

**A Psychological and Empirical Investigation of the
Online Echo Chamber Phenomenon**

Dissertation zur Erreichung des Grades “Dr. phil.”

Vorgelegt an der Universität Passau

Von Robert Luzsa

Erstgutachterin: Prof. Dr. Susanne Mayr

Zweitgutachter: Prof. Dr. Detlef Urhahne

Bearbeiter: Robert Luzsa

robert@luzsa.de und robert.luzsa@uni-passau.de

Königstr. 13

94113 Tiefenbach

Passau, den 04.06.2019

Preface

This dissertation is made up of three research papers that have either been already published in scientific journals or whose submission is in preparation, accompanied by introductory and summary texts. The papers are:

- Chapter 2: Luzsa, R. & Mayr, S. (2019). Links between users' online social network homogeneity, ambiguity tolerance, and estimated public support for own opinions. *Cyberpsychology, Behavior and Social Networking*, 22(5), 235-239.
- Chapter 3: Luzsa, R. & Mayr, S. (submission in preparation). False Consensus in the Echo Chamber: Experimental Exposure to Favorably Biased Social Media News Feeds Leads to Higher Perceived Public Support for Own Opinions.
- Chapter 4: Luzsa, R. (submission in preparation). Intergroup Bias in Online Echo Chambers: Evidence from Word Frequency Analysis of Attitudinally Homogeneous and Heterogeneous Reddit Groups.

Due to copyright reasons, original texts of the initially submitted pre-review versions are used. Therefore, chapters 2 through 4 are not identical to the final published articles and are not suitable for citing (e.g. the title of the paper in chapter 2 was shortened during peer-review). Moreover, the papers have been or will be submitted to different journals with different manuscript guidelines. Therefore, inconsistencies in formatting, layout and bibliographical citation style occur. For the same reason, figures and tables are not numbered continuously across chapters. As copyright allows only unaltered submitted papers to be used, these inconsistencies could not be eliminated. Also, in chapter 4, trailing zero for p -values are given as demanded by the journal intended for publication.

Acknowledgements

I would like to thank Prof. Dr. Susanne Mayr and Prof. Dr. Detlef Urhahne for their scientific support and supervision of this doctoral thesis. I also want to thank my current and former colleagues (especially Dr. Malte Möller and Stephanie Schmitt-Rüth) for many interesting and beneficial discussions, and my friends and family for their support.

Abstract

In the public debate it is often assumed that communication in so-called “Echo Chambers” - online structures in which like-minded people share mostly messages that confirm their mutual, shared attitudes - can lead to negative outcomes such as increased societal polarization between groups holding opposing beliefs. This thesis aimed to examine this assumption from a psychological perspective and substantiate it empirically. First, based on existing research and psychological theories, a working definition of Echo Chambers was formulated, that highlights two key factors: Selective Exposure to attitudinally congruent messages and communication in homogeneous networks. Then, three studies were conducted to test links between these factors and two individual-level outcomes that are associated to subjects’ actual behavior: Their False Consensus, that is, how strongly subjects perceive the public in agreement with their own attitudes, and their Intergroup Bias, which reflects to which degree subjects’ identify as members of an in-group that is in conflict with negatively perceived out-groups. The studies employed questionnaire-based, experimental, as well as real-world data driven approaches. Overall, they confirm that exposure to Echo Chamber-like online structures can indeed lead to a more favorably distorted perception of public opinions and to more signs of Intergroup Bias in subjects’ communicational style. Thus, the thesis provides first psychologically founded empirical evidence for effects of online Echo Chamber exposure on behavior-related individual-level outcomes. The results can serve as a basis for further research as well as for the discussion of possible strategies to counter negative effects of online Echo Chambers.

Table of contents

| | |
|--|-----|
| Preface | 2 |
| Acknowledgements | 3 |
| Abstract..... | 4 |
| Table of contents | 5 |
| 1. Introduction: A psychological perspective on online Echo Chambers | 6 |
| 1.1. Background and motivation | 6 |
| 1.2. Theoretical framework | 8 |
| 1.2.1. Echo Chambers and selective exposure..... | 8 |
| 1.2.2. Echo Chambers and network homogeneity | 11 |
| 1.2.3. A working definition of Echo Chambers..... | 12 |
| 1.3. Individual-level outcomes affected by Echo Chambers | 13 |
| 1.3.1. False Consensus..... | 14 |
| 1.3.2. Intergroup Bias | 15 |
| 1.3.3. Effects via central and peripheral route | 18 |
| 2. Overview of studies | 21 |
| 2.1. Overview of study one..... | 21 |
| 2.2. Overview of study two | 23 |
| 2.3. Overview of study three | 27 |
| 3. Study one: Network homogeneity and False Consensus..... | 30 |
| 4. Study two: Selective exposure and False Consensus | 49 |
| 5. Study three: Intergroup Bias in Echo Chamber message texts..... | 95 |
| 6. Conclusions and general discussion | 118 |
| 6.1. Summary and interpretation | 118 |
| 6.2. Methodological considerations and future directions..... | 121 |
| 6.3. Practical implications | 127 |
| References | 130 |
| Versicherung (gem. §4 Abs. 3 Satz 1 Nr. 5 PromO):..... | 147 |

1. Introduction:

A psychological perspective on online Echo Chambers

1.1. Background and motivation

Over the last decade there has been a drastic change in how public and science evaluate the impact of online social media on the individual and society: In the early 2010s, the opportunities for connecting with others and free sharing of information that services like Facebook (founded in 2004) and Twitter (founded in 2006) offered were often regarded quite positively. For example, after wide-spread citizen protests in the Arab world (“Arab Spring”, 2011-2012), news outlets (e.g. Kassim, 2012) as well as research (e.g. Hussain & Howard, 2012; Lotan et al., 2011) often emphasized how Facebook and Twitter allowed for the uncensored spread of democratic ideas under repressive regimes and how they helped activists to communicate and organize. Social media thus appeared as tools of empowerment and social change (Ali, 2011). Governments' attempts to influence network providers to limit users' freedom of communication and information were criticized as censorship (Youmans & York, 2012).

More recently, this positive notion gave way to a more critical view: With headlines like "How Social Media Echo Chambers Drown Out the Voices in the Middle" (Mosley, 2018) or "Your Filter Bubble is Destroying Democracy" (El-Bermawy, 2015), popular media warn that Internet and social media may influence public opinion formation and lead to increased societal polarization and strengthening of extremist positions. At the core of this criticism, which is also voiced by researchers (e.g. Grömping, 2014; Williams, McMurray, Kurz, & Lambert, 2015), are the two closely related concepts of “Echo Chambers” and “Filter Bubbles”. Both terms share the assumption that Internet and social

media facilitate selectivity in information consumption, making it easy for subjects to surround themselves with like-minded others and with congruent information that support their own views, while keeping away deviant, incongruent opinions. Existing research mostly examines whether users of social media websites really exhibit a tendency to preferably connect with like-minded others and consume and share congruent information (e.g. Barberá, Jost, Nagler, Tucker, & Bonneau, 2015; Colleoni, Rozza, & Arvidsson, 2014). However, one of the central questions still lacks an adequate answer: Does exposure to such Echo Chambers really affect users in regard to their perceptions and behavior, that is, in a psychologically meaningful and relevant way?

While lots of research on Echo Chambers originates in the disciplines of political science and network analysis (e.g. Colleoni et al., 2014; Goldie, Linick, Jabbar, & Lubienski, 2014; Williams et al., 2015), the question whether there are effects of Echo Chambers on the individual is a deeply psychological one as it relates to topics such as heuristics and biases in human decision making and psychological effects of media exposure. Thus, a psychological examination of the effects of online Echo Chambers on individual users appears necessary. The aim of this thesis is to undertake such a psychologically and empirically founded examination.

To achieve this goal, first an overview of existing definitions and research on Echo Chambers as well as relevant psychological theories and constructs is given and used to formulate a working definition of Echo Chambers (chapter 1.2). Based on this, with False Consensus and Intergroup Bias, two constructs in which effects of Echo Chamber exposure are assumed to occur are introduced (chapter 1.3). Chapter 2 then gives an overview of three substudies that were conducted to test the assumed effects. The studies are fully reported in chapters 3 to 5. Finally, chapter 6 summarizes the studies' findings

and discusses future directions in research as well as practical implications, with an focus on possible interventions to reduce negative effects of Echo Chambers.

1.2. Theoretical framework

To better understand the term Echo Chamber and reach a satisfying working definition, this chapter will examine the terms' origins as well as its use in research. It will be shown that different research traditions use the term quite differently, but common motives can be uncovered. Two main views of Echo Chambers will be identified that are related to two psychological constructs: Selective exposure and homophily. Both perspectives are then integrated into this thesis' working definition of Echo Chambers.

1.2.1. Echo Chambers and selective exposure

An early (non-scientific) mention of Echo Chambers is found in a document of tobacco company Phillip Morris in 1998 (Scruggs, 1998). In this, the term is positively used to describe a lobbying strategy by which law makers should be influenced to support tobacco-friendly policies. To achieve this, sources from whom lawmakers receive information and with whom they interact frequently are identified (e.g. local and national media, colleagues, lobbyists). Then, pro-tobacco information is spread among as many of these as sources possible. Thus, that the law maker is selectively surrounded by and exposed to pro-tobacco information that circulates or “echoes” amongst his or her network of sources, leading to a strong effect of the information on him or her.

This general idea - subjects are surrounded by and selectively exposed to information that expresses one perspective on a topic - is remarkably similar to the term's later use in science: Early scientific investigations of Echo Chambers were made by

Sunstein (2001). The author analyzed communication and discussion among and between supporters of the American Democrat and Republican parties during the 2000 US presidential elections and the impeachment of President Clinton. Sunstein found that members of each faction tended to mostly communicate with and confirm each other in their views, that inter-faction discussion (i.e. between Democrats and Republicans) hardly occurred, and that arguments of the other faction were ignored. Thus, supporters of both parties were isolated in their respective Echo Chambers and mostly exposed to selective information that supported their pre-existing attitudes.

From this, the first key component of online Echo Chambers can be deduced: In an Echo Chamber, subjects are selectively exposed to one-sided information (e.g. messages written by other users of an online group), that are congruent to their own attitudes and pre-existing views.

Selective exposure as a basic human tendency

This concept of selective exposure is, however, not unique to online contexts. Research in media psychology has long established that subjects generally prefer attitudinally consistent information and avoid contradictory information (for an overview see Knobloch-Westerwick, 2014). Studies in selective exposure find this preference regarding various kinds of media, from traditional newspapers and television (e.g. Stroud, 2008) to digital online media (e.g. Knobloch-Westerwick & Jingbo Meng, 2009). Often the theory of cognitive dissonance (Festinger, 1957) is cited as explanation for this preference: Subjects want to avoid the negative feelings that arise from confrontation with information that questions their existing beliefs and therefore prefer information that is congruent to these beliefs. Furthermore, selective exposure can also be attributed to

subjects' use of heuristics (Tversky & Kahneman, 1974), that is, mental shortcuts which often allow for quick and efficient decision making but can also lead to misjudgments, so called biases. Such biases might further strengthen selective exposure: For example, subjects often use strategies of information acquisition that focus on confirming existing knowledge and impressions (confirmation bias; Wason, 1968), and their perceptions of ideas and information that they are repeatedly exposed to tend to get more positive, leading them to seek out further similar, congruent information (mere exposure effect, Zajonc, 1968).

Selective exposure in online contexts

These findings illustrate that a preference for selective exposure reflects general human tendencies and is not new or unique to online environments. However, social media appear to make being selective more easy than traditional media:

Users can “follow” news outlets and other users that support their own views on platforms like Facebook and twitter, thereby surrounding themselves with a highly personalized and permanently available stream of attitudinally congruent information. In case of traditional media, subjects may also select journals or television channels that are generally in agreement with their own views, but this selectivity is less personalized, less complete, and less constantly available.

Furthermore, providers of social media platforms like Facebook or Twitter use algorithms to analyze users' behavior (e.g. what links they click on, which messages they “like”) in order provide them with contents that fit their interests and usage patterns. Of course, if users already demonstrate tendencies to prefer attitudinally congruent information, these mechanisms will strengthen this selective exposure and further isolate

users from incongruent, alternative information and opinions. Thus, the users' endogenous tendency to selective exposure is complemented by exogenous selectivity, which they might not even be aware of.

Pariser (2011) introduced the term of "Filter Bubble" to describe this algorithm-driven selective exposure. However, it appears quite obvious that the endogenous and exogenous selective exposure are not unrelated concepts and not separable from each other. Instead, they are merely two simultaneously active and complementary mechanisms that lead to the same outcome: Isolation of the user in an online Echo Chamber made up of congruent information. Therefore, differentiating Filter Bubbles and Echo Chambers does not provide additional insights for the current thesis' aims, and the more general term Echo Chamber, encompassing both the effects of endogenous and exogenous selective exposure, is used from now on.

1.2.2. Echo Chambers and network homogeneity

While the previously described perspective on Echo Chambers highlights the role of selective exposure to attitudinally congruent information, many researchers focus on another aspect: Homophily and network homogeneity (e.g. McPherson, Smith-Lovin, & Cook, 2001), that is, subjects tendency to preferentially connect to others that are similar to themselves. In this tradition, Echo Chambers can be understood as highly homogeneous networks, in which most users are similar to each other, for example regarding their attitudes toward a topic but also regarding factors like social or ethnical backgrounds (Colleoni et al., 2014; Quattrociocchi, Scala, & Sunstein, 2016; Schmidt et al., 2017; Zollo et al., 2017).

Most research that takes this perspective on Echo Chambers investigates how users connect to each other, often by analyzing large social media data sets (e.g. Barberá et al., 2015; Grömping, 2014). In general, studies report differently strong tendencies to form homogeneous networks, with most signs of homogeneous Echo Chambers for highly political topics (Barberá et al., 2015).

Of course, this notion of online Echo Chambers as networks of users similar to each other is somewhat connected to the notion of Echo Chambers characterized by selective exposure to mostly attitudinally-congruent messages: If users are surrounded by others similar to themselves, they should also more likely receive messages that are congruent to their own attitudes, than when they are surrounded by a more heterogeneous network.

However, it is important to be aware of these two notions and to take both of them into account in an empirical investigation, as the perspective taken strongly influences what factors are highlighted in research: Studies that take the first perspective (selective exposure and congruence of information) will most likely analyze metrics like the number of attitudinally congruent messages users receive, while studies in the second tradition (homogeneous networks) will try to measure network homogeneity and examine ties between users. However, to get a comprehensive understanding of effects of Echo Chambers on individuals, studies should take both perspectives into account. This is what the present study strived to achieve.

1.2.3. A working definition of Echo Chambers

Based on the outlined considerations and perspectives, the following working definition of Echo Chambers can be formulated and will be used as a basis for this thesis' studies:

The term Echo Chamber refers to all kind of online communication that is characterized (a) by users being surrounded by homogeneous networks of others that are similar to themselves regarding attitudes or other traits, and/or (b) by users being selectively exposed to mostly attitudinally congruent contents (e.g. messages written by others) that express agreement to their own attitudes. Both aspects - selective exposure and network homogeneity - can be endogenous, that is, result from users' own active selectivity, but to some degree also exogenous, that is, enhanced by algorithms selectively suggesting contents and contacts to users.

The studies reported in this thesis address both aspects of this definition. However, before they can be reported, it must be discussed how Echo Chambers may actually influence the individual exposed to them.

1.3. Individual-level outcomes affected by Echo Chambers

In order to test possible effects of online Echo Chambers on individuals, a theoretical framework is needed that allows to choose relevant outcomes for further investigation. The Theory of Reasoned Action (Fishbein & Ajzen, 1975; for a critical review see Liska, 1984) provides such a framework: It states that two factors influence individuals' behavioral intention and actual behavior (e.g. buying organic food): Their attitude towards the behavior (e.g. their beliefs about organic food being more healthy) and their perceived social norm regarding the behavior (e.g. whether they think it is common and accepted by their peers to prefer organic-food). Both factors can either increase or decrease the individuals' likelihood to show the actual behavior.

Based on this model, two mechanisms can be identified via which exposure to online Echo Chambers may influence subjects' behavior: Via a shift in perceived social

norm or via a shift in attitudes. From these two mechanisms, the two outcomes examined in this thesis - False Consensus and Intergroup Bias - are extrapolated.

1.3.1. False Consensus

If subjects participate in attitudinally homogeneous online Echo Chambers, they should observe that many other users share their own attitudes and write and send messages that support these attitudes. Subjects can be assumed to deduce from this that their own attitudes are common and widely accepted, and that it is the social norm to act upon them. For example, if subjects with xenophobic views are connected to other users in social media that have similar views, they will mostly receive messages that support xenophobia. Therefore, they will tend to perceive public opinion in a favorably distorted way, and believe that large parts of the population agree with their world view. In other words: They get the impression that their own points of view are in accordance to the social norms, and that it is acceptable to act upon them. This favorable distortion in perceived social norm may, in turn, decrease their inhibition to act upon their beliefs, and increase their willingness to, for the example of xenophobia, insult or degrade foreigners.

Therefore, this thesis examined the link between exposure to Echo Chambers and perceived social norm or public opinion as a behavior-related outcome, which was operationalized via the so-called False Consensus Effect.

The False Consensus Effect (Ross, Greene, & House, 1977) addresses the fact, that humans in general tend to perceive public opinion not objectively but favorably distorted: It finds that subjects tend to "[...] see their own behavioral choices and judgements as relatively common and appropriate to existing circumstances while viewing alternative responses as uncommon, deviant, or inappropriate" (Ross et al., 1977).

False Consensus is usually measured by asking subjects in favor of a statement (e.g. "Death penalty should be reintroduced") and subjects opposed to it what percentage of the population they think is in favor of the statement. It is found that subjects in favor of the statement give a higher estimate of public support for it than subjects opposed to it. For example, subjects who support the death penalty may estimate that 30% of the population are in favor of it, while subjects who oppose death penalty may estimate only 10% of the public favoring it. In this case a False Consensus Effect of 20% would be measured. The False Consensus Effect is very well documented and often replicated (e.g. Marks & Miller, 1987; Mullen et al., 1985)

It seems logical to expect that the above described assumed effect of Echo Chambers on perceived social norm is reflected in subjects' False Consensus: Subjects exposed to an Echo Chamber should display stronger False Consensus than subjects exposed to more heterogeneous non-Echo Chamber online communication. Indeed, first studies exist that suggest links between exposure to attitudinally congruent and heterogeneous media contents and a higher False Consensus (Bauman & Geher, 2002; Wojcieszak, 2008). It has also been demonstrated that False Consensus actually influences subjects' behavioral intention (Bauman & Geher, 2002).

However, no studies exist that systematically examine how exposure to online Echo Chambers influences False Consensus. Therefore, the first two studies of this thesis tested this link with correlative (*study one*) and experimental (*study two*) approaches.

1.3.2. Intergroup Bias

According to the Theory of Reasoned Action (Fishbein & Ajzen, 1975), a second factor that determines behavior are subjects' attitudes. It can be assumed that exposure to

an Echo Chamber affects these attitudes: If subjects participate in an online Echo Chamber group that is homogeneous regarding a certain topic (e.g. strongly opposed to immigration), their attitudes towards the topic should be influenced by the other group members' presence and the messages they post. As the other members and their messages mostly embrace subjects' own attitudes and, most likely, provide them with even more reasons and arguments why they are correct, these attitudes should display a shift in the Echo Chamber's general ideological direction and get more extreme.

Research on cognitive biases supports the assumption of such an effect of the group on individuals' attitudes: For example, it is known that mere repeated exposure to persons, objects, or ideas can lead to an increase in sympathy for them (Zajonc, 1968). Moreover, according to group conformity theory (Asch, 1961), subjects in attitudinally homogeneous groups tend to support the group consensus due to fear of social rejection or due to getting genuinely convinced of the majority's point of view. This, in turn, can lead to attitudes of individual group members and the group as a whole getting more extreme (Myers, 1975; Myers & Lamm, 1976).

Thus, repeated exposure to an online Echo Chamber group should lead to a shift in users' attitudes in the direction of the views expressed in the majority of messages shared in the group.

However, empirically investigating such a shift, for example via experimentally exposing subjects for a longer period of time to an attitudinally homogeneous artificial Echo Chamber and measuring signs of radicalization of their attitudes, appears as methodologically and ethically challenging. Moreover, such an approach would require focusing on one attitude towards one specific topic that is discussed in an Echo Chamber

(e.g. regarding immigration laws, regarding stricter emission controls etc.). Even if such an effect was found, it would be open for debate whether the effect is specific to the concrete topic, or whether it could be generalized to other topics and Echo Chambers.

Therefore, it appears more promising to examine a more generalizable and also more manageable attitudinal outcome that can be assumed to show effects in different kinds of Echo Chambers and for different topics. This thesis therefore examines the so-called Intergroup Bias as one attitudinal outcome in which Echo Chamber effects are expected.

The term Intergroup Bias refers to the fact that feeling as a member of a group has an effect on attitudes both towards the other members of this so-called in-group as well as towards members of other, rival out-groups: Subjects who are members of a group perceive the other members in their in-group more favorably than members of out-groups (Tajfel, 1970). For example, they ascribe more positive traits to in-group, more negative to out-group members, perceive both in- and out-group as rather homogeneous regarding these traits (Otten & Moskowitz, 2000), and they prefer in-group members when it comes to the distribution of limited resources (Tajfel, 1970). This attitudinal effect of group membership on perception of in- and out-groups is so strong that it even occurs when subjects are randomly and anonymously assigned to non-informatively labelled groups and even when no interaction within or between groups takes place (Tajfel, 1970). A reason for these effects can be that subjects try to achieve self-esteem, a positive social identity and a feeling of distinctiveness by focusing on and emphasizing (real or hypothetical) differences between in- and out-group (Social Identity Theory; Hogg, Terry, & White, 1995; Tajfel, 1970; Trepte & Loy, 2017).

It appears logical to assume that an Intergroup Bias also occurs in attitudinally homogeneous online Echo Chambers that serve as in-groups for their members: If users participate in an EC, they should tend to view the social world in group categories, differentiating others in categories of “we” (in-group members who share the Echo Chambers’ attitudes) versus “they” (out-group members who are in opposition to these attitudes). Moreover, they should try to strengthen and enhance their positive in-group identity by devaluating and attacking out-groups while expressing affirmation with the in-group worldview. This attitudinal shift in regard to perception of others might then, according to the Theory of Reasoned Action, lead to changes in actual behavior, for example, a higher likelihood to act in accordance with in-groups/Echo Chambers’ shared attitudes, or more aggression towards persons who hold opposing views.

The final study of this thesis (*study three*) examined whether such a link between participation in an online Echo Chamber and signs of Intergroup Bias can be found.

1.3.3. Effects via central and peripheral route

Finally, it must be discussed which components of information in online Echo Chambers can affect subjects’ False Consensus and Intergroup Bias and should therefore be addressed or manipulated in empirical studies: What role do the contents of messages shared between users play, and what additional factors have to be taken into account?

An answer to this question can be found in the Elaboration Likelihood Model (Petty, 1986). This model identifies two routes via which persuasive information may affect subjects: Information may be processed either via the central route that involves a high level of elaboration - the individual forms an opinion by processing the arguments presented and critically evaluating them based on prior knowledge - or via the peripheral

route that requires low elaboration - the individual forms an opinion based rather superficial attributes of the arguments, e.g. if the source of the argument is a well known person or not, and without deeper reflection. What route is chosen depends on the individuals' level of involvement, for example whether the topic is important for them.

By applying the Elaboration Likelihood Model on the common design of many social media platforms, two pathways via which Echo Chambers may influence subjects can be identified, which reflect central and peripheral processing routes: Message contents and popularity cues.

The central route: Message contents

One attribute of Echo Chambers may be the existence of a large number of message contents that are consistent with the attitudes of the users. For example, left-wing users may join a Facebook group or a sub-reddit on Reddit that discusses politics from a left-wing perspective. They will then mostly see messages that are written by other left-wing users and reflect left-wing views. By processing the message contents, which support their own attitudes, they are assured of the correctness of their own views.

Thus, message contents can be seen as central contents that influence subjects via the central, more elaborative route (e.g. shifting their attitude by giving them additional arguments that their point of view is correct).

The peripheral route: Popularity cues

In most social networks there are also social context cues that indicate the popularity of a content, for example the amount of “likes” a message has received by other users on Facebook, or the score of a post on Reddit. These so-called popularity cues (Haim, Kümpel, & Brosius, 2018; Porten-Cheé, Haßler, Jost, Eilders, & Maurer, 2018)

can be understood as peripheral attributes of the messages that influence users via the peripheral route. For example, when a message that supports their own attitudes has many “likes”, subjects may take this as an indicator of high endorsement for their attitudes by others, and feel convinced of the correctness of their own views.

Moderation of effects by subjects' involvement

It can be assumed that both routes influence False Consensus as well as Intergroup Bias. According to the Elaboration Likelihood Model, subjects' level of involvement, that is, their interest in a topic, should determine the strength of each route's influence: The effect of popularity cues such as numbers of “likes” should be strongest when subjects view messages regarding topics they have little interest in. On the other hand, the messages actual contents (e.g. how many messages agree or disagree with the subjects' own attitudes) should exert a strong effect when subjects are highly interested in a topic.

Therefore, this thesis will consider subjects' interest in topics as a moderating variable.

2. Overview of studies

Three empirical/quantitative studies were conducted to test the assumed links between exposure to online Echo Chambers and subjects' False Consensus (*studies one and two*) and Intergroup Bias (*study three*). The methodological strategies employed in the studies took into account the different elements of Echo Chambers identified in the thesis' introduction:

Study one focused on the aspect of network homogeneity in Echo Chambers via an correlative, questionnaire-based approach.

