

<https://doi.org/10.5281/zenodo.19262853>

The Role of Social Media in Political Polarization in Pakistan



Hufsa Ali	MS Political Science, Central European University, Vienna, Austria hufsa.adnan@outlook.com
-----------	--

Dania Malik	MS Public Policy, Central European University, Vienna, Austria daniamalik44@gmail.com
-------------	--

Abstract: *This study examines how social media contributes to political polarization through the formation of “echo chambers,” where selective exposure limits individuals’ engagement with opposing viewpoints. Social networking platforms such as Facebook and Twitter allow users to curate their information environments and interactions, potentially reinforcing existing beliefs and promoting ideological extremity. Addressing ongoing debates in the literature, this research investigates whether exposure to reinforcement messages within such environments intensifies political polarization. Using a survey-based experimental design focused on the #MeToo movement, participants were randomly assigned to treatment and placebo groups. The treatment group received belief-congruent messages, while the placebo group was exposed to unrelated content. Pre- and post-surveys measured changes in attitudes. Employing Ordinary Least Squares (OLS) regression and a difference-in-difference approach, the findings reveal statistically significant effects, showing that reinforcement messages strengthen prior beliefs and push individuals toward more extreme positions. The study contributes to understanding the causal role of social media in shaping political attitudes in Pakistan and highlights implications for future research on digital communication and polarization.*

Keywords: Political polarization; Social media; Echo chambers; Confirmation bias; Reinforcement messages

Introduction

Political polarization, or the extremism of political attitudes or ideological radicalism, is an essential aspect of today’s politics. It manifests itself as increasing differences between opposing political groups and diminishing common political ground, where people with conflicting opinions can mutually agree. Moreover, it intensifies the differences between opposing narratives, thus giving way to radicalism and extremism. (Carothers, 2019). It also has implications for political stability in a diverse national context because of the

inevitable resentment between the opposing groups. Both extremes reinforce and abide by their own beliefs, thereby refusing to bind together and come to a mutual agreement. This increasing polarization is a worldwide concern, and despite substantial research on the topic, no consensus exists on the factors that contribute to political polarization in communities (O'Donohue and Carothers, 2019). However, substantial research suggests that social media plays an important role in promoting political radicalism. On the other hand, this is also true that technological advancements have made this world a global village. In this era of technology,

where social media has emerged as an effective tool for political communication and information dissemination (Barbera, 2015), it is also seen as a massive polarization machine because of its feature to let people interact with like-minded people and the formation of groups who defend their beliefs (Prior, 2013; Sunstein, 2002; Webster, 2005). It has been suggested that social media contributes to political polarization by creating an echo chamber environment, defined by selective exposure, which shields people from opposing viewpoints on specific issues (Bail et al., 2018). Facebook and Twitter, for example, may minimize contact with opposing and challenging perspectives and can further manipulate the formation of communities where like-minded individuals interact with each other, creating and reinforcing a shared view (Cinelli et al., 2020). However, it has been argued that this impact is sometimes exaggerated. According to Ngyuen and Vu (2019), there is no difference in polarization between those who get their news from social media sites and those who get their information from other sources.

Following the widespread concern regarding polarization and mixed findings in the literature, this paper attempts to find out the impact of social media on political polarization. I focus on the case of Pakistan because of its low literacy rate and people have less political knowledge. The interactive element of social media has particularly attracted youth who are not typically interested in political debates. They have begun to pay attention to political topics and engage in debate with others who are politically active and hold solid ideological beliefs. (Amjad, 2020) There is high dependence on social media for news and it is speculated to be the cause of increasing polarization in society. Henceforth, I will investigate

- *What is the impact of echo chambers, where people are exposed to arguments which reinforce their beliefs due to interaction with like-minded individuals?*
- *Do the exposure to reinforcement messages make individuals move away from moderate positions towards the extreme positions?*

- *Do the reinforcing messages shared through echo chambers on social media impact the minds of the networked public sphere and lead to a more politically polarized society in Pakistan?*

The strategy will test the 'echo chamber' effect with the specific focus on the impact of reinforcement messages and the *confirmation bias* theory. The term 'confirmation bias' implies that people often pay more attention to the content that aligns with their preexisting beliefs than the one that challenges their preexisting perspectives. (Stroud, 2008)

Literature Review

The Discourse of Political Polarization

Political polarization is one of the most heated debates and discussions amongst scholars and academics because of its adverse consequences on political stability (Leo, 2018). It is because when individuals align themselves strongly with a specific political ideology, they start binding themselves according to certain identities and political affiliations; and this world becomes a space with coexisting ideological groups ([Green et al., 2002](#), [Iyengar et al., 2012](#)). The strong biases towards one particular identity can consequently lead to inter-group conflicts, based on the inclination towards the 'us', and the hostility towards the 'them' group ([Turner et al., 1987](#)). The extremism of political affiliations can restrict individuals from having a civil discussion with the other groups and prevent them from coming and agreeing to a mutual ground (Lee, 2018). Thus, making political polarization an area of concern, and a crucial matter for academic debate.

Digital Intervention: Bridging Gaps or Intensifying Polarization?

Notably, in the last few years, because of the innovative intervention of digital technologies and active use of social media sites, the debates regarding the impact of social media on society's political polarization have started to get momentum. This argument can be traced to Van Alstyne and Brynjolfsson (1996), who argued that information technologies could result in 'cyber balkanization'. It means that technology

may "shrink geographic distances", henceforth forming a global village; however, it was also hypothesized that technology could lead to a split in the societies as a result of more balkanized interactions. It could be because "the internet makes it easier to find like-minded individuals, strengthening far-flung communities which have the same ideology and are scattered geographically"; and once like-minded people align with each other, "their subsequent interactions can further polarize their views or even ignite calls-to-action".

Additionally, Tornberg (2018) claims that despite the earlier optimism regarding this virtual world that seemed to be bridging gaps, the online world is now beginning to seem less and less likely like a standard "table" that aims to "gathers us together" so that open discussions can take place. On the contrary, the virtual world seems to bring forth the worst kind of human instincts; it clusters humans together in groups that essentially interact with other individuals for the reaffirmation of their beliefs so that they can protect themselves from disagreements. It is like "echo chambers" that are in place to primarily reinforce the existing narratives and foster confirmation biases. (Bessi, 2015) At the same time, it is also argued that social media can support diverse discourses and narratives; however, there is a widespread concern that platforms like Twitter do not have a mechanism that can result in the communication amongst contradictory narratives. (Shore et al., 2018).

Moreover, Sunstein (2018) claims that the rise of the biased news channel and biased social-media sites can intensify the biases, further polarizing people's opinions, thereby driving them towards an extreme ideological discourse. This claim is based on the premise that due to the extensive accessibility and availability of information on social media, the individuals opt for selective exposure where they choose to receive the information they agree with and avoid the ones they disagree with. Thus, the argument states that higher exposure to the same content reinforces and strengthens one's beliefs, and eventually drives them towards extreme positions ([Stroud, 2010](#)).

Social Networking Sites and The Trends of

Political Polarization

Despite this straightforward relationship between social media sites and political polarization, the consensus in the literature is not very clear. Boxell et al. (2017) acknowledge the fact political polarization has recently increased, the surge is also witnessed for non-active social media users. According to their research, "the internet explains a small share of the recent growth in polarization". Nguyen and Vu (2019) also reinstated a similar conclusion after analyzing the survey data from 28 European countries. The result implied that there is no evident difference in the polarization level of the ones who rely mainly on social media for their source of information, and the ones who rely on their sources for the information. Interestingly, the research also reflects that those individuals who engage in cross-cutting discussions on social media have a lower level of political polarization (Heatherly, Lee, and Lu 2017).

