Questions:

-I refer to the concepts “hate speech”, “positive speech”, “neutral speech”. What do you think about it. Does it make sense while reading the paper, or would it be better to call them “negative bias”/“hate bias”, “positive bias”, “hate bias”?

-How to refer to search engine and politican? “Information source”/ “communication channel”?

-Do you think I need to expand on general algorithmic bias and the threat to minorities and democracy? Which other part should I cut in order to stay in the word limit?

Pre-registration: If I share the link to my pre-registration in EGAP in the manuscript, the review process is not anonymous anymore. Do you know any common practise to avoid this?

-Do you think it makes sense to keep the analysis on click behavior, that I explore in the end of the paper? Does it fit into the paper?

-Journal: I would first like to submit it to a Political Psychology Journal. Do you think it suits good to a Political Psychology Journal or should I try a Political Communication Journal/ or a more broader Journal?

- Is everything understandable or are there any parts that could be explained in a better way?

-Alternative title ideas:

“Adopt or backfire? The impact of hate speech and positive speech by search engines and politicians on political” //

“Adopt or backfire? The impact of hate speech and positive speech about refugees by search engines and politicians on political attitudes”//

“Adoption or backfire effects of hate speech and positive speech about refugees? The impact of popular information sources search engine and politicians on political attitudes.”

“Embrace it or backfire on hate speech and positive speech? The impact of the popular information sources search engines and politicians on political attitudes.”//

“Take over or refire hate speech and positive speech? The impact of the popular information sources search engines and politicians on political attitudes.”//

WORK IN PROGESS: Please do not circulate or cite without permission.

**The impact of hate speech on political attitudes – Evidence from an online experiment**

# Abstract

Search engines play an important role for political information acquisition and are generally trusted, yet their impact on political attitudes remains unknown. I investigate how online users respond when search engines present non-neutral search suggestions in the context of characteristics of a minority group in society. On the basis of an online experiment, I assess the effects of hate speech about refugees on trust and policy preferences. The experiment varies both the tone of the suggestions (positive, neutral, negative) as well as the source of the suggestions (search engine and politician). The study examines the role of trust in the information, political ideology and populist attitudes. It highlights hitherto unknown effects of search engines on political attitudes. The results show that search engines are generally trusted to a much higher degree than politicians. However, when the search content is positive or negative biased, trust in the search engine declines to a level comparable to trust in politicians as the content becomes politicized. Individuals become more critical towards asylum policy after being exposed to hate speech about refugees. Moreover, the trust into the content about refugees is the highest when it covers neutral content about refugees. Trust in content about is lower for positive content and lowest for negative content about refugees. Importantly, the political orientation and populist attitudes affect the trust into the search engine content about refugees. Individuals who trust into hate speech are more hostile in terms of the immigration and asylum policy compared to those who do not trust into the content. The effect of hate speech is particularly apparent among people with a right-wing political identity and strongly populist attitudes. However, a positive speech about refugees in search engines can also lead to a backfire effect by provoking a more hostile attitude towards asylum policy. Critically, individuals having a right political identity and those with stronger populist attitudes, were also more likely to click on hate speech search and less likely to click on positive ones than their counterparts. Individuals with an extreme right political identity reinforced their restrictive attitudes towards immigration and asylum policy after being exposed to hate speech in search engines.

# Keywords

Search engines, hate speech, positive speech, attitudes towards refugees, asylum policy, polarization, extremism, trust

# Introduction

Nowadays, we are confronted with the absence of neutrality in technologies and algorithms which can lead to a new form of discrimination and social inequality. Especially search engines are frequently used in our everyday life, they are generally trusted, but they can also reflect derogatory content about certain groups. Research indicates that hate speech in general can affect political behavior. For example, 75 percent of European citizens who follow or participate in debates have experienced hate speech and about 50 percent indicate that it makes them hesitate to participate in debates (Special Eurobarometer 452 2016). But research on hate speech in search engines and its potential impact on political attitudes is scarce. Prior research provided first insights into differences in the portrayal of certain groups, such as sexist, racist (Noble 2018) or anti-Semitic content (Bar‐Ilan 2006) in search engines. However, effects of such underlying derogatory contents on political attitudes is unknown. On the other hand, how is the content processed if the sentiment of the content about refugees is positive (i.e. positive speech)?