Study two examined the effects of selective exposure to attitudinally congruent messages on participants' False Consensus via experimental investigation.

To ensure reliability of the acquired results, both studies were replicated.

Study three, finally, extended the previous studies with an innovative, data driven approach that promised high ecological validity. It tested whether messages written by users in attitudinally homogeneous online fora, that is, Echo Chambers, displayed more linguistic signs of Intergroup Bias than messages in more heterogeneous non-Echo Chamber fora.

All studies are reported in full extent in chapters 3 to 5. Prior to this, in the following, an overview of each study is given and the studies' interconnections are explained.

2.1. Overview of study one

Study one set out to examine whether the homogeneity of the online contact networks (e.g. friends on Facebook, frequent e-mail and chat partners) of users is

correlated with their extent of False Consensus Effect. This assumption was made as a more homogeneous network should lead to users receiving more messages that are congruent to their own attitudes, which in turn should lead to higher False Consensus.

To test this assumption, 380 participants (181 in the first substudy, 199 in the replication study) completed an online questionnaire. Measures included subjects' False Consensus Effect in regard to twenty current societal topics, and their network homogeneity, operationalized via an existing scale (Zúñiga, Jung, & Valenzuela, 2012) that measures how similar users' contacts are to each other (e.g. whether a users mostly communicates with people with the same social, ethnic, religious background). Additionally, a scale for subjects Ambiguity Tolerance (McLain, 2009), that is, how well subjects are able to cope with ambiguous, unclear and contradicting information, was administered. This was done because low Ambiguity Tolerance might lead subjects to prefer homogeneous networks that do not questions their beliefs by providing them with ambiguous information. Thus, to ensure that correlations between False Consensus and network homogeneity were genuine and could not be attributed to Ambiguity Tolerance, Ambiguity Tolerance was measured and statistically controlled for.

The results of *study one* confirmed that subjects with a more homogeneous online contact network display a significantly higher False Consensus Effect, that is, believe that more people in the general population share their own points of view. In addition, subjects with higher Ambiguity Tolerance displayed a lower False Consensus Effect. The effect of network homogeneity was still significant when taking into account Ambiguity Tolerance's influence. Both effects could be replicated.

Subjects who surround themselves online with like-minded and similar others thus appear in higher danger of overestimating how strongly the public endorses their own attitudes. On the other hand, being able to tolerate ambiguous, contradicting information appears as a protective factor that allows people to make more realistic judgements of other's attitudes. As both network homogeneity and Ambiguity Tolerance exhibit genuine and independent effects on False Consensus, it can be concluded that the negative effects of network homogeneity can, to some degree, be alleviated if subjects are otherwise open to ambiguous, contradicting information.

Thus, an empirical link between network homogeneity and False Consensus could be established.

2.2. Overview of study two

Study two examined the role of selective exposure in Echo Chambers. It tested whether exposure to Echo Chamber-like online environments in which most or all messages support users' opinions (e.g. an homogeneous Facebook group of very left- or right-wing oriented users), leads to a higher False Consensus.

Based on the Elaboration Likelihood Model and the two cognitive pathways of influence that it states (central and peripheral), the following assumptions were made:

Firstly, subjects who receive only or mostly online messages that agree with their own opinions, that is, congruent messages, display a higher False Consensus Effect than subjects who receive only or mostly incongruent messages that question their views. This hypothesis reflected the central, informational pathway in the Elaboration Likelihood Model.

Secondly, it was hypothesized that the number of “likes” which a message displays, that is, how strongly it is endorsed by other users, also influences False Consensus: If congruent messages display higher likes than incongruent ones, subjects should infer strong public support for their own opinions, leading to higher False Consensus. Similarly, if incongruent messages display higher likes, False Consensus should be lower. This assumption reflected the peripheral pathway of the Elaboration Likelihood Model.

Finally, subjects interest in the messages’ topic was expected to determine how strongly their False Consensus is affected by message contents and numbers of “likes”: High interest in a topic should lead to high involvement and therefore a stronger effect of the message contents, while low interest should lead to low involvement and subjects being more influenced by “likes”.

Two experiments with 331 and 207 participants, respectively, were conducted to test these hypotheses by experimentally inducing an Echo Chamber and then measuring False Consensus.

Both experiments used the following paradigm: Subjects were exposed to news feed similar to the news feeds that commonly serve as front pages in social media platforms like Facebook or Twitter. They saw messages that expressed either agreement or disagreement with a common topic (e.g. “Drugs like marihuana should be legalized”). The experimental manipulation consisted of varying the number of messages that expressed the same opinion as the participant (which was measured prior to exposure). Thus, an Echo Chamber could be induced by showing subjects mostly or only messages that shared their opinion. Each message consisted of a headline, a short teaser text that served as preview for a full news article, and a number of “likes”. The messages pertained

to nine topics. Topics were selected based on their success in eliciting both a strong False Consensus Effect and a rather balanced distribution of agreement and disagreement in the previous questionnaire studies (*study one*). The message texts were acquired from German online news websites, homogenized in regard to their length and stylistic properties and then pre-tested ($n = 15$) to select only messages that subjects perceived to state a clear (either pro or contra) opinion.



Figure 1. Example of a news feed used in the first experiment of *study two*.

In the first experiment, subjects were first asked to state their own opinions and interest regarding the nine topics. Then, for each topic, they saw four messages on one page (see fig. 1). Two factors were manipulated: Three, that is, the majority of messages was either congruent or incongruent to their own opinion. And either congruent or incongruent messages displayed higher numbers of “likes”. Subjects were asked to select the message whose full article they would like to read the most. After exposure to all nine pages, False Consensus was measured.

In the second experiment, the paradigm was slightly altered: Opinion and interest were measured like in *study one*, and subjects again saw four messages regarding the each

of the nine topics. However, now only one message per screen was displayed and subjects were asked to indicate on a Likert-scale how like they would click on it (see fig. 2). Furthermore, now all four messages were either congruent or incongruent to the subjects' opinions, and no deviant message was included. This was done to induce a "full" Echo Chamber and to eliminate possible effects which the single deviant message per topic might have had in experiment one. False Consensus was again measured after exposure.

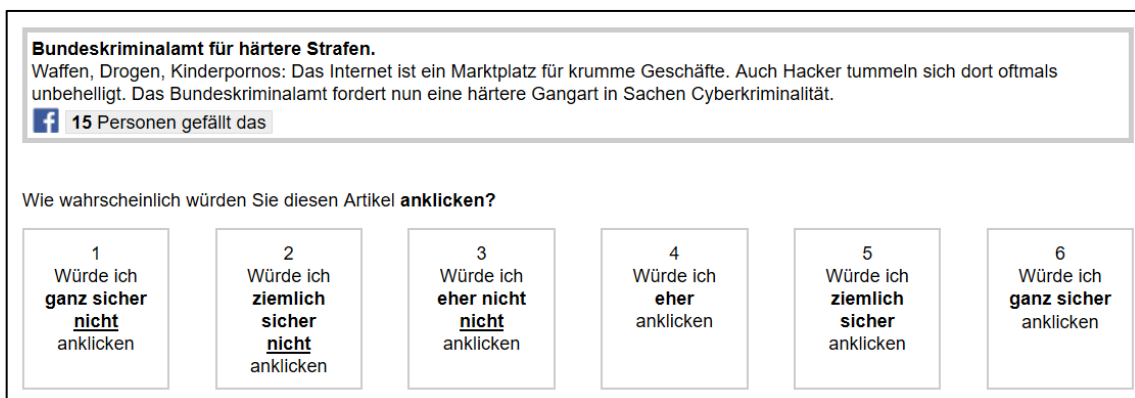


Figure 2. Example of a message presented in the second experiment of *study two*.

The results of both experiments confirmed that participants' estimate of public support for their own opinions, that is, their False Consensus Effect, is influenced by the level of agreement they encounter in online messages:

Subjects in an Echo Chamber who mostly or only saw messages supporting their own opinions displayed a higher False Consensus Effect than subjects who saw messages that mostly or only contradict their views. Regarding messages' numbers of "likes", the results were more differentiated and illustrated that subjects appear skeptical of the validity of "likes": As expected, if subjects reported little interest in a topic, high "likes" for attitudinally congruent messages lead to a high False Consensus Effect. However, if when subjects were highly interested in a topic, the effect reversed, and high likes for attitudinally congruent messages lead to a low False Consensus. This reversal was

unexpected but could still be interpreted in light of the Elaboration Likelihood Model and discussed as an effect of reactance. The pattern of effects is described in full extent in chapter 4.

Study two thus successfully identified attitudinal congruency of messages as one mechanism by which exposure to online Echo Chambers can influence users: If users participate in Echo Chamber groups, in which most of the messages written by other users are congruent to their own points of view, they can be expected to conclude from this agreement to a high level of agreement by the general population. The consequences of these findings for future research and practical interventions are discussed in detail in chapter 4 and in the general discussion in chapter 6.

As shown, *study one* and *study two* succeeded in confirming links between Echo Chambers and False Consensus. Therefore, the subsequent *study three* focused on the second outcome assumed to be influenced by Echo Chamber exposure: Intergroup Bias.

2.3. Overview of study three

While the previous studies employed traditional questionnaire-based and experimental approaches, *study three* aimed to provide an expanded perspective by using an innovative data-driven approach that promises high ecological validity. The study tested the assumption that subjects who participate in an online Echo Chamber, that is, an attitudinally homogeneous online group, display a stronger Intergroup Bias than subjects participating in more heterogeneous online groups.

Based on Social Identity Theory (Hogg et al., 1995; Tajfel, 1970; Trepte & Loy, 2017), it was assumed that users of Echo Chambers exhibit a strong group identity, that

is, perceive themselves as a member of a (positive) in-group that is in conflict with out-groups that are viewed negatively and devaluated. Previous research suggests that key elements of intergroup bias - salience of group vs. self identity and signs of intergroup conflict - are reflected in the words subjects use (Arguello et al., 2006; Chung & Pennebaker, 2007; Ruscher & Hammer, 1996; Schnake & Ruscher, 1998; Vingerhoets, Bylsma, & Vlam, 2013). Therefore it was assumed that in Echo Chambers, messages that display more signs of Intergroup Bias receive more endorsement by other members of the Echo Chamber, for example by getting more up-votes or “likes”. Three hypotheses were made:

H1) The more group references, that is, first (e.g. “we”, “us”) and third person plural pronouns (e.g. “they”, “them”), a message employs, the more endorsement it receives by other users.

H2) Messages that display more signs of intergroup conflict and out-group using negative emotion words and swearing receive more endorsement.

H3) More expressions of affirmation and agreement with the group (e.g. “yes”, “exactly”) and a less tentative style (e.g. less use of “maybe”, “perhaps”) should also lead to more endorsement.

These hypotheses were tested by analyzing a large data set of real messages written by 78,531 users of discussions groups - so-called sub-reddits - on the social website Reddit. Three homogeneous sub-reddits that explicitly demand users to agree with their general ideological views and ban divergent opinions were taken as self-proclaimed Echo Chambers (e.g. r/latestagecapitalism, which states in its community rules: “Defense of capitalism and of the parties or ideologies that support it [...] are strictly prohibited.

Likewise, opposition to socialism [...] is strictly prohibited.”). They were contrasted to three heterogeneous non-Echo Chamber sub-reddits that highlight open discussion and invite divergent opinions (e.g. r/neutralpolitics, which according to its community rules is “[...] dedicated to evenhanded, empirical discussion of political issues. Based on facts and respectful discussion.”). In these non-Echo Chambers, no effects similar to those assumed in the Echo Chamber groups were expected.

Scale values for word frequencies were extracted with the software LIWC 2015 (Pennebaker, Boyd, Jordan, & Blackburn, 2015). Endorsement of messages was operationalized via their score, that is, how many up-votes they received by users minus the number of down-votes. Prior to analysis, all data were aggregated on level of authors in order to increase reliability. The hypotheses were then tested via Bonferroni-Holm alpha-adjusted Spearman correlations between authors’ scores and the LIWC scales reflecting their usage of group-pronouns, emotion and swearing, and affirmative style.

The first two hypotheses could be confirmed: In Echo Chamber groups, authors who used more pronouns referring to in- and out-groups, more negative emotion words and more swearing received more higher scores for their messages. No similar systematic correlative patterns were found in non-Echo Chamber neutral groups, and no systematic significant results regarding hypothesis three occurred.

Overall, the results thus supported the thesis’ second assumption that participation in online Echo Chambers can be linked to increased Intergroup Bias. The results, practical implications, directions for further research as well as methodological limitations of the study are discussed in detail in chapter 4, and, together with the thesis’ other studies’ results, in chapter 6.

3. Study one: Network homogeneity and False Consensus

Original paper:

Luzsa, R. & Mayr, S. (2019). Links between users' online social network homogeneity, ambiguity tolerance, and estimated public support for own opinions. *Cyberpsychology, Behavior and Social Networking*, 22(5), 235-239.

Please note that the following chapter is the unedited pre-review version of the published article and does not reflect changes made during the peer-review process. For citing, please refer to the published paper.

Always on my side? Links between users' online social network homogeneity, ambiguity tolerance, and estimated public support for own opinions.

Keywords: Echo chambers, social networking, false consensus

Abstract

False consensus effect (FCE) refers to a cognitive bias of relative over-estimation of public support for one's own opinion. FCE has been linked to selective social interaction with like-minded people as well as to selective exposure to attitude consistent information. While previous studies tested these links mostly in offline settings, it is assumed that FCE is also affected by the homogeneity of users' online contact network and their intensity of online social network (OSN) usage as well as their individual tendency to avoid ambiguous information. Two online studies with a total of 380 participants aged 18 - 35 were conducted to test these hypotheses via a multilevel modeling approach. In Study 1 participants with a more homogeneous online network, longer daily OSN usage time, and lower ambiguity tolerance displayed significantly higher FCE. The effects of network homogeneity and ambiguity tolerance could be replicated in Study 2. The implications of these findings are interpreted in the context of prior studies on FCE as well as the notion of OSN as "echo chambers".

Introduction

Online Echo Chambers and False Consensus

The effects of online social networks (OSN) like Facebook or Twitter on users' news consumption and opinion formation is subject of controversial public and scientific debate. The term "echo chamber" has been coined to characterize online communication biased towards interaction with like-minded persons and sharing of information consistent with own beliefs^{1,2}. This tendency towards homogeneity and selective exposure has been linked to negative consequences, such as spread of misinformation³, radicalization, and increased societal polarization between groups holding opposing beliefs^{4,5}.

This study addresses another possible consequence of homogeneity related to the "Law of Small Numbers", a cognitive bias⁶ describing the tendency to overestimate the representativeness of small samples: Users who are members of homogeneous OSN groups perceive themselves as surrounded by people that agree with them and they may therefore generalize that a large percentage of the population share their opinions, despite their OSN not reflecting true public opinion. Humans are known to show a general tendency to perceive public opinion as favorably distorted towards their own beliefs. This tendency is called the false consensus effect (FCE)⁷: Given two options (e.g. "Yes" or "No" regarding approval to the statement "Smoking in public places should be banned.") the percentage of the population in favor of one option is overestimated by subjects in favor of this option as compared to subjects in favor of the other option, and vice versa. This biased perception of others' opinions has been linked to biased perception of social norms and, as a consequence, to radicalization of actual behavior⁸: For example, people holding radical political views may interpret a high number of similar minded people in

their biased OSN as the public supporting their views and endorsing acting upon them, thus reducing their inhibition to (violently) do so. It can therefore be assumed that OSN usage behavior characterized by a tendency to homogeneity leads to higher FCE, more favorably distorted perceived social norms and, in turn, negative outcomes like polarization and radicalization. Although some studies already investigated FCE in radical online groups^{9,10}, general links between characteristics of OSN usage (namely intensity and homogeneity) and FCE have not been empirically tested yet. However, several characteristics of online echo chambers have been found to be correlated with FCE and support such a link: For example, higher availability of divergent opinions^{8,11} and more interaction with out-group members¹² have been linked to smaller FCEs.

Hypotheses

Based on these findings the present studies test the hypothesis that a higher homogeneity of a user's network of online contacts – that is, consisting of people with a high degree of similarity in traits like social or ethnical background or opinions, as opposed to higher variety in traits – is associated with a higher level of FCE displayed by the user (H1).

Moreover, several studies report that subjects use OSN¹³⁻¹⁵ and other online media like blogs¹⁶ or online news¹⁷ in a selective way, preferring information that is consistent with their own opinions. Therefore, it is assumed that the intensity of OSN usage and reliance on online sources for opinion formation should be associated with FCE. Thus, two more hypotheses were tested: FCE for a topic is positively correlated with the extent to which users rely on online sources to get information about the topic (H2) as well as with their over-all daily OSN usage time (H3).

Finally, the study takes into account that users' cognitive style may also influence the extent of FCE: Ambiguity tolerance¹⁸⁻²⁰ refers to an individual's trait of perceiving ambiguous situations as desirable or threatening. While there is no literature linking ambiguity tolerance directly to FCE or intensity of OSN usage, ambiguity tolerance has been found to be positively correlated with measures like curiosity²¹ that can be assumed to influence a subject's information seeking and appraisal behavior. For example, higher curiosity should lead subjects to seek more information from different sources, thus being confronted with more divergent beliefs and, in turn, getting a more realistic picture of public opinion distribution which should lead to a lower FCE. Therefore the study assumes that users with lower ambiguity tolerance tend to avoid divergent opinions and information in OSN, thus displaying a higher FCE (H4).

Materials and methods

Participants

A first online study with 181 German participants^a (112 female, aged 18 to 35 years, $M = 23.40$; $SD = 4.19$, primarily students recruited on University of Passau campus) was conducted to test the hypotheses. A second online study with 199 participants^a (143

^a Raw samples sizes were 203 (Study 1) and 239 (Study 2). Obvious cases of invalid data (e.g. participants invariably selecting one endpoint in every scale) were excluded. Further cleaning was based on completion time which originally displayed great variance (218s to 46h), hinting at possible distortions due to superficial answering or long pauses. Therefore, cases with extremely high (> 60 min.) or low (< 5 min.) completion times were initially excluded. Then only cases with completion times between 2SD below and above mean completion time were kept (final completion time: Study 1: $M = 16.89$; $SD = 6.43$ min., Study 2: $M = 16.44$ min.; $SD = 6.14$ min.).

female, aged 18 to 35, $M = 24.58$; $SD = 3.91$) recruited via German student social media groups was intended to replicate the findings. A post-hoc power analysis with G*Power²² showed that with 180 participants an FCE on group level of size $d = 0.50$ (equaling a medium-sized effect²³) could be found with a statistical power of $1 - \beta = .96$, given $\alpha = .05$, one-sided testing, and equal group sizes.

Materials

FCE was measured for 20 topics^b, eight of them adapted from the literature⁸ and the remaining based on recent public debates in Germany. The standard FCE paradigm⁷ was used. Participants were shown a statement (e.g. “There should be a referendum about EU membership of Germany”), then asked to estimate the percentage of population in favor of it, and finally asked whether the participants themselves were in favor of or opposed the statement.

Homogeneity of participants’ OSN was operationalized via the mean of a translated 12-item scale for online discussion network attributes²⁴ that has been shown to be correlated with measures like civic and political participation²⁴. The scale does not directly ask participants about their OSN homogeneity - which may lead to answers distorted by social desirability - but instead measures how often participants interact with different kinds of people online, answered on a 6-point Likert scale with endpoints ‘very seldom’ and ‘very

^b Topics included: Euthanasia, social welfare, animal testing, adoption by homosexual persons, legalization of marihuana, public smoking ban, traditional family values, strict punishment for crime, immigration, protection of environment vs. economic growth, European unification, capital punishment, foreign aid, foreign cultural infiltration, video surveillance on public places, ankle monitoring of terror suspects, halal meals in public cafeterias, ban on diesel vehicles in town centers, dual citizenship

often' (items^c included e.g. "friends and family" and "people who disagree with me"). Despite this not being a direct measure of homogeneity, lower values, indicating less diverse online contacts, can be interpreted as higher homogeneity. Intensity of OSN usage was measured via an overall estimate of daily OSN usage times in minutes. Topic specific internet usage was measured by asking how often the user read online information about the topic during the last week on a 6-point Likert scale (end-points "very seldom" and "very often").

Ambiguity tolerance was measured via the 13-item Multiple Stimulus Types Ambiguity Tolerance Scale-II MSTAT-II²⁵, which was chosen because it assesses global AT without reference to specific content like working life that may not apply to the student population. MSTAT-II was translated into German and backward-translated by a bilingual speaker. Based on feedback after Study 1 two items were slightly rephrased. A short scale for social desirability was used as a control variable²⁶.

Procedure

In both studies participants first answered questions on intensity of OSN usage and network homogeneity, then completed FCE questions followed by their estimate of topic-specific internet usage and finally answered the ambiguity tolerance and social desirability scales.

Data analysis

Data analysis was conducted with GNU R 3.5.0.

^c The original scale included 11 items, an additional item "people who know as much about a topic as myself" was added after pre-testing.

Overall FCE was calculated for each topic via standard⁷ group mean comparison: The mean estimated percentage of the population in favor of a statement was compared between participants who themselves agreed with or opposed the statement via one-tailed Bonferroni-Holm corrected *t*-tests or Wilcoxon rank sum tests, depending on normality of data.

Internal reliabilities of the ambiguity tolerance and network homogeneity scales were calculated via Cronbachs α .

For testing of hypotheses, an individual-level measurement of FCE was necessary. Different calculations for this have been proposed in the literature^{27,28}. This study, however, utilizes an alternative linear multilevel modeling approach (using the *lme4*²⁹ package for estimation of models and *lmerTest*²⁹ for significance testing, employing Satterthwaite's approximation to determine the degrees of freedom of predictors).

The model predicts the subject's estimation of the population percentage in favor of a statement. This outcome is first predicted by two random intercepts of the crossed random factors³⁰ *subject* and *topic*, thus taking into account baseline differences of estimations between participants (subjects may show idiosyncratic tendencies to estimate high or low public support, independent of topic) and topics (topics may consistently induce high or low estimates of public support across all subjects). The FCE is then modeled as the effect of the subject's *own opinion* concerning a topic (dichotomous with 0 = "opposed to" and 1 = "in favor of") on the outcome (i.e. the subject's estimate of approval of the statement in the population) with the regression weight indicating mean strength of FCE.

The hypothesized effects of *intensity of OSN usage*, *topic-specific internet usage*, and *ambiguity tolerance* as well as the control variables *social desirability*, *gender*, and *age*

on FCE are modeled as interactive terms between each variable and the dichotomous predictor *own opinion*. The interactions' regression weights thus display the effect of each predictor on extent of FCE. Due to differently scaled predictors all variables were standardized via z-transformation.

As the predictors *own opinion* and *topic-specific internet usage* were measured for each level of the factor *topic* they could be potentially modeled as random factors, thus taking into account variance in their effects between the topics. However, the predictors' small intra-class correlations (calculated as ratio of inter-topic variance to total variance)³¹ ranging between .00 and .02 illustrate that hardly any inter-topic variance in their effects exists. They were therefore included as fixed factors.

Results

Scales

Ambiguity tolerance showed good internal reliability after exclusion of one item ("I dislike ambiguous situations") in both studies ($\alpha = .86$ and $\alpha = .83$). Network homogeneity also showed good consistence ($\alpha = .89$ and $\alpha = .91$). Thus, scale means were calculated. Mean AT in Study 1 was 3.91 ($SD = .71$) and 3.87 ($SD = .63$) in Study 2, and homogeneity also showed little difference between studies ($M = 3.02$, $SD = .99$ and $M = 2.77$, $SD = 1.02$). Regarding intensity of OSN usage, participants reported a daily mean of 99.53 minutes ($SD = 120.95$) in Study 1 and 105.01 ($SD = 89.94$) in Study 2.

Overall false consensus effect

On group level, using standard group mean comparison both studies revealed an overall significant FCE for 18 (Study 1) and 19 (Study 2) out of 20 topics with strength of FCE varying between 23.94% ($W = 3000, p < .001$) and 8.21% ($W = 5300, p < .001$).

Multilevel Modelling

The results of the multilevel modelling are shown in table 1, with model 0 as random-intercept-only reference model, model 1 testing only control variables and model 2 testing all predictors. In both studies model 2 shows significantly better fit than model 1 (Study 1: $\chi^2(8) = 103.00, p < .001$; Study 2: $\chi^2(8) = 53.60, p < .001$), therefore results from model 2 are reported.

A significant FCE was found in both studies: If participants were in favor of a statement they estimated the percentage of the population in favor of it 10.81% (Study 1: $t(3513) = 10.95, p < .001$) or 12.04% (Study 2: $t(3865) = 11.10, p < .001$) higher than if they were opposed to it.

Significant interaction effects of network homogeneity and ambiguity tolerance with own opinion were also found and replicated: A more homogeneous online network increased the effect of own opinion, that is, the FCE (Study 1: $\beta = 2.03, t(3488) = 3.41, p < .001$, and Study 2: $\beta = 1.13, t(3841) = 2.04, p = .04$). Higher ambiguity tolerance was associated with a smaller FCE (Study 1: $\beta = -1.68, t(3480) = -2.79, p = .01$, and Study 2: $\beta = -1.11, t(3855) = 2.04, p = .049$). In the first study overall intensity of OSN usage also displayed a positive effect on FCE ($\beta = 1.20, t(3520) = 2.03, p = .04$), which, however, was not replicated in Study 2 ($\beta = -0.39, t(3828) = -0.71, p = .48$).

