous group for each independent variable from the categories that we mentioned above. For each group, we calculated for what percentage of questions the majority picked the right answer, as our measure of "crowd wisdom". We also calculated the confidence interval of crowd wisdom as the 2.5th percentile (lower bound) and 97.5th percentile (upper bound) across all 1,000 simulated trials. We tested for a group difference in crowd wisdom between a homogeneous and a heterogeneous group by evaluating the z-score of the value 0 in the distribution of the difference in crowd wisdom between the two groups. The p-value corresponding to this z-score represents the chance that a homogeneous group of 12 subjects drawn from our subject population outperforms a heterogeneous group of 12 students drawn from the same population.

3.3 Variables

In this part, we will describe the variables that were used for our analysis. Our unit of analysis is a group of twelve subjects that we assigned them in the groups that we made in order to conduct our analysis.

3.3.1 Dependent variable

The dependent variable is crowd wisdom. We consider a crowd "wise" with regard to a given question if a majority of members answered that question correctly. This variable has two values, namely: "0" if the majority gave the wrong answer and "1" if it gave the right answer.

3.3.2 Independent variables

The following variables were measured by a questionnaire that was given to the participants at the end of every experimental session. The questionnaire had both closed and open questions.

To begin with, gender was measured by the question: "*Are you male or female?"*. The participants had to choose between the two given options: *"Male"* or *"Female"*. The majority of the respondents were females with 66,3%. The number of males in our sample is 33 and the females 65 (N = 98). For the construction of our groups, we made 1,000 groups with twelve females only and 1,000 groups with twelve males only for the homogeneous category and 1,000 groups with six males and six females for the heterogeneous.

The age (M= 23.51, SD= 0.12) of the participants was measured by the following question: *"What is your age?"*. Participants had to fill in their age. The minimum age in our sample was eighteen years old and the maximum sixty-six years old. In order to make artificial groups we divided respondents is two categories: young people and old people. We defined as young people those who were eighteen to twenty-one years old and as old those who were twenty-two and above. Thus, in our sample, we have 47 young people and 51 old people (N = 98). We made 1,000 homogeneous group of twelve young people and 1,000 with twelve old people and then we combined these two categories in order to make 1,000 heterogeneous group with six participants from each category.

Another open question that the questionnaire had was about the ethnicity of the participants. This variable was measured by the question: *"What is your nationality?"*. The majority of the participants were Dutch (61,22%) and thus we used them as a reference point in the construction of our groups. In our sample, we have 59 Dutch participants and 39 non-Dutch participants (N = 98). In the synthetic groups that we made, we used twelve Dutch people for 1,000 homogeneous groups and 1,000 times a mix from other nationalities for the heterogeneous. For this variable, there were two types of heterogeneous groups: one with six Dutch and six individuals of different nationalities and one with twelve individuals with different nationalities. However, the 1,000 groups with twelve participants of different nationalities could be considered as heterogeneous groups, we will treat them as homogeneous in order to be consistent with our analysis.

Regarding the educational background of the participants, first they had to answer the following question: *"Are you a student?"* with possible answers *"Yes"* or *"No"*. Then there was a follow-up question: *"If yes, what do you study?"*, in which respondents had to state their educational background. For the purpose of this study, we randomly picked sociologists as the reference point of the homogeneous groups and again for heterogeneous we made a blend of participants from other sectors. The number of sociologists and non-sociologists in our sample is both 20 (N=40). Similar to ethnicity, we made 1,000 homogeneous group with twelve sociologists, 1,000 heterogeneous groups with six sociologists and six participants with a random educational background and 1,000 groups with twelve participants with random educational background. Again, we will treat the 1,000 groups with the respondents with a diverse educational background as homogeneous group.

3.4 Descriptive statistics

In table 1. the dependent variable, the sessions of the experiment and the independent variables are presented. The frequencies of the correct and the wrong answers refer to the overall answers that individuals gave in the experiment. Thus, it includes the answers to all thirty questions within the eight sessions. Furthermore, the frequencies of the independent variables refer to the number of participants that were in the sample that we used for our analysis. Either of the two groups that every independent variable has are the categories from which we made our synthetic groups.

Table 1. Descriptive statistics of the experiment (N answers = 2.940, N participants = 98)

Dependent variable	Frequency	Percentage
Answer given by a subject		
Correct	1.839	62.55
Wrong	1.101	37.45
	Number of participants	Percentage
Session		
1	12	12.24
2	14	14.29
3	14	14.29
4	11	11.22
5	12	12.24
6	12	12.24
7	11	11.22
8	12	12.24
Gender		
Female	65	66.33
Male	33	33.67
Age		
Young	47	47.96
Old	51	52.04
Nationality		
Dutch	59	60.20
Non-Dutch	39	39.80
Educational Background		
Sociologists	20	50.00
Non-sociologists	20	50.00