I build on research on group priming to investigate how hate speech and positive speech in search engines influence political attitudes. Therein, I examine how trust, political ideology and populist attitudes influence the information processing of the derogatory content related to a minority group of our society. Specifically, the study focuses on derogatory content, but also approving content, about refugees because this group has been increasingly visible since the so-called “refugee crisis”. They have been frequently portrayed in the European press with narratives promoting suspicion, hate speech and hostility (Georgiou and Zaborowski 2017). Moreover, while recent refugees come most often from countries in which the Islam is practiced, the religion is mainly portrayed as being violent (Ahmed and Matthes 2017) and Muslims have been in several debates in the media (Amir-Moazami 2005). At the same time, there are also voices that draw attention to the positive aspects of refugees and benefits of immigration (Philo et al. 2013). In the digital age, an important information source is the internet, where people inform themselves increasingly, for example Google is the most visited website globally and people are increasingly informing themselves online (Dutton et al. 2017; Pradel 2020). But politicians are also an important information source on political issues, whose expressions find both direct access to the voter via social media such as Twitter (Wagner et al. 2017; Mertens et al. 2020) and indirect access through media reports (Philo et al. 2013).

By comparing search engines and politicians, I will provide insight into the importance of the communication channel that provides individuals with derogatory or positive content about refugees. I will focus on political orientation because previous research showed that the effects of group-related content can vary with individuals’ political orientation (Costello et al. 2019). Individuals may reinforce their political attitudes when confronted with derogatory content that could lead to a political polarization. But how is the content about refugees is processed if it is positive? Do individuals adopt the positive attitude or does it in fact backfire by leading to more hostile attitudes towards refugees? One of the most important findings is that hate speech can indeed lead to more restrictive attitudes towards asylum policy, which is particularly noticeable among persons who identify themselves as politically right-wing and have strong populist attitudes. As important, positive speech can have the unintended effect to backfire and to lead to even more hostile attitudes towards asylum policy. It is crucial to note that the study implies that search engines are more trusted than politicians, which underlines the importance of studying search engines and their influence on political attitudes. Although there is research investigating the existence of negative content about minorities in search engines, so far there has been no research focusing on the effects of such content on voters and the underlying mechanisms that shape the information processing.

# Biases in new media content about marginalized groups // Biases in traditional, new media and search enginesand in the case of Germany

Different opinions clash on certain topics, especially when it comes to refugees, like when surfing the internet, at WhatsApp or in offline discussions. Not only important politicians play a role here, but also new technologies that are increasingly becoming part of our everyday lives and provide us with information. However, there is still little research on how positive, neutral and negative expressions are processed by individuals with different characteristics, and especially how comparatively neutral perceived technologies like search engines play a role.

Prior research supports the assumption that individuals use stereotypes and categorize individuals into different groups to simplify the information processing (Allport 1958; Tajfel 1970). This can lead to intergroup bias and discrimination as individuals gain self-esteem by favouring people from their social group (ingroup), i.e. people with similar characteristics to themselves, and by derogating people outside this group (outgroup) (Tajfel 1970; Tajfel/Turner 1971; Billig 1973.). This indicates that they tend to see individuals of the group they identify as positive and those outside the group as negative (Billig 1973). This study will take a deeper look into the effects of hate speech about refugees – a form of derogation of an outgroup – in the new media platform search engines.