Topic-specific internet usage did not moderate the effect of own opinion (Study 1: $\beta = -0.63$, $t(3509) = -1.08$, $p = .28$, and Study 2: $\beta = 0.06$, $t(3848) = 0.11$, $p = .92$), yet showed a positive main effect: More topic-specific internet usage was associated with a higher estimation of public opinion being in favor of a topic, independent of the participants own opinion (Study 1: $\beta = 3.23$, $t(3567) = 7.06$, $p < .001$, and Study 2: $\beta = 2.20$, $t(3822) = 4.94$, $p < .001$).

No other effects were consistently found across the two studies.

Table 1. Multi-level linear regression on estimated percentage of population in favor of a statement

| | Study 1 | | | Study 2 | | |
|---|----------|----------|----------|----------|----------|----------|
| | Model 0 | Model 1 | Model 2 | Model 0 | Model 1 | Model 2 |
| <i>Random Effects (Std. Dev.)</i> | | | | | | |
| Participant | 7.77 | 7.41 | 7.01 | 7.43 | 6.74 | 6.70 |
| Topic | 10.79 | 9.32 | 8.97 | 10.89 | 9.52 | 9.25 |
| Residual | 17.75 | 16.95 | 16.74 | 17.41 | 16.76 | 16.64 |
| <i>Intercept</i> | 42.60*** | 36.79*** | 36.81*** | 42.61*** | 34.96*** | 34.99*** |
| <i>Fixed Main Effects (Std. β)</i> | | | | | | |
| Own opinion (1 = in favor) | | 10.15*** | 10.81*** | | 11.77*** | 12.04*** |
| Age | | 0.29 | 0.54 | | -1.04 | -0.86 |
| Gender (1 = Female) | | -0.69 | -0.50 | | 2.47 | 2.56 |
| Social desirability | | 0.09 | 0.41 | | -0.57 | -0.56 |
| Ambiguity tolerance | | | -0.58 | | | -0.10 |
| Network homogeneity | | | 0.08 | | | 1.10 |
| Daily OSN usage time | | | 0.69 | | | 0.07 |
| Topic-specific internet usage | | | 3.23*** | | | 2.20*** |
| <i>Fixed Interaction Effects with own opinion (Std. β)</i> | | | | | | |
| Age | | -1.86** | -1.94** | | -0.03 | -0.13 |
| Gender (1 = Female) | | 3.46** | 1.88 | | -0.40 | -1.06 |
| Social desirability | | 0.24 | 0.24 | | -1.45** | -1.24* |
| Ambiguity tolerance | | | -1.68** | | | -1.11* |
| Network homogeneity | | | 2.03*** | | | 1.13* |
| Daily OSN usage time | | | 1.20* | | | -0.39 |
| Topic-specific internet usage | | | -0.63 | | | 0.06 |
| <i>-2LogLikelihood</i> | 31464 | 31128 | 31026 | 34426 | 34102 | 34048 |
| $\chi^2(\Delta -2LogLikelihood)$ | | 336.0*** | 103.0*** | | 326.0*** | 53.6*** |

*** $p < .001$; ** $p < .01$; * $p < .05$; N(Study 1) = 181; N(Study 2) = 199

Discussion

Both studies confirm the assumption that a more selective and homogeneous online contact network is associated with a higher estimation of public support for own opinions: The less diverse a subject's self-reported online contacts, the higher the subject's FCE. Although this effect is only correlative it appears plausible to assume that diversity affects FCE instead of the reverse. Thus, subjects appear to rely on their online contact network when judging public opinion, and a biased network, for example by mostly participating in attitudinally homogeneous OSN groups, should lead to biased estimation of public opinion. With estimated public opinion as indicator of perceived social norm⁸ and social norm linked to behavioral intention³², the study could thus identify a potential mechanism in which homogeneous online networks – “echo chambers” – may influence actual behavior.

Evidence for an effect of ambiguity tolerance on FCE could also be found and replicated, with subjects with higher ambiguity tolerance displaying smaller FCE, which is significant even when controlling for effects of online usage characteristics. This allows for several interpretations: On the one hand, lower ambiguity tolerance may lead to more selective consumption of information via other, non-online sources, for example traditional media or social groups, which in turn leads to higher FCE. On the other hand, there might be a cognitive effect of lower ambiguity tolerance leading to a preferred recall of information consistent with own attitudes when making judgments about public opinion, independent of the real availability of information.

Ambiguity tolerance as well as network homogeneity displayed lower effect sizes in Study 2 as compared with Study 1 ($\beta = 1.13$ vs. $\beta = 2.03$ for AT, and $\beta = -1.11$ vs.

$\beta = -1.68$ for homogeneity) and were closer to the boundary of significance (especially ambiguity tolerance with $p = .049$). Furthermore, a positive effect of overall intensity of OSN usage on FCE in Study 1 could not be replicated in Study 2. These differences might be attributed to sample and situational characteristics: The first study relied mostly on students of one university, while the second recruited users of student OSN groups. Moreover, Study 1 was conducted in spring 2017 and Study 2 in September 2017, parallel to German general elections, so participants in Study 2 might have been more sensitized towards the societal and political topics used to measure FCE. Additionally, the public debate about online “echo chambers”, “filter bubbles”, and “fake news” and their possible negative effects gained momentum during the year 2017, which may have led participants in the replication study to avoid answering in a way that let them appear as biased. We found higher social desirability associated with lower FCE in Study 2 but not Study 1 which provides some support for this explanation.

It should be noted that, while inclusion of predictors led to significant reduction in unexplained variance compared to an intercept-only model without predictors (see table 2), this reduction was numerically small and much variance remained unexplained. This hints at other predictors of FCE not accounted for in this study. Future studies should try to assess and discriminate more offline and online variables influencing FCE. One such variable could be offline contact network homogeneity, whose effects could be compared to those of online network homogeneity.

The assumed positive correlation of topic-specific internet usage and FCE was not found in any of the two studies. However, a positive topic-independent main effect of usage on general estimation of percentage of population in favor of a statement was found and replicated. This unexpected finding is difficult to explain and may hint at a low validity

of the self-report measure “topic-specific internet usage” which also displayed small variance with most subjects reporting low topic-specific usage.

While the reported studies provide new insights into possible links between online “echo chambers”, users’ cognitive style and behavior, they exhibit weaknesses typical of correlative studies: Online usage, OSN characteristics, and ambiguity tolerance were measured via self-report and may have been influenced by memory, response biases, or traits like diligence, leading to reduced objectivity and validity. Moreover, both studies can only report correlative relationships yet no causal effects. Further experimental studies are needed to assess causal links between OSN characteristics and user-side behavioural outcomes, as well as to counteract distortions due to social desirability in survey-based studies caused by increased public discussion of negative effects of biased online information.

Author Disclosure Statement

No competing financial interests exist.

References

1. Sunstein CR. (2001) *Echo chambers: Bush v. Gore, impeachment, and beyond*. Princeton, Oxford: Princeton University Press.
2. Gilbert E, Bergstrom T, Karahalios K. (2009) Blogs Are Echo Chambers. Proceedings of the 42nd Hawaii International Conference on System Sciences 2009.
3. Del Vicario M, Bessi A, Zollo F, et al. The spreading of misinformation online. Proceedings of the National Academy of Sciences 2016; 113:554–559.
4. Grömping M. ‘Echo Chambers’: Partisan Facebook Groups during the 2014 Thai Election. Asia Pacific Media Educator 2014; 24:39–59.
5. Williams HTP, McMurray JR, Kurz T, et al. Network analysis reveals open forums and echo chambers in social media discussions of climate change. Global Environmental Change 2015; 32:126–138.
6. Tversky A, Kahneman D. Belief in the law of small numbers. Psychological Bulletin 1971; 76:105–110.
7. Ross L, Greene D, House P. The “false consensus effect”: An egocentric bias in social perception and attribution processes. Journal of experimental social psychology 1977; 13:279–301.
8. Bauman KP, Geher G. We Think You Agree: The Detrimental Impact of the False Consensus Effect on Behavior. Current Psychology 2002; 21:293–318.
9. Wojcieszak M. False Consensus Goes Online: Impact of Ideologically Homogeneous Groups on False Consensus. Public Opinion Quarterly 2008; 72:781–791.

10. Wojcieszak ME. Computer-Mediated False Consensus: Radical Online Groups, Social Networks and News Media. *Mass Communication and Society* 2011; 14:527–546.
11. Wojcieszak M, Price V. What Underlies the False Consensus Effect? How Personal Opinion and Disagreement Affect Perception of Public Opinion. *International Journal of Public Opinion Research* 2009; 21:25–46.
12. Bosveld W, Koomen W, van der Pligt J. Selective exposure and the false consensus effect: The availability of similar and dissimilar others. *British Journal of Social Psychology* 1994; 33:457–466.
13. Quattrociocchi W, Scala A, Sunstein CR. Echo chambers on facebook. *Harvard John M. Olin Discussion Paper Series* 2016: 877.
14. Schmidt AL, Zollo F, Del Vicario M, et al. Anatomy of news consumption on Facebook. *Proceedings of the National Academy of Sciences* 2017; 114:3035–3039.
15. Zollo F, Bessi A, Del Vicario M, et al. Debunking in a world of tribes. *PLOS ONE* 2017; 12:e0181821.
16. Johnson TJ, Bichard SL, Zhang W. Communication Communities or “CyberGhettos?”: A Path Analysis Model Examining Factors that Explain Selective Exposure to Blogs. *Journal of Computer-Mediated Communication* 2009; 15:60–82.
17. Knobloch-Westerwick S, Johnson BK, Westerwick A. Confirmation Bias in Online Searches: Impacts of Selective Exposure Before an Election on Political Attitude Strength and Shifts. *Journal of Computer-Mediated Communication* 2015; 20:171–187.
18. Frenkel-Brunswik E. Intolerance of Ambiguity as an Emotional and Perceptual Personality Variable. *Journal of Personality* 1949; 18:108–143.

19. Budner S. Intolerance of ambiguity as a personality variable¹. *Journal of Personality* 1962; 30:29–50.
20. Norton RW. Measurement of Ambiguity Tolerance. *Journal of Personality Assessment* 1975; 39:607–619.
21. Litman JA. Relationships between measures of I- and D-type curiosity, ambiguity tolerance, and need for closure: An initial test of the wanting-liking model of information-seeking. *Personality and Individual Differences* 2010; 48:397–402.
22. Faul F, Erdfelder E, Lang A-G, et al. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods* 2007; 39:175–191.
23. Cohen J. (1988) *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale: Lawrence Erlbaum Associates.
24. Zúñiga HG de, Jung N, Valenzuela S. Social Media Use for News and Individuals' Social Capital, Civic Engagement and Political Participation. *Journal of Computer-Mediated Communication* 2012; 17:319–336.
25. McLain DL. Evidence Of The Properties Of An Ambiguity Tolerance Measure: The Multiple Stimulus Types Ambiguity Tolerance Scale-II (MSTAT-II). *Psychological Reports* 2009; 105:975–988.
26. Lück H, Timaeus E. Soziale Erwünschtheit (SDS-CM). *Zusammenstellung sozialwissenschaftlicher Items. ZIS Version 15.00*.
27. de la Haye A-M. A methodological note about the measurement of the false-consensus effect. *European Journal of Social Psychology* 2000; 30:569–581.

28. Galesic M, Olsson H, Rieskamp J. (2013) False consensus about false consensus. In: Knauff M, Pauen M, Sebanz N, et al., eds. *Proceedings of the 35th Annual Conference of the Cognitive Science Society*, pp. 472–476.
29. Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 2017; 82.
30. Baayen RH, Davidson DJ, Bates DM. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 2008; 59:390–412.
31. Hox JJ. (2002) *Multilevel analysis: techniques and applications*. Mahwah, N.J: Lawrence Erlbaum Associates.
32. Ajzen I. The theory of planned behavior. *Organizational Behavior and*

4. Study two: Selective exposure and False Consensus

Original paper:

Luzsa, R. & Mayr, S. (submission in preparation). False Consensus in the Echo Chamber:

Exposure to Favorably Biased Social Media News Feeds Leads to Increased

Perception of Public Support for Own Opinions.

Please note that the following chapter is the unedited pre-review version of the article and does not reflect possible changes made during the peer-review process.

False Consensus in the Echo Chamber: Exposure to Favorably Biased Social Media News Feeds Leads to Increased Perception of Public Support for Own Opinions

Abstract

Studies have found that users of online social networking sites often preferably connect with like-minded others, leading to homogenous “Echo Chambers” in which attitudinally congruent information circulates. However, little is known about how Echo Chambers affect individuals’ perceptions and behavior. This study aimed to experimentally induce an Echo Chamber and to measure subjects’ perception of public opinion as a behavior-related outcome. It was predicted that subjects’ estimate of public support for their own attitudes is affected by the level of agreement they encounter in social media news feeds, with high agreement leading to a higher estimate of public support than low agreement. In two online experiments ($n_1 = 331$ and $n_2 = 207$) subjects were exposed to nine news feeds, each containing four messages. Two factors were manipulated: Agreement expressed in message texts (all but one [Exp.1] / all [Exp.2] messages were congruent or incongruent to subjects’ attitudes) and endorsement of congruent messages by other users (messages congruent with subjects’ attitudes displayed either higher or lower numbers of “likes” than attitudinally incongruent messages). Additionally, based on Elaboration Likelihood Theory, subjects’ interest in a topic was considered as a moderating variable. Both studies confirmed that subjects infer public support for their own attitudes from the degree of agreement they encounter in online messages, yet are skeptical of the validity of “likes”, especially if their interest in a topic

is high. Based on the findings, possible interventions to reduce adverse effects of Echo Chambers are discussed.

The authors do not have any conflicts of interest to report.

Introduction

There is growing debate in media, society, and science whether online social networks like Facebook or Twitter facilitate biased information consumption and opinion formation, and, in turn, give rise to negative consequences like misinformation (Del Vicario et al., 2016), radicalization, and societal polarization (Grömping, 2014; Williams, McMurray, Kurz, & Lambert, 2015). Indeed, studies have demonstrated that users of online social networks tend to display confirmation bias (Wason, 1968) and homophily: They often prefer information that is consistent to their own attitudes (e.g. Knobloch-Westerwick, 2015) and form ties with like-minded others (Del Vicario et al., 2016; Quattrociocchi, Scala, & Sunstein, 2016; Schmidt et al., 2017; Zollo et al., 2017), especially in regard to highly political topics (Barberá, Jost, Nagler, Tucker, & Bonneau, 2015). Online communication that shows these tendencies has been characterized as an “Echo Chamber” (Gilbert, Bergstrom, & Karahalios, 2009; Quattrociocchi et al., 2016; Sunstein, 2001), as similar attitudes and information supporting these attitudes “echo” between users, and deviant opinions are excluded.

At first glance, it is not surprising that users show these tendencies, as humans have long been known to prefer selective exposure to attitudinally congruent information (Knobloch-Westerwick, 2014; Sears & Freedman, 1967) in order to reduce cognitive dissonance (Cotton & Hieser, 1980; Hart et al., 2009). It is also known that social ties are often formed based on perceived similarity (McPherson, Smith-Lovin, & Cook, 2001). However, there are several reasons why the effects of online Echo Chambers on individuals’ behavior and experience appear as qualitatively new phenomena and require further examination:

Firstly, it can be argued that being selective is much easier with social media than with traditional media. By following news outlets and other users that support their own views on social media, individuals can easily surround themselves with a highly personalized and permanently available stream of attitudinally congruent information. With traditional media, subjects may also select TV channels or newspapers that they know to support their own views, but this selectivity is less personalized, less complete, and more limited in scope.

Secondly, in online social networks, algorithms analyze users' online behavior and provide them with content based on their interests and prior usage patterns. Thus, the users' endogenous tendency to selective exposure is complemented by exogenous selectivity, which they might not even notice. It has been argued that this further isolates users from contradicting information and strengthens their biases (Pariser, 2011).

Thirdly, contents on social media are usually not subject to editorial review and fact checking, allowing for the spread of highly subjective or even intentionally manipulative contents, the latter often referred to as "Fake News" (Allcott & Gentzkow, 2017; DiResta et al., 2018; Howard, Ganesh, Liotsiou, Kelly, & François, 2018; Tandoc, Lim, & Ling, 2018).

In summary, social media appear to facilitate selective exposure to one-sided information, biased in favor of users' pre-existing attitudes. While it seems conceivable to assume that this bias leads to the aforementioned negative outcomes such as radicalization and polarization, existing research into Echo Chambers usually does not focus such links directly. Instead, studies often employ large (social media) datasets to identify Echo Chamber-like structures and behavior by analyzing how information is

shared between users (e.g. Barberá, Jost, Nagler, Tucker, & Bonneau, 2015; Grömping, 2014). This approach allows for high ecological validity, however, it cannot substantiate causal relationships between exposure to Echo Chambers and behavior-related outcomes on an individual level.

To this end, controlled experimental studies are necessary which expose subjects to online environments with either high or low Echo Chamber characteristics and then measure their effects on the individual. To the authors' best knowledge, no such studies exist. This study therefore reports two experiments that exposed subjects to artificially created online Echo Chambers and measured an outcome related to actual behavior: False Consensus, that is, subjects' perception of public support for their own opinions.

Links between Echo Chambers and Perceived Public Support for Own Opinions

It is known that the perception of public opinion or social norm is an important predictor of actual behavior (Ajzen & Fishbein, 1977; Fishbein, 1976; Liska, 1984): For example, individuals who hold radical and aggressive political views but perceive the public in disagreement with these views may not act upon them. However, if they perceive the public in support of their views, their inhibitions to show hostile or aggressive behavior may be reduced.

Research has also demonstrated that subjects do not estimate public opinion objectively. Instead, they show a general tendency to perceive public opinion as favorably distorted towards their own beliefs. This tendency is called the False Consensus Effect (FCE, Ross, Greene, & House, 1977): Given two options (e.g. "Yes" or "No" regarding approval to the statement "Marihuana should be legalized.") the percentage of the population in favor of one option is overestimated by subjects in favor of this option as

compared to subjects in favor of the other option, and vice versa. For example, subjects with a positive stance towards marijuana legalization might estimate that 45% of the population also have a positive stance, while subjects strongly opposed to legalization might estimate that only 25% favor legalization. The difference of 20% reflects the extent of the FCE, that is, how strongly the estimate of public opinion depends on subjects' own opinion.

Previous studies suggest that the extent of FCE is correlated with subjects' selective exposure to attitudinally congruent information and personal network homogeneity: For example, Bauman and Geher (2002) measured subjects' FCE after exposure to packages of information (e.g. brochures, pamphlets or video-taped discussions). The authors found that subjects who were exposed to balanced packages which contained both information supporting and questioning their own opinions displayed a lower FCE than control group members without any exposure. This can be interpreted as an effect of selective exposure: In their everyday life, subjects prefer congruent information from which they infer high public support for their opinions. However, when subjects are exposed to different views on a topic, they are also forced to process incongruent information that they would not normally seek out, leading to a lower FCE.

That an effect of selective exposure on FCE might also occur by participating in homogenous online groups is illustrated by Wojcieszak (2008) who examined influences on the FCE for members of neo-Nazi and radical environmentalist online forums. In the case of the neo-Nazi forum, the study found FCE positively correlated with subjects' degree of forum participation as well as their level of extremism (however, in case of the environmentalist forum, only extremism, not forum participation correlated with FCE).

Given these findings, it stands to reason that the perceived public support for one's own opinion should also be influenced by how strongly an user's network of online contacts resembles an Echo Chamber: If subjects surround themselves online with others that mostly share their attitudes, they should receive more messages that contain congruent and less incongruent information and they should also experience less interactions with others holding opposing beliefs. This should lead to a stronger FCE. Indeed, in a prior correlative questionnaire study (AUTHORS, 2019), such a link was found: The more homogenous subjects described their online social network, the stronger their FCE turned out to be, measured for twenty current political topics. Due to the correlative nature of that study, however, the cause for this stronger FCE could not be unambiguously identified. The assumption that FCE is influenced by selective exposure to attitudinally congruent information shared in subjects' networks still requires testing. For this, an experimental approach is necessary that examines how attributes of messages shared in online social media affect FCE.

Message Attributes in Online Echo Chambers

Typical for online social media are (individual) news feeds consisting of messages that other connected users have shared. Each message consists of central content (e.g. headlines of news articles or personal commentaries), accompanied by so-called popularity cues, that is, numeric indicators which illustrate how many other users positively evaluated or endorsed the content, for example by giving it a "like" (Haim, Kümpel, & Brosius, 2018; Porten-Cheé, Haßler, Jost, Eilders, & Maurer, 2018). While there are also additional message attributes, such as the name of the sender, their avatar or included images, this study will focus on popularity cues, in addition to message contents. This focus is made because the influence of popularity cues on users' attention

(Dvir-Gvirsman, 2019) as well as on users' selection and appraisal of messages (Chang, Yu, & Lu, 2015; Haim et al., 2018; Messing & Westwood, 2014) is well documented.

It can be argued that message content and other users' endorsement via "likes" are key attributes that differentiate Echo Chambers from more balanced, heterogeneous online communication: Firstly, the news feed of users with homogenous networks should express *agreement* with their own attitudes, that is, consist mostly of attitudinally congruent messages. Secondly, these attitudinally congruent contents should display strong *endorsement* by their network, while occasionally occurring attitudinally incongruent contents should show less endorsement. In contrast, if users' networks are heterogeneous there should be neither dominance of attitudinally congruent messages nor higher endorsement for congruent than incongruent ones. In fact, depending on topic and network, users might even encounter mostly incongruent messages as well as popularity cues that display low endorsement for their own attitudes – for example in communities that emphasize controversial discussions (Guest, 2018).

Hypotheses

Based on these considerations two hypotheses regarding the effect of Echo Chamber news feeds on FCE are formulated:

Effects of agreement: Subjects that are exposed to a news feed made up of messages mostly congruent to their own attitudes will display a stronger FCE, compared to subjects exposed to mostly incongruent messages.

Effects of endorsement: If messages congruent to subjects' own attitudes display higher endorsement by others than incongruent ones, subjects will exhibit a stronger FCE,

compared to the situation in which incongruent messages display higher endorsement than congruent messages.

In addition, it is important to note that both factors – *agreement* expressed in message texts and *endorsement* of attitudinally congruent messages expressed by “likes” – can vary independently of each other: For example, users might read a feed in which 90% of the messages are congruent to their own attitudes. A positive Echo Chamber effect on FCE should occur. However, what happens if the remaining 10% of incongruent messages display significantly higher numbers of “likes”? Will this reduce the positive effect of message *agreement* on FCE? The opposite situation can also be imagined, with users reading a feed with mostly incongruent messages. This should lead to a weaker FCE. However, if the few attitudinally congruent messages display the highest *endorsement*, will users interpret this as a “silent majority” agreeing with them, and therefore display a stronger FCE?

Previous research does not allow for a clear assumption whether *agreement* or *endorsement* is the pivotal factor and whether there will be an interactive effect. Therefore, an open research question is formulated:

Interactive effect: Is there an interactive effect between *agreement* and *endorsement* on FCE?

Finally, in the tradition of the Elaboration Likelihood Model (Petty, 1986), popularity cues such as “likes” can be conceived as a peripheral message cue, and message content as the central cue. The model states that when subjects’ involvement is high, they will be mostly affected by central cues, while in case of low involvement peripheral cues will have a stronger impact.

Transferred to the current study, this implies that subjects' *interest* in a topic might moderate the effects of *agreement* as well as *endorsement* on FCE: Regarding the role of *agreement*, if subjects have little *interest* in a topic they might put less effort in processing messages related to it, with message content showing little effect on FCE. However, if *interest* is high, subjects might more readily put effort in reading and evaluating messages, leading to a stronger effect of message content than when *interest* is *low*. Regarding the role of *endorsement*, subjects who have little *interest* in a topic might focus on available popularity cues such as “likes” as an easy and effortless way to estimate public opinion. In contrast, the effect of “likes” on FCE should turn out weaker in case of *high interest*, because possibly higher interest with a topic might lead subjects to be more skeptical regarding the representativeness of displayed popularity cues.

From this follows the final hypothesis:

Moderation by interest: The effects of *agreement* as well as of *endorsement* on FCE are moderated by subjects' *interest* in a topic. The effect of *agreement* will turn out stronger in case of *high interest* than when *interest* is *low*. The effect of *endorsement* will turn out stronger in case of *low interest* than when *interest* is *high*.

These questions and hypotheses were examined in two online experiments.

Experiment 1

Experiment 1 adapted selective exposure paradigms (Lee & Yoon Jae Jang, 2010; Messing & Westwood, 2014; Peter, Rossmann, & Keyling, 2014) in a way that allows the short-term induction of an Echo Chamber: Subjects were exposed to several news feeds, each consisting of four simultaneously presented messages, that is, short news headlines regarding one topic. Each message was accompanied by a number of “likes”.

Subjects' task was to select the message whose linked full article they preferred to read. No full articles were displayed afterwards. The task was merely given to ensure that subjects read and processed all the messages.

The two aforementioned attributes of Echo Chamber news feeds – *agreement* expressed in messages and *endorsement* by “likes” – were independently manipulated: Subjects were exposed to either three messages congruent to their own attitudes and one incongruent (condition *high agreement*) or to three incongruent and one congruent (condition *low agreement*). Similarly, either the congruent messages had high and the incongruent messages low numbers of likes (condition *high endorsement*) or vice versa (condition *low endorsement*). The news feeds reflecting these conditions were created during runtime of the experiment based on the initially assessed own attitudes of the subjects.



Figure 1. Example of a news feed used in Experiment 1.

An example: If a subject favors the legalization of marijuana, figure 1 illustrates a news feed that reflects *low agreement* (most messages are incongruent as three of four

highlight the dangers of marihuana) and *low endorsement* for his or her own attitude (the three anti-marihuana messages have higher numbers of “likes” than the one in favor of legalization).