Chapter 4 – Results

4.1 Gender

To begin with, the 1,000 mixed-gender groups on average answered 80% of the thirty questions with a correct majority, with a 95% confidence interval spanning 67% to 90%. In addition, the groups with only females answered 77% of the questions with a correct majority, with a 95% confidence interval running from 67% to 87%. The groups with only males answered on average 83% of the questions with a correct majority, with a 95% confidence interval running from 73% to 90%. As said, in order to test if the percentages of the questions answered by a correct majority in heterogeneous groups is significantly higher than in homogeneous groups we calculate the z score for the value 0 in the distribution of the difference in crowd wisdom between the groups. First, we will compare the heterogeneous groups with the groups that had only females:

$$z = \frac{(.8 - .77)}{\sqrt{(.05804 + .05547)}} = .09$$

The present value is only .09, so there is no significant difference in the group wisdom between groups of six women and six men and groups of only twelve females.

For heterogeneous groups and groups with males only is the following:

$$z = \frac{(.8 - .83)}{\sqrt{(.05804 + .05158)}} = -.09$$

Again, there is no significant difference in the wisdom between mixed gender groups and groups with males only, since the value is -.09. Thus, since there are no significant differences between heterogeneous and homogeneous groups, the hypothesis H1 is rejected.

4.2 Age

In terms of age, the 1,000 mixed age groups on average answered 80% of the thirty questions with a correct majority, with a 95% confidence interval running from 67% to 90%. In addition, the groups with only young answered 73% of the questions with a correct majority, with a 95% confidence interval running from 63% to 87%. The groups with only old answered on average 83% of the questions with a correct majority, with a 95% confidence interval running from 70% to 90%. For the comparison involving the heterogeneous groups with the groups that had only young people the z score is given by:

$$z = \frac{(.8 - .73)}{\sqrt{(.05803 + .05518)}} = .21$$

The present value is .21, hence there is no significant difference in the group wisdom between groups of six young and six old and groups of only twelve young people.

For heterogeneous groups and groups with old people:

$$z = \frac{(.8\ -\ .83)}{\sqrt{(.05803 +\ .05207)}} =\ -.09$$

There are no significant differences in the wisdom of the crowd of heterogeneous groups and homogeneous groups with old people. Our second hypothesis H2 is rejected as heterogeneous groups did as well in the decision-making process as homogeneous groups with young people.

4.3 Ethnicity

Regarding ethnicity, the 1,000 groups with participants who had different nationality, the majority on average answered 77% of the thirty questions correct, with a 95% confidence interval running from 67% to 90%. Groups with Dutch people only answered 80% of the thirty questions with a 95% confidence interval running from 70% to 90%. Again, we conduct a z test in order the check if the percentages of the questions answered by a correct majority in heterogeneous group is significantly higher than homogeneous.

$$z\ = \frac{(.77 -\ .8)}{\sqrt{(.05740 +\ .05383)}} = -.09$$

The value is -.09, so the correct majority of homogeneous groups was not significantly higher than homogeneous groups. Therefore, hypothesis H3 is rejected. In terms of nationality, homogeneous groups performed as well in the decision-making process as heterogeneous.

4.4 Educational background

The 1,000 heterogeneous groups consist of people with a different educational background, the on average answered 97% of the thirty questions with a correct majority, with a 95% confidence interval running from 90% to 100%. The groups of sociologists answered 77% of the thirty questions with a correct majority, with a 95% confidence interval running from 67% to 87%. The groups with non-sociologists answered 80% of the thirty questions with a correct majority with 95% confidence interval running from 73% to 87%. We check if the correct majority of heterogeneous groups was significantly higher. The z test for the heterogeneous groups with both sociologists and non-sociologists and homogeneous groups of only sociologists is:

$$z\ = \frac{(.97 -\ .77)}{\sqrt{(.033065 +\ .051422)}} =\ .69$$

The value is .69, hence there is not a significant difference in the group wisdom between groups of six sociologists and six non-sociologists and groups of only sociologists. Hypothesis H4 is rejected.

Chapter 5 – Conclusion

5.1 Conclusion

In order to investigate if diversity could be an adequate mechanism to prevent fake news on social media, we created two research questions: "*Is majority opinion typically correct?*" and "*Are majority opinions in heterogeneous groups more often correct than in homogeneous groups?*", that we managed to answer through our analysis. Through the following results, we will answer the above research questions.

To begin with, we expected that mixed gender groups would achieve better results in the decision-making process than groups consisted of the same gender. Our results showed that in terms of gender diversity, groups with both females and males did slightly better than homogeneous groups with females only. However, groups with males only were a little bit better than heterogeneous groups. A possible explanation for this is that men and women tend to be better at different categories (Song et al., 2015). Taking into account that only have five different categories of questions, we can assume that males tend to be better at these categories.