Furthermore, Bakshy et al. (2015) and Barberá (2015) share that Facebook and Twitter users are also being observed to have a surprisingly high level of diverse perspectives. Barberá has also made an effort to observe if the virtual communication takes the form of an "echo chamber" or a "national conversation". In his research, he found out that the discussions on online platforms are highly issue-dependent and that political polarization greatly depends on the topic of the discussion. According to his research, the universalization of the virtual world does not restrict the cross-cutting engagement and discussion within the networked public spheres. This is particularly true when it comes to nonpolitical ideological discussion. However, when political matters are discussed, there is indeed a visible inclination towards the information received from the ideologically similar resources compared to the ideologically opposing resources. (Barbera et al. 2015)

In an article, "Network structure and patterns of information diversity on Twitter", it was reported that on average, Twitter users post more centrist content than the content they receive on their feeds; thereby, undermining the prevailing narrative of social networking sites being echo

chambers. (Schmidt et al., 2017) Most probably, the widespread view of such polarization might result from a network paradox where the behavior of nodes can be mistaken to be typical (Feld, 199). According to his findings, the members of the social network core do show cross-sectional polarization. They are responsible for most of the tweets received due to their active status and popularity that could also explain the polarization on social networking sites.

Echo Chamber, Selective Exposure and Ingroup Homophily

Sunstein (2002) defines group polarization as “the members of a deliberating group that intentionally move towards a more extreme position indicated by the member’s deliberation tendencies” In this scenario, the echo chamber is perceived as a mechanism that reinforces the existing ideologies within a specific group, and thus moving the individuals towards the extreme position. The effect of echo chambers has been witnessed in diverse digital platforms like blogs (Gilbert et al., 2009), forums (Edward, 2013), and social networking sites (Barbera, 2015). Literature also highlights the impact of an echo chamber on human tendencies and behavior like selective exposure and ingroup homophily (Baumann, 2020; Sunstein, 2001). Selective exposure entails that humans tend to view and perceive the information that aligns with their existing beliefs ([Prior 2002](#)). The homophily implies that individuals like to surround themselves with like-minded individuals who share similar characteristics, such as gender, socio-economic background, and political orientation. ([McPherson et al. 2001](#)). Both selective exposure and homophily are such aspects of human behaviorism that can further prompt the echo chambering- because the continuous exposure of similar information and being surrounded by like-minded individuals strengthens the existing narratives of human beings. (Garimella et al., 2018; Sunstein, 2001) Despite the concurring research on the functioning of echo chambers, the impact and the presence of echo chambers have been questioned in the last few years. (Dubois, 2018; Barbera, 2015; Bruns, 2017)

Theoretical framework

Echo Chamber Effect

The echo chamber is defined as a technique for reinforcing an existing group's perspective with increased exposure to posts containing reinforcement messages, with the intention that the entire group will gravitate towards extreme positions. No one can read every post on the internet or connect with every user; instead, individuals must choose where to focus their attention. Because of the human tendency toward homophily (McPherson, Smith-Lovin, and Cook, 2001) and confirmation bias (Nickerson, 1998), social media users are more likely to follow those who share their views. In terms of the theoretical approach, it has been suggested that increased access to information on social media may lead to selective exposure to ideologically congenial content, creating an "echo chamber" setting that could foster social extremism and political polarization (Adamic & Glance, 2005; Prior, 2007; Iyengar & Hahn, 2009). In current literature, the related idea is defined as an "echo chamber," which mirrors one's own opinions back to reinforce established beliefs (Bail et al. 2018). An *echo chamber* is defined in this network as users who share two characteristics: *opinion* and *network polarization*. When it comes to a particular issue, *opinion polarization* indicates that they are more likely to hold similar opinions. The term "network polarization" refers to the fact that they are more densely connected than outside networks (Tornberg, 2018). In other words, an echo chamber is a network of nodes that are more likely to share a common opinion on a particular subject. In this study, the echo chamber is defined as a technique for reinforcing an existing group view by increased exposure to posts containing reinforcement messages, hoping that the entire group will gravitate towards extreme positions.

Confirmation Bias

Another theory that dominates the literature on this subject is the ‘confirmation bias concept, which was popularized by Raymond Nickerson (1998) in psychology and is now embraced by other authors regarding social media and

political polarization (Geschke et al. 2019). It suggests that, rather than questioning their inherent preconceptions, social networking pays more attention to arguments that are likely to affirm their preexisting biases (Sanders, 2020). Polarization can be induced by introducing "stubborn" agent's who want to stick to their initial beliefs rather than interacting with and learning from their peers (Acemoglu, 2010). This technique is similar to confirmation bias. Overall, while the literature shows that some notion of "bias" in networks is necessary to reproduce realistic dynamics of opinion formation, it remains challenging to provide a cohesive framework that accounts for information aggregation, polarization, and learning (Sikder, 2020). This research aims to naturally capture the influence of confirmation bias on social interaction through an experimental investigation and examine how it can radically change the way a networked, decentralized society processes information and leads to political division. This research builds on confirmation bias research, intending to test a hypothesis based on the concept of an "echo chamber."

Research design

Sampling

Random sampling is opted for this experiment because the research aims to keep the selection biases to a minimum. The survey aims to give equal representation and reach out to the population from diverse socio-cultural and political backgrounds through randomized selection. Additionally, randomized selection also helps in collecting a representative sample of youth active on social media. The advertisement is posted on Facebook for the sampling procedure. The Advertisement was available to all Facebook users in Pakistan between 18 and 40, as the target group is the young generation. Through the Facebook algorithm, the survey link is randomly exposed to the target group, and they were asked to consent to participate in the survey.

Furthermore, to maintain the authenticity and validity of the survey, financial incentives were also be provided to the participants. The

economic incentives play a pivotal role in ensuring that participants maintain their focus throughout the survey, concentrate on questions, and answer the questions with utmost authenticity. Another critical factor in ensuring the research's validity is keeping respondents engaged until the end of the experiment and convince them to read the content shared with them. In this research, financial incentives are used to encourage the participants to take part in the survey and increase the representation.

The respondents are provided with the opportunity to win (up to Euros 200). This incentive was included in the Facebook advertisement, and only those respondents who completed the experiment are eligible for the lucky draw. The poll was completely anonymous, the data was used in aggregate, and participation was entirely voluntary. The survey was created using Qualtrics, and the link to the survey was be distributed to participants via advertising.

Data Collection

The data is collected in an artificial setting. The treatment and placebo groups are assigned at random; the former one gets the treatment, and efficacy is measured compared to the placebo group. For this experiment, Facebook stimulation through an artificial setting is constructed because, in an artificial environment, the information accessed by the respondents can be controlled in a manner that a causal relationship between the information and the outcome polarization can be gauged. On the contrary, in a natural environment, there is the influence of external factors that impact the information accessed by the participants, the unpredictability of human behavior, uncertainty about the content to be shared in the upcoming posts, and other factors like the social and cultural background of the respondent. For this reason, one cannot conclude that the relationship between the variables is a causal one and not a correlation. However, in an artificial setting, complete information accessible by the respondents minimizes the uncertainty and unpredictability during the experimentation, and randomization minimizes the impact of omitted variable bias.

Experiment Design and Survey Flow

The study employs a survey experiment with a sample of 270 participants, who were randomly assigned to treatment and placebo groups. Both groups were initially asked a series of questions to assess their ideological positions and views on the #MeToo movement. These baseline responses were used to determine participants' initial stance on the issue.

Following the pre-survey, the treatment group was exposed to reinforcing posts aligned with their existing beliefs, while the placebo group received content unrelated to the #MeToo movement. The purpose of this design was to isolate the effect of reinforcement messages on participants' attitudes. After exposure, both groups were asked the same set of questions again to capture any changes in their positions. To simulate a realistic social media environment, participants were placed in an artificial Facebook-like setting where they viewed posts and indicated their likely reactions (e.g., like, comment, share, ignore). This design allowed the study to replicate typical user interactions on social networking platforms while maintaining control over the content exposure.

The key objective was to compare participants' responses before and after exposure to treatment. Changes in their ideological positions were used to evaluate the extent to which reinforcement messages influenced their beliefs. The ideological difference between treatment and placebo groups was quantified, and the level of polarization was assessed.

To estimate causal effects, the study employed a **difference-in-difference approach**, comparing pre- and post-treatment responses across both groups. This method enabled the identification of changes attributable specifically to the intervention (reinforcement messages), rather than external factors.

Measurement of Political Polarization

Political polarization was operationalized through participants' responses to a set of survey questions. The central measure was a self-reported scale assessing support for the #MeToo

movement. Participants were asked to indicate their position on a scale from 1 to 10, where:

- 1 represented strong opposition, and
- 10 represented strong support

Based on this response, participants were categorized into two groups:

- Opponents (≤ 5)
- Supporters (≥ 6)

To ensure the validity of this classification, additional follow-up questions were included addressing related issues such as women's rights and gender equality. These questions served as robustness checks to confirm participants' ideological alignment and strengthen the reliability of the measurement.