Method

Sample. A self-administered online experiment was conducted with 388 German subjects, recruited on University of Passau campus and from the authors’ volunteer database. Implausible cases were excluded based on completion times: First, subjects who took less than 5 or longer than 60 minutes to complete the experiment were dropped. Then, only cases within 2 SD of the resulting mean completion time were kept. Final mean completion time was 12.03 minutes ($SD = 4.57$; $Min = 5.15$; $Max = 28.52$). The remaining sample comprised 331 subjects (231 female; age between 18 and 35 years with $M = 22.36$ and $SD = 0.27$). A-priori power analysis with G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) determined that at least 195 subjects were necessary to achieve a power of $1 - \beta = .95$, given $\alpha = .05$ and a medium sized effect of $f^2 = 0.15$ in the later described model used for hypothesis testing¹.

Design. The experiment implemented a 2 x 2 between-subjects design with the factors *agreement* and *endorsement*: Subjects were consistently exposed to either mostly

¹ The power refers to an F -test evaluating whether the later described full model explains significantly more outcome variance than a control variables only model. As the estimated linear mixed effects model uses random intercepts but only fixed slopes, power is approximated via power calculation for traditional OLS regression. This serves as conservative lower-boundary approximation of power, as in case of fixed slopes, mixed effect models generally have more power than OLS (Snijders, 2005).

attitudinally congruent (*high agreement*) or incongruent messages (*low agreement*), and congruent messages consistently displayed either higher (*high endorsement*) or lower (*low endorsement*) numbers of “likes” than incongruent ones. Subjects’ estimate of public opinion for nine topics was measured to calculate FCE as dependent variable.

Materials. The topics of the news feeds were selected in accordance with results of a prior study (AUTHORS, 2019) which had measured subjects’ own attitudes and FCE regarding twenty current political topics. From these, nine topics were selected which had elicited a strong FCE yet had also displayed some variance of subjects’ own attitudes, that is, which had not evoked unanimous assent or dissent. Topics included: Legalization of marihuana, traditional family values, strict punishment for crime, European unification, video surveillance on public places, foreign cultural infiltration, animal testing of drugs, dual citizenship, and ban on diesel vehicles in city centers.

News feed messages were based on headlines and teaser texts found on social media accounts and websites of German news outlets (e.g. “Der Spiegel”, “Die Welt”). For each topic, 16 texts were selected, eight of which expressed consent regarding the topic’s statement (e.g. highlighted the advantages of legalization of marihuana), while the remaining eight stated dissent (e.g. emphasizing the dangers of marihuana). In order to eliminate confounding variables, the assenting and dissenting messages’ characteristics (e.g. word count, stylistic features such as use of exclamation marks) were balanced by rephrasing messages when necessary. Example messages are presented in figure 1. In a pre-test, 15 subjects rated each message regarding two aspects: How strong is the message’s assenting or dissenting stance towards the topic (6-item Likert scale from "Strongly agrees" to "Strongly disagrees")? How likely would subjects themselves click on the message in order to read the full article, if they saw it in their own social media

news feed (6-item Likert scale from “Would certainly not click on it” to “Would certainly click on it”)? For the experiment, four assenting and four dissenting messages for each topic were selected which had been rated as voicing strong opinions and as moderately likely to be clicked on. The number of words in assenting ($M = 29.33$) and dissenting messages ($M = 28.47$) was approximately equal.

The numbers of “likes” displayed to indicate low or high endorsement of messages were similar to those used by Messing and Westwood (2014), which were based on actual average numbers of “likes” of (American) Facebook messages: Low endorsement was expressed by a number of 100 to 500 “likes”, high endorsement by a number between 6000 and 19000.

Measures. Prior to news feed presentation, subjects indicated their own stance towards each topic (e.g. agreement to the statement “Marihuana should be legalized.”) on a 6-point Likert scale (“Strongly agree” to “Strongly disagree”). To ensure that the initial attitude measurements were stable and a valid basis for FCE calculation, attitudes were measured again at the end of the experiment. Due to the short re-test interval, the initial statements were not reused. Instead, two additional items per topic were formulated (mostly based on existing publicly available questionnaires, e.g. “Personal possession of marihuana should not be criminalized.”) and confirmed in a pre-test to be consistent with the initial item (all Cronbach’s $\alpha > .80$). The means of the two items were then used for post-exposure attitude measurement. In addition to own stance towards the topics, subjects indicated how interested they were in each topic on a 6-point Likert scale (“very interested” to “not at all interested”).

During news feed presentation, it was recorded how often subjects chose attitudinally congruent or incongruent messages and how often messages with high or low endorsement were selected.

Finally, to allow for FCE calculation, perceived public opinion was assessed after news feed presentation: For each topic, subjects were shown the statements previously used for attitude measurement (e.g. “Marihuana should be legalized”) and were asked to estimate the percentage of the population in favor of the statement via numerical input (0 to 100%).

Procedure. First, subjects were first informed about voluntariness of participation and the possibility to cancel at any time. Then, they gave their consent regarding data privacy. After stating their own attitudes and interest regarding the topics, they were presented with the nine news feeds in accordance with their randomly assigned experimental condition. The order of presentation of feeds as well as the order of messages in each feed were randomized. Each feed was displayed until the subject selected a message. Afterwards, subjects estimated the percentage of the population with a positive stance regarding each topic. Then, they answered the post-exposure attitude items and gave basic demographic data. Finally, subjects were debriefed and informed about the experimental manipulation.

Analysis. Data analysis was conducted with GNU R 3.5.2.

Stability of attitudes. To ensure that the initially measured attitudes were stable and could be used for the further analysis, correlations with the post-exposure attitude items were calculated.

Overall False Consensus Effect. First, it was tested whether the topics used in this experiment successfully elicited an overall FCE (independent of experimental manipulation). As subjects' attitude towards topics was measured with scales and not with traditional dichotomous questions, FCE was conceptualized as the correlation between subjects' own stance towards a topic and their estimated percentage of the public with positive stance towards the topic. Positive correlations were expected (i.e., the more positive subjects' attitude towards a topic, the larger the estimated percentage of the public with a positive attitude towards the topic).

Effects on False Consensus. For hypothesis testing, an approach that had already been successfully employed in a previous study (AUTHORS, 2019) was used: Traditionally, FCE is operationalized by between-groups comparison (Ross et al., 1977). However, for the current experiment, an individual-level measurement appeared suitable, as individual-level factors such as subjects' own interest in a topic needed to be considered. Several approaches for an individual-level measurement of FCE exist but they are debated controversially (de la Haye, 2000; Galesic, Olsson, & Rieskamp, 2013). Therefore, an alternative linear mixed effects modelling approach was employed, using the "lme4" and "lmerTest" R-packages (Kuznetsova, Brockhoff, & Christensen, 2017) for model estimation and significance testing².

The model predicted subjects' estimate of population percentage that has a positive stance towards a topic. First, this outcome was predicted by two random intercepts of the crossed random factors (Baayen, Davidson, & Bates, 2008) *subject* and *topic*. By this,

² In "lmerTest", for significance testing, degrees of freedom are approximated via Satterthwaite's approximation.

baseline differences of estimates between subjects (subjects may display idiosyncratic tendencies to give high or low estimates, independent of topic) and topics (topics may generally lead to higher or lower estimates) were taken into account. Then, FCE was modeled as the effect of subjects' own *attitude* towards a topic on the outcome, that is, the estimated population percentage in accordance with the own attitude. The regression weight of *attitude* indicates the mean strength of the FCE, that is, how strongly subjects' estimates of public opinion are biased in favor of their own attitudes.

The effects of the manipulated factors *agreement* and *endorsement* on FCE were then operationalized as interactive terms of the dummy-coded factors (with -1 indicating low agreement/endorsement and 1 high agreement/endorsement, respectively) and subjects' own attitudes. For example, subjects were hypothesized to display stronger FCE in the *high agreement* than in the *low agreement* condition. Therefore, an interaction between the factor *agreement* and subjects' *attitude* on FCE should be found, meaning that in the *high agreement* condition the estimate of public opinion should be more strongly biased in favor of subjects' own attitudes.

Finally, subjects' *age* and *gender* as well as the news feed *presentation order* were included as control variables. All predictors entered were centered around population means in order to reduce variance inflation due to the included interactive terms.

Message selection. The task to click on a message during news feed exposition was primarily given to ensure that subjects read the messages, and no hypotheses regarding message selection were formulated. Nonetheless, an exploratory analysis of possible effects of the experimental manipulation on message selection was conducted. For this, first, the number of attitudinally congruent message choices of each subject (e.g. for 3 of

9 topics) was taken. Then, the number expected assuming random selection (e.g. $9 * 3/4 = 6.75$ if three of four presented messages were congruent) was subtracted. The resulting value (e.g. -3.75) was compared between experimental conditions via 2 x 2 ANOVA.

Results

Stability of attitudes. For 8 of 9 topics, strong positive correlations (r from .66 to .89, $p < .001$) between attitudes at the beginning and end of the experiment were found. Only the topic “EU integration” displayed a moderately positive correlation ($r = .42$, $p < .001$). Attitudes were therefore considered as stable and were used to calculate FCE. However, to eliminate possible confounding, hypotheses were additionally examined excluding the moderately stable topic “EU integration”. As results turned out virtually identical, results for all topics are reported.

Overall False Consensus Effect. An overall FCE was found for all topics: The more positive subjects’ own attitude towards a topic, the higher they estimated the percentage of the population with positive attitude towards it, with correlations varying between .14 (topic "Ban on diesel vehicles in city centers") and .42 (topic "Strict punishment for crime"). For example, subjects who strongly opposed stricter punishment for crime (answer 1 on a 6-point Likert-scale) estimated that 31.67% of the population support stricter punishment. This estimate increased to 66.31% if subjects reported to somewhat support stricter punishment (answer 4), and further to 77.58% if they strongly supported it (answer 6).

Table 1. Linear mixed effects regression on estimated percentage of population in favor of a statement

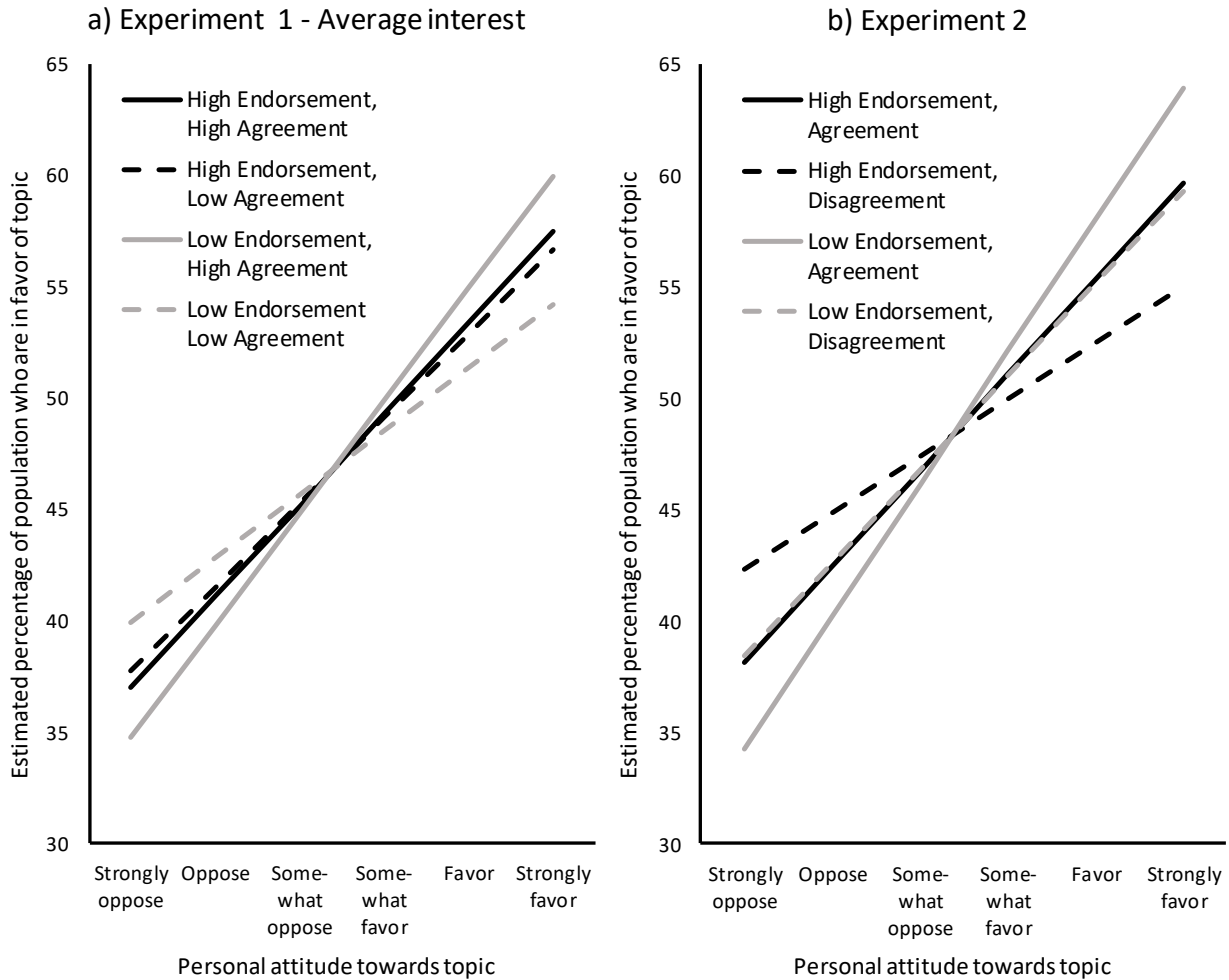
| | Experiment 1 | | | Experiment 2 | | |
|--|--------------|----------|----------|--------------|----------|----------|
| | Model 0 | Model 1 | Model 2 | Model 0 | Model 1 | Model 2 |
| <i>Random Effects (Std. Dev.)</i> | | | | | | |
| Subject | 6.38 | 5.87 | 5.87 | 8.60 | 7.74 | 7.40 |
| Topic | 8.27 | 7.21 | 7.14 | 8.90 | 7.12 | 7.04 |
| Residual | 17.72 | 17.04 | 16.92 | 18.61 | 17.77 | 17.64 |
| Intercept (β_0) | 47.40*** | 46.69*** | 46.68*** | 48.21*** | 48.42*** | 48.34*** |
| <i>Fixed Main Effects (Centered β)</i> | | | | | | |
| Own attitude | | 3.63*** | 3.95*** | | 4.33*** | 4.24*** |
| Age | | 0.13 | 0.13 | | -0.05 | -0.01 |
| Gender | | 1.02 | 1.05 | | 0.45 | 0.17 |
| Presentation order | | 0.18 | 0.20 | | 0.06 | 0.03 |
| Interest in topic | | | 0.51 | | | 0.27 |
| Agreement (1 = high, -1 = low) | | | -0.02 | | | -1.08 |
| Endorsement (1 = high, -1 = low) | | | -0.24 | | | -1.07 |
| Agreement x Endorsement | | | 0.32 | | | 0.18 |
| Agreement x Interest | | | -0.21 | | | -0.60 |
| Endorsement x Interest | | | -0.04 | | | 0.11 |
| Agree. x Endors. x Interest | | | 0.17 | | | 0.45 |
| <i>Fixed Interaction Effects with own attitude (Centered β)</i> | | | | | | |
| Age | | 0.00 | 0.01 | | -0.17 | -0.12 |
| Gender | | 0.73 | -0.01 | | 0.17 | 0.05 |
| Presentation order | | 0.00 | -0.02 | | 0.07 | 0.07 |
| Interest in topic | | | -0.45*** | | | -0.12 |
| Agreement | | | 0.63*** | | | 0.88** |
| Endorsement | | | 0.34 | | | -0.82** |
| Agreement x Endorsement | | | -0.47* | | | 0.13 |
| Agreement x Interest | | | 0.07 | | | -0.13 |
| Endorsement x Interest | | | -0.47*** | | | 0.19 |
| Agree. x Endors. x Interest | | | 0.07 | | | 0.05 |
| <i>-2LogLikelihood</i> | 25874 | 25626 | 25586 | 16437 | 16245 | 16208 |
| $\chi^2(\Delta -2LogLikelihood)$ | | 248*** | 40*** | | 191*** | 37*** |

Note: All variables are centered around population means. Thus, regression weights illustrate effects of a predictor when all other predictors display their respective means. Dichotomous variables employ sum contrasts, their regression weights therefore indicate the difference between levels, with -1 indicating low and +1 indicating high agreement/endorsement.

*** $p < .001$; ** $p < .01$; * $p < .05$; $N(\text{Experiment 1}) = 331$; $N(\text{Experiment 2}) = 207$

Effects on False Consensus. The left side of table 1 displays the results of the linear mixed effects modelling for Experiment 1, with model 0 as random-intercept-only reference model, model 1 including only control variables, and model 2 as full model

with all predictors. Fit of the full model was significantly better than fit of model 1 ($\chi^2(14) = 40, p < .001$), therefore results of the full model are stated.



Figures 2a and 2b. Strength of the False Consensus Effect in Experiments 1 and 2 as a function of the factors *agreement* and *endorsement*, illustrated as regression lines.

Note: Values on the vertical axis are values predicted by the regression model, not observed values).

Firstly, the model confirmed a general FCE by finding *estimated public opinion* positively correlated with subjects' own *attitude* ($\beta = 3.95, t(2909) = 10.44, p < .001$). Figure 2a visualizes FCE by plotting this correlation as regression lines for all experimental conditions.

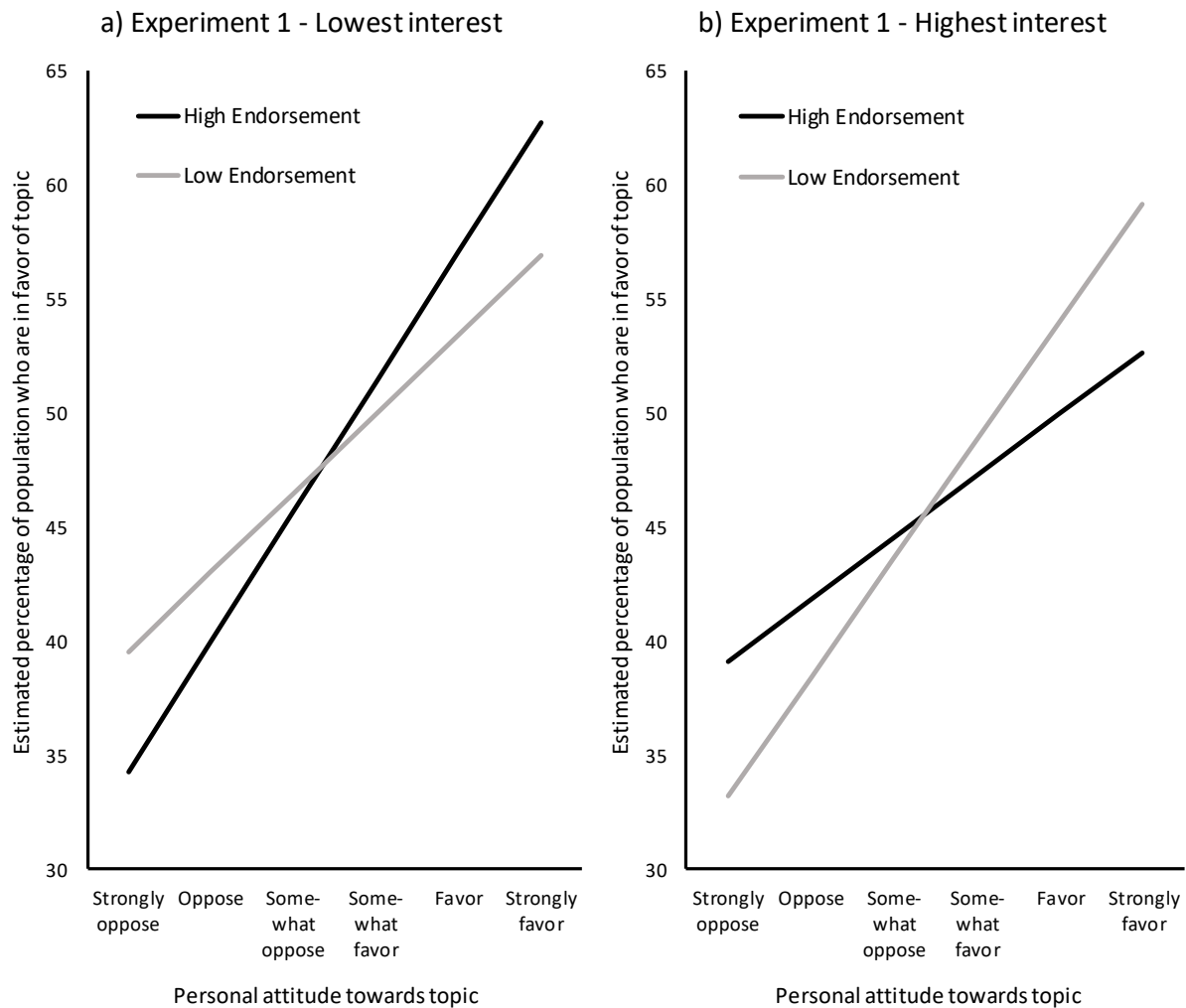
The positive correlation between own attitude and estimated public opinion, that is, the FCE, was stronger when subjects were exposed to mostly congruent messages (*high*

agreement) than when exposed to mostly incongruent ones (*low agreement*, $\beta = 0.63$, $t(2865) = 2.94$, $p = .004$). In figure 2a, this is indicated by the solid lines, representing *high agreement*, being steeper than the dashed lines that represent *low agreement*.

The extent of this effect was moderated by *endorsement* ($\beta = -0.47$, $t(2865) = -2.16$, $p = .031$): If congruent messages displayed *low endorsement*, the effect of *agreement* was clearly evident, and subjects who saw mostly congruent messages (*high agreement*) displayed higher FCE than subjects who saw mostly incongruent messages (*low agreement*). However, if congruent messages displayed *high endorsement*, the FCE was not affected by the number of congruent messages, that is, the factor *agreement* (see the difference in steepness between solid and dashed lines in case of *high* and *low agreement*, represented by black and grey lines, respectively, in fig. 2a).

Additionally, an interactive effect of *endorsement* and subjects' *interest* in a topic was found ($\beta = -0.47$, $t(2908) = -2.93$, $p = .003$). To understand this interactive effect, the differences in FCE between the two *endorsement* conditions when subjects' *interest* was lowest vs. highest are visualized in figures 3a and 3b, respectively: If subjects had *low interest* in a topic, *high endorsement*, that is, higher numbers of "likes" for congruent than incongruent messages, led to a higher FCE than *low endorsement*. However, if they showed *high interest* in a topic, this effect reversed, and *high endorsement* led to a weaker FCE than *low endorsement*. In contrast to the interactive effect of *endorsement* and subjects' *interest* on FCE, there was no interactive effect of *agreement* and *interest*.

Message Selection. Regarding message selection, subjects displayed a novelty or oddity effect (Berlyne & Ditkofksy, 1976) by preferentially clicking on the one message that voices a deviant opinion ($F(1;327) = 370.91$, $p < .001$).



Figures 3a and 3b. Strength of the False Consensus Effect in Experiment 1 as a function of the factor endorsement in case of lowest (3a) and highest (3b) interest values.

Note: The factor agreement did not enter into an interactive effect with interest and is therefore not shown (i.e., graphs reflect the effects of interest and endorsement). Values on the vertical axis are values predicted by the regression model, not observed values.

Discussion

Interpretation of results. The experiment could mostly confirm the expected effects of agreement expressed in news feed messages and endorsement by “likes” on subjects’ perceived public support for their own opinions. Moreover, it sheds additional light upon the interplay of message contents, numbers of “likes”, and subjects’ interest in a topic.

It was assumed that exposure to news feeds with mostly attitudinally congruent messages (*high agreement*) would lead subjects to estimate a higher percentage of the population to share their views than exposure to mostly incongruent messages (*low agreement*). Indeed, subjects displayed a stronger FCE in the *high agreement* condition. This confirms that subjects' estimate of public opinion is influenced by the level of agreement they encounter in online messages.

However, the factor *agreement* played a major role mostly when the congruent messages had less "likes" than the incongruent ones (condition *low endorsement*). In contrast, when congruent messages displayed *high endorsement*, the factor *agreement*, that is, the number of congruent messages, had no effect on FCE.

Two explanations for this interaction appear conceivable: When attitudinally congruent messages consistently display high numbers of "likes", subjects might get skeptical of the validity of the numbers and suspect an attempt of manipulation. This might lead them to be more critical towards the overall news feed and the messages it contains. Thus, the effect of the number of congruent messages is reduced or disappears.

The results, however, could also be explained as a result of confirmation bias, that is, subjects' need to experience agreement by others and at the same time avoid conflicting information: In the condition *high endorsement*, the number of "likes" of the first displayed message always indicated high endorsement of subjects' own attitude, either by stating high endorsement for the subjects' own point of view or by showing low endorsement for the opposing point of view. For the subjects, this might have been a pleasant realization, which they were not motivated to question later on, even when they encountered a majority of disagreeing messages (*low agreement*) in the remaining news

feed. Instead, the initially perceived endorsement might have served as anchor for their estimate of public support, independent of *agreement*.