In terms of age, heterogeneous groups outperformed homogeneous groups with young people but not the homogeneous groups with old people. This could be explained by the fact older people have more experiences than younger people and they outweigh more careful their decisions (Bruine de Bruin, Parker and Fischhoff, 2012; Dror, Katona and Mungur, 1998). Thus, they tend to think more and be more cautious about their answers, resulting in a greater correct majority.

Regarding nationality, the homogeneous groups were better than both of the heterogeneous groups. According to McLeod, Lobel and Cox (1996) culturally diverse groups have a superior performance than homogeneous groups in tasks that require creativity and knowledge of different cultures. Taking into account that the questions of our experiment did not require creativity and knowledge of different cultures, we could justify our results. Furthermore, as we mentioned above literacy is closely related with IQ and vice versa (Marks, 2010). Thus, we can presume that Dutch people which the reference point of the homogeneous groups were acquired higher literacy than the other ethnicities and had better knowledge on the five categories of questions that were asked compared to participants from other nationalities.

Concerning the educational background diversity, the correct majority of heterogeneous groups was higher than the correct majority of homogeneous groups, but again there were no significant differences. An explanation for this result is that diversity in the educational background could work as a positive factor only when there are group processes and not in independent decision-makers. Bantel and Jackson (1989) reported that the different perspectives, enhance group discussions and lead to a superior quality of decisions. Furthermore, since our questions were across five categories, we can presume that our participants did not have strong knowledge in these sectors.

Overall, we could report that the majority opinion is typically correct, as the correct majority of all of our groups scored really high in both homogeneous and heterogeneous groups. Diversity does not work as an improving factor in the crowd wisdom. That could be explained from the fact that diversity could work more effectively in a group performance when it involves group discussions and processes as they examine the different perspectives that emerge and end up in the best possible solution (Bantel and Jackson, 1989). Nevertheless, crowd wisdom proved to be a positive factor both in terms of homogeneity and heterogeneity as the multiple voices of a group can bring it closer to the truth.

Regarding the fake news, diversity could not be used as an effective mechanism that prevents or limits the fake news but a crowd with independent decision-makers could be proved as a beneficial factor for the reduction of fake news. In addition, we assumed that demographic homogeneity and "filter bubbles" render individuals more vulnerable in the spread of fake news, something that it was not proved in our analysis. Hence, homogeneity and the isolation of individuals from diverse contents do not consist a crucial factor in boosting the spread of misinformation. Other factors such as *social influence* which refers to the phenomenon whereby individuals change their behaviors under the influence of others (Peng, Yang, Cao, Yu and Xie, 2016), should be taken into account for further research.

Chapter 6 - Discussion

In this part of the thesis, it is evaluated to what extent the study succeeded in developing an accurate understanding of how to handle the problem at hand and what were the strong and the weak points.

To begin with, theory gave as a formal basis to support or analysis as there is a mathematical proof that heterogeneous groups have a greater probability of having a correct majority compared to homogeneous. Thus, our analysis was based on a proved argument. Furthermore, through our analysis, we were able to answer all of the research questions that we set in the beginning. In this way, this thesis is able to reach the research objectives successfully and conclude that diversity is not an effective mechanism that could limit or even prevent misinformation on social media.

A limitation of the present analysis was that our set of questions consisted of five categories: visual, art, equations, history, and geometry. Even though that is a set that includes a variety of different sectors, a broader set could have given a better indicator of the wisdom of the crowd as it could provide more likelihoods of success to participants that have restricted knowledge and information of these specific categories. Furthermore, we randomly decided to make groups of twelve participants and hence maybe bigger groups (e.g. groups of thirty participants) could have more statistical power. Thus, a future study could research diversity and wisdom of crows across multiple sectors and in bigger groups.

6.1 Policy advice

In the present study, we wanted to find an adequate mechanism to prevent fake news on social media. By focusing on the explanation that fake news spread easily among the homogeneous individuals who are trapped in their filter bubbles and echo chambers and hence they collectively consume fake news, we research if diversity could work as an adequate mechanism to limit or even prevent the spreading of misinformation. We showed that a diverse crowd with independent decision-makers is not wiser than a crowd consisting of people with the same abilities and knowledge. More specifically, the correct majority of both homogeneous and heterogeneous groups it can correct the individual errors. Thus, diversity could not be used as an effective mechanism in social media.

Even if we do not find diversity as a significant mean for the reduction of fake news, the scores of correct majorities were really high and thus the median judgment of independently deciding individuals could be considered as a positive component on social media. Therefore, we can report that a wise crowd with independent decision-makers could help in the limitation of fake news as it brings them closer to the truth.