DEFINITION AND MAIN CONSEQUENCES: There is still controversy over the definition of hate speech and incivility. Much of the research defines hate speech as , while hate speech is defined as . In the following I will talk about positive, negative and neutral speech. By negative speech I mean ..., by positive speech and by neutral speech .... Research on the effects of hate speech and incivility towards minorities in a society shows that hate speech attitudes can reinforce negative prejudices and even aggressiveness (?) towards the minority. It can have devastating consequences for the minority group, because hate speech can lead to negative psychological consequences (e.g., depression, decreased self-esteem, social isolation), to work-related problems (Klaßen/Geschke 2019), political consequences (e.g., lower participation, shifts in policy preferences) (XXXX) and social consequences (e.g., lower willingness to donate to a refugee aid organization) (Ziegele).

Research showed that activated minority-related prejudices to social groups can have an impact on political attitudes. NOT NEW, EXPLANATION (ÜBERGANG FEHLT): EFFEKTE – von prejudice primen: For example, Valentino’s research (1999) showed that priming of racial attitudes with crime news led to decreasing support for the political candidate Clinton. This study showed that individuals’ political decisions are driven by simplifying shortcuts partly influenced by stereotypes and prejudices. Thus, activated minority-related prejudices to social groups can have an impact on political attitudes. Other research found that even subtle racial priming (statement having a typical Black name) increased criticisms about Barack Obama (Pyszczynski et al. 2010). Related work (Gilliam and Iyengar 2000) also revealed that minority-related primes can change attitudes towards a social group and the support for certain policies like punitive crime policy agenda. Furthermore, a study revealed that primed stereotypes like prejudices about African Americans and females can affect the interpretation of related and unrelated media events (Power, Murphy, and Coover 1996). Moreover, priming stereotypes can lead to increases in intergroup conflict (Hsueh, Yogeeswaran, and Malinen 2015). Additionally, hate speech in the online environment matters. Hateful online user comments provoked negative attitudes towards donating money to a refugee aid organization, but positive ones towards donating it to homeless people (Ziegele, Koehler, and Weber 2018).

SEARCH ENGINES AS A SPECIAL CASE of hate speech; Search engines are a new and frequently used medium that can confront users with stereotypes. However, previous research investigating the influence of web search engines on political attitudes and behavior focused mainly on selective exposure and personalization of search engines (Edgerly et al. 2014; Epstein and Robertson 2015; Muddiman 2013). There are few studies on hate speech and stereotypical information in search engines. These studies found that search engines are biased towards certain groups and provide stereotypical information.

Noble (2018) found several sexist and racist Google search results, images, image-labeling and images or even maps locations. For example, racial slurs redirected users to the White House during Obama's presidency and Google Photos categorized a photo of Blacks with “gorillas”. Other research indicates that search engines contained anti-Semitic (Bar-Ilan 20016), gender-stereotypcial (Otterbacher 2017-competent) and stereotypical and/or negative content for certain social groups (Baker and Potts 2013). Crucially, users having more stereotypes (i.e. gender stereotypes) about a group are also less likely to notice these biases (e.g. Otterbacher et al. 2018) and they even perceived the reality in a more stereotypical way (Kay, Matuszek, and Munson 2015).

While the precedent outline of related studies suggest that minority-related hate speech can affect individuals and their political attitudes, search engines as gatekeeper of information can play a crucial part when they confront individuals with minority-related hate speech. This might be especially problematic since research showed that negative stimuli attract and affect individuals in a larger extent (Soroka and McAdams 2015). Although search engines are frequently used, prior research showed the existence of biases towards minorities and suggests users’ attraction to it, there is no research that investigates how they affect political attitudes.

In this work, I will focus on this gap and define hate speech generally as derogating, hostility based language on features attributed to a group as a whole (Koltsova et al. 2017), i.e. negative stereotypes towards refugees.