In contrast, in the condition *low endorsement*, the first message always expressed low endorsement of subjects' own attitude: The message was either congruent and displayed a low number of "likes", or it was incongruent and displayed a high number of "likes". It can be assumed that this perception conflicted with subjects' motivation to experience approval by others, thus leading to cognitive dissonance. In order to reduce this dissonance, subjects might have looked for additional information that allowed them to discard the initial impression of disapproval. Possibly, this additional information was provided by the level of *agreement*: If most of the messages were congruent, subjects could focus on this *high agreement* as indicator for public support and disregard the *low numerical endorsement*. However, if both *agreement* and *endorsement* were *low*, subjects appeared to give in and lower their estimate of public support.

While there was no main effect of *endorsement* on FCE per se, the experiment found a second interactive effect, involving *endorsement* and subjects' *interest* in a topic: Based on Elaboration Likelihood Theory (Petty, 1986), "likes" were conceived as peripheral cues which should have a stronger effect when involvement/interest was low. Thus, in case of *low interest* in a topic, subjects were expected to take numbers of "likes" as indicators of public opinion, resulting in a positive effect of *endorsement* on FCE. This effect was confirmed in the experiment. In case of *high interest* in a topic, however, subjects were expected to pay less attention to "likes". Thus, the effect of *endorsement* should turn out weaker or disappear altogether. In fact, the experiment found that the effect did not merely disappear but actually even reversed: When *interest* in a topic was

high, high endorsement of congruent messages by means of many “likes” led to a weaker FCE than *low endorsement*.

These findings are still in accordance with the Elaboration Likelihood explanation: As described, in case of *low interest*, subjects appear to accept numbers of “likes” as valid indicators of public opinion and put little effort in processing of messages. However, when they have *high interest* in a topic, they elaborate the messages and the numbers of “likes” more thoroughly and more critically. They appear not to take the validity of endorsement via “likes” for granted but question the objectivity and representativeness of “likes”. Moreover, subjects might even suspect that the numbers are intentionally manipulated in order to influence them, for example by social bots or so-called click-farms (De Cristofaro, Friedman, Jourjon, Kaafar, & Shafiq, 2014; Ikram et al., 2017) which popular media occasionally report on. It is known that subjects who suspect a persuasive intent can react with resistance and reactance (van Noort, Antheunis, & van Reijmersdal, 2012). The fact that *high endorsement* led to a lower FCE for *high interest* topics could have resulted from such reactance: Subjects might have assumed that the high numbers of “likes” were intended to manipulate them. Therefore, they adjusted their estimate of public opinion downwards significantly, in order not to fall victim to this supposed manipulative attempt. Of course, this explanation is post-hoc and therefore speculative. Further research is necessary to test whether the assumed processes are indeed responsible for the observed effect and to examine why only the effect of *endorsement* but not of *agreement* was moderated by subjects’ *interest*.

Methodological considerations. Overall, the experiment confirmed the assumption that *agreement* and *endorsement* encountered in social media news feeds influence subjects’ perception of public opinion. However, a replication experiment

appears necessary to test the robustness and generalizability of the found effects. To achieve this, several aspects of the paradigm used to induce an Echo Chamber should be reevaluated:

Firstly, the paradigm implemented only one possible kind of online Echo Chamber, that is, one in which there is a majority view on the topic and one message with a deviant opinion. However, it is well known that exposure to a mostly homogenous group in which one member states a deviant opinion has less impact on subjects' judgements than exposure to a completely unanimous group (Asch, 1961). A similar effect might have occurred in the present experiment, where the Echo Chamber always included one deviant opinion. Therefore, it appears worthwhile to examine the effects on FCE when there is unanimity in the Echo Chamber, that is, no message pursuing a deviant point of view.

Secondly, the experiment used a forced-choice paradigm for news feed exposure that required subjects to explicitly click on only one of four messages. While this was based on existing paradigms, it can be criticized for having low ecological validity (Knobloch-Westerwick, 2014): When browsing real social media news feeds, users are not forced to follow only one link, but might open several links one after another. Thus, it cannot be ruled out that the current experiments' specific task might have led subjects to process and evaluate the messages and popularity cues differently than in a more naturalistic setting. A replication experiment could avoid this by instead letting subjects indicate for each message how likely they are to click upon it.

Finally, the experiment used numbers of "likes" based on actual numbers of Facebook posts of large US-American news outlets. The subjects, however, were German and can be expected to read mostly messages of German news outlets in their everyday

lives. Due to their language and scope, these messages can be expected to have a smaller target audience and less “followers”, resulting in lower numbers of “likes”. Thus, the high numbers of “likes” in the experiment could have appeared exaggerated to the subjects. This might have contributed to the assumed perception of a persuasive intent. Therefore, a follow-up experiment should use more realistic and ecologically valid numbers of “likes”.

These issues were addressed in Experiment 2.

Experiment 2

Experiment 2 aimed at replicating the findings of Experiment 1 regarding effects of agreement and endorsement on FCE³, thereby testing the robustness of the findings under changed modes of news feed presentation. The paradigm was altered as follows: Subjects were again exposed to messages that were either congruent or incongruent to their own attitudes, and messages again displayed either high or low *endorsement*. However, subjects were exposed to only one message per screen and were required to indicate how likely they would click on (and read) this message if it appeared in their own social media news feed (see figure 4). Four messages per topic were shown in sequence. In contrast to Experiment 1, no attitudinally deviant message was included: In the *agreement* condition, all messages were in accordance with subjects’ point of view, in the *disagreement*

³ Experiment 2 examined participants’ willingness to share attitudinally congruent messages with their social network contacts as an additional dependent variable. However, no effects of experimentally manipulated factors were found. The measurement took place at the end of the experiment and did not interfere with the other measurements. As the findings are not informative with respect to the central purpose of Experiment 2, they are not discussed in further detail.

condition all messages disagreed with subjects' opinion. *Endorsement* was again manipulated via displayed number of “likes”: In case of *low endorsement*, congruent messages had low or incongruent high numbers of “likes”, and vice versa for *high endorsement*. As all messages were either congruent or incongruent, subjects always saw either high or low numbers of likes.

Because of the omission of a deviant message with differing numbers of “likes”, subjects cannot judge whether the numbers of “likes” they see are high or low (in Experiment 1 there were always messages with high and low numbers in one feed, allowing for comparison). They first need to form an anchor (Furnham & Boo, 2011; Kahneman, 1992) to which they can compare the later seen numbers. Such an anchor was provided in the beginning of Experiment 2 by presenting four irrelevant messages (topic “housing costs”), two of them displaying high, two low numbers of “likes”.

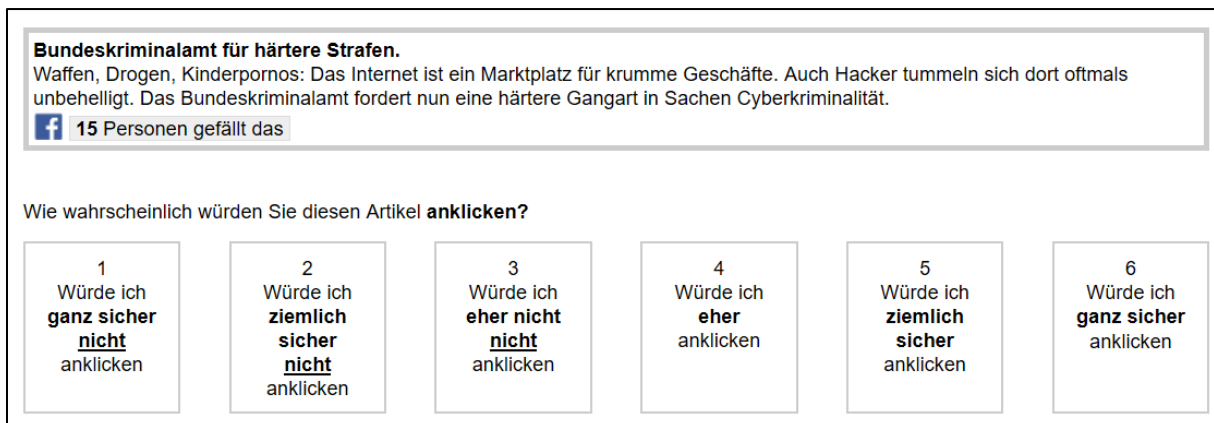


Figure 4. Example of a message presented in Experiment 2.

Method

Sample. A sample of 236 subjects was recruited via student social media groups and on University of Passau campus. After data cleaning similar to Experiment 1 and exclusion of subjects who already took part in Experiment 1, 207 valid cases remained

(146 female; age between 18 and 35 years with $M = 21.69$ and $SD = 2.75$). Comparable to Experiment 1, the criterion of 195 cases for a power of $1 - \beta = .95$, $f^2 = 0.15$ and $\alpha = .05$ in the regression model was met.

Materials. Topics and news feed texts were identical to Experiment 1. The numbers of “likes” to express *high* or *low endorsement* were based on actual numbers that were observed for Facebook messages of several major German news outlets (e.g. “Der Spiegel”, “Tagesschau”) tracked during a two-day period. Observed values mostly ranged between 10 and 1000. Therefore, *low endorsement* was expressed by 10 to 50 “likes” and *high endorsement* by 300 to 999 “likes”.

The four irrelevant messages displayed before news feed exposure to set an anchor for the numbers of “likes” were taken from German news articles. Their topic was “housing prices”, which is unrelated to the other topics used in the experiment. Two of them displayed *low endorsement* numbers, two *high endorsement* numbers.

Measures. Subjects’ own attitudes and interest regarding the topics as well as their estimate of public opinion were measured as in Experiment 1. Message selection was operationalized as likelihood to click on each message as stated by subjects on a 6-point Likert scale (“Would certainly click” to “Would certainly not click”).

Design. As in Experiment 1, a 2 x 2 between-subjects design with factors *agreement* and *endorsement* was used: Subjects either saw only congruent (*agreement*) or only incongruent messages (*disagreement*), and *endorsement* was either *high* (congruent messages displaying high or incongruent low numbers of “likes”) or *low* (incongruent messages displaying low or congruent high numbers of “likes”). FCE was again the dependent variable, calculated based on estimated public opinion.

Procedure. As in Experiment 1, subjects first were informed about voluntariness of participation, gave consent and then stated own attitudes and interest in topics. Afterwards, the four neutral messages were displayed, two with high, two with low numbers of likes. Then, subjects were exposed to the biased news feed, with four messages for each of the nine topics displayed sequentially, one per screen. The presentation orders of topics and of messages for each topic were randomized, but all messages for one topic were displayed blocked. After exposure, subjects gave their estimates of public opinion, then answered demographic questions and were debriefed.

Analysis. Existence of a general FCE was examined by correlating own attitude and estimated public opinion as in Experiment 1. Similarly, the effects of experimental conditions on FCE were analyzed with the same model used in Experiment 1. Also, an exploratory evaluation of subjects' message selection was conducted. As message selection was now indicated as likelihood to click on an article on a 6-point Likert-scale, a linear mixed effect regression model identical to that used for FCE analysis was employed to evaluate effects on message selection.

Results

Overall False Consensus. As in Experiment 1, an overall FCE was found for all topics, with own *attitude* positively correlated with estimate of public opinion for all topics. Correlations ranged from .15 (topic "EU integration") to .50 (topic "Animal testing").

Effects on False Consensus. The effects of *agreement* and *endorsement* on FCE are illustrated on the right side of table 1 and in figure 2b. Results are similar to Experiment 1 but do not reveal the same complex interactive structure: The overall FCE

was again reflected by the effect of own *attitude* on *estimated public opinion* ($\beta = 4.24$, $t(1819) = 12.39$, $p < .001$). Also comparable to Experiment 1, FCE turned out stronger in the *agreement* condition, that is, when only congruent messages were shown than when subjects only saw incongruent messages (i.e., *disagreement* condition, $\beta = 0.88$, $t(1787) = 3.08$, $p = .002$). *High endorsement*, however, led to a weaker FCE compared to *low endorsement* ($\beta = -0.82$, $t(1783) = -2.83$, $p = .005$), independent of *interest* in topic and other factors. No further effects were significant.

Message Selection. Subjects indicated a higher likelihood to click on messages when the opinion voiced in the message was congruent to their own ($\beta = 0.19$, $t(203) = 4.30$, $p < .001$), when the message had a high number of likes ($\beta = 0.10$, $t(203) = 2.26$, $p = .025$), and when their *interest* in the messages topic was high ($\beta = 0.04$, $t(1614) = 2.48$, $p = .013$).

Discussion

Experiment 2 successfully replicated the effect of *agreement* on subjects' FCE: Similar to Experiment 1, if subjects saw congruent messages, they displayed higher FCE than when exposed to incongruent messages. The size of this effect was independent of *endorsement*, that is, the numbers of "likes" displayed, while in Experiment 1 it was stronger when "likes" expressed *low endorsement* for the subjects' attitude.

The effect of *endorsement* on FCE, however, could only be partially replicated: In Experiment 1, *endorsement* interacted with subjects' *interest* in a topic, with *high endorsement* increasing FCE when *interest* was low, but decreasing FCE when *interest* was high. In Experiment 2, subjects' FCE was weaker in the *high endorsement* condition than in the *low endorsement* condition, independent of *interest* in topic.

Thus, in Experiment 2, the assumption of higher endorsement leading to higher likes has to be discarded, and instead, an inversed effect occurred. An explanation for the reversal of the effect could be that in the second experiment, subjects might have always been skeptical of “likes” and suspected a persuasive, manipulative intention. The reason for this more critical view of *endorsement* might be the unanimity displayed by the numbers of “likes”: In Experiment 1, there was always one deviant message that displayed a number of “likes” strongly different from the other numbers. For subjects with *low interest* in a topic, this observable variance in numbers of “likes” might have been enough to accept them as unbiased and as a valid indicator of popular opinion, while highly involved subjects perceived them as biased. In Experiment 2, “likes” were continuously high or low in comparison to the initially displayed anchor values. Therefore, even subjects with *low interest* might have realized that the numbers show little variance and are biased, leading them to also react with skepticism and lower their estimates of public support, thus weakening FCE. This possible explanation is also in accordance with prior findings (Haim et al., 2018) that indicate, that the interpretation of popularity cues such as “likes” strongly depends on both context of presentation and subjects’ experiences with popularity cues.

General Discussion

Evidence for Effects of Online Echo Chambers

The experiments reported in this paper aimed at examining whether so-called online Echo Chambers can influence behavior-related outcomes on the level of the individual. This was tested by exposing subjects to biased news feeds and by experimentally manipulating the two factors *agreement* and *endorsement*: Most/all messages in the feed

were either congruent (*high agreement*) or incongruent (*low agreement*) to subjects' own attitudes. Higher numbers of "likes" were displayed for congruent (*high endorsement*) or incongruent messages (*low endorsement*). As behavior-related outcome subjects' False Consensus effect (FCE) was measured, that is, how strongly they overestimated public support for their own opinions.

Both experiments confirm that, indeed, exposure to biased online news feeds leads to a more favorably distorted perception of public opinions: When subjects encounter *high agreement*, that is, (mostly) attitudinally congruent messages in their news feed, they will display a stronger FCE. This occurs independent of subjects' interest for a topic and regardless of mode of presentation and whether there is one message with a deviant opinion or complete unanimity. Such an effect was expected, given previous findings regarding links between availability of attitudinally (in)congruent information and FCE (Bauman & Geher, 2002; Wojcieszak, 2008). However, the current experiments verify for the first time that this effect also occurs when information comes in the shape of very short social media news messages. They demonstrate that subjects' perception of public opinion can get distorted by exposure to a biased news feed, even if no further interaction (e.g. clicking on links and reading longer texts) occurs. Moreover, the effect of *agreement* suggests that subjects tend to accept information expressed in message contents as basis for their estimate of public opinion, and that they are less suspicious of biases in texts than of biases in numbers of "likes".

These findings highlight that warnings of negative effects of online Echo Chambers are at least somewhat substantiated: Users of online social networks often tend to surround themselves preferably with like-minded others (Del Vicario et al., 2016; Quattrociocchi et al., 2016; Schmidt et al., 2017; Zollo et al., 2017) and thus receive

mostly messages that agree with their own attitudes. The current study suggests that they will most likely conclude from this selective exposure to attitudinally congruent messages to high public support for their own points of view. This perception of public support and favorably distorted social norm might, in turn, make them more willing to act upon their attitudes, giving rise to possibly harmful consequences. For example, subjects who hold racist or otherwise hostile attitudes might conclude from the approval by their biased social network that large parts of the population agree with them, and they might therefore be more likely to put their beliefs into (violent) action.

However, the present study also allows for more optimistic conclusions. While subjects appear not to suspect biases in *agreement* expressed in message texts, they appear critical of *endorsement* indicated by the numbers of “likes”: They do not universally interpret high numbers of “likes” for messages congruent to their own attitudes as valid evidence for the public agreeing with them, especially if their interest in a topic is high. Instead, they lower their estimate of public agreement. Thus, subjects appear to be wary of biases in numbers of “likes” and should be somewhat resistant towards attempts to influence their perception of public opinion via manipulated numbers.

Strategies to Counter Negative Effects of Echo Chambers

Based on the present findings, some strategies to counter negative effects of Echo Chambers can be discussed. The central conclusion is simply that exposure to Echo Chambers, that is, to very one-sided and biased online information and communication, should be avoided if one wants to maintain a realistic estimate of public opinion. This might, however, not be easy, as it means tackling basic human tendencies such as conformation bias, homophily and reduction of cognitive dissonance.

Some researcher have suggested technological means, for example by designing browser add-ons that inform users about biases in their news consumption (Munson, Lee, & Resnick, 2013) or by visualizing biases in the composition of users' social network (Nagulendra & Vassileva, 2014). However, it can be assumed that this will only work if subjects are willing to reduce their bias by using and trusting in such technologies. If they do not perceive a problem in their bias, they will not use technologies that promise reducing it. Moreover, empirical evidence regarding the actual effects of such technological solutions is needed. For example, it could be experimentally tested whether subjects who are informed about biases in their online network via an indicator actually start to include more diverse contacts, or if such warnings are ignored. Technological solutions should only be suggested if such tests turn out positive.

Instead of focusing on technological aides, interventions could also try to strengthen individuals' traits that alleviate negative effects of Echo Chambers: Experiment 1 demonstrated that interest in a topic can make subjects more critical towards biased numbers of "likes". Furthermore, a previous correlative study (AUTHORS, 2019) found that higher ambiguity tolerance, that is, more openness towards ambiguous and contradicting information, leads to weaker FCE. Both factors – interest/involvement and ambiguity tolerance – thus appear to increase subjects' resilience towards negative effects of Echo Chambers. Strengthening this resilience by supporting citizens' involvement with societal topics and encouraging open-mindedness and media literacy is, of course, a task for society as a whole and its (educational) institutions.

Methodological Considerations

Some limitations of both reported experiments must be addressed:

Firstly, it is unclear how stable the effects of Echo Chamber induction on FCE are, as the latter was only measured directly after news feed exposure. A long-term effect appears as highly plausible, given that in real life individuals are not exposed to news feeds once but repeatedly, whenever they visit a social networking platform. To test this empirically, a repeated measurement design would be necessary.

Secondly, due to the experimental approach, the current study focused on only the two manipulated message attributes *agreement* and *endorsement*. Real social media messages, however, display additional attributes that might influence subjects, such as images, sender names, and avatars. Their effects should be addressed in future experiments. Special attention could be paid to other types of popularity cues (e.g. how often messages were shared or commented on) and whether these display similar effects as “likes”. Moreover, in addition to subjects’ *interest*, future studies could also address further individual-level variables. One variable could be subjects’ topic-related knowledge, which has been argued to influence how strongly users of websites rely on popularity cues (Porten-Cheé et al., 2018).

Thirdly, the current study used messages based on news articles as stimuli. This was done as FCE was measured towards topics reported in news and because news play an important role in social media (Bastos, 2015; Swart, Peters, & Broersma, 2018). However, interaction in networks like Facebook is mostly made up of personal messages shared between individuals. Thus, it should be tested whether the current findings can be replicated if instead of news texts, more subjectively phrased personal messages are used.

Finally, both experiments compared conditions in which the majority of messages either voiced opinions congruent or incongruent to subjects' opinions. No balanced condition with an equal number of congruent and incongruent messages was included. This decision was made in order to maximize power of the experimental manipulation. However, future studies might include such a neutral condition to establish a baseline of FCE values for their respective samples.

Despite these limitations, the present study offers valuable insights into the effects of online Echo Chambers on the individual level, that is, individuals' actual perceptions and behavior. Furthermore, the study illustrates that experimental approaches that modify existing paradigms can be successfully used to investigate Echo Chambers and uncover their causal effects on outcomes like FCE.

References

- Ajzen, I., & Fishbein, M. (1977). Attitude-Behavior Relations: A Theoretical Analysis and Review of Empirical Research. *Psychological Bulletin*, *5*(84), 888–918.
- Allcott, H., & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives*, *31*(2), 211–236.
<https://doi.org/10.1257/jep.31.2.211>
- Asch, S. E. (1961). Effects of group pressure upon the modification and distortion of judgments. In M. Henle (Ed.), *Documents of Gestalt Psychology* (pp. 222–236). Berkeley, Los Angeles: University of California Press.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>
- Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., & Bonneau, R. (2015). Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological Science*, *26*(10), 1531–1542.
- Bastos, M. T. (2015). Shares, Pins, and Tweets: News readership from daily papers to social media. *Journalism Studies*, *16*(3), 305–325.
<https://doi.org/10.1080/1461670X.2014.891857>
- Bauman, K. P., & Geher, G. (2002). We Think You Agree: The Detrimental Impact of the False Consensus Effect on Behavior. *Current Psychology*, *21*(4), 293–318.
- Berlyne, D. E., & Ditkofksy, J. (1976). Effects of novelty and oddity on visual selective attention. *British Journal of Psychology*, *67*(2), 175–180.

- Chang, Y.-T., Yu, H., & Lu, H.-P. (2015). Persuasive messages, popularity cohesion, and message diffusion in social media marketing. *Journal of Business Research*, 68(4), 777–782. <https://doi.org/10.1016/j.jbusres.2014.11.027>
- Cotton, J. L., & Hieser, R. A. (1980). Selective exposure to information and cognitive dissonance. *Journal of Research in Personality*, 14(4), 518–527. [https://doi.org/10.1016/0092-6566\(80\)90009-4](https://doi.org/10.1016/0092-6566(80)90009-4)
- De Cristofaro, E., Friedman, A., Jourjon, G., Kaafar, M. A., & Shafiq, M. Z. (2014). Paying for Likes? Understanding Facebook Like Fraud Using Honeypots. *ArXiv:1409.2097 [Physics]*. Retrieved from <http://arxiv.org/abs/1409.2097>
- de la Haye, A.-M. (2000). A methodological note about the measurement of the false-consensus effect. *European Journal of Social Psychology*, 30, 569–581.
- Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., ... Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, 113(3), 554–559. <https://doi.org/10.1073/pnas.1517441113>
- DiResta, R., Shaffer, K., Ruppel, B., Sullivan, D., Matney, R., Fox, R., ... Johnson, B. (2018). *The Tactics & Tropes of the Internet Research Agency*. Retrieved from <https://www.newknowledge.com/articles/the-disinformation-report/>
- Dvir-Gvirsman, S. (2019). I like what I see: studying the influence of popularity cues on attention allocation and news selection. *Information, Communication & Society*, 22(2), 286–305. <https://doi.org/10.1080/1369118X.2017.1379550>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical

sciences. *Behavior Research Methods*, 39(2), 175–191.

<https://doi.org/10.3758/BF03193146>

Fishbein, M. (1976). A Behavior Theory Approach to the Relations between Beliefs about an Object and the Attitude Toward the Object. In U. H. Funke (Ed.), *Mathematical Models in Marketing: A Collection of Abstracts* (pp. 87–88).

https://doi.org/10.1007/978-3-642-51565-1_25

Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The Journal of Socio-Economics*, 40(1), 35–42.

<https://doi.org/10.1016/j.socec.2010.10.008>

Galesic, M., Olsson, H., & Rieskamp, J. (2013). False consensus about false consensus. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 472–476).

Retrieved from

<http://csjarchive.cogsci.rpi.edu/Proceedings/2013/papers/0109/paper0109.pdf>

Gilbert, E., Bergstrom, T., & Karahalios, K. (2009). Blogs Are Echo Chambers: Blogs Are Echo Chambers. *42nd Hawaii International Conference on System Sciences, 2009. HICSS'09*. Retrieved from

<http://ieeexplore.ieee.org/abstract/document/4755503/>

Grömping, M. (2014). ‘Echo Chambers’: Partisan Facebook Groups during the 2014 Thai Election. *Asia Pacific Media Educator*, 24(1), 39–59.

<https://doi.org/10.1177/1326365X14539185>

Guest, E. (2018). (Anti-)Echo Chamber Participation: Examining Contributor Activity Beyond the Chamber. *Proceedings of the 9th International Conference on Social*

Media and Society - SMSociety '18, 301–304.

<https://doi.org/10.1145/3217804.3217933>

Haim, M., Kümpel, A. S., & Brosius, H.-B. (2018). Popularity cues in online media: A review of conceptualizations, operationalizations, and general effects. *Studies in Communication | Media*, 7(2), 186–207. <https://doi.org/10.5771/2192-4007-2018-2-58>

Hart, W., Albarracín, D., Eagly, A. H., Brechan, I., Lindberg, M. J., & Merrill, L. (2009). Feeling validated versus being correct: A meta-analysis of selective exposure to information. *Psychological Bulletin*, 135(4), 555–588. <https://doi.org/10.1037/a0015701>

Howard, P. N., Ganesh, B., Liotsiou, D., Kelly, J., & François, C. (2018). *The IRA, Social Media and Political Polarization in the United States, 2012-2018* (p. 47).