Another aspect that we should take into consideration is that in the earlier experiment of the dataset conducted by Frey & van de Rijt (2018), they showed that social influence undermines the crowd wisdom. Furthermore, Becker, Brackbill and Centola (2017) have shown that social influence could undermine the wisdom of the crowd because when it leads to correlated errors the diversity and the independence are reduced, something that puts in at risk the veracity of the groups' judgment. Statistical explanations for this particular phenomenon argue that group precision relies on the estimates taken from groups where individuals' errors are either uncorrelated or negatively correlated, thereby preserving the diversity of opinions in a population (Becker, Brackbill and Centola, 2017). Thus, through social influence, a majority belief could spread in social media and become dominant. Combining these findings with our findings, we can suggest that if we limit social influence that undermines wisdom of the crowd and makes individuals more vulnerable to spread a false belief and enhance more the crowd wisdom which is proved to bring people closer to the veracity, it could be established as an effective way to restrict the spread of fake news. The wisdom of the crowd could be used as a filter that decides the veracity of a claim. In that way, if something is perceived as fake news, it will not further spread on the network. Hence, the popularity of the misinformation will significantly decrease or even disappear before it spreads all around.

Literature

Acemoglu, D., Ozdaglar, A., & ParandehGheibi, A. (2010). Spread of (mis)information in social networks. *Games and economic behavior*, *70*(2), 194–227. doi:10.1016/j.geb.2010.01.005

Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs: General and Applied*, *70*(9), 1–70. doi:10.1037/h0093718

Aymanns, C., Foerster, J., & Georg, C.-P. (2017). Fake News in Social Networks. arXiv:1708.06233

Bantel, K. A. & Jackson, S. E., & (1989). Top management and innovations in banking: Does the composition of the top team make a difference? *Strategic Management Journal*, *10*(S1), 107–124. doi:10.1002/smj.4250100709

Becker, J., Brackbill, D., & Centola, D. (2017). Network dynamics of social influence in the wisdom of crowds. *Proceedings of the National Academy of Sciences*, *114*(26). doi:10.1073/pnas.1615978114

Berend, D., & Paroush, J. (1998). When is Condorcet's Jury Theorem valid? *Social Choice and Welfare, 15*(4), 481–488. doi:10.1007/s003550050118

Boland, P. J. (1989). Majority systems and the Condorcet Jury Theorem. *Journal of the Royal Statistical Society*, *38*(3), 181-189. doi: 10.2307/2348873

Bruine de Bruin, W., Parker, A. M., & Fischhoff, B. (2012). Explaining adult age differences in decision-making competence. *Journal of Behavioral Decision Making*, *25*(4), 352–360. doi:10.1002/BDM.712

Chatman, J. A., Polzer, J. T., Barsade, S. G., & Neale, M. A. (1998). Being different yet feeling similar: The influence of demographic composition and organizational culture on work processes and outcomes. *Administrative Science Quarterly*, *43*(4), 749–780. doi: 10.2307/2393615

Chen, J., Ren, Y., & Riedl, J. (2010). The effects of diversity on group productivity and member withdrawal in online volunteer groups. *Proceedings of the 28th International Conference on Human Factors in Computing Systems* (CHI '10), 821. doi:/10.1145/1753326.1753447

de Oliveira, S., & Nisbett, R. E. (2018). Demographically diverse crowds are typically not much wiser than homogeneous crowds. *Proceedings of the National Academy of Sciences*, (16). doi:10.1073/pnas.1717632115

Dufwenberg, M., & Muren, A. (2006). Gender composition in teams. *Journal of Economic Behavior and Organization*, *61*(1), 50–54. doi:10.1016/j.jebo.2005.01.002

Dror, I. E., Katona, M., & Mungur, K. (1998). Age differences in decision making: To take a risk or not? *Gerontology*, *44*(2), 67–71. doi:10.1159/000021986

Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly*, *80*(Special issue 1), 298–320. Doi:10.1093/poq/nfw006

Frey, V. & van de Rijt, A. (2018). Social influence undermines crowd wisdom in sequential decision-making. Unpublished manuscript.

Garrett, R. K., & Weeks, B. E. (2013). The promise and peril of real-time corrections to political misperceptions. *Proceedings of the 2013 Conference on Computer Supported Cooperative Work - CSCW '13*, 1047. doi:10.1145/2441776.2441895

Hoeffding, W. (1956). On the distribution of the number of success in independent Trials. *The Annals of Mathematical Statistics, 3*, 295–312. doi:10.2307/24304959

Hogg, M. A. (2016). Social identity theory. In S. McKeown, R. Haji, & N. Ferguson (Eds.), *Understanding peace and conflict through social identity theory: Contemporary global perspectives* (pp. 3-17). Cham: Springer International Publishing.