# The importance of biased content about refugees in Germany

Germany offers a particularly good case to investigate the effects of hate speech, positive speech and neutral speech about refugees in search engines. Since the so-called "refugee crisis" refugees and migration as a topic have been in focus in the public sphere (Franzman et al 2019). It was the beginning of the rise of the right-wing populist party, which by comparison was strongly opposed to a culture of welcome and thus found support among parts of the population (e.g. Charles Lee's 2018; Mader/Schoen 2019)). An increasing anti-Islamic mood could be observed and there were public demonstrations by PEGIDA and AfD against immigration from Syria and other countries (Charles Lee; Deutsche Welle 2017). However, the electorate was rather divided on immigration, i.e. the anti-Islamic and anti-migration mood of the electorate was opposed to other parts which were more pro-immigration (Franzmann et al. 2019). The importance of the topic is particularly evident from the fact that since 2015 it has been mentioned by citizens as one of the most important problems in Germany (Forschungsgruppe Wahlen 2020). Thus, biased content on refugees in search engines should impact political attitudes of German citizens.

# Expected Effects of Hate Speech and Positive Speech on Political Attitudes

Building on the social identity theory and research on intergroup conflict (Tajfel 1970; Tajfel et al. 1971; Tajfel and Turner 1979), this research investigates how individuals respond when information in search engines contain derogating – compared to positive and neutral – information about minority groups. More precisely, the study allows to examine how content containing derogating – compared to positive and neutral – information about minority groups impact preferences for political policies (i.e. immigration and asylum policies). The social identity theory (Tajfel 1970; Tajfel et al. 1971; Tajfel and Turner 1979) was proposed to explain intergroup behavior. According to the theory, individuals are belonging to different social groups. Individuals categorize themselves and are also categorized by others into social groups. The theory further assumes that individuals derive their self-esteem by enhancing their belonging to a social group and by distancing themselves from out-group members. One proposition is that individuals favor members of their social group sharing similar characteristics such as ethnicity and gender (in-group favoritism) and engage in derogating others (out-group derogation). Individuals may adopt an negative stereotypical attitude about out-group members. Therefore, I expect individuals become more critical toward immigration policy (H1a) and asylum policy (H2a) after being exposed to hate speech.

On the other hand, it is also plausible that hate speech is not always processed in the same way. Rather, I expect the participants’ political identity to be crucial for how they process hate speech. Since liberal individuals tend to have more positive attitudes towards refugees and more positive attitudes towards immigration policies, it is plausible that they do not adopt the negative stereotypes to the same extent. I expect the negative effect of hate speech on attitudes towards the policies weaker for individuals with a liberal political orientation. Put differently, I hypothesize that individuals who place themselves more on the right on the political spectrum strengthen their preexisting attitudes towards immigration (H1b) and asylum policy (H2b) more strongly. Also, I expect the effect of hate speech on immigration (H1c) and asylum policy (H2c) to be stronger, when individuals indicate stronger populist attitudes. Thus, I expect that individuals tend to adapt hate speech towards a minority group and consequently develop more restrictive attitudes towards minority-related policies.

Similarly, I expect positive expressions related to the minority groups can be harmful too by provoking a backfire effect. Worldviews that are deviating from the own – for example, the world- view full of hate speech one might see in search engine suggestions – can be considered as a threat which provokes coping strategies (Bassett et al. 2015; Castano et al. 2011; Proulx and Major 2013). For example, one study showed that a threat caused by a violent conflict lead participants behave either pro-socially or antisocially depending on their preexisting meaning frameworks (Rovenpor et al. 2016). Positive expressions about refugees may provoke stress for individuals, who have a rather critical attitude towards refugees. Positive expressions about refugees could, therefore, cause negative effects on attitudes towards immigration (H3a) and asylum policy (H4a). Once again, I expect the political identity to be a crucial moderator. I hypothesize that positive expressions about a minority group provoke more stress and thus, more need for coping strategies. In line with the theory, I expect that individuals do this by strengthening their preexisting attitudes more strongly. Hence, I hypothesize that the effect of positive expressions about refugees on immigration (H3b) and asylum policy (H4b) is stronger, the more individuals’ place themselves on the right of the political spectrum. Similarly, I expect the effect of positive expressions on immigration (H3c) and asylum policy (H4c) to be stronger, the more individuals place themselves on the right of the political spectrum.