Treatment and Placebo Conditions

After the initial classification, participants were exposed to different types of content depending on their group assignment.

The treatment group received posts specifically designed to reinforce their preexisting beliefs. Participants who supported the #MeToo movement were shown content that framed the movement positively, while those who opposed it were exposed to content that reinforced opposing views. This ensured that reinforcement operated in both ideological directions.

The placebo group, on the other hand, was exposed to neutral and unrelated content on general social, political, and economic issues. This was done to ensure that any observed changes in the treatment group could be attributed to the reinforcing nature of the content rather than mere exposure to information. Importantly, the structure and format of content presentation were kept consistent across both groups. This helped control for potential biases arising from differences in content delivery.

Outcome Variable

The primary outcome variable in this study is political polarization, defined as the movement of individuals toward more extreme ideological positions. Polarization was measured by calculating the change in participants' responses between the pre- and post-survey. A shift toward

higher or lower values on the scale indicated movement toward stronger support or opposition, respectively.

Additionally, group-level polarization was assessed by examining the distance between the mean positions of supporters and opponents. An increase in this distance indicated higher polarization, while a decrease suggested convergence toward moderate positions.

Group Classification and Analytical Framework

Participants in both treatment and placebo conditions were further categorized into four subgroups based on their initial responses:

- Treatment Supporting
- Treatment Opposing
- Placebo Supporting
- Placebo Opposing

This classification enabled a more nuanced analysis of how reinforcement messages affected different segments of the population. The study employed a difference-in-difference analytical framework to estimate the impact of treatment. This involved calculating:

- The change in polarization within each group (pre vs. post)
- The difference between treatment and placebo groups

This approach allowed for the identification of the causal effect of reinforcement messages. The experimental design is based on the premise that exposure to ideologically consistent information reinforces existing beliefs. By comparing changes in attitudes between treatment and placebo groups, the study isolates the effect of reinforcement messages.

The expectation is that individuals exposed to reinforcing content will exhibit greater movement toward ideological extremes, thereby increasing polarization. In contrast, individuals in the placebo group are expected to show minimal or no systematic change. Overall, the experimental setup provides a controlled framework to test the echo chamber

effect and the role of confirmation bias in shaping political attitudes. By simulating a social media environment and systematically varying exposure to reinforcement messages, the study offers a robust approach to understanding how social media contributes to political polarization.

Analysis and discussion

Descriptive Statistics

As explained earlier, Facebook advertisements were used to recruit the participants. The initial responses were 293. However, some respondents did not complete the survey. The final number of responses that contributed to this analysis was 270. The responses were randomly assigned to the treatment group using Qualtrics in-built randomization mechanism. The randomization minimized the impact of omitted variable bias. In order to analyze the causal impact of treatment, the placebo and control groups were matched for different demographic factors. Since the support of the #Meetoo Movement can vary due to the Education level, Region, Marital status, Employment, etc., the statistical summary was drawn for different factors to ensure that the two groups are similar.

Summary Statistics

intervention Variable	N	Placebo Mean	SD	N	Treatment Mean	SD	Test
Treatment	133	0	0	136	1	0	F=1.777965065322 2e+31***
Gender	133			136			X2=2.255
... Female	69	51.9%		68	50%		
... Male	61	45.9%		60	44.1%		
... Prefer not to say	3	2.3%		8	5.9%		
Education	133			136			X2=2.98
... Bachelors	47	35.3%		45	33.1%		
... Intermediate	20	15%		25	18.4%		
... Masters/ M.Phil	56	42.1%		61	44.9%		
... Matric	6	4.5%		4	2.9%		
... Other	4	3%		1	0.7%		
Region	133			136			X2=1.889
... Khyber Pakhtunkwa	25	18.8%		30	22.1%		
... Other	14	10.5%		11	8.1%		
... Punjab	68	51.1%		62	45.6%		
... Sindh	26	19.5%		33	24.3%		
Marital_Status	133			136			X2=0.867
... Divorced/ Separated	2	1.5%		3	2.2%		
... Married	50	37.6%		52	38.2%		
... Prefer not to say	9	6.8%		6	4.4%		
... Single, Never Married	72	54.1%		75	55.1%		
Employment	133			136			X2=2.714
... Employed	41	30.8%		45	33.1%		
... Intern	4	3%		2	1.5%		
... Prefer not to say	3	2.3%		5	3.7%		
... Student	40	30.1%		32	23.5%		
... Unemployed	45	33.8%		52	38.2%		
Socialmedia_usage	133			135			X2=3.542
... 1- 2 hours	39	29.3%		40	29.6%		
... 2 - 3 hours	29	21.8%		36	26.7%		
... 30 to 60 minutes	17	12.8%		17	12.6%		
... Less than 30 minutes	7	5.3%		2	1.5%		
... More than 3 hours	41	30.8%		40	29.6%		
News_source	133			136			X2=3.289
... Directly from people (family members, friends, co-workers, etc.)	2	1.5%		6	4.4%		
... Print (newspaper, magazine)	5	3.8%		2	1.5%		
... Social Media (e.g. Facebook, T witter, Whatsapp, YouTube, Instagram, Telegram, etc.)	111	83.5%		112	82.4%		
... Television	15	11.3%		16	11.8%		
Reliousness	113	7.08	1.905	118	7.144	2.039	F=0.061
Religious practices	104	6.663	2.56	113	6.487	3.12	F=0.206
Traditionalist	121	7.339	2.064	127	7.417	2.28	F=0.081

Statistical significance markers: * p<0.1; ** p<0.05; *** p<0.01

Figure 1: The statistical summary of the data.

The above table shows the balance of different factors in placebo and treatment groups. The chi-square test shows the statistical significance, and it can be seen that none of the variables has the statistically significant value for the difference.

The visual representation below also shows the comparison and the distribution of respondents according to different factors which might impact their position regarding the movement.

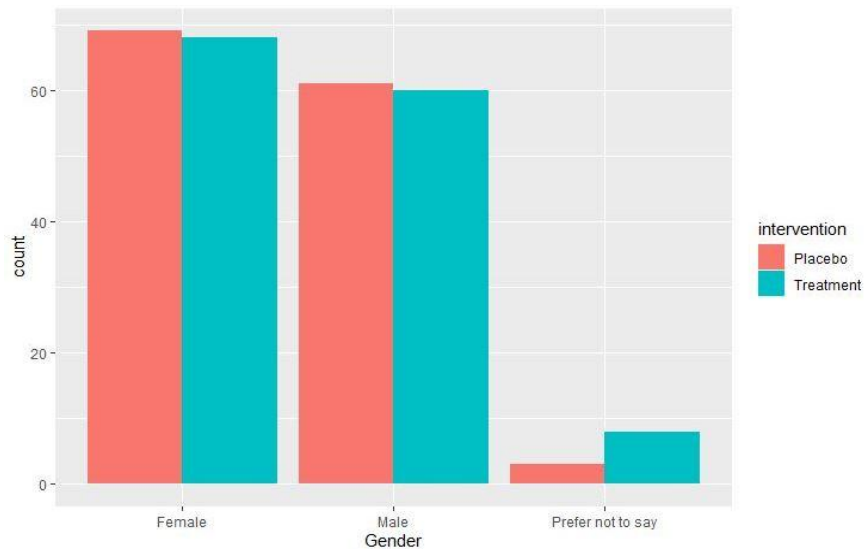


Figure 2: The statistical distribution of respondents on the basis of gender.