Hoffman, L. R., & Maier, N. R. F. (1961). Quality and acceptance of problem solutions by members of homogeneous and heterogeneous groups. *Journal of Abnormal and Social Psychology*, *62*(2), 401–407. doi:10.1037/h0044025

Howell, L. (2013) Digital wildfires in a hyperconnected world. *WEF Report 2013*. Retrieved from: http://reports.weforum.org/global-risks-2013/risk-case-1/digital-wildfires-in-a-hyperconnected-world/

Jehn, K. A., Northcraft, G. B., & Neale, M. A. (1999). Why differences make a difference: A field study of diversity, conflict, and performance in workgroups. *Administrative Science Quarterly, 44*(4), 741. doi:10.2307/2667054

Kanazawa, S. (1998). A brief note on a further refinement of the Condorcet Jury Theorem for heterogeneous groups. *Mathematical Social Sciences, 35*(1), 69–73. doi:10.1016/S0165-4896(97)00028-0

Kim, J., Tabibian, B., Oh, A., Schölkopf, B., & Gomez-Rodriguez, M. (2017). Leveraging the crowd to detect and reduce the spread of fake news and misinformation, 324–332. doi:10.1145/3159652.3159734

Larrick, R. P., & Soll, J. B. (2006). Intuitions About Combining Opinions: Misappreciation of the averaging principle. *Management Science, 52*(1), 111–127. doi:10.1287/mnsc.1050.0459

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., … Zittrain, J. L. (2018). The science of fake news. *Science*, *359*(6380), 1094–1096. doi:10.1126/science.aao2998

Lount, R. B., & Phillips, K. W. (2007). Working harder with the out-group: The impact of social category diversity on motivation gains. *Organizational Behavior and Human Decision Processes*, *103*(2), 214–224. doi:10.1016/j.obhdp.2007.03.002

Marks, D. F. (2010). IQ Variations across time, race, and nationality: An artifact of differences in literacy skills. *Psychological Reports, 106*(3), 643–664. doi:10.2466/pr0.106.3.643-664

Marsiske, M., & Willis, S. L. (1995). Dimensionality of everyday problem solving in older adults. *Psychology and Aging, 10*(2), 269–283. doi:10.1037/0882-7974.10.2.269

McLeod, P. L., Lobel, S. A. and Cox, T. H., Jr. (1996). Ethnic diversity and creativity in small groups. *Small Group Research*, *27*(2), 248 – 264. doi:10.1177/1046496496272003

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology, 27*(1), 415–444. doi:10.1146/annurev.soc.27.1.415

Nguyen, T. T., Hui, P.-M., Harper, F. M., Terveen, L., & Konstan, J. A. (2014). Exploring the filter bubble. *Proceedings of the 23rd International Conference on World Wide Web,* 677–686. doi:10.1145/2566486.2568012

Peng, S., Yang, A., Cao, L., Yu, S., & Xie, D. (2017). Social influence modeling using information theory in mobile social networks. *Information Sciences*, *379*, 146–159. doi:10.1016/j.ins.2016.08.023

Pennycook, G., & Rand, D. G. (2017). Who falls for fake news? The roles of analytic thinking, motivated reasoning, political ideology, and bullshit receptivity. doi:10.2139/ssrn.3023545

Song, H. G., Restivo, M., van De Rijt, A., Scarlatos, L., Tonjes, D., & Orlov, A. (2015). The hidden gender effect in online collaboration: An experimental study of team performance under anonymity. *Computers in Human Behavior, 50,* 274–282. doi:10.1016/j.chb.2015.04.013

Surowiecki, J. (2004). *The Wisdom of Crowds. How Collective Wisdom Shapes Business, Economies, Societies and Nations*. New York: Doubleday.

Tajfel, H. (1974). Social identity and intergroup behaviour. *Social Science Information, 13*(2), 65–93. doi:10.1177/053901847401300204

Tsui, A. S., Egan, T. D., & O'Reilly, C. A. (1992). Being different: Relational demography. *Administrative Science Quarterly*, *37*(4), 549-579. doi:org/10.2307/2393472

van Knippenberg, D., De Dreu, C. K. W., & Homan, A. C. (2004). Work group diversity and group performance: An integrative model and research agenda. *Journal of Applied Psychology*, *89*(6), 1008–1022. doi:10.1037/0021-9010.89.6.1008

Watson, W. E., Kumar, K., & Michaelsen, L. K. (1993). Cultural diversity's impact on interaction process and performance: Comparing homogeneous and diverse task groups. *Academy of Management*, *3*(3), 590–602. doi:10.5465/256593

Williams, K., & O'Reilly, A. C. (1998). Demography and diversity in organizations: A review of 40 years of research. *Academy of Management Journal*, *35*, 91–121.

Zenger, T. R., & Lawrence, B. S. (1989). Organizational demography: The differential effects of age and tenure distributions on technical communication. *The Academy of Management Journal, 32*(2), 353–376. doi: 10.2307/256366