Besides the political identity, the source that provides individuals with hate speech may be an important component for the information processing. It may be central that search engines are generally perceived as neutral and their information is trusted. By comparing hate speech coming from another communication channel (i.e. a politician), the relative importance of a web technology compared to a human for the effects of hate speech can be estimated. Compared to the communication channel search engine, a political candidate may be perceived as less objective and less trusted. Research indicates that the technology sector and search engines are generally trusted, while there is a mistrust in politicians and the government (Edelman 2019). Other research showed that users implicitly tend to trust search engines results by using eye-tracking technology (Joachims et al. 2017). Individuals may intuitively perceive the search engines as neutral compared to a politician and trust their results. Consequently, they may adopt the negative stereotypes and react on it more strongly. Thus, I expect that content provided by a search engine is more trusted than the same content provided by a politician (H5a). I expect, therefore, that there is a stronger effect of hate speech (H5b) and positive expressions (H5c) among participants being assigned to a search engine compared to a politician that confronts individuals with the content.

To investigate these assumptions, I conduct an online experiment in which I prime participants in the experimental group with non-neutral (positive and negative) search engine suggestions and neutral suggestions (control group). I compare these experimental groups with other groups, in which I claim that the same expressions have been expressed by a political candidate.

# Procedure, Treatment Groups and Measures

# Design, Procedure and Measures

Before running the experiment, the ethic commission of the Faculty of Management, Economics and Social Sciences at the University of Cologne approved the study (approval number: 19020FP) and has been pre-registered on EGAP.[[1]](#footnote-2) I validated all presented expressions of the prime stimuli by human coders who rated a set of randomly presented expressions into either a neutral, negative or positive category.

***Participants***: 1200 participants (female=607, male=593) with a minimum age of 18 years have been recruited by an online panel (Lucid). The platform is comparable to Amazon Mechanical Turk and provides reliable and valid data for online experiments (Coppock and McClellan 2019). The recruited participants are representative for German citizens in terms of gender, age and education. I included an attention test by asking participants to click on a certain category to have a test of data quality. Participants who dud not answer correctly to the attention test have been excluded from the data analyses.

***Measures*** *[alternatively Materials]:*

*Sociodemographic information:* The first questionsasked participants about their age, gender and nationality.

*Political Orientation:* Based on threat-related research (Bassett et al. 2015; Castano et al. 2011; Proulx and Major 2013), the political orientation (or similar political identity-based attitudes (Rovenpor et al. 2016)) should be an important moderator when the derogation of the minority group is salient. I used a 10-point scale (1 = left; 10 = right) to measure the political orientation with a left-right self-placement, which is commonly used in surveys in Germany (Breyer 2015).

*Populist Attitudes:* Participants evaluated 9 items of the German version of the new populist attitudes scale (Silva et al. 2018) ranging from 1 (“strongly disagree”) to 10 (“strongly agree”). Two items were excluded because of their low correlation with the sub-dimension and other items of the scale (Q10\_8\_re muss raus, Q10\_5\_re). The scale includes items such as “*Politicians should always listen closely to the problems of the people*”, “*Politicians don’t have to spend time among ordinary people to do a good job”*, “*The government is pretty much run by a few big interests looking out for themselves*” and “*You can tell if a person is good or bad if you know their politics*”. The items have been summarized to a factor score (alpha=0.65) after recoding the reversed items.