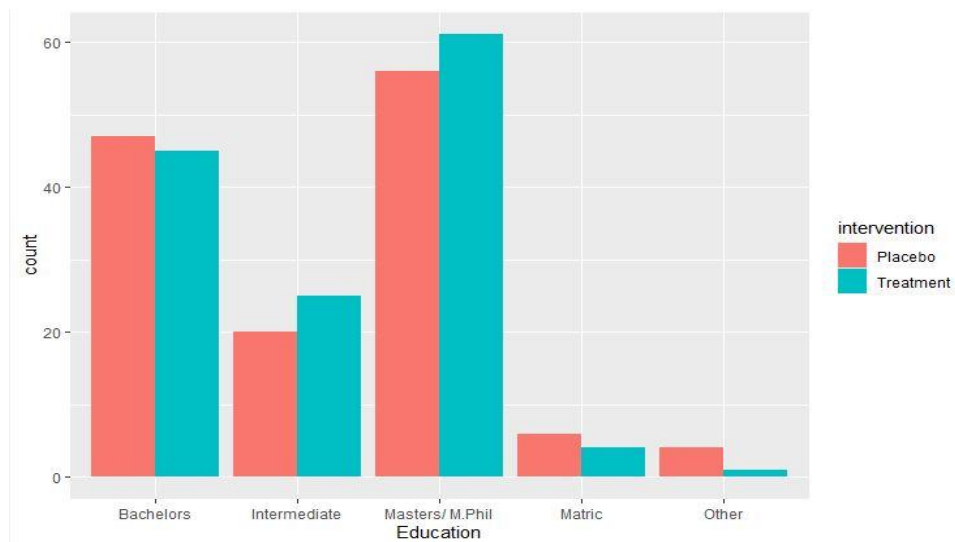


Figure 3: The statistical distribution of respondents on the basis of education

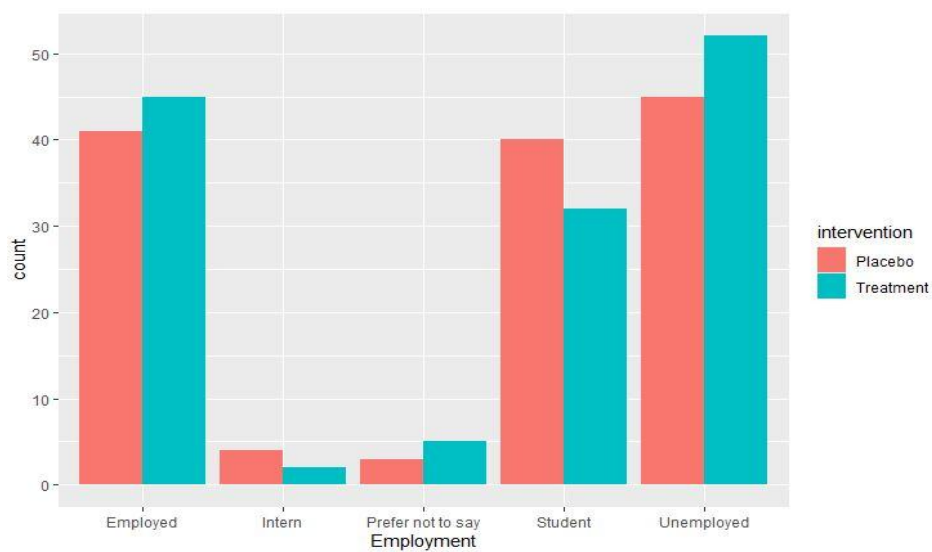


Figure 4: The statistical distribution of respondents on the basis of employment.

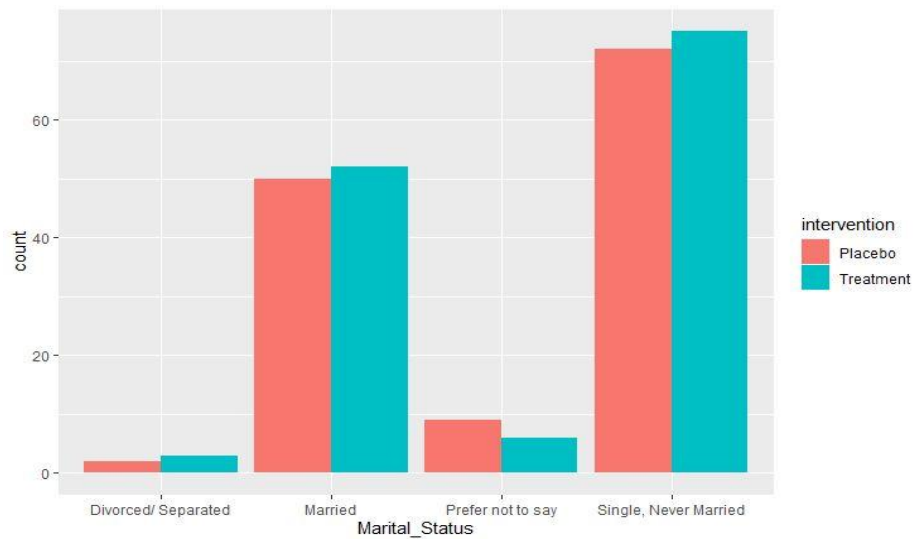


Figure 5: The statistical distribution of respondents on the basis of marital status.

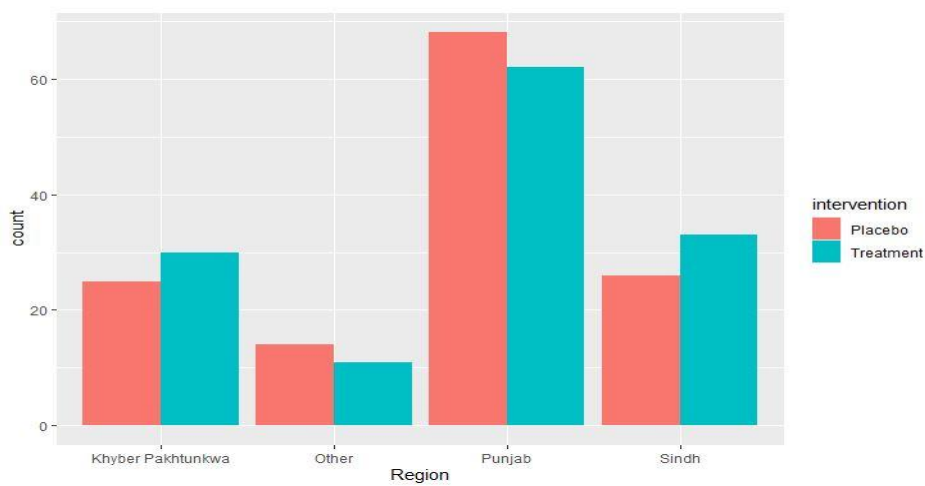


Figure 6: The statistical distribution of respondents on the basis of region.

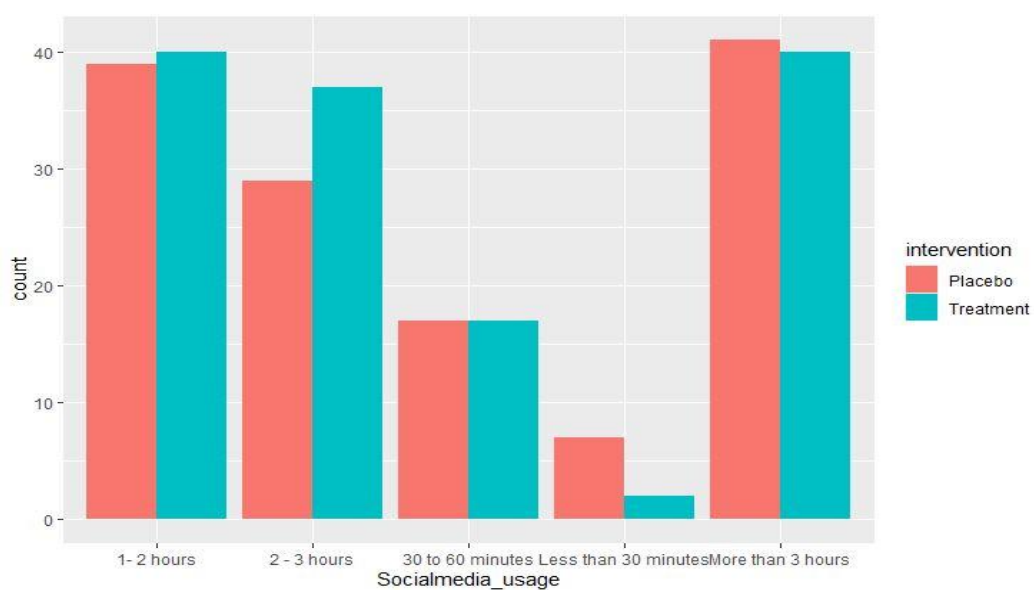


Figure 7: The statistical distribution of respondents on social media usage.