Appendix A – Analysis code

```
// Individual-level relationship between gender and wisdom
encode Sex, gen (Sex_n)
recode Sex_n (2=1)(1=0), gen(Sex_group)
tab Sex_group
ttest corr, by(Sex_group)
gen sessqid = session * 100 + q
xtset sessqid
xtlogit corr Sex_group, fe
forvalues q = 1(1)30 {
    logit corr Sex_group if q == `q'
}

// Actual group-level relationship between gender composition and wisdom
bys session: egen percg = total(Sex_group * (q == 1) / gs)
bys sessqid: gen sessqtag = _n == 1
bys session: sum percg if sessqtag
```

```
. // Simulated group-level relationship between gender composition and crowd wisdom
. preserve

.          mata: M = J(1, 1000, .)

.          forvalues it = 1(1)1000 {
  2.                   quietly {
  3.                   gen x = rnormal()
  4.                   sort Sex_group q x
  5.                   by Sex_group: gen sampled = _n <= 6
  6.                   gsort id -sampled
  7.                   by id: replace sampled = sampled[1]
  8.                   bys q: egen perccorr = total((corr * sampled) / 12)
  9.                   bys q: gen qtag = _n == 1
 10.                   gen mcorr = perccorr >= .5
 11.                   sum mcorr if qtag
 12.                   local rmean = r(mean)
 13.                   mata: M[1,`it'] = `rmean'
 14.                   drop x sampled perccorr qtag mcorr
 15.                   }
 16.          }

.         mata: Y = sort(M',1)

.         mata: Y[25];Y[500];Y[975]
  .6666666667
  .8
  .9

.         mata: y = sqrt(variance(Y))

.         mata: y
  .0580496319
. restore
```

```
. // Only Sex == 0 (females)
. preserve

.         keep if Sex_group == 0
(990 observations deleted)

.         mata: B = J(1, 1000, .)

.         forvalues it = 1(1)1000 {
  2.                 quietly {
  3.                 gen x = rnormal()
  4.                 sort q x
  5.                 gen sampled = _n <= 12
  6.                 gsort id -sampled
  7.                 by id: replace sampled = sampled[1]
  8.                 bys q: egen perccorr = total((corr * sampled) / 12)
  9.                 bys q: gen qtag = _n == 1
 10.                 gen mcorr = perccorr >= .5
 11.                 sum mcorr if qtag
 12.                 local rmean = r(mean)
 13.                 mata: B[1,`it'] = `rmean'
 14.                 drop x sampled perccorr qtag mcorr
 15.                 }
 16.         }

.         mata: Y = sort(B',1)

.         mata: Y[25];Y[500];Y[975]
  .6666666667
  .7666666667
  .8666666667

.         mata: y = sqrt(variance(Y))

.         mata: y
  .0554724504
. restore
```

```
. // Only Sex == 1 (males)
. preserve

.         keep if Sex_group == 1
(1,950 observations deleted)

.         mata: G = J(1, 1000, .)

.         forvalues it = 1(1)1000 {
  2.                 quietly {
  3.                 gen x = rnormal()
  4.                 sort q x
  5.                 gen sampled = _n <= 12
  6.                 gsort id -sampled
  7.                 by id: replace sampled = sampled[1]
  8.                 bys q: egen perccorr = total((corr * sampled) / 12)
  9.                 bys q: gen qtag = _n == 1
 10.                 gen mcorr = perccorr >= .5
 11.                 sum mcorr if qtag
 12.                 local rmean = r(mean)
 13.                 mata: G[1,`it'] = `rmean'
 14.                 drop x sampled perccorr qtag mcorr
 15.                 }
 16.         }

.         mata: Y = sort(G',1)

.         mata: Y[25];Y[500];Y[975]
  .7
  .8333333333
  .9333333333

.         mata: y = sqrt(variance(Y))

.         mata: y
  .051588481
. restore
```

```
// Individual-level relationship between age and crowd wisdom
recode Age (min/21=0)(22/max=1), gen(Age_group)
tab Age_group
ttest corr, by(Age_group)
gen sessqid = session * 100 + q
xtset sessqid
xtlogit corr Age_group, fe
forvalues q = 1(1)30 {
    logit corr Age_group if q == `q'
}
//Actual group-level relationship between gender composition and crowd wisdom
bys session: egen percg = total(Age_group * (q == 1) / gs)
bys sessqid: gen sessqtag = _n == 1
bys session: sum percg if sessqtag
```

```
. // Simulated group-level relationship between age composition and crowd wisdom
. preserve

.         mata: M = J(1, 1000, .)

.         forvalues it = 1(1)1000 {
                  quietly {
                  gen x = rnormal()
                  sort Age_group q x
                  by Age_group: gen sampled = _n <= 6
                  gsort id -sampled
                  by id: replace sampled = sampled[1]
                  bys q: egen perccorr = total((corr * sampled) / 12)
                  bys q: gen qtag = _n == 1
                  gen mcorr = perccorr >= .5
                  sum mcorr if qtag
                  local rmean = r(mean)
                  mata: M[1,`it'] = `rmean'
                  drop x sampled perccorr qtag mcorr