Ikram, M., Onwuzurike, L., Farooqi, S., De Cristofaro, E., Friedman, A., Jourjon, G., ... Shafiq, M. Z. (2017). Measuring, Characterizing, and Detecting Facebook Like Farms. *ArXiv:1707.00190 [Cs]*. Retrieved from <http://arxiv.org/abs/1707.00190>

Kahneman, D. (1992). Reference points, anchors, norms, and mixed feelings. *Organizational Behavior and Human Decision Processes*, 51(2), 296–312. [https://doi.org/10.1016/0749-5978\(92\)90015-Y](https://doi.org/10.1016/0749-5978(92)90015-Y)

Knobloch-Westerwick, S. (2014). *Choice and Preference in Media Use : Advances in Selective Exposure Theory and Research*. New York, London: Routledge.

Knobloch-Westerwick, S. (2015). The Selective Exposure Self- and Affect-Management (SESAM) Model: Applications in the Realms of Race, Politics, and Health. *Communication Research*, 42(7), 959–985. <https://doi.org/10.1177/0093650214539173>

- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13).
<https://doi.org/10.18637/jss.v082.i13>
- Lee, E.-J., & Yoon Jae Jang. (2010). What Do Others' Reactions to News on Internet Portal Sites Tell Us? Effects of Presentation Format and Readers' Need for Cognition on Reality Perception. *Communication Research*, 37(6), 825–846.
<https://doi.org/10.1177/0093650210376189>
- Liska, A. E. (1984). A Critical Examination of the Causal Structure of the Fishbein/Ajzen Attitude-Behavior Model. *Social Psychology Quarterly*, 47(1), 61. <https://doi.org/10.2307/3033889>
- 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.
- Messing, S., & Westwood, S. J. (2014). Selective exposure in the age of social media: Endorsements trump partisan source affiliation when selecting news online. *Communication Research*, 41(8), 1042–1063.
- Munson, S. A., Lee, S. Y., & Resnick, P. (2013). Encouraging Reading of Diverse Political Viewpoints with a Browser Widget. *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media*, 419–428.
Retrieved from <https://www.aaai.org/Library/ICWSM/icwsm13contents.php>
- Nagulendra, S., & Vassileva, J. (2014). Understanding and controlling the filter bubble through interactive visualization: a user study. *Proceedings of the 25th ACM Conference on Hypertext and Social Media - HT '14*, 107–115.
<https://doi.org/10.1145/2631775.2631811>

- Pariser, E. (2011). *The Filter Bubble: What The Internet Is Hiding From You*. London: Penguin.
- Peter, C., Rossmann, C., & Keyling, T. (2014). Exemplification 20: Roles of direct and indirect social information in conveying health messages through social network sites. *Journal of Media Psychology: Theories, Methods, and Applications*, 26(1), 19–28. <https://doi.org/10.1027/1864-1105/a000103>
- Petty, R. E. (1986). The Elaboration Likelihood Model of Persuasion. *Advances in Experimental Social Psychology*, 19, 123–205.
- Porten-Che e, P., Ha ler, J., Jost, P., Eilders, C., & Maurer, M. (2018). Popularity cues in online media: Theoretical and methodological perspectives. *Studies in Communication | Media*, 7(2), 208–230. <https://doi.org/10.5771/2192-4007-2018-2-80>
- Quattrociocchi, W., Scala, A., & Sunstein, C. R. (2016). *Echo chambers on facebook*. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2795110
- Ross, L., Greene, D., & House, P. (1977). The “false consensus effect”: An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology*, 13(3), 279–301.
- Schmidt, A. L., Zollo, F., Del Vicario, M., Bessi, A., Scala, A., Caldarelli, G., ... Quattrociocchi, W. (2017). Anatomy of news consumption on Facebook. *Proceedings of the National Academy of Sciences*, 114(12), 3035–3039. <https://doi.org/10.1073/pnas.1617052114>
- Sears, D. O., & Freedman, J. L. (1967). Selective Exposure to Information: A Critical Review. *Public Opinion Quarterly*, 31(2), 194. <https://doi.org/10.1086/267513>

- Snijders, T. A. B. (2005). Power and sample size in multilevel modeling. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of Statistics in Behavioral Science* (pp. 1570–1573). Chichester etc.: Wiley.
- Sunstein, C. R. (2001). *Echo chambers: Bush v. Gore, impeachment, and beyond*. Princeton, Oxford: Princeton University Press.
- Swart, J., Peters, C., & Broersma, M. (2018). Sharing and Discussing News in Private Social Media Groups: The social function of news and current affairs in location-based, work-oriented and leisure-focused communities. *Digital Journalism*, 1–19. <https://doi.org/10.1080/21670811.2018.1465351>
- Tandoc, E. C., Lim, Z. W., & Ling, R. (2018). Defining “Fake News”: A typology of scholarly definitions. *Digital Journalism*, 6(2), 137–153. <https://doi.org/10.1080/21670811.2017.1360143>
- van Noort, G., Antheunis, MarjolijnL., & van Reijmersdal, EvaA. (2012). Social connections and the persuasiveness of viral campaigns in social network sites: Persuasive intent as the underlying mechanism. *Journal of Marketing Communications*, 18(1), 39–53. <https://doi.org/10.1080/13527266.2011.620764>
- Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, 20(3), 273–281. <https://doi.org/10.1080/14640746808400161>
- Williams, H. T. P., McMurray, J. R., Kurz, T., & Lambert, H. F. (2015). Network analysis reveals open forums and echo chambers in social media discussions of climate change. *Global Environmental Change*, 32, 126–138. <https://doi.org/10.1016/j.gloenvcha.2015.03.006>

Wojcieszak, M. (2008). False Consensus Goes Online: Impact of Ideologically Homogeneous Groups on False Consensus. *Public Opinion Quarterly*, 72(4), 781–791. <https://doi.org/10.1093/poq/nfn056>

Zollo, F., Bessi, A., Del Vicario, M., Scala, A., Caldarelli, G., Shekhtman, L., ... Quattrociocchi, W. (2017). Debunking in a world of tribes. *PLOS ONE*, 12(7), e0181821. <https://doi.org/10.1371/journal.pone.0181821>

5. Study three: Intergroup Bias in Echo Chamber message texts

Original paper:

Luzsa, R. (submission in preparation). Intergroup Bias in Online Echo Chambers: Evidence from Word Frequency Analysis of Attitudinally Homogeneous and Heterogeneous Reddit Groups.

Please note that the following chapter is the unedited pre-review version of the article and does not reflect possible changes made during the peer-review process.

**Intergroup Bias in Online Echo Chambers: Evidence from Word Frequency
Analysis of Attitudinally Homogeneous and Heterogeneous Reddit Groups**

Keywords: Echo chambers, social networking, intergroup bias, Reddit

Abstract

Communication in attitudinally and ideologically homogeneous online groups – online Echo Chambers (EC) – is subject to controversial public and scientific debate and often associated with increasing societal polarization. However, little is known about how EC affect the way group members communicate, and if their communication displays signs of polarization. Based on Social Identity Theory and findings on intergroup bias, EC users can be assumed to perceive themselves as members of an in-group that is in conflict with negatively perceived out-groups. This should lead to differences in word use in user-contributed texts – so-called submissions – between EC and more heterogeneous, neutral non-EC groups: Submissions in EC should display more references to in- and out-groups, more negative emotion and swearing, and a more affirmative style with more expressions of assent and less tentativeness. To test this, submissions made in three attitudinally homogeneous EC groups and three heterogeneous non-EC groups on the social website Reddit were examined. It was hypothesized that in EC groups, submissions displaying more linguistic signs of intergroup bias receive more endorsement by other users, that is, have a higher “score”. By contrast, no similar relations between word use and score were expected in non-EC groups. Linguistic properties of submissions were extracted with the software LIWC and correlated with submission scores. The overall pattern of correlations was in accordance with the hypotheses and supported the assumption of intergroup bias in online EC. Thus, the results provide empirical evidence for linking EC with polarization.

Introduction

Intergroup bias in online Echo Chambers

The evaluation of the impact of social networking sites (SNS) like Facebook, Twitter, or Reddit on the individual and society as a whole has changed significantly over the last years: Initially, SNS were often seen as a tool for free and uncensored spread of information, and research often highlighted and examined this empowering role, for example regarding democratic change in developing countries¹⁻³.

Recently, a more critical view has gained prominence, highlighting the tendency of SNS users to form attitudinally homogeneous groups – so called “echo chambers” (EC) – in which information is preferably shared with like-minded others while exchange with those holding opposing views is rendered unlikely^{4,5} (e.g. very left or right wing political groups on Facebook). This tendency is sometimes cited as a reason for increased societal polarization, that is, conflict between societal groups, and the rise of radical political positions^{6,7}. However, existing research mostly focuses on whether users indeed show a tendency to connect or share information with like-minded others^{8,9}. Whether communication in attitudinally homogeneous EC actually displays signs of polarization remains largely unknown.

From a psychological perspective, signs of polarization can be expected: Even random and anonymous assignment to non-informative groups can induce a so-called intergroup bias¹⁰, with subjects’ perceptions of in-group members turning out to be more favorable than those of out-groups members, even when no interaction within or between groups takes place. For example, more positive traits are ascribed to in-group, more

negative to out-group members, and in- and out-group are perceived as homogeneous regarding these traits¹¹.

According to Social Identity Theory^{10,12,13}, this intergroup bias can be explained by subjects' desire to maintain self-esteem, a positive social identity, and a feeling of distinctiveness. To achieve this, they partake in intergroup discrimination by focusing on (real or hypothetical) differences between in- and out-group. Thus, categorization as a group member becomes salient and meaningful to subjects and affects their actual behavior, for example by favoring in-group over out-group members when it comes to the distribution of limited resources¹⁰.

Joining an attitudinally homogeneous EC group on SNS can be seen as an act of self-categorization, which should lead to increased salience of group membership and intergroup bias: If users participate in an EC, they should tend to view the social world in group categories, differentiating others in categories of “we” (in-group) versus “they” (out-group). Moreover, they should try to enhance the in-group identity by devaluating and attacking out-groups while expressing affirmation with in-group members.

This study tests these assumptions by examining whether EC-like attitudinally homogeneous groups in SNS display stronger signs of intergroup bias than attitudinally heterogeneous non-EC groups. Finding such differences could substantiate the assumption that EC enhance polarization between societal groups. Moreover, it could pave the way for the detection of potentially problematic groups in SNS.

Echo Chambers on Reddit

This study examines EC and non-EC groups on the social website Reddit. Reddit is a popular online service that allows users to create and join thematically ordered fora (so-

called sub-reddits) in which self-written texts, links to news articles, images, and videos can be submitted and shared. This structure has made Reddit attractive for research, with studies often focusing on usage patterns in different sub-reddits^{14,15}

It is of high relevance for the present purpose that moderators of a sub-reddit may enact rules that apply to its users. For example, sub-reddits might require submissions to agree with the group's general stance (e.g. being pro-socialist and anti-capitalist in the group *r/lateststagecapitalism*) while banning divergent opinions. Such sub-reddits can be regarded as EC as their rules will lead to attitudinally homogeneous communication patterns. Non-EC groups, on the other hand, explicitly invite open debate and divergent opinions (e.g. *r/changemyview*).

Moreover, users on Reddit usually write anonymously, using pseudonyms. According to the SIDE model of deindividuation^{16,17}, anonymity and the lack of individual, not group-related information about others make the group identity more salient. Therefore, differences between EC and non-EC should be especially prevalent when users post anonymously, and, as a consequence, analyses of Reddit groups appear more promising than analyses on SNS requiring use of real names.

Finally, Reddit offers a mechanism that is central for the current study: Users can up- or down-vote submissions to indicate their endorsement, and the resulting scores determine how prominently submissions are displayed on sub-reddits' front pages. High scoring submissions are displayed on the top where they are likely to be seen¹⁸, while low scoring submissions are less visible. Thus, submission scores are popularity cues^{19,20} that indicate (a) to which degree the submissions reflects the preferences and general stance

of the group, and (b) how strong the submissions' impact on the group is (by being seen by either few or many users).

As EC members can be expected to react positively to messages that strengthen their group identity and positive self-perception by displaying intergroup bias, signs of intergroup bias in messages should lead to more endorsement of these messages by users than, for example, messages that voice balanced or even deviant views. Thus, this study assumes that linguistic signs of intergroup bias in messages positively correlate with message scores in EC groups but not in non-EC groups.

Psycholinguistic markers of intergroup bias

To determine a submissions' degree of intergroup bias, this study examines the frequency of several categories of words serving as psycholinguistic markers^{21,22}:

Firstly, subjects' use of pronouns has been related to their focus of attention and the salience of cognitive categories, with first-person singular pronouns (e.g. "I", "me") indicating attention to the self, and plural pronouns (e.g. "we", "us") indicating higher salience of the group identity²³. Additionally, the use of first person plural pronouns can be interpreted as a sign of in-group solidarity, while third personal plural pronouns (e.g. "they", "them") may highlight differences between in- and out-group²⁴. Thus, an intergroup bias should be reflected in subjects' use of pronouns:

Hypothesis 1: In EC groups, authors' scores correlate positively with the frequency of first and third person plural pronouns and negatively with the frequency of first person singular pronouns.

Secondly, studies demonstrate that group members strengthen their positive in-group identity by focusing on negative views and stereotypes of the out-group^{25,26}.

Therefore, in EC sub-reddits, successful submissions should propagate this negative view by using more negative emotion words (e.g. “bad”, “hate”, “ugly”). More swearing (e.g. “damn”) should also be used, partly as expression of intergroup conflict and partly because swearing is a common pro-social strategy to enhance in-group bonding²⁷. In EC, these tendencies should lead to associations between the frequency of negative words and swearing and message scores:

Hypothesis 2: In EC groups, authors’ scores correlate positively with the frequency of negative emotion words and swearing.

To control for the possibility that a generally more emotional style might lead to higher endorsement also in non-EC groups²⁸, positive emotion words are also analyzed, with a weaker effect expected in EC.

Finally, users in EC groups should employ a more affirmative style, with more assent (e.g. “agree”) and less tentativeness (e.g. “maybe”) to express solidarity and identification with the in-group and its point of view:

Hypothesis 3: In EC groups, the frequency of assentive words is positively, the frequency of tentative words negatively correlated with authors’ scores.

In non-EC groups, no such correlations are expected.

Method

Sample

The hypotheses were tested based on six sub-reddits. Three of them were EC groups that explicitly demand submissions to agree with the group’s general stance and ban deviant views: *r/latestagecapitalism*, a left-wing group (excerpt from the community’s

rules: “Defense of capitalism and of the parties or ideologies that support it [...] are strictly prohibited. Likewise, opposition to socialism [...] is strictly prohibited.”); *r/the_donald*, a right-wing group of supporters of American President Donald Trump (“No Cucks or Leftists. This forum is for Trump supporters only.”); and *r/conservative*, a politically conservative group (“We provide a place on Reddit for conservatives, both fiscal and social, to read and discuss political and cultural issues from a distinctly conservative point of view.”).

These groups were compared to three neutral, non-EC groups: *r/neutralpolitics*, a political discussion group that highlights openness and neutrality (“[...] dedicated to evenhanded, empirical discussion of political issues. Based on facts and respectful discussion.”); *r/casualconversation*, a sub-reddit for open-topic discussions (“The friendlier part of reddit. Have a fun conversation about anything that is on your mind”); and *r/changemyview*, an “Anti-Echo Chamber”²⁹ whose authors post their opinions and explicitly invite others to question their views (“A place to post an opinion you accept may be flawed, in an effort to understand other perspectives on the issue.”).

From these groups all message texts of submissions made between January 2017 and August 2018 were acquired via the pushshift API³⁰, a web service archiving all publicly available Reddit submissions. The temporal limitations were necessary as changes in Reddit’s algorithm³¹ in December 2016 make scores before and after that date incomparable, and because a ban of several radical political sub-reddits in September 2018 might have impacted the verbal conduct in the examined groups.

Only original submissions containing text were used, that is, no image-only or empty submissions. To increase representativity and robustness, texts written by authors

with deleted accounts and by the most prolific authors, that is, the 1% of authors with exceptionally many submissions, who were also unordinary in other regards (e.g. a higher median score of 35.7 versus 10.5 in the remaining 99%), were excluded. Only top-level submissions were acquired but not comments made to these submissions by other users.

The final dataset encompassed 146,315 submissions by 78,513 authors, with median message lengths varying between 48 and 279 words (see table 1).

Table 1. Description of examined sub-reddits.

| | <i>Submissions</i> | <i>Authors</i> | <i>Words per message</i> | <i>Score</i> |
|----------------------------|--------------------|----------------|--------------------------|----------------------|
| <i>latestagecapitalism</i> | 2490 | 2189 | 61.0 (23.0 - 148.0) | 8.7 (1.0 - 26.0) |
| <i>the_donald</i> | 72886 | 32212 | 48.0 (22.3 - 103.0) | 19.9 (11.0 - 42.0) |
| <i>conservative</i> | 1971 | 1587 | 82.0 (44.0 - 163.0) | 5.0 (1.0 - 15.0) |
| <i>casualconversation</i> | 52708 | 30730 | 79.0 (46.5 - 135.0) | 5.0 (2.0 - 11.0) |
| <i>changemyview</i> | 15532 | 11223 | 279.0 (192.5 - 423.0) | 7.0 (1.0 - 23.5) |
| <i>neutralpolitics</i> | 728 | 572 | 158.0 (101.2 - 251.0) | 174.0 (36.2 - 451.8) |

Note: Values are based on the data included in the analysis. Words per message and score are median values. Values in brackets indicate the interquartile range between first and third quartile.

Measures

Linguistic properties were extracted with LIWC 2015²² which is often used in psycholinguistic research. LIWC offers a dictionary with 93 psychometrically validated³² categories that represent psychologically meaningful linguistic properties. LIWC counts the occurrence of words from each category and outputs scale values for each linguistic property. For example, if 9% of the words in a text come from the “anger” dictionary (e.g. “hate”), then the text receives a value of 0.09 on the “anger” scale.

To test the hypotheses regarding use of pronouns, emotion words, and affirmative style, ten scale values of message texts were extracted which are illustrated in table 2.

Table 2. Used LIWC scales and example words

| <i>LIWC scale</i> | <i>Example words</i> |
|--------------------------|----------------------|
| <i>Pronouns</i> | |
| 1st Person Singular | I, me, mine |
| 1st Person Plural | we, us, our |
| 3rd Person Plural | they, their, they'd |
| <i>Emotion</i> | |
| Anxiety | worried, fearful |
| Anger | hate, kill, annoyed |
| Sadness | crying, grief, sad |
| Swearing | fuck, damn, shit |
| Positive Emotion | love, nice, sweet |
| <i>Affirmative style</i> | |
| Tentativeness | maybe, perhaps |
| Assent | agree, OK, yes |

Note: Example words are taken from the LIWC 2015 manual³²

Analysis

Data analysis was conducted with GNU R 3.5.3.

Before analysis, all data were aggregated on author level, so that each record reflected the average LIWC scale values of all submissions made by an author and his or her mean score. This was done to reduce situational noise (e.g. a single submission's score may strongly benefit from the time point the submission was published), and to alleviate the fact that single submissions' LIWC scale values are often zero (i.e., when submissions contain none of the scale's words).

Hypotheses were tested via non-parametric Spearman correlation between authors' LIWC scale values and their mean scores because both variables were non-normally distributed. To account for repeated testing, *p*-values were Bonferroni-Holm³³ adjusted.

Table 3. Correlations between authors' LIWC scale values and post scores.

| | <i>latestagecapitalism</i> (n = 2189) | <i>the_donald</i> (n = 32212) | <i>conservative</i> (n = 1587) | <i>casualconversation</i> (n = 30730) | <i>changemyview</i> (n = 11223) | <i>neutralpolitics</i> (n = 572) |
|--------------------------|--|----------------------------------|-----------------------------------|--|------------------------------------|-------------------------------------|
| <i>Pronouns</i> | | | | | | |
| 1st Person Singular | 0.02 | -0.02** | -0.02 | 0.02** | 0.04*** | -0.04 |
| 1st Person Plural | 0.06* | 0.08*** | 0.06 | 0.04*** | 0.00 | 0.05 |
| 3rd Person Plural | 0.10*** | 0.09*** | 0.04 | 0.04*** | 0.02 | -0.06 |
| <i>Emotion</i> | | | | | | |
| Anxiety | 0.09*** | 0.11*** | 0.05 | 0.05*** | 0.04*** | -0.03 |
| Anger | 0.07* | 0.10*** | 0.06 | 0.02*** | 0.02 | 0.07 |
| Sadness | 0.04 | 0.10*** | 0.09*** | 0.05*** | 0.03* | 0.03 |
| Swearing | 0.11*** | 0.12*** | 0.07* | 0.04*** | 0.00 | 0.09 |
| Positive emotion | -0.05 | 0.03*** | -0.03 | 0.10*** | 0.02 | -0.07 |
| <i>Affirmative style</i> | | | | | | |
| Tentativeness | -0.01 | -0.07*** | -0.03 | -0.03*** | 0.02 | -0.07 |
| Assent | -0.02 | 0.08*** | 0.01 | 0.06*** | 0.03* | 0.02 |

Note: Values are Spearman correlation coefficients. Effects of message length are partialled out to control for possible correlations between message length and scores.

* p < 0.05; ** p < 0.01; *** p < 0.001. Significances are Bonferroni-Holm adjusted to reduce alpha-error inflation due to repeated testing. Observations (n) indicate numbers of authors.

Results

An initial exploratory descriptive analysis of baseline word frequencies between sub-reddits revealed several differences, for example most “I”-references in *r/casualconversation* [$Md = 8.6$, $IQR = 6.0 - 10.9$] and fewest in *r/neutralpolitics* [$Md = 0.7$, $IQR = 0.0 - 1.6$]. These, however, most likely reflect different topical foci (e.g. small talk versus fact-based discussion). Therefore, they are not informative regarding the study’s hypotheses and are not further discussed.

Table 3 states the correlations between authors’ LIWC scales and submission scores for each sub-reddit. The frequency of plural pronouns was positively correlated with score magnitude in the EC groups *r/lateststagecapitalism* (first person plural: $r = 0.06$, $p < 0.05$; third person plural: $r = 0.10$, $p < 0.001$) and *r/the_donald* (first person plural: $r = 0.08$, $p < 0.001$; third person plural: $r = 0.09$, $p < 0.001$). First person singular pronouns were weakly negatively correlated with score in *r/the_donald* ($r = -0.02$, $p < 0.01$). In the non-EC group *r/casualconversation*, all pronoun frequencies correlated weakly positively with score ($r = 0.02$ for first person singular, $p < 0.01$; $r = 0.04$ for first and third person plural, $p < 0.001$), and in *r/changemyview*, more first person singular pronouns were associated with higher scores ($r = 0.04$, $p < 0.001$). In *r/neutralpolitics* and *r/conservative*, no correlations were significant.

Differences between EC and non-EC were also found regarding negative emotion words: In the EC groups *r/lateststagecapitalism* and *r/the_donald*, use of swearing and negative emotion words (except sadness in *r/lateststagecapitalism*) was significantly associated with higher scores, with correlations ranging between $r = 0.07$, $p < 0.05$ and $r = 0.12$, $p < 0.001$. In *r/conservative*, similar numeric tendencies occurred, but were only

significant for sadness ($r = 0.09, p < 0.001$) and swearing ($r = 0.07, p < 0.05$). Weaker positive correlations of score with all negative emotion words were found in the non-EC group *r/casualconversation* (from $r = 0.02$ for anger to $r = 0.05$ for anxiety and sadness, all $p < 0.001$) and with anxiety ($r = 0.04, p < 0.001$) and sadness ($r = 0.03, p < 0.05$) in *r/changemyview*. Positive emotion correlated positively with score in *r/casualconversation* ($r = 0.10, p < 0.001$) and numerically weaker in *r/the_donald* ($r = 0.03, p < 0.001$).

Regarding tentativeness and assent, no clear pattern occurred: Less tentative words and more assent in messages were associated with higher scores in *r/the_donald* ($r = -0.07, p < 0.001$; $r = 0.08, p < 0.001$) but also in the non-EC group *r/casualconversation* ($r = -0.03, p < 0.001$; $r = 0.06, p < 0.001$), and other effects were weaker, non-systematic or non-significant.

Discussion

This study aimed to find evidence for intergroup bias in attitudinally homogeneous EC groups on Reddit via analysis of word frequencies in user-submitted texts. It was expected that in EC, authors who use more references to in- and out-group, more negative emotion words and swearing, and a more affirmative style would receive higher scores, that is, more endorsement by other users. No such effect was expected in attitudinally heterogeneous non-EC groups.

The results mostly confirm the assumptions regarding pronouns and negative emotion: In the EC groups *r/the_donald* and *r/lateststagecapitalism*, authors received higher scores when they used more first-person plural and third-person plural pronouns and a more negative style with more anger, anxiety, sadness, and swearing. As expected,

the correlative patterns in non-EC groups were less distinctive and consistent, and significant correlations, e.g. positive effects of negative but also of positive emotions in *r/casualconversation*, might reflect users' general preference for more emotional messages²⁸.

In the EC group *r/conservative*, most effects were non-significant. This might be attributed to its smaller sample size ($n = 572$) but also to its less explicit and vehement way to demand adherence to its ideological preferences (see the group's rules cited in the sample description above). Thus, *r/conservative* could be understood as a less exclusive EC, with less identification of its members with the group and less intergroup bias.

The assumption of a more affirmative style leading to higher scores only in EC groups could not be confirmed. This might be attributed to this study's focus on original submissions, in which authors post new content but usually do not directly reply to other users' statements. Therefore, affirmation expressed in original submissions might have little influence on upvoting behavior, and effects of an affirmative style might more likely be visible regarding other users' comments on submissions.