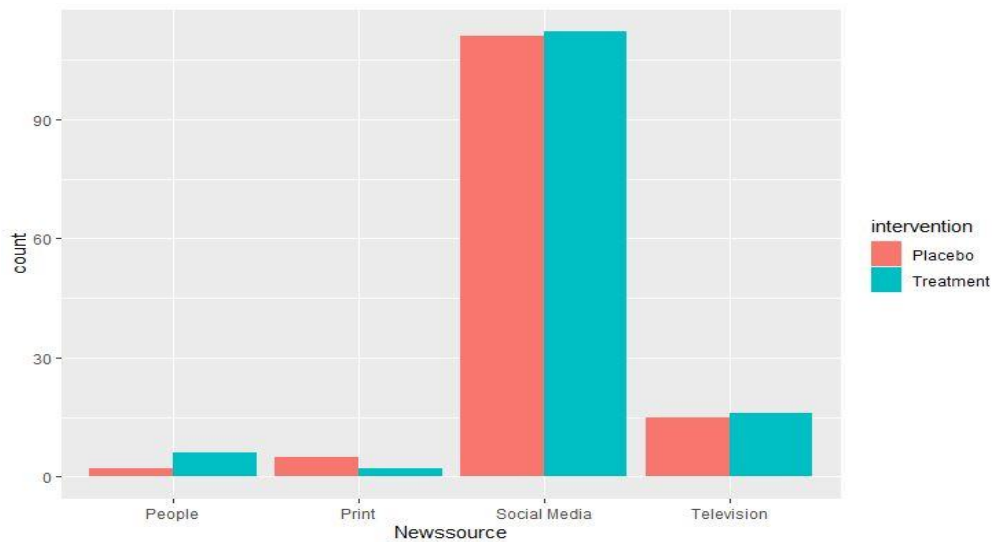


Figure 8: The statistical distribution of respondents on the basis of news source.

Box plot was used to analyze categorical variables (Gender, Region, Marital Status, etc.), whereas continuous variables like age were analyzed using the box plot. All the plots showed a similar distribution of participants and

showed that the groups are balanced across various demographic factors. The distributions according to religious inclination and age are given as:

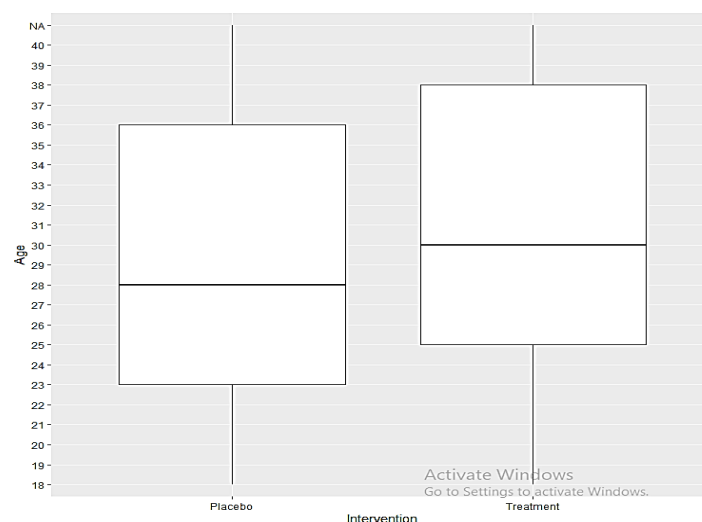


Figure 9: The distribution of respondents according to age.

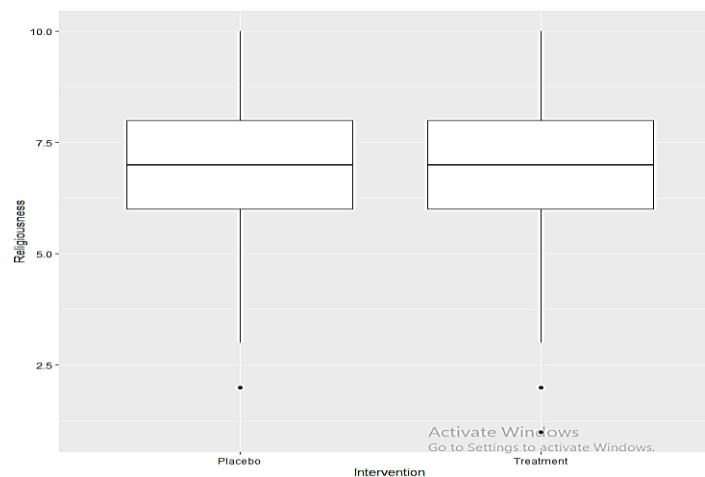


Figure 10: The distribution of respondents according to religiousness.

Other than balancing the groups on demographic factors, the initial position of the respondents is also crucial. The pre-treatment position of supporting groups in the placebo group was compared with that of in treatment groups to ensure that both the groups are comparable. The

change towards the end of the experiment is the causal change as an effect of the treatment. In the same way, the comparison was made for the placebo and treatment groups for initial opposing behaviour. The visual representation of the behaviour is given as:

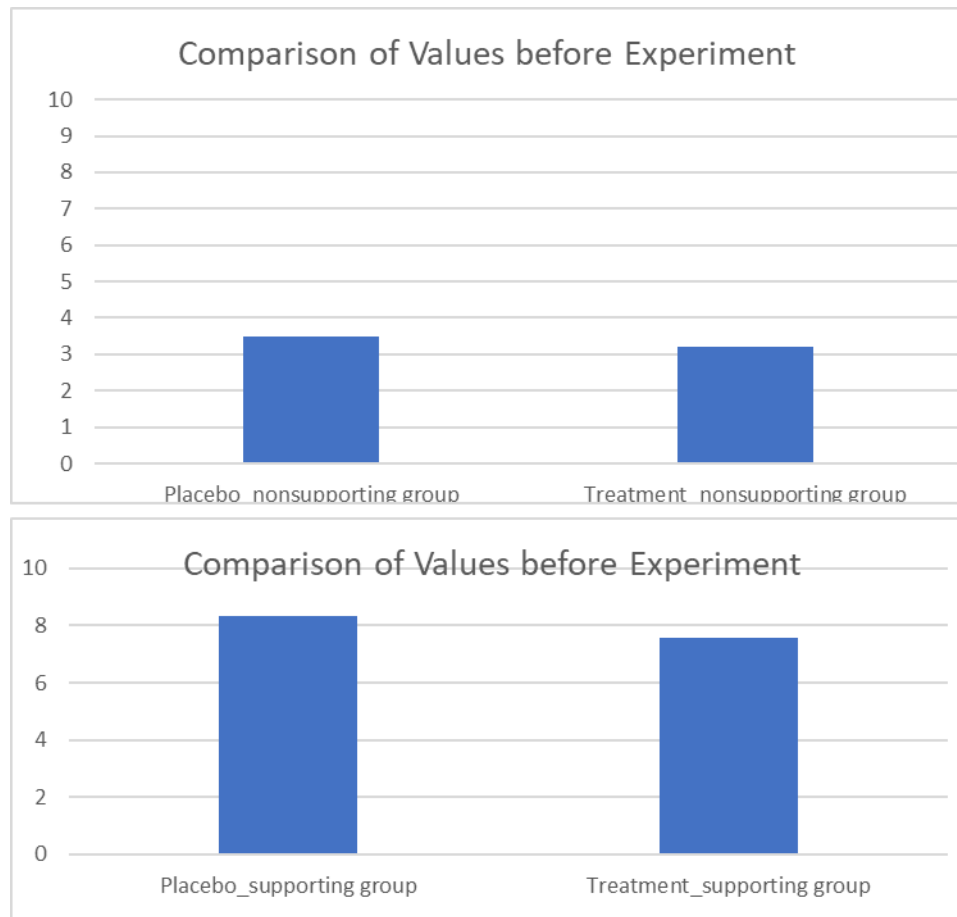


Figure 11: Comparison of values of two groups for initial position

Statistical Analysis

In order to see the impact of the intervention on the supporting and opposing behaviors of the treatment groups, regression analyses were being conducted. Individual polarization of the groups is calculated by the difference of the values take before and after the treatment. This individual polarization was the outcome variable in these analyses. Thus, the dependent variable was support change, and the explanatory variable was a dummy variable for the treatment, showing whether a respondent has got the treatment or not. For both the supporting and

opposing groups, linear regression was being run between change in support (dependent variable) and treatment (an independent variable that indicates whether a person was in a placebo or treatment group). Control for variables like gender, region, education, marital status etc., were also included in the regression analyses to estimate the causal effect of a treatment on an outcome. It showed whether supporters become more supportive when they were treated with reinforcing messages. In another regression model, the whole process was repeated for the opposing group.