                  }
          }

.         mata: Y = sort(M',1)

.         mata: Y[25];Y[500];Y[975]
  .6666666667
  .8
  .9

.         mata: y = sqrt(variance(Y))

.         mata: y
  .0580323854
. restore
```

```
. // Only Age_group == 0 (young)
. preserve

.         keep if Age_group == 0
(1,530 observations deleted)

.         mata: B = J(1, 1000, .)

.         forvalues it = 1(1)1000 {
  2.                 quietly {
  3.                 gen x = rnormal()
  4.                 sort q x
  5.                 gen sampled = _n <= 12
  6.                 gsort id -sampled
  7.                 by id: replace sampled = sampled[1]
  8.                 bys q: egen perccorr = total((corr * sampled) / 12)
  9.                 bys q: gen qtag = _n == 1
 10.                 gen mcorr = perccorr >= .5
 11.                 sum mcorr if qtag
 12.                 local rmean = r(mean)
 13.                 mata: B[1,`it'] = `rmean'
 14.                 drop x sampled perccorr qtag mcorr
 15.                 }
 16.         }

.         mata: Y = sort(B',1)

.         mata: Y[25];Y[500];Y[975]
  .6333333333
  .7333333333
  .8666666667

.         mata: y = sqrt(variance(Y))

.         mata: y
  .0551839322
. restore
```

```
. // Only Age_group == 1 (old)
. preserve

.         keep if Age_group == 1
(1,410 observations deleted)

.         mata: G = J(1, 1000, .)

.         forvalues it = 1(1)1000 {
  2.                 quietly {
  3.                 gen x = rnormal()
  4.                 sort q x
  5.                 gen sampled = _n <= 12
  6.                 gsort id -sampled
  7.                 by id: replace sampled = sampled[1]
  8.                 bys q: egen perccorr = total((corr * sampled) / 12)
  9.                 bys q: gen qtag = _n == 1
 10.                 gen mcorr = perccorr >= .5
 11.                 sum mcorr if qtag
 12.                 local rmean = r(mean)
 13.                 mata: G[1,`it'] = `rmean'
 14.                 drop x sampled perccorr qtag mcorr
 15.                 }
 16.         }

.         mata: Y = sort(G',1)

.         mata: Y[25];Y[500];Y[975]
  .7
  .8333333333
  .9

.         mata: y = sqrt(variance(Y))

.         mata: y
  .0520730197
. restore
```

```
// Individual-level relationship between nationality and wisdom
encode Nationality, gen (Nationality_n)
recode Nationality_n (6/8=0) (19/23=0) (35=0) (37=0) (1/5=1) (9/18=1) (24/34=1) (36=1) (38/39=1), gen(Nationality_group)
tab Nationality_group
ttest corr, by(Nationality_group)
gen sessqid = session * 100 + q
xtset sessqid
xtlogit corr Nationality_group, fe
forvalues q = 1(1)30 {
    logit corr Nationality_group if q == `q'
}
//Actual group-level relationship between nationality composition and wisdom
bys session: egen percg = total(Nationality_group * (q == 1) / gs)
bys sessqid: gen sessqtag = _n == 1
bys session: sum percg if sessqtag
```

```
. // Simulated group-level relationship between nationality composition and crowd wisdom
. preserve

.         mata: M = J(1, 1000, .)

.         forvalues it = 1(1)1000 {
                  quietly {
                  gen x = rnormal()
                  sort Nationality_group q x
                  by Nationality_group: gen sampled = _n <= 6
                  gsort id -sampled
                  by id: replace sampled = sampled[1]
                  bys q: egen perccorr = total((corr * sampled) / 12)
                  bys q: gen qtag = _n == 1
                  gen mcorr = perccorr >= .5
                  sum mcorr if qtag
                  local rmean = r(mean)
                  mata: M[1,`it'] = `rmean'
                  drop x sampled perccorr qtag mcorr
                  }
          }

.         mata: Y = sort(M',1)

.         mata: Y[25];Y[500];Y[975]
  .6666666667
  .7666666667
  .9

.         mata: y = sqrt(variance(Y))

.         mata: y
  .0585544833
. restore
```

```
. // Only Nationality_group == 0 (Dutch)
. preserve

.          keep if Nationality_group == 0
(1,170 observations deleted)

.          mata: B = J(1, 1000, .)

.          forvalues it = 1(1)1000 {
  2.                 quietly {
  3.                 gen x = rnormal()
  4.                 sort q x
  5.                 gen sampled = _n <= 12
  6.                 gsort id -sampled
  7.                 by id: replace sampled = sampled[1]
  8.                 bys q: egen perccorr = total((corr * sampled) / 12)
  9.                 bys q: gen qtag = _n == 1
 10.                 gen mcorr = perccorr >= .5
 11.                 sum mcorr if qtag
 12.                 local rmean = r(mean)
 13.                 mata: B[1,`it'] = `rmean'
 14.                 drop x sampled perccorr qtag mcorr
 15.                 }
 16.         }

.          mata: Y = sort(B',1)

.          mata: Y[25];Y[500];Y[975]
  .7
  .8
  .9

.          mata: y = sqrt(variance(Y))