Thus, while this study focused on original submissions that make up sub-reddits' front pages and thus characterize them, future endeavors could also examine comments made in response to submissions. By this, comments elicited by non-text submissions (e.g. images or videos) could be included, and additional analyses made possible: For example, linguistic markers of intergroup bias could be associated with more complex measures of authors' success like responsiveness³⁴ or "virality" (e.g. number of re-posts or reactions elicited by a comment) or with different user roles³⁵ (e.g. "opinion leaders" that predominantly post new contents vs. "followers" that mostly comment).

Moreover, instead of using LIWC's word frequency scales, further research could try to take into account contexts of word use (e.g. if negative words describe the out- and positive the in-group, or if irony is used). Such approaches could build on tools used in affective computing and sentiment analysis³⁶⁻³⁸, and also avoid the problem that LIWC values of shorter texts are often left-skewed and non-normally distributed and therefore only allow for non-parametric statistical analysis.

Future studies might also examine more differentiated sub-reddits. To ensure robust operationalization of constructs, this study only examined groups whose community rules allowed for clear classification as EC or non-EC. The study's findings could be replicated and enhanced by including other sub-reddits that lack such self-imposed rules but also cater to attitudinally homogeneous groups (e.g. by hosting politically extreme content that frightens away all but those in agreement) as well as additional neutral groups (e.g. sub-reddits concerned with non-political topics like hobbies). Given such a wider and more representative database, algorithms (e.g. random forest^{39,40} or binary elastic net models^{41,42}) could be trained to identify groups that show signs of intergroup bias and require intervention by moderators.

Finally, while this study's analysis of real user-written submissions offers high ecological validity, it can only establish correlative but not causal relationships. For example, it remains unclear if really the homogeneous nature of EC groups makes users endorse messages that display intergroup bias, or if subjects who generally use a negative and confrontational conversational style tend to gather in such groups. One approach to identify causalities could be longitudinal analysis of users' word use after joining an EC: If the frequency of markers of intergroup bias increases over time, this could be seen as

evidence for a polarizing effect of the group on the individual. Similar analyzes have been successfully conducted in depression-related online groups⁴³.

Despite these limitations, the current study provides first insights into the relationship between the words used in Reddit submissions and their score and, by this, substantiates the assumption of intergroup bias and polarization in online Echo Chambers.

Author Disclosure Statement

No competing financial interests exist.

References

1. Ali AH. The Power of Social Media in Developing Nations: New Tools for Closing the Global Digital Divide and Beyond. *Harvard Human Rights Journal* 2011; 24:185–219.
2. Hussain MM, Howard PN. (2012) Opening Closed Regimes. In: Anduiza E, Jensen MJ, Jorba L, eds. *Digital Media and Political Engagement Worldwide*. Cambridge: Cambridge University Press, pp. 200–220.
3. Lotan G, Graeff E, Ananny M, et al. The Revolutions Were Tweeted: Information Flows During the 2011 Tunisian and Egyptian Revolutions. *International Journal of Communication* 2011; 5:1375–1405.
4. Sunstein CR. (2001) *Echo chambers: Bush v. Gore, impeachment, and beyond*. Princeton, Oxford: Princeton University Press.
5. Bruns A. (2017) Echo Chamber? What Echo Chamber? Reviewing the Evidence. In: *6th Biennial Future of Journalism Conference (FOJ17)*. Cardiff: (Unpublished).
6. Grömping M. ‘Echo Chambers’: Partisan Facebook Groups during the 2014 Thai Election. *Asia Pacific Media Educator* 2014; 24:39–59.
7. Williams HTP, McMurray JR, Kurz T, et al. Network analysis reveals open forums and echo chambers in social media discussions of climate change. *Global Environmental Change* 2015; 32:126–138.

8. Barberá P, Jost JT, Nagler J, et al. Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological Science* 2015; 26:1531–1542.
9. Colleoni E, Rozza A, Arvidsson A. Echo Chamber or Public Sphere? Predicting Political Orientation and Measuring Political Homophily in Twitter Using Big Data: Political Homophily on Twitter. *Journal of Communication* 2014; 64:317–332.
10. Tajfel H. Experiments in Intergroup Discrimination. *Scientific American* 1970; 223:96–102.
11. Otten S, Moskowitz GB. Evidence for Implicit Evaluative In-Group Bias: Affect-Biased Spontaneous Trait Inference in a Minimal Group Paradigm. *Journal of Experimental Social Psychology* 2000; 36:77–89.
12. Hogg MA, Terry DJ, White KM. A Tale of Two Theories: A Critical Comparison of Identity Theory with Social Identity Theory. *Social Psychology Quarterly* 1995; 58:255–269.
13. Trepte S, Loy LS. (2017) Social Identity Theory and Self-Categorization Theory. In: Rössler P, Hoffner CA, van Zoonen L, eds. *The International Encyclopedia of Media Effects*. Hoboken, NJ, USA: John Wiley & Sons, Inc., pp. 1–13.
14. Choudhury MD, De S. (2014) Mental Health Discourse on reddit: Self-disclosure, Social Support, and Anonymity. In: *Eighth International AAAI Conference on Weblogs and Social Media*. Michigan: Association for the Advancement of Artificial Intelligence.

15. Mills RA. Pop-up political advocacy communities on reddit.com: SandersForPresident and The Donald. *AI & SOCIETY* 2018; 33:39–54.
16. Reicher SD, Spears R, Postmes T. A Social Identity Model of Deindividuation Phenomena. *European Review of Social Psychology* 1995; 6:161–198.
17. Lee E-J. Deindividuation Effects on Group Polarization in Computer-Mediated Communication: The Role of Group Identification, Public-Self-Awareness, and Perceived Argument Quality. *Journal of Communication* 2007; 57:385–403.
18. Glenski M, Pennycuff C, Weninger T. Consumers and Curators: Browsing and Voting Patterns on Reddit. *IEEE Transactions on Computational Social Systems* 2017; 4:196–206.
19. Haim M, Kümpel AS, Brosius H-B. Popularity cues in online media: A review of conceptualizations, operationalizations, and general effects. *Studies in Communication | Media* 2018; 7:186–207.
20. Porten-Cheé P, Haßler J, Jost P, et al. Popularity cues in online media: Theoretical and methodological perspectives. *Studies in Communication | Media* 2018; 7:208–230.
21. Pennebaker JW, Mehl MR, Niederhoffer KG. Psychological Aspects of Natural Language Use: Our Words, Our Selves. *Annual Review of Psychology* 2003; 54:547–577.
22. Tausczik YR, Pennebaker JW. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology* 2010; 29:24–54.

23. Chung C, Pennebaker J. (2007) The Psychological Functions of Function Words. In: Fiedler K, ed. *Social Communication*. pp. 343–359.
24. Arguello J, Butler BS, Joyce L, et al. (2006) Talk to me: foundations for successful individual-group interactions in online communities. In: *Proceedings of the SIGCHI conference on Human Factors in computing systems - CHI '06*. Montreal, Quebec, Canada: ACM Press, pp. 1–10.
25. Schnake SB, Ruscher JB. Modern Racism as a Predictor of the Linguistic Intergroup Bias. *Journal of Language and Social Psychology* 1998; 17:484–491.
26. Ruscher JB, Hammer EY. Choosing to Sever or Maintain Association Induces Biased Impression Formation. *Journal of Personality and Social Psychology* 1996; 70:701–712.
27. Vingerhoets AJJM, Bylsma LM, Vlam C de. Swearing: A biopsychosocial perspective. *Psychological Topics* 2013; 22:287–304.
28. Stieglitz S, Dang-Xuan L. Emotions and Information Diffusion in Social Media-Sentiment of Microblogs and Sharing Behavior. *Journal of Management Information Systems* 2013; 29:217–248.
29. Guest E. (2018) (Anti-)Echo Chamber Participation: Examining Contributor Activity Beyond the Chamber. In: *Proceedings of the 9th International Conference on Social Media and Society - SMSociety '18*. Copenhagen, Denmark: ACM Press, pp. 301–304.
30. Baumgartner J. (2019) pushshift.io.<https://pushshift.io/> (accessed May29 2019).

31. KeyserSosa. (2016) Scores on posts are about to start going up.https://www.reddit.com/r/announcements/comments/5gvd6b/scores_on_posts_are_about_to_start_going_up/ (accessed May29 2019).
32. Pennebaker JW, Boyd RL, Jordan K, et al. *The Development and Psychometric Properties of LIWC2015*. Austin: University of Texas at Austin.
33. Holm S. A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics* 1979; 6:65–70.
34. Choi D, Han J, Chung T, et al. (2015) Characterizing Conversation Patterns in Reddit: From the Perspectives of Content Properties and User Participation Behaviors. In: *Proceedings of the 2015 ACM on Conference on Online Social Networks - COSN '15*. Palo Alto, California, USA: ACM Press, pp. 233–243.
35. Buntain C, Golbeck J. (2014) Identifying social roles in reddit using network structure. In: *Proceedings of the 23rd International Conference on World Wide Web - WWW '14 Companion*. Seoul, Korea: ACM Press, pp. 615–620.
36. Cambria E, Speer R, Havasi C, et al. (2010) SenticNet: A Publicly Available Semantic Resource for Opinion Mining. In: *Commonsense Knowledge: Papers from the AAAI Fall Symposium*. Arlington, Virginia: Association for the Advancement of Artificial Intelligence, pp. 14–18.
37. Cambria E, Hussain A. (2015) *Sentic Computing: A Common-Sense-Based Framework for Concept-Level Sentiment Analysis*. Cham, Heidelberg, New York, Dordrecht, London: Springer.

38. Wilson T, Wiebe J, Hoffmann P. (2005) Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. In: *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP)*. Vancouver, BC: Association for Computational Linguistics, pp. 347–354.
39. Breiman L. Random Forests. *Machine Learning* 2001; 45:5–32.
40. Liaw A, Wiener M. Classification and Regression by randomForest. *R News* 2002; 2:18–22.
41. Algamal ZY, Lee MH. Applying Penalized Binary Logistic Regression with Correlation Based Elastic Net for Variables Selection. *Journal of Modern Applied Statistical Methods* 2015; 14:168–179.
42. Zou H, Hastie T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2005; 67:301–320.
43. Park A, Conway M. Longitudinal Changes in Psychological States in Online Health Community Members: Understanding the Long-Term Effects of Participating in an Online Depression Community. *Journal of Medical Internet Research* 2017; 19.

6. Conclusions and general discussion

6.1. Summary and interpretation

The goal of this thesis was to test whether subjects' exposure to online Echo Chambers can be empirically linked to negative outcomes that can be seen as indicators of societal polarization and radicalization. This experimental investigation was motivated by the fact that such links have been proposed in the political and public debate but were not sufficiently tested before.

A review of existing conceptions of Echo Chambers and related psychological theories and findings (chapter 1) identified two interweaved key components that can differentiate Echo Chambers from more heterogeneous and more open online communication: Network homogeneity, that is, to which degree users' online networks are composed of others that are similar to themselves, and selective exposure, meaning whether most messages shared between and presented to users support are congruent to their own attitudes and points of view.

Three studies were conducted to test whether Echo Chamber exposure is associated with two negative outcomes: Firstly, subjects' False Consensus was examined, that is, how strongly they perceive the public to support their own attitudes. This outcome was chosen as, according to the Theory of Planned Behavior, the perception of public opinion or social norm influences subjects' actual behavior (*studies one and two*). Secondly, it was investigated whether messages written by users display linguistic signs of Intergroup Bias, that is, to which degree they appear to identify as a members of an in-group that is in conflict with out-groups that are attacked and devaluated using negative emotional language (*study three*).

In general, all studies succeeded in establishing empirical links:

Study one focused on the aspect of homophily and demonstrated via online questionnaire that subjects with a more homogeneous online contact network display a stronger False Consensus Effect, that is, estimate higher public support for their own opinions.

Study two then focused on selective exposure in Echo Chambers and used an experimental approach. It confirmed that exposure to mostly or only attitudinally congruent messages - messages that express agreement with the subjects' attitudes - leads to stronger False Consensus than exposure to mostly or only incongruent, opposing messages. Thus, effects of the two assumed components of online Echo Chambers on subjects' perception of public opinion could be found.

Study three, finally, found that the the use of verbal signs of Intergroup Bias in user-written messages on the website Reddit (more references to in- and out-groups, more negative emotion and swearing) was positively associated with endorsement of the message by other group members. As the correlations in non-Echo Chamber groups did not reveal a similar, distinct pattern, *study three's* findings can be interpreted as first evidence for a higher Intergroup Bias in Echo Chambers.

Taking everything into account, the goal of the thesis could be accomplished, and the assumption that Echo Chambers lead to societal polarization could be empirically substantiated: Exposure to online communication structures that display attributes of Echo Chambers leads, indeed, to higher estimation of public support for one's own opinions. In other words: If subjects surround themselves mostly with others who are similar to themselves, and when they mostly receive messages that support their own pre-

existing attitudes, for example by interacting in topically homogeneous online groups or reading online media with clear and exclusive ideological agendas, they will come to the conclusion that large parts of the population share their views.

This biased perception of public opinion may lead to various negative consequences: For one thing, the belief in their attitude being endorsed by large parts of the population might make subjects even less willing to accept other, divergent opinions and statements, because these appear to subjects as an unimportant minority's view. Such a lack of communication and interchange between people with different views can be considered as highly detrimental for democratic societies that require free, open and non-violent discussion and interchange of ideas to function properly. For another thing, a favorably biased perception of public opinion and social norm may, according to the Theory of Planned Behavior, increase the subjects' likelihood to act upon his or her beliefs. If their beliefs are negative, hateful or violent, violence may be the result. For example, someone who holds prejudice against certain minorities might be more likely to verbally attack and insult members of the majority because he or she estimates that a majority of the population or the bystanders (silently) share his or her prejudices and will endorse the act. Again, the impacts on society are negative.

With regard to these negative consequences, it appears necessary to discuss how the tendency to form online Echo Chambers can be reduced. However, before strategies to achieve this are discussed in the concluding chapter 6.3, the methodological limitations of the reported studies and possible directions for future research must be addressed.

6.2. Methodological considerations and future directions

Studies one and two

The main limitation of the questionnaire-based *study one* was its correlative nature due to which it could only establish correlative but not causal relationships between the examined variables. *Study two* avoided this limitation by employing an experimental paradigm to test the effects of agreement with subjects' opinions expressed in messages and endorsement of subjects' opinions expressed via numbers of "likes" on False Consensus.

While the results of *study two* confirmed that selective exposure to agreement in message texts lets subjects estimate higher public support for their own opinions, the patterns of findings in the two sub-experiments of were complex, and interactions between agreement, level of endorsement and subjects' interest in topics occurred. These effects demand further replication and investigation.

Of high interest is the question whether future studies can replicate the finding that in case of high interest in a topic, higher numbers of "likes" for attitudinally congruent messages do not lead to a higher but instead a lower estimate of public opinion. In *study two*, this effect was interpreted as a sign of reactance and subjects questioning the validity of the numbers of "likes". However, existing research on effects of numbers of "likes" on other outcomes like message selection mostly does not reveal such an effect of reactance (Haim et al., 2018; Porten-Che   et al., 2018). Perhaps this could mean that subjects did not generally doubt the validity of "likes" but were suspicious of their concrete use in the present experiment, in which message contents and numbers of "likes" were the single

two stimulus features and therefore, perhaps, identifiable as manipulated by the experiments.

One solution for these problems could be the use of more ecologically valid and more realistic stimuli as well as experimental settings that better reflect real online communication and discussion environments. For instance, a future study could ask subjects to participate in an specifically created online discussion community for a longer period of time (e.g. several weeks). This community could be designed similar to real social networking sites, offering discussion groups, news feeds, user profiles and so on. The intended experimental manipulations could be implemented in this environment, for example by greeting users with a personalized news feed that, like in study two, exhibits either high or low agreement and either high or low endorsement of the users' own attitudes (measured in the projects' beginning via questionnaire). After being required to read through the news feed and select the most interesting messages for later reading, the users could interact with the community in other ways. Thus, while basically offering the same manipulation as study two, the environment would appear much more realistic to the user, and the fact, that the news feed is experimentally manipulated, would be much less overt. Moreover, such a mid- or long-term endeavor would allow for not only single (e.g. after news feed exposure) but repeated measurement of False Consensus, thus providing the opportunity to test the stability of effects.

Of course, creating such a more realistic setting requires effort and resources both on the side of experimenter and participants. However, this effort should quickly pay off as such an community could serve as an environment for additional studies, for example for network analysis of interactions between participants. It could also provide the basis for multidisciplinary cooperation, for example with political scientists who might use the

community for testing of own research questions (e.g. simulated campaigns and elections).

Independent of whether future studies employ such more realistic or instead traditional experimental settings, they should also address a final weakness of *study two*: In order to maximize the experimental manipulation, *study two* contrasted conditions in which subjects encountered either mostly/only agreement or mostly/only disagreement. A true neutral condition in which subjects were exposed to an equal number of agreeing and disagreeing messages or to only topically irrelevant messages was not included. Such a neutral condition should be included to measure baseline False Consensus and to determine how strongly exposure to agreement vs. disagreement increases or decreases False Consensus in regard to the baseline.

Study three

Study three was intended to enhance the perspectives of the previous studies with an innovative and real-world data driven approach. Its overall findings can be interpreted to support the assumed link between Echo Chambers and Intergroup Bias. However, its pattern of results is less clear and distinct than in the previous questionnaire-based and experimental studies. While this can be seen as a necessary consequence of the use of real-world data that is subject to numerous influences outside of the examined variables, it also offers a basis for further studies that try to improve upon it. Future research may try to implement optimizations both regarding the measures used, as well as the Reddit discussion groups that served as data base:

The software LIWC was employed to analyze the frequency of several categories of words that can be associated with intergroup bias. While LIWC has the advantage of

offering psychometrically validated scales, its scope is limited: For one thing, it scans a text only for words that are in its dictionary. If words that may be relevant to the research questions (e.g. non-traditional swearing terms or jargon that is unique to a community) are not in its dictionary, they are not counted. For another thing, this dictionary-based approach can not identify more subtle verbal cues like negations (e.g. “not good”), irony or sarcasm. Moreover, it also can not identify grammatical dependencies, for example whether negative emotion words and swearing are used in reference to out-group members, and positive emotion words in reference to in-group members. As LIWC only counts the overall frequencies of words, assuming such dependencies can currently be only speculative.

These weaknesses regarding measures could be addressed in several ways:

The problem of limited dictionaries could be easily addressed by creating own additional dictionaries which can be easily integrated into LIWC. To achieve this, first the word use in a target community has to be (manually) analyzed, for example regarding idioms and relevant phrases unique to the community. Then, these have to be collected and sorted into theory-based scales (e.g. an additional swearing scale that features terms commonly used by a community to derogate the out-group) which should be pre-tested to ensure their internal consistency. After integration in LIWC, these new scales could be used similar to the existing ones.

To address the problem of grammatical dependencies, negations and irony, alternative analytical approaches appear necessary. For example, several tools exist that implement linguistics based algorithm for detecting emotional valence while also taking into account negations, irony and so on (Cambria & Hussain, 2015; Cambria, Speer,

Havasi, & Hussain, 2010; Wilson, Wiebe, & Hoffmann, 2005). These tools, however, usually lack the psychological and psychometrical foundations that LIWC offers. Independent of the path that future research takes, it appears most promising to try to integrate more perspectives and competencies from the fields of (psycho-)linguistics to tackle the described challenges.

Moreover, future endeavors should try to enhance the data base of user messages that were analyzed: In study three, only self-proclaimed Echo Chamber groups that explicitly forbid divergent opinions were analyzed. This allowed for a clear operationalization of constructs but limited the representativity of analyzed groups. Future studies could try to identify more groups that could also be conceived as Echo Chambers (for example because they provide a place to voice rather extreme opinions that drive away all users that do not agree). Also, more neutral, non-Echo Chamber groups should be analyzed in order to get a broader baseline of communication and word use to which the Echo Chamber groups could be compared. Given such a wider data base, attempts to train machine learning algorithm could be made in order to identify groups that display Echo Chamber behavior, polarization and radicalization and that might profit from tempering intervention by moderators.

Finally, instead of message scores, further studies could also examine other, potentially more robust measures of popularity: Scores, resulting from up- and down-votes given by others users, can be influenced by numerous factors, pertaining to the actual message (contents, linguistic style) but also to external, more or less random factors (e.g. messages published during times of day when many users visit Reddit may receive more attention and thus higher scores). *Study three* tried to reduce this random noise by aggregating data on author-level, but the noise could most likely be not fully eliminated

and may be one reason why the observed correlations were rather weak. A more robust and eclectic measure of popularity could, for example, be seen in the “virality” of a message (Choi et al., 2015), that is, how many overall reactions it elicited by other users, encompassing score but also the number of comments written in answer to it, how often it has been shared or reposted, and so on.

Alternative approaches

In addition to the described directions for further research, it should also be discussed what alternative methodological approaches could be used to further investigate links between online Echo Chambers and possible negative outcomes.

One alternative outcome of Echo Chamber exposure that could be focused is subjects’ actual attitude towards a topic: To which degree does it get distorted in the direction of the Echo Chamber’s general ideology? Can signs of radicalization be observed in participants of Echo Chambers?

It can be assumed that actual radicalization of attitudes can only be observed after long term exposure, which makes the creation of paradigms difficult and also ethically challenging. This is one main reason why this thesis did not focus on measuring attitudes but instead measured False Consensus. However, one approach to test for signs of radicalization could be longitudinal analysis of messages written by users that initially were not members of an Echo Chamber group on a social media platform but at some point joined such a group and started to actively participate. Given a valid way to measure the degree of radicalization that a message exhibits (e.g. via LIWC with an custom extended dictionary), it could be tested if messages after joining the Echo Chamber are characterized by more radical attitudes than messages before. Similar approaches have

been used to examine, for example, beneficial effects of depression-related online groups on subjects' use of negative emotional language (Park & Conway, 2017).

An experimental way to test for signs of radicalization could be based on an online discussion community that is created and used by experimenters as already described above: Users that are invited to participate in such a community could be seemingly randomly assigned to groups and instructed to discuss topics with each other over a longer period of time. However, the actual assignment to groups would not be random but based on the participants' attitudes regarding the topics discussed, with some groups being homogeneous, that is, made up of participants with similar attitudes, and some groups being heterogeneous, that is, made up of participants with different attitudes. The homogeneous groups thus serve as Echo Chambers, the heterogeneous groups as neutral comparison groups. After several episodes of interaction (e.g. a couple of weeks with daily discussions and/or sharing of messages between group members), relevant outcomes, such as attitudes regarding the topics but also False Consensus or additional measurements like positive and negative perceptions of in-group and out-group members, could be measured. Differences between the homogeneous and heterogeneous groups could be interpreted as effects of exposure to an (albeit small-scale) Echo Chamber. Of course, a thorough debriefing of participants to avoid actual radicalization would be necessary afterwards.

6.3. Practical implications

Finally, it shall be discussed which practical interventions to reduce negative effects of online Echo Chambers appear feasible, given this thesis' empirical evidence as well as prior research.

There exist already several technological tools that try to warn users when their online usage behavior exhibits signs of Echo Chambers, for example when they mostly visit websites that are biased in one political direction, or when they mostly connect to other users that share their own opinions (Munson, Lee, & Resnick, 2013; Nagulendra & Vassileva, 2014). These tools visualize biases via icons in a browser plug in or by plotting users' networks. As already argued in chapter 4, these tools share one central weakness: Users have to decide to use them and to follow their recommendations, and, of course, only users that are actively want to avoid biases and to consume more balanced information will make this decision. Users who are not aware of the possibility of biases in information exposure and network, or users who simply do not care about such biases, will most likely not use these tools. Thus, the user groups who are most likely to fall victim to negative effects of Echo Chambers are not reached. Furthermore, the effectiveness of such tools is not empirically proven. Therefore, before attempting to create any further technological means that try to make users aware of biases or distortions in their online behavior, it should be experimentally tested if such information has positive impact on the user and does not, instead, lead to reactance.

However, in general, it appears more promising not to tackle the negative effects of already existing Echo Chambers but instead to deal with the roots of the problem and address the reasons why people feel the necessity to form Echo Chambers.

As outlined in chapter 1, selective exposure is a basic human tendency that predates the Internet. Therefore, eliminating it can not be a realistic goal. Yet, as *study one* has demonstrated, there exist personality traits that might alleviate the negative effects of Echo Chambers, one of which is subjects' tolerance for ambiguous information and situations: The higher subjects' ambiguity tolerance, the lower their False Consensus,

even if they describe their online networks as homogeneous. Thus, ambiguity tolerance appears as a protective factor that leads to more realistic estimates of public opinion, most likely because subjects who are tolerant of ambiguous information are more open to contradicting opinions and information, even when their network is somewhat biased.

Interventions should therefore address and strengthen positive traits like ambiguity tolerance, for example by educating students early on about the importance of getting information from different sources, about the intrinsic value of pluralism and diversity in opinions and attitudes, and also about competent use of (online) media.

References

This list of references includes all references used in the individual studies and the framework text. References that are cited only in individual studies but not the framework text are indicated with a leading asterisk.

- * Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50, 179–211. [https://doi.org/10.1016/0749-5978\(91\)90020-T](https://doi.org/10.1016/0749-5978(91)90020-T)
- Ajzen, I., & Fishbein, M. (1977). Attitude-Behavior Relations: A Theoretical Analysis and Review of Empirical Research. *Psychological Bulletin*, 5(84), 888–918.
- * Algamal, Z. Y., & Lee, M. H. (2015). Applying Penalized Binary Logistic Regression with Correlation Based Elastic Net for Variables Selection. *Journal of Modern Applied Statistical Methods*, 14(1), 168–179. <https://doi.org/10.22237/jmasm/1430453640>
- Ali, A. H. (2011). The Power of Social Media in Developing Nations: New Tools for Closing the Global Digital Divide and Beyond. *Harvard Human Rights Journal*, 24, 185–219.
- * Allcott, H., & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives*, 31(2), 211–236. <https://doi.org/10.1257/jep.31.2.211>
- Arguello, J., Butler, B. S., Joyce, L., Kraut, R., Ling, K. S., & Wang, X. (2006). Talk to me: foundations for successful individual-group interactions in online communities. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '06*, 1–10. <https://doi.org/10.1145/1124772.1124916>

- Asch, S. E. (1961). Effects of group pressure upon the modification and distortion of judgments. In M. Henle (Ed.), *Documents of Gestalt Psychology* (pp. 222–236). Berkeley, Los Angeles: University of California Press.
- * Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>
- Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., & Bonneau, R. (2015). Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological Science*, *26*(10), 1531–1542.
- * Bastos, M. T. (2015). Shares, Pins, and Tweets: News readership from daily papers to social media. *Journalism Studies*, *16*(3), 305–325. <https://doi.org/10.1080/1461670X.2014.891857>
- Bauman, K. P., & Geher, G. (2002). We Think You Agree: The Detrimental Impact of the False Consensus Effect on Behavior. *Current Psychology*, *21*(4), 293–318.
- * Baumgartner, J. (2019, May 12). pushshift.io. Retrieved May 29, 2019, from pushshift.io website: <https://pushshift.io/>
- * Berlyne, D. E., & Dittkofsky, J. (1976). Effects of novelty and oddity on visual selective attention. *British Journal of Psychology*, *67*(2), 175–180.
- * Bosveld, W., Koomen, W., & van der Pligt, J. (1994). Selective exposure and the false consensus effect: The availability of similar and dissimilar others. *British Journal of Social Psychology*, *33*, 457–466. <https://doi.org/10.1111/j.2044-8309.1994.tb01041.x>
- * Breiman, L. (2001). Random Forests. *Machine Learning*, *45*, 5–32.

- * Bruns, A. (2017). Echo Chamber? What Echo Chamber? Reviewing the Evidence. *6th Biennial Future of Journalism Conference (FOJ17)*. Presented at the 6th Biennial Future of Journalism Conference (FOJ17), Cardiff. Retrieved from <https://eprints.qut.edu.au/113937/>
- * Budner, S. (1962). Intolerance of ambiguity as a personality variable1. *Journal of Personality*, 30(1), 29–50. <https://doi.org/10.1111/j.1467-6494.1962.tb02303.x>
- * Buntain, C., & Golbeck, J. (2014). Identifying social roles in reddit using network structure. *Proceedings of the 23rd International Conference on World Wide Web - WWW '14 Companion*, 615–620. <https://doi.org/10.1145/2567948.2579231>
- Cambria, E., & Hussain, A. (2015). *Sentic Computing: A Common-Sense-Based Framework for Concept-Level Sentiment Analysis*. Cham, Heidelberg, New York, Dordrecht, London: Springer.
- Cambria, E., Speer, R., Havasi, C., & Hussain, A. (2010). SenticNet: A Publicly Available Semantic Resource for Opinion Mining. *Commonsense Knowledge: Papers from the AAAI Fall Symposium*, 14–18. Arlington, Virginia: Association for the Advancement of Artificial Intelligence.
- * Chang, Y.-T., Yu, H., & Lu, H.-P. (2015). Persuasive messages, popularity cohesion, and message diffusion in social media marketing. *Journal of Business Research*, 68(4), 777–782. <https://doi.org/10.1016/j.jbusres.2014.11.027>
- Choi, D., Han, J., Chung, T., Ahn, Y.-Y., Chun, B.-G., & Kwon, T. T. (2015). Characterizing Conversation Patterns in Reddit: From the Perspectives of Content Properties and User Participation Behaviors. *Proceedings of the 2015 ACM on Conference on Online Social Networks - COSN '15*, 233–243. <https://doi.org/10.1145/2817946.2817959>

- * Choudhury, M. D., & De, S. (2014). Mental Health Discourse on reddit: Self-disclosure, Social Support, and Anonymity. *Eighth International AAAI Conference on Weblogs and Social Media*. Presented at the Eighth International AAAI Conference on Weblogs and Social Media, Michigan. Retrieved from <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/index>
- Chung, C., & Pennebaker, J. (2007). The Psychological Functions of Function Words. In K. Fiedler (Ed.), *Social Communication* (pp. 343–359).
- * Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale: Lawrence Erlbaum Associates.
- Colleoni, E., Rozza, A., & Arvidsson, A. (2014). Echo Chamber or Public Sphere? Predicting Political Orientation and Measuring Political Homophily in Twitter Using Big Data: Political Homophily on Twitter. *Journal of Communication*, 64(2), 317–332. <https://doi.org/10.1111/jcom.12084>
- * Cotton, J. L., & Hieser, R. A. (1980). Selective exposure to information and cognitive dissonance. *Journal of Research in Personality*, 14(4), 518–527. [https://doi.org/10.1016/0092-6566\(80\)90009-4](https://doi.org/10.1016/0092-6566(80)90009-4)
- * De Cristofaro, E., Friedman, A., Jourjon, G., Kaafar, M. A., & Shafiq, M. Z. (2014). Paying for Likes? Understanding Facebook Like Fraud Using Honeypots. *ArXiv:1409.2097 [Physics]*. Retrieved from <http://arxiv.org/abs/1409.2097>
- * de la Haye, A.-M. (2000). A methodological note about the measurement of the false-consensus effect. *European Journal of Social Psychology*, 30, 569–581.

- * Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., ... Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, 113(3), 554–559.
<https://doi.org/10.1073/pnas.1517441113>
- * DiResta, R., Shaffer, K., Ruppel, B., Sullivan, D., Matney, R., Fox, R., ... Johnson, B. (2018). *The Tactics & Tropes of the Internet Research Agency*. Retrieved from <https://www.newknowledge.com/articles/the-disinformation-report/>
- * Dvir-Gvirzman, S. (2019). I like what I see: studying the influence of popularity cues on attention allocation and news selection. *Information, Communication & Society*, 22(2), 286–305. <https://doi.org/10.1080/1369118X.2017.1379550>
- El-Bermawy, M. M. (2015, November 18). Your Filter Bubble is Destroying Democracy. *Wired*. Retrieved from <https://www.wired.com/2016/11/filter-bubble-destroying-democracy/>
- * Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford: Stanford University Press.
- * Fishbein, M. (1976). A Behavior Theory Approach to the Relations between Beliefs about an Object and the Attitude Toward the Object. In U. H. Funke (Ed.), *Mathematical Models in Marketing: A Collection of Abstracts* (pp. 87–88).
https://doi.org/10.1007/978-3-642-51565-1_25

Fishbein, M., & Ajzen, I. (1975). *Belief, Attitude, and Behavior: An Introduction to Theory and Research*. Reading, Mass.: Addison Wesley.

* Frenkel-Brunswik, E. (1949). Intolerance of Ambiguity as an Emotional and Perceptual Personality Variable. *Journal of Personality*, 18(1), 108–143.

<https://doi.org/10.1111/j.1467-6494.1949.tb01236.x>

* Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The Journal of Socio-Economics*, 40(1), 35–42. <https://doi.org/10.1016/j.socec.2010.10.008>

* Galesic, M., Olsson, H., & Rieskamp, J. (2013). False consensus about false consensus. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 472–476). Retrieved from <http://csjarchive.cogsci.rpi.edu/Proceedings/2013/papers/0109/paper0109.pdf>

* Gilbert, E., Bergstrom, T., & Karahalios, K. (2009). Blogs Are Echo Chambers: Blogs Are Echo Chambers. *42nd Hawaii International Conference on System Sciences, 2009. HICSS'09*. Retrieved from <http://ieeexplore.ieee.org/abstract/document/4755503/>

* Glenski, M., Pennycuff, C., & Weninger, T. (2017). Consumers and Curators: Browsing and Voting Patterns on Reddit. *IEEE Transactions on Computational Social Systems*, 4(4), 196–206. <https://doi.org/10.1109/TCSS.2017.2742242>

Goldie, D., Linick, M., Jabbar, H., & Lubienski, C. (2014). Using bibliometric and social media analyses to explore the “echo chamber” hypothesis. *Educational Policy*, 28(2), 281–305.

- Grömping, M. (2014). 'Echo Chambers': Partisan Facebook Groups during the 2014 Thai Election. *Asia Pacific Media Educator*, 24(1), 39–59.
<https://doi.org/10.1177/1326365X14539185>
- * Guest, E. (2018). (Anti-)Echo Chamber Participation: Examining Contributor Activity Beyond the Chamber. *Proceedings of the 9th International Conference on Social Media and Society - SMSociety '18*, 301–304. <https://doi.org/10.1145/3217804.3217933>
- Haim, M., Kümpel, A. S., & Brosius, H.-B. (2018). Popularity cues in online media: A review of conceptualizations, operationalizations, and general effects. *Studies in Communication | Media*, 7(2), 186–207. <https://doi.org/10.5771/2192-4007-2018-2-58>
- * Hart, W., Albarracín, D., Eagly, A. H., Brechan, I., Lindberg, M. J., & Merrill, L. (2009). Feeling validated versus being correct: A meta-analysis of selective exposure to information. *Psychological Bulletin*, 135(4), 555–588. <https://doi.org/10.1037/a0015701>
- Hogg, M. A., Terry, D. J., & White, K. M. (1995). A Tale of Two Theories: A Critical Comparison of Identity Theory with Social Identity Theory. *Social Psychology Quarterly*, 58(4), 255–269. <https://doi.org/10.2307/2787127>
- * Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics*, 6(2), 65–70.
- * Howard, P. N., Ganesh, B., Liotsiou, D., Kelly, J., & François, C. (2018). *The IRA, Social Media and Political Polarization in the United States, 2012-2018* (p. 47).
- * Hox, J. J. (2002). *Multilevel analysis: techniques and applications*. Mahwah, N.J: Lawrence Erlbaum Associates.

- Hussain, M. M., & Howard, P. N. (2012). Opening Closed Regimes. In E. Anduiza, M. J. Jensen, & L. Jorba (Eds.), *Digital Media and Political Engagement Worldwide* (pp. 200–220). <https://doi.org/10.1017/CBO9781139108881.011>
- * Ikram, M., Onwuzurike, L., Farooqi, S., De Cristofaro, E., Friedman, A., Jourjon, G., ... Shafiq, M. Z. (2017). Measuring, Characterizing, and Detecting Facebook Like Farms. *ArXiv:1707.00190 [Cs]*. Retrieved from <http://arxiv.org/abs/1707.00190>
- * Johnson, T. J., Bichard, S. L., & Zhang, W. (2009). Communication Communities or “CyberGhettos?”: A Path Analysis Model Examining Factors that Explain Selective Exposure to Blogs. *Journal of Computer-Mediated Communication*, *15*, 60–82. <https://doi.org/10.1111/j.1083-6101.2009.01492.x>
- * Kahneman, D. (1992). Reference points, anchors, norms, and mixed feelings. *Organizational Behavior and Human Decision Processes*, *51*(2), 296–312. [https://doi.org/10.1016/0749-5978\(92\)90015-Y](https://doi.org/10.1016/0749-5978(92)90015-Y)
- Kassim, S. (2012, June 3). Twitter Revolution: How the Arab Spring Was Helped By Social Media. *Mic*. Retrieved from <https://mic.com/articles/10642/twitter-revolution-how-the-arab-spring-was-helped-by-social-media#.LrlG8X8Pv>
- * KeyserSosa. (2016, December 6). Scores on posts are about to start going up. Retrieved May 29, 2019, from Reddit website: https://www.reddit.com/r/announcements/comments/5gvd6b/scores_on_posts_are_about_to_start_going_up/
- Knobloch-Westerwick, S. (2014). *Choice and Preference in Media Use : Advances in Selective Exposure Theory and Research*. New York, London: Routledge.

- * Knobloch-Westerwick, S. (2015). The Selective Exposure Self- and Affect-Management (SESAM) Model: Applications in the Realms of Race, Politics, and Health. *Communication Research*, 42(7), 959–985. <https://doi.org/10.1177/0093650214539173>
- Knobloch-Westerwick, S., & Jingbo Meng. (2009). Looking the Other Way: Selective Exposure to Attitude-Consistent and Counterattitudinal Political Information. *Communication Research*, 36(3), 426–448. <https://doi.org/10.1177/0093650209333030>
- * Knobloch-Westerwick, S., Johnson, B. K., & Westerwick, A. (2015). Confirmation Bias in Online Searches: Impacts of Selective Exposure Before an Election on Political Attitude Strength and Shifts. *Journal of Computer-Mediated Communication*, 20, 171–187. <https://doi.org/10.1111/jcc4.12105>
- * Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13). <https://doi.org/10.18637/jss.v082.i13>
- * Lee, E.-J. (2007). Deindividuation Effects on Group Polarization in Computer-Mediated Communication: The Role of Group Identification, Public-Self-Awareness, and Perceived Argument Quality. *Journal of Communication*, 57(2), 385–403. <https://doi.org/10.1111/j.1460-2466.2007.00348.x>
- * Lee, E.-J., & Yoon Jae Jang. (2010). What Do Others' Reactions to News on Internet Portal Sites Tell Us? Effects of Presentation Format and Readers' Need for Cognition on Reality Perception. *Communication Research*, 37(6), 825–846. <https://doi.org/10.1177/0093650210376189>
- * Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News*, 2(3), 18–22.

- Liska, A. E. (1984). A Critical Examination of the Causal Structure of the Fishbein/Ajzen Attitude-Behavior Model. *Social Psychology Quarterly*, 47(1), 61.
<https://doi.org/10.2307/3033889>
- * Litman, J. A. (2010). Relationships between measures of I- and D-type curiosity, ambiguity tolerance, and need for closure: An initial test of the wanting-liking model of information-seeking. *Personality and Individual Differences*, 48, 397–402.
<https://doi.org/10.1016/j.paid.2009.11.005>
- Lotan, G., Graeff, E., Ananny, M., Gaffney, D., Pearce, I., & Boyd, D. (2011). The Revolutions Were Tweeted: Information Flows During the 2011 Tunisian and Egyptian Revolutions. *International Journal of Communication*, 5, 1375–1405.
- * Lück, H., & Timaeus, E. (n.d.). Soziale Erwünschtheit (SDS-CM). *Zusammenstellung Sozialwissenschaftlicher Items. ZIS Version 15.00*.
- Marks, G., & Miller, N. (1987). Ten Years of Research on the False-Consensus Effect: An Empirical and Theoretical Review. *Psychological Bulletin*, 102(1), 72–90.
- McLain, D. L. (2009). Evidence Of The Properties Of An Ambiguity Tolerance Measure: The Multiple Stimulus Types Ambiguity Tolerance Scale-II (MSTAT-II). *Psychological Reports*, 105(3), 975–988. <https://doi.org/10.2466/PR0.105.3.975-988>
- 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.
- * Messing, S., & Westwood, S. J. (2014). Selective exposure in the age of social media: Endorsements trump partisan source affiliation when selecting news online. *Communication Research*, 41(8), 1042–1063.

- * Mills, R. A. (2018). Pop-up political advocacy communities on reddit.com: SandersForPresident and The Donald. *AI & SOCIETY*, 33(1), 39–54.
<https://doi.org/10.1007/s00146-017-0712-9>
- Mosley, T. (2018, November 5). How Social Media Echo Chambers Drown Out the Voices in the Middle. *KQED*. Retrieved from <https://www.kqed.org/news/11703717/how-social-media-echo-chambers-drown-out-the-voices-in-the-middle>
- Mullen, B., Atkins, J. L., Champion, D. S., Edwards, C., Hardy, D., Story, J. E., & Vanderklok, M. (1985). The false consensus effect: A meta-analysis of 115 hypothesis tests. *Journal of Experimental Social Psychology*, 21(3), 262–283.
- Munson, S. A., Lee, S. Y., & Resnick, P. (2013). Encouraging Reading of Diverse Political Viewpoints with a Browser Widget. *Proceedings of the Seventh International AAI Conference on Weblogs and Social Media*, 419–428. Retrieved from <https://www.aaai.org/Library/ICWSM/icwsm13contents.php>
- Myers, D. G. (1975). Discussion-Induced Attitude Polarization. *Human Relations*, 28(8), 699–714.
- Myers, D. G., & Lamm, H. (1976). The Group Polarization Phenomenon. *Psychological Bulletin*, 83(4), 602–627.
- Nagulendra, S., & Vassileva, J. (2014). Understanding and controlling the filter bubble through interactive visualization: a user study. *Proceedings of the 25th ACM Conference on Hypertext and Social Media - HT '14*, 107–115.
<https://doi.org/10.1145/2631775.2631811>

- * Norton, R. W. (1975). Measurement of Ambiguity Tolerance. *Journal of Personality Assessment*, 39(6), 607–619.
- Otten, S., & Moskowitz, G. B. (2000). Evidence for Implicit Evaluative In-Group Bias: Affect-Biased Spontaneous Trait Inference in a Minimal Group Paradigm. *Journal of Experimental Social Psychology*, 36(1), 77–89. <https://doi.org/10.1006/jesp.1999.1399>
- Pariser, E. (2011). *The Filter Bubble: What The Internet Is Hiding From You*. London: Penguin.
- Park, A., & Conway, M. (2017). Longitudinal Changes in Psychological States in Online Health Community Members: Understanding the Long-Term Effects of Participating in an Online Depression Community. *Journal of Medical Internet Research*, 19(3). <https://doi.org/10.2196/jmir.6826>
- Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The Development and Psychometric Properties of LIWC2015*. Austin: University of Texas at Austin.
- * Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological Aspects of Natural Language Use: Our Words, Our Selves. *Annual Review of Psychology*, 54(1), 547–577. <https://doi.org/10.1146/annurev.psych.54.101601.145041>
- * Peter, C., Rossmann, C., & Keyling, T. (2014). Exemplification 20: Roles of direct and indirect social information in conveying health messages through social network sites. *Journal of Media Psychology: Theories, Methods, and Applications*, 26(1), 19–28. <https://doi.org/10.1027/1864-1105/a000103>
- Petty, R. E. (1986). The Elaboration Likelihood Model of Persuasion. *Advances in Experimental Social Psychology*, 19, 123–205.

- Porten-Che , P., Ha ler, J., Jost, P., Eilders, C., & Maurer, M. (2018). Popularity cues in online media: Theoretical and methodological perspectives. *Studies in Communication Media*, 7(2), 208–230. <https://doi.org/10.5771/2192-4007-2018-2-80>
- Quattrociocchi, W., Scala, A., & Sunstein, C. R. (2016). *Echo chambers on facebook*. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2795110
- * Reicher, S. D., Spears, R., & Postmes, T. (1995). A Social Identity Model of Deindividuation Phenomena. *European Review of Social Psychology*, 6(1), 161–198. <https://doi.org/10.1080/14792779443000049>
- Ross, L., Greene, D., & House, P. (1977). The “false consensus effect”: An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology*, 13, 279–301.
- Ruscher, J. B., & Hammer, E. Y. (1996). Choosing to Sever or Maintain Association Induces Biased Impression Formation. *Journal of Personality and Social Psychology*, 70(4), 701–712.
- * Schmidt, A. L., Zollo, F., Del Vicario, M., Bessi, A., Scala, A., Caldarelli, G., ... Quattrociocchi, W. (2017). Anatomy of news consumption on Facebook. *Proceedings of the National Academy of Sciences*, 114(12), 3035–3039. <https://doi.org/10.1073/pnas.1617052114>
- Schnake, S. B., & Ruscher, J. B. (1998). Modern Racism as a Predictor of the Linguistic Intergroup Bias. *Journal of Language and Social Psychology*, 17(4), 484–491.
- Scruggs, J. (1998). *The “Echo Chamber” Approach to Advocacy*. Philip Morris.

- * Sears, D. O., & Freedman, J. L. (1967). Selective Exposure to Information: A Critical Review. *Public Opinion Quarterly*, 31(2), 194. <https://doi.org/10.1086/267513>
- * Snijders, T. A. B. (2005). Power and sample size in multilevel modeling. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of Statistics in Behavioral Science* (pp. 1570–1573). Chichester etc.: Wiley.
- * Stieglitz, S., & Dang-Xuan, L. (2013). Emotions and Information Diffusion in Social Media-Sentiment of Microblogs and Sharing Behavior. *Journal of Management Information Systems*, 29(4), 217–248. <https://doi.org/10.2753/MIS0742-1222290408>
- Stroud, N. J. (2008). Media Use and Political Predispositions: Revisiting the Concept of Selective Exposure. *Political Behavior*, 30, 341–366.
- Sunstein, C. R. (2001). *Echo chambers: Bush v. Gore, impeachment, and beyond*. Princeton, Oxford: Princeton University Press.
- * Swart, J., Peters, C., & Broersma, M. (2018). Sharing and Discussing News in Private Social Media Groups: The social function of news and current affairs in location-based, work-oriented and leisure-focused communities. *Digital Journalism*, 1–19. <https://doi.org/10.1080/21670811.2018.1465351>
- Tajfel, H. (1970). Experiments in Intergroup Discrimination. *Scientific American*, 223(5), 96–102. <https://doi.org/10.1038/scientificamerican1170-96>
- * Tandoc, E. C., Lim, Z. W., & Ling, R. (2018). Defining “Fake News”: A typology of scholarly definitions. *Digital Journalism*, 6(2), 137–153. <https://doi.org/10.1080/21670811.2017.1360143>

- * Tausczik, Y. R., & Pennebaker, J. W. (2010). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology, 29*(1), 24–54. <https://doi.org/10.1177/0261927X09351676>
- Trepte, S., & Loy, L. S. (2017). Social Identity Theory and Self-Categorization Theory. In P. Rössler, C. A. Hoffner, & L. van Zoonen (Eds.), *The International Encyclopedia of Media Effects* (pp. 1–13). <https://doi.org/10.1002/9781118783764.wbieme0088>
- * Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76*(2), 105–110.
- Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science, 185*(4157), 1124–1131.
- * van Noort, G., Antheunis, Marjolijn L., & van Reijmersdal, Eva A. (2012). Social connections and the persuasiveness of viral campaigns in social network sites: Persuasive intent as the underlying mechanism. *Journal of Marketing Communications, 18*(1), 39–53. <https://doi.org/10.1080/13527266.2011.620764>
- Vingerhoets, A. J. J. M., Bylsma, L. M., & Vlam, C. de. (2013). Swearing: A biopsychosocial perspective. *Psychological Topics, 22*(2), 287–304.
- Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology, 20*(3), 273–281. <https://doi.org/10.1080/14640746808400161>
- Williams, H. T. P., McMurray, J. R., Kurz, T., & Lambert, H. F. (2015). Network analysis reveals open forums and echo chambers in social media discussions of climate change. *Global Environmental Change, 32*, 126–138. <https://doi.org/10.1016/j.gloenvcha.2015.03.006>

- Wilson, T., Wiebe, J., & Hoffmann, P. (2005). Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP)*, 347–354. Vancouver, BC: Association for Computational Linguistics.
- Wojcieszak, M. (2008). False Consensus Goes Online: Impact of Ideologically Homogeneous Groups on False Consensus. *Public Opinion Quarterly*, 72(4), 781–791. <https://doi.org/10.1093/poq/nfn056>
- * Wojcieszak, M. E. (2011). Computer-Mediated False Consensus: Radical Online Groups, Social Networks and News Media. *Mass Communication and Society*, 14(4), 527–546. <https://doi.org/10.1080/15205436.2010.513795>
- * Wojcieszak, M., & Price, V. (2009). What Underlies the False Consensus Effect? How Personal Opinion and Disagreement Affect Perception of Public Opinion. *International Journal of Public Opinion Research*, 21(1), 25–46. <https://doi.org/10.1093/ijpor/edp001>
- * Youmans, W. L., & York, J. C. (2012). Social Media and the Activist Toolkit: User Agreements, Corporate Interests, and the Information Infrastructure of Modern Social Movements. *Journal of Communication*, 62(2), 315–329. <https://doi.org/10.1111/j.1460-2466.2012.01636.x>
- Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, 9(2p2), 1.
- Zollo, F., Bessi, A., Del Vicario, M., Scala, A., Caldarelli, G., Shekhtman, L., ... Quattrociocchi, W. (2017). Debunking in a world of tribes. *PLOS ONE*, 12(7), e0181821. <https://doi.org/10.1371/journal.pone.0181821>

* Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net.

Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67(2), 301–320. <https://doi.org/10.1111/j.1467-9868.2005.00503.x>

Zúñiga, H. G. de, Jung, N., & Valenzuela, S. (2012). Social Media Use for News and Individuals' Social Capital, Civic Engagement and Political Participation. *Journal of Computer-Mediated Communication*, 17, 319–336. <https://doi.org/10.1111/j.1083-6101.2012.01574.x>

Versicherung (gem. §4 Abs. 3 Satz 1 Nr. 5 PromO):

Ich versichere hiermit

- an Eides statt, dass ich die Dissertation selbständig angefertigt, außer den im Schriftenverzeichnis sowie den Anmerkungen genannten Hilfsmitteln keine weiteren benutzt und die Herkunft der Stellen, die wörtlich oder sinngemäß aus anderen Werken übernommen sind, bezeichnet habe,
- dass ich die Dissertation nicht bereits in derselben oder einer ähnlichen Fassung an einer anderen Fakultät oder einer anderen Hochschule zur Erlangung eines akademischen Grades eingereicht habe.

Robert Luzsa, Passau, den 04.06.2019