Table 1: OLS regression model for Treatment effect

	Model 1	Model 2
(Intercept)	0.32 (0.81)	0.65 (0.61)
Treatment Supporting	1.63*** (0.42)	
Gender. Male	-0.41 (0.43)	-0.00 (0.30)
Education. Intermediate & Below	0.11 (0.55)	0.61 (0.45)
Education. Masters/ M.Phil.	-0.94 (0.50)	-0.14 (0.33)
Region. South Region	-0.05 (0.45)	0.33 (0.33)
Marital_ Status. Single	0.53 (0.49)	-0.47 (0.34)
Employment. Student	-0.99 (0.63)	-0.52 (0.42)
Employment. Unemployed	-0.87 (0.51)	-0.65 (0.35)
Social Media Usage. More than 2 hours	0.05 (0.43)	-0.27 (0.29)
News Source. Social Media	-0.40 (0.58)	0.43 (0.43)
Treatment Opposing		-0.91** (0.30)

R ²	0.17	0.15
Adj. R ²	0.10	0.08
Num. obs.	129	140

*** p < 0.001; ** p < 0.01; * p < 0.05

The regression table shows that the 'Treatment supporting' coefficient is significant and has a positive value, which shows the positive association between the treatment (reinforcement messages) and supportive group.

Similarly, 'Treatment opposing' also showed statistically significant results but in the opposite direction. The effect of the treatment is also apparent in the regression plot.

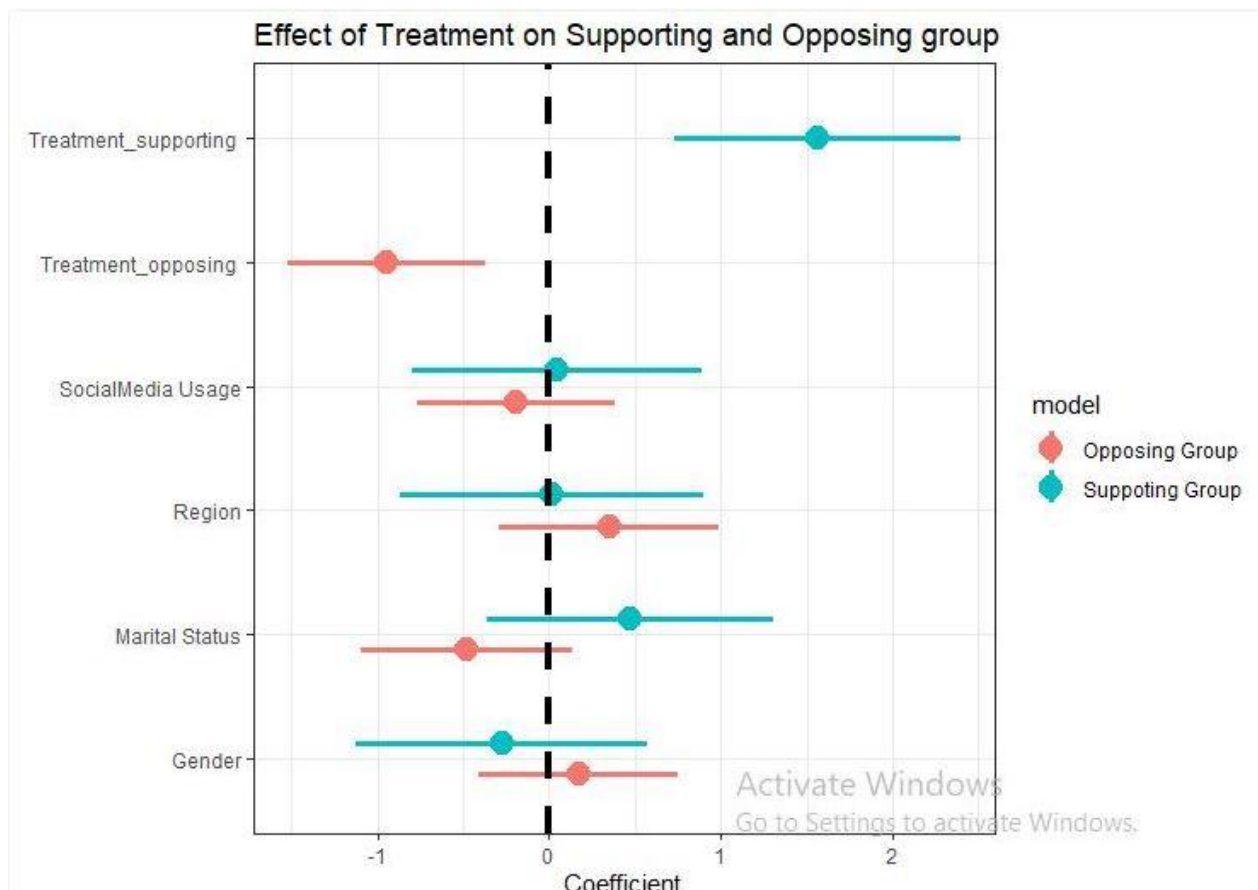


Figure 12: Treatment effect on Supporting and Opposing Groups

The figure summarizes the effect of the intervention on the treatment groups. The first hypothesis assumes that the positive reinforcing me too movement messages would lead to stronger support of the Metoo movement. As evident through the linear regression the treatment of supporting group has a value of 1.63 and is statistically significant ($p < 0.001$). This impact is also apparent in the figure 13, which shows that supporting treatment has a strong effect when it comes to change in position for the supporting groups. Hence the experiment provides evidence that the first Hypothesis (H1) is supported.

In the same way, the second hypothesis assumes that negative reinforcing me too movement messages would lead to stronger opposition of Metoo movement. The regression analysis shows a negative value of 0.91 with statistical significance ($p < 0.01$). The negative value indicates that the respondents change their position in the negative (as opposing to the supporter's) direction. Figure 13 also indicates the strong effect of the treatment on the opposing

group, hence providing evidence in support of the second hypothesis (H2).

These changes before and after are observed through the difference between the observations of the treatment and the placebo group. This difference explains the polarization as the polarization is the distance between positions on the spectrum. In this scenario, the polarization is calculated by the mean difference of the score of the opposing and supportive groups based on the stance of participants on the Metoo movement. The change in polarization will be calculated by the difference of the polarization taken before and after the treatment (for both the placebo and the placebo treatment groups separately).

The mean value calculated for the placebo-supporting and non-supporting groups before the experiment were 8.3 and 3.5, respectively, and after the experiment were 7.4 and 3.8, respectively. The placebo groups' polarisation before and after the treatment is calculated by finding the difference of values of supporting and opposing groups (before and after the treatment). The polarization values for the

placebo group before and after the treatment are given as

:

Polarization of placebo group before = mean value of placebo supporting before – mean value of placebo opposing before
 $= 8.3 - 3.5$
 $= 4.84$

Polarization of placebo group after = mean value of placebo supporting after – mean value of placebo opposing after
 $= 7.4 - 3.8$
 $= 3.67$

Similarly, the mean values calculated for the position of treatment supporting and non-supporting group before the experiment were 7.5 and 3.2, respectively, and after the experiment were 8.3 and 2.6, respectively. The treatment groups' polarisation before and after the

treatment is calculated by finding the difference of values of supporting and opposing groups (before and after the treatment). The polarization values for the treatment group before and after the treatment are given as:

Polarization of treatment group before = mean value of treatment supporting group before – mean value of treatment opposing group before
 $= 7.57 - 3.2$
 $= 4.37$

Polarization of treatment group after = mean value of treatment supporting group after – mean value of treatment opposing group after
 $= 8.3 - 2.6$
 $= 5.73$

Once the values of polarization for all the four groups are obtained, the polarization for both (Placebo and treatment) groups can be

calculated by calculating the difference of the values before and after the treatment.