.          mata: y
  .0538331187
. restore
```

```
. // Only Nationality_group == 1 (non-Dutch)
. preserve

.         keep if Nationality_group == 1
(1,770 observations deleted)

.         mata: G = J(1, 1000, .)

.         forvalues it = 1(1)1000 {
  2.                  quietly {
  3.                  gen x = rnormal()
  4.                  sort q x
  5.                  gen sampled = _n <= 12
  6.                  gsort id -sampled
  7.                  by id: replace sampled = sampled[1]
  8.                  bys q: egen perccorr = total((corr * sampled) / 12)
  9.                  bys q: gen qtag = _n == 1
 10.                  gen mcorr = perccorr >= .5
 11.                  sum mcorr if qtag
 12.                  local rmean = r(mean)
 13.                  mata: G[1,`it'] = `rmean'
 14.                  drop x sampled perccorr qtag mcorr
 15.                  }
 16.         }

.         mata: Y = sort(G',1)

.         mata: Y[25];Y[500];Y[975]
  .6333333333
  .7666666667
  .8666666667

.         mata: y = sqrt(variance(Y))

.         mata: y
  .0574003903
. restore
```

```
// Individual-level relationship between study and wisdom
tab Study
encode Study, gen (Study_n)
tab Study_n
recode Study_n (1=0) (4=0) (15/18=0) (52/53=0) (2/3=1) (6/14=1) (19/25=1) (nonmissing=.), gen(Study_group)
tab Study_group
ttest corr, by(Study_group)
gen sessqid = session * 100 + q
xtset sessqid
xtlogit corr Study_group, fe
forvalues q = 1(1)30 {
    logit corr Study_group if q == `q'
}
//Actual group-level relationship between study composition and wisdom
bys session: egen percg = total(Study_group * (q == 1) / gs)
bys sessqid: gen sessqtag = _n == 1
bys session: sum percg if sessqtag
```

```
. // Simulated group-level relationship between study composition and crowd wisdom
. preserve

. mata: M = J(1, 1000, .)

.         forvalues it = 1(1)1000 {
  2.                 quietly {
  3.                 gen x = rnormal()
  4.                 sort Study_group q x
  5.                 by Study_group: gen sampled = _n <= 6
  6.                 gsort id -sampled
  7.                 by id: replace sampled = sampled[1]
  8.                 bys q: egen perccorr = total((corr * sampled) / 12)
  9.                 bys q: gen qtag = _n == 1
 10.                 gen mcorr = perccorr >= .5
 11.                 sum mcorr if qtag
 12.                 local rmean = r(mean)
 13.                 mata: M[1,`it'] = `rmean'
 14.                 drop x sampled perccorr qtag mcorr
 15.                }
 16.         }

.         mata: Y = sort(M',1)

.         mata: Y[25];Y[500];Y[975]
  .9
  .9666666667
  1

.         mata: y = sqrt(variance(Y))

.         mata: y
  .0330653896
. restore
```

```
. // Only Study_group == 0 (Sociologist)
. preserve

.           keep if Study_group == 0
(2,340 observations deleted)

.           mata: B = J(1, 1000, .)

.           forvalues it = 1(1)1000 {
                        quietly {
                        gen x = rnormal()
                        sort q x
                        gen sampled = _n <= 12
                        gsort id -sampled
                        by id: replace sampled = sampled[1]
                        bys q: egen perccorr = total((corr * sampled) / 12)
                        bys q: gen qtag = _n == 1
                        gen mcorr = perccorr >= .5
                        sum mcorr if qtag
                        local rmean = r(mean)
                        mata: B[1,`it'] = `rmean'
                        drop x sampled perccorr qtag mcorr
                        }
                }

.           mata: Y = sort(B',1)

.           mata: Y[25];Y[500];Y[975]
  .6666666667
  .7666666667
  .8666666667

.           mata: y = sqrt(variance(Y))

.           mata: y
  .0514220424
. restore
```

```
. // Only Study_group == 1 (non-sociologist)
. preserve

.         keep if Study_group == 1
(2,340 observations deleted)

.         mata: G = J(1, 1000, .)

.         forvalues it = 1(1)1000 {
  2.                 quietly {
  3.                 gen x = rnormal()
  4.                 sort q x
  5.                 gen sampled = _n <= 12
  6.                 gsort id -sampled
  7.                 by id: replace sampled = sampled[1]
  8.                 bys q: egen perccorr = total((corr * sampled) / 12)
  9.                 bys q: gen qtag = _n == 1
 10.                 gen mcorr = perccorr >= .5
 11.                 sum mcorr if qtag
 12.                 local rmean = r(mean)
 13.                 mata: G[1,`it'] = `rmean'
 14.                 drop x sampled perccorr qtag mcorr
 15.                 }
 16.         }

.         mata: Y = sort(G',1)

.         mata: Y[25];Y[500];Y[975]
  .7333333333
  .8
  .8666666667

.         mata: y = sqrt(variance(Y))

.         mata: y
  .0386995821
. restore
```