Polarization for the placebo group (Difference for placebo) = Polarization of placebo group after - Polarization of placebo group before
 $= 3.67 - 4.84$
 $= -1.17$

Polarization of treatment group (Difference for treatment) = Polarization of treatment group after - Polarization of treatment group before
 $= 5.73 - 4.37$
 $= 1.36$

The polarization value for the treatment group is 1.36, which shows that polarization has increased by 1.36 points due to the intervention. At the same time, the value of polarization for the placebo group is -1.17. The negative value indicates the decrease in polarization level for

this group.

In order to analyse the impact of the treatment (reinforcement messages), the difference in difference method is followed, and calculated polarization of the placebo group is subtracted

from that of the treatment group.

$$\begin{aligned}
 \text{Difference-in-difference} &= \text{Difference for treatment group} - \text{Difference for placebo group} \\
 &= 1.36 - (1.17) \\
 &= 2.53
 \end{aligned}$$

The number 2.53 is a causal change, and it shows that the intervention (reinforcement messages) has caused an increase in polarization as assumed by the third hypothesis (H3). The echo chamber effect is created in the lab experiment and as expected the reinforcement messages made the treatment supporting and opposing groups change their positions towards the extreme ends. The increase of the distance between the mean positions of both the groups increases the polarization of the group meeting the theoretical expectations. But for the placebo group, results came out to be a bit unexpected. The supporting and opposing groups were expected to retain their position in before and after conditions as they were not

exposed to any statement regarding #metoo movement. But their position was still changed and they moved towards the center (reducing the distance). Hence a decrease in polarization was observed for this group. This is just a speculation that time has an impact in such a case, and that with time question gets sinks in respondents' minds and they respond differently when they are asked the same question the next time. Whatever the reason is, one aspect is worth noting here, that condition would hold for the treatment group here as well. This is clearly shown in the graphical representation of difference-in-difference estimation in the figure below.

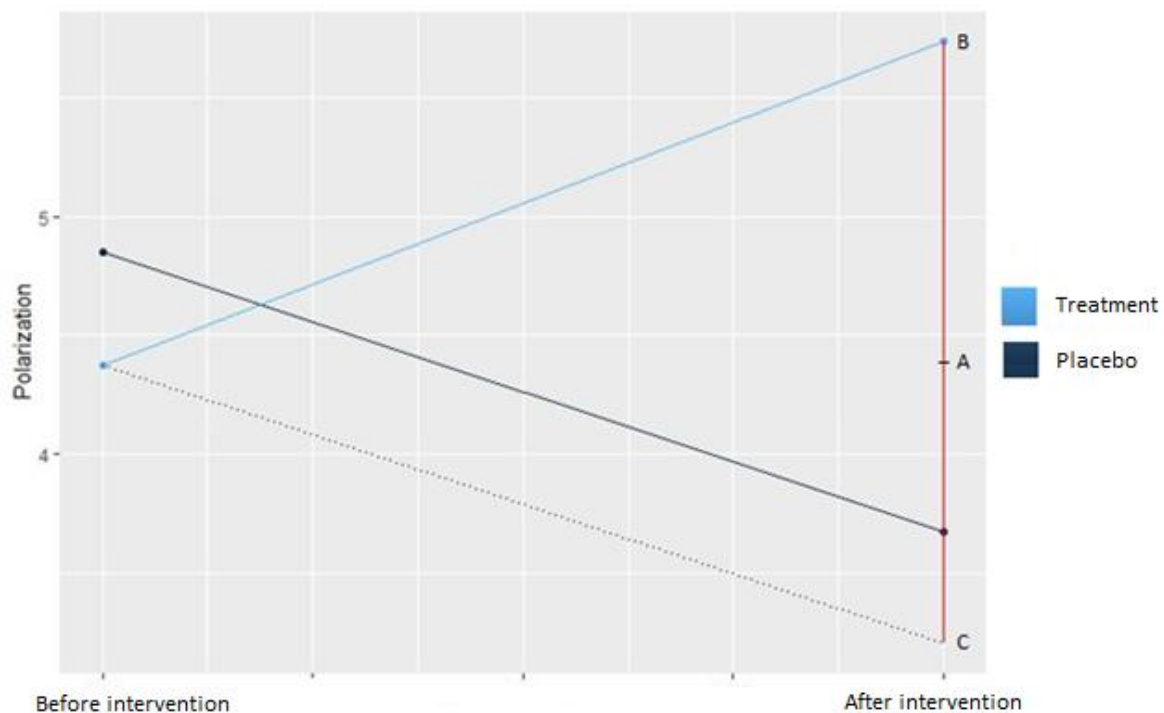


Figure 13: The graphical representation of difference-in-difference estimation.

The difference in difference analysis also relies on a parallel assumption that the unobserved differences between treatment and control groups are constant over time in the absence of the treatment. In the absence of the treatment,

the expected trend is shown by the dotted line in the image. The blue line represents the actual trend for the treatment group, and the red line shows the difference in trend due to intervention. Hence the treatment effect is really strong in this

case and is even more than what we see apparently from point A to B (the difference of polarization before and after = 1.36). In fact, it is measured from where the treatment group would have been in the absence of the treatment to where it actually is (Point C to B). This distance is indicated by the vertical red line in the figure above (value = 2.53). It supports the assumption made by the third hypothesis that reinforcement messages increase the polarization. The size of polarization in treatment group is greater than the placebo group. Though it can't be said that null hypothesis is rejected, the evidence suggests that the third hypothesis (H3) is supported.

Conclusion

This study demonstrates that social media contributes to political polarization through the formation of echo chambers, where selective exposure and reinforcement messaging intensify individuals' pre-existing beliefs while limiting engagement with opposing viewpoints. Addressing a key gap in the literature, the research extends the debate beyond Western contexts by providing experimental evidence from Pakistan, a setting marked by rapid digital adoption and comparatively low media literacy. Using a survey-based experimental design with random assignment to treatment and placebo groups, the findings show that exposure to belief-congruent content particularly around a polarizing issue such as the #MeToo movement significantly strengthens ideological positions and pushes individuals toward greater extremity. The results, supported by OLS regression and difference-in-difference analysis, confirm the theoretical expectation that echo chamber dynamics and confirmation bias reinforce polarization. While the study establishes causal relationships, its generalizability is limited due to the controlled experimental setting, relatively small sample size, and focus on a specific age group. Nonetheless, it makes an important contribution by highlighting how socio-cultural context, limited access to diverse information, and lower literacy levels can amplify the polarizing effects of social media in developing societies. The findings underscore the need for future research in more naturalistic settings, with

broader samples and longer exposure periods, as well as deeper investigation into time effects, differences between positive and negative reinforcement, and the unexpected depolarization observed in placebo groups. Overall, the study emphasizes the growing influence of algorithmically mediated content in shaping political attitudes and calls for greater attention to policy and platform-level interventions to mitigate polarization.

References

- Acemoglu, Daron, Asuman Ozdaglar, and Ali ParandehGheibi. "Spread of (mis) information in social networks." *Games and Economic Behavior* 70, no. 2 (2010): 194-227.
- Adamic, Lada A., and Natalie Glance. "The political blogosphere and the 2004 US election: divided they blog." In *Proceedings of the 3rd international workshop on Link discovery*, pp. 36-43. 2005.
- Amjad, Kinza, Muzammil Saeed, Farahat Ali, and Muhammad Awais. "Social Media Use and Political Polarization: Political Engagement as a Mediator." *The Journal of Social Sciences Research* 6, no. 8 (2020): 804-810.
- Baumann, Fabian, Philipp Lorenz-Spreen, Igor M. Sokolov, and Michele Starnini. "Modeling echo chambers and polarization dynamics in social networks." *Physical Review Letters* 124, no. 4 (2020): 048301.
- Bail, Christopher A., Lisa P. Argyle, Taylor W. Brown, John P. Bumpus, Haohan Chen, MB Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky. "Exposure to opposing views on social media can increase political polarization." *Proceedings of the National Academy of Sciences* 115, no. 37 (2018): 9216-9221.
- Bakshy, Eytan, Solomon Messing, and Lada A. Adamic. (2015). "Exposure to Ideologically Diverse News and

- Opinion on Facebook." *Science* 348 (6239): 1130-1132.
- Barberá, Pablo, John T. Jost, Jonathan Nagler, Joshua A. Tucker, and Richard Bonneau. "Tweeting from left to right: Is online political communication more than an echo chamber?." *Psychological science* 26, no. 10 (2015): 1531-1542.
- Bessi, Alessandro, Fabiana Zollo, Michela Del Vicario, Antonio Scala, Guido Caldarelli, and Walter Quattrociocchi. "Trend of narratives in the age of misinformation." *PloS one* 10, no. 8 (2015): e0134641.
- Boxell, Levi, Matthew Gentzkow, and Jesse M. Shapiro. "Greater Internet use is not associated with faster growth in political polarization among US demographic groups." *Proceedings of the National Academy of Sciences* 114, no. 40 (2017): 10612-10617.
- Bruns, Axel. "Echo chamber? What echo chamber? Reviewing the evidence." In *6th Biennial Future of Journalism Conference (FOJ17)*. 2017.
- Carothers, Thomas, and Andrew O'Donohue, eds. *Democracies divided: The global challenge of political polarization*. Brookings Institution Press, 2019.
- Carothers, Thomas, and Andrew O'Donohue. "How to Understand the Global Spread of Political Polarization." *Carnegie Endowment for International Peace* (2019).
- Cinelli, Matteo, Gianmarco De Francisci Morales, Alessandro Galeazzi, Walter Quattrociocchi, and Michele Starnini. "The echo chamber effect on social media." *Proceedings of the National Academy of Sciences* 118, no. 9 (2021).
- Del Vicario, Michela, Antonio Scala, Guido Caldarelli, H. Eugene Stanley, and Walter Quattrociocchi. "Modeling confirmation bias and polarization." *Scientific reports* 7, no. 1 (2017): 1-9.
- Dubois, Elizabeth, and Grant Blank. "The echo chamber is overstated: the moderating effect of political interest and diverse media." *Information, communication & society* 21, no. 5 (2018): 729-745.
- Edwards, Arthur. "(How) do participants in online discussion forums create 'echo chambers'?": The inclusion and exclusion of dissenting voices in an online forum about climate change." *Journal of Argumentation in Context* 2, no. 1 (2013): 127-150.
- Garimella, Kiran, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. "Political discourse on social media: Echo chambers, gatekeepers, and the price of bipartisanship." In *Proceedings of the 2018 World Wide Web Conference*, pp. 913-922. 2018.
- Gilbert, Eric, Tony Bergstrom, and Karrie Karahalios. "Blogs are echo chambers: Blogs are echo chambers." In *2009 42nd Hawaii International Conference on System Sciences*, pp. 1-10. IEEE, 2009.
- Geschke, Daniel, Jan Lorenz, and Peter Holtz. "The triple-filter bubble: Using agent-based modelling to test a meta-theoretical framework for the emergence of filter bubbles and echo chambers." *British Journal of Social Psychology* 58, no. 1 (2019): 129-149.
- Green, Donald, Bradley Palmquist, and Eric Schickler. *Partisan hearts and minds*. Yale University Press, 2008.
- Guess, Andrew, Brendan Nyhan, Benjamin Lyons, and Jason Reifler. "Avoiding the echo chamber about echo chambers." *Knight Foundation* 2 (2018).
- Heatherly, Kyle A., Yanqin Lu, and Jae Kook Lee. "Filtering out the other side? Cross-cutting and like-minded discussions on social networking sites." *New Media & Society* 19, no. 8 (2017): 1271-1289.
- Herne, Kaisa, Henrik Serup Christensen, and

- Kimmo Grönlund. "The influence of political knowledge on opinion polarization in citizen deliberation." *Political Research Exchange* 1, no. 1 (2019): 1-23.
- Iyengar, Shanto, and Kyu S. Hahn. "Red media, blue media: Evidence of ideological selectivity in media use." *Journal of communication* 59, no. 1 (2009): 19-39.
- Jarral, K. (2018). Pakistan Election: the role of social media. Asia Research Institute: Available:
<https://theasiadialogue.com/2018/10/15/pakistan-elections-the-role-of-social-media/>
- Khan, Mamona Yasmin, and Asma Kashif Shahzad2 Ghulam Yasin Mahar. "New Technologies and Digital Literacy in Education: A Shifting Paradigm." (2018).
- Lee, Changjun, Jieun Shin, and Ahreum Hong. "Does social media use really make people politically polarized? Direct and indirect effects of social media use on political polarization in South Korea." *Telematics and Informatics* 35, no. 1 (2018): 245-254.
- Lorenz, Jan. "Continuous opinion dynamics under bounded confidence: A survey." *International Journal of Modern Physics C* 18, no. 12 (2007): 1819-1838.
- Nickerson, Raymond S. "Confirmation bias: A ubiquitous phenomenon in many guises." *Review of general psychology* 2, no. 2 (1998): 175-220.
- Nguyen, An, and Hong Tien Vu. "Testing popular news discourse on the "echo chamber" effect: Does political polarisation occur among those relying on social media as their primary politics news source?." *First Monday* 24, no. 5 (2019).
- Pariser, Eli. "The filter bubble: What the internet is hiding." *Policy Perspectives* () (2012).
- Prior, Markus. "Media and political polarization." *Annual Review of Political Science* 16 (2013): 101-127.
- Qureshi, Israr, Babita Bhatt, Samrat Gupta, and Amit Anand Tiwari. "Causes, Symptoms and Consequences of Social Media Induced Polarization (SMIP)."
- Sanders, Audrey C. "Developing a Conceptual Framework for Research on Social Media and Political Polarization." (2020).
- Shami, Savera, Sana Naveed Khan, and Ayesha Ashfaq. "Crafting political images on Twitter: Analysis of public relations strategy of politicians of Pakistan." *Journal of the Research Society of Pakistan* 56, no. 2 (2019): 105.
- Stieglitz, S. and Dang-Xuan, L., 2012. "Political communication and influence through microblogging--An empirical analysis of sentiment in Twitter messages and retweet behavior." In 2012 45th Hawaii International Conference on System Sciences. IEEE. pp. 3500-09
- Shore, Jesse, Jiye Baek, and Chrysanthos Dellarocas. "Network structure and patterns of information diversity on Twitter." *arXiv preprint arXiv:1607.06795* (2016).
- Sikder, Orowa, Robert E. Smith, Pierpaolo Vivo, and Giacomo Livan. "A minimalistic model of bias, polarization and misinformation in social networks." *Scientific reports* 10, no. 1 (2020): 1-11.
- Stroud, Natalie Jomini. "Polarization and partisan selective exposure." *Journal of communication* 60, no. 3 (2010): 556-576.
- Sunstein, Cass. 2001. Republic.Com. Princeton, New Jersey: Princeton University Press.
- Sunstein, Cass R. 2002. "The Law of Group Polarization." *Journal of Political Philosophy* 10(2):175.

- Sunstein, Cass, and Cass R. Sunstein. #*Republic*. Princeton university press, 2018.
- Sunstein, Cass R. *Going to extremes: How like minds unite and divide*. Oxford University Press, 2009.
- Tajfel, Henri, John C. Turner, William G. Austin, and Stephen Worchel. "An integrative theory of intergroup conflict." *Organizational identity: A reader* 56, no. 65 (1979): 9780203505984-16.
- Tankovska, H. "Number of Global Social Network Users 2017-2025." *Statistics* (2021).
- Turner, John C., Michael A. Hogg, Penelope J. Oakes, Stephen D. Reicher, and Margaret S. Wetherell. *Rediscovering the social group: A self-categorization theory*. Basil Blackwell, 1987.
- Törnberg, Petter. "Echo chambers and viral misinformation: Modeling fake news as complex contagion." *PloS one* 13, no. 9 (2018): e0203958.
- Van Alstyne, Marshall, and Erik Brynjolfsson. "Electronic communities: Global village or cyberbalkans." In *Proceedings of the 17th International Conference on Information Systems*. New York: Wiley, p. 32. 1996.
- Webster, James G. "Beneath the veneer of fragmentation: Television audience polarization in a multichannel world." *Journal of communication* 55, no. 2 (2005): 366-382.
- Wojcieszak, Magdalena E., and Diana C. Mutz. "Online groups and political discourse: Do online discussion spaces facilitate exposure to political disagreement?." *Journal of communication* 59, no. 1 (2009): 40-56.
- Zahid, Muhammad, Arshad Ali, and Sami Ullah. "Effects of Media Exposure on the Political Polarization Patterns of Students in Pakistan."