*Priming the tone of the information:* (originally in German, see Figure 1 and 2): One third of the participants were randomly assigned to negative stereotypical expressions in German about refugees in order to make hate speech towards a minority group salient. All expressions have been validated by human coders: *“Refugees are criminal”*, *“Refugees are currently in the debate”*, *“Refugees are a danger”*, *“Refugees are less intelligent”* and *“Refugees cannot be integrated”*. The positive treatment group included expressions such as *“Refugees are peaceful”*, *“Refugees are currently in the debate”*, *“Refugees are a cultural enrichment”*, *“Refugees are intelligent”*, *“Refugees can be integrated”*. Expressions of the neutral treatment group were *“Refugees are in Germany”*, *“Refugees are currently in the debate”*, *“Refugees are a group of people”*, *“Refugees are diverse”*, *“Refugees are in Europe”*. All treatment groups include one expression being more neutral “*Refugees are currently in the debate*” to make the suggestions appearing more natural since in reality the suggestions are mixed and rarely solely negative nor positive. Moreover, the experiment contained an additional neutral group that was not refugee-related, rather it was more general to avoid priming a specific political issue. Participants were asked to imagine that a search engine (vs. a politician) provided content related to a political topic.

*Priming the communication source:*Half of participants were randomly assigned to a search engine as source for the political information. The expressions were introduced by *“Please imagine that a known search engine suggests the following content related to refugees:”*. The other participants were assigned to a hypothetical political candidate functioning as communication source. Here, the expressions were introduced by *“Please imagine that an elected person from the Bundestag addresses the following content related to refugees”*. Participants in the neutral experimental groups (not refugee-related) have been shown either *“Please imagine that a known search engine is suggesting various content related to a political topic”* or *“Please imagine that an elected person from the Bundestag is addressing various content related to a political topic”* without further information.

[Outcome Variables]

*Trust*: Participants indicated whether they trust into the information on a 10-point scale ranging from “not at all trustworthy” to “entirely trustworthy”.

*Attitudes toward immigration policy:* I took over items from the European Social Survey (ESS) measuring attitudes toward immigration policies (Jowell et al. 2020) to construct an sum index (alpha=0.9): “*Allow many/few immigrants of same race/ethnic group as majority*”, “A*llow many/few immigrants of different race/ethnic group from majority*”, “*Allow many/few immigrants from poorer countries in Europe*”, “*Allow many/few immigrants from poorer countries outside Europe*”.

*Attitudes toward asylum policy:* Participants answered to more questions (taken over from ESS (Prinz and Glöckner-Rist 2009) about attitudes towards the asylum policy, such as “*While their applications for refugee status are being considered, people should be allowed to work in Germany*”, “*The government should be generous in judging people’s applications for refugee status*”, “*While their cases are being considered, the German government should give financial support to applicants*”, “*Refugees whose applications are granted should be entitled to bring in their close family members*”, “*Germany has more than its fair share of people applying for refugee status*”, “*Most applicants for refugee status aren’t in real fear of persecution in their own countries*”, “*While their cases are being considered, applicants should be kept in detention centers*”. After recoding the reversed items, all seven items have been summarized to a sum score (alpha=0.81).

*Perceived polarization*: Participants then answered a question about the perceived divide in the German society.

*Internet skills:* Moreover, for exploratory reasons, additional questions were tapping internet skills, usage of internet in general and search engines, partly taken over or adapted from the German Internet Panel (Blom et al. 2018).

*Perception of treatment groups and media climate:* Finally, participants answered question about the treatment groups and how they perceive them, which of the suggestions they would like to click on to get the corresponding search results, and about the media climate in Germany. This has been done to explore the underlying mechanisms of information processing further.

Click-behaviour: Participants were asked at the end of the survey experiment which search suggestion they would like to click on to see in order to see the corresponding search engine results (randomized list of positive, neutral and negative search suggestions).

*Manipulation Check*: I used two different kinds of manipulation checks. First, the treatment groups have been validated before the experiment by human coders who categorized all single expressions of the treatment groups (randomized) into either negative, positive or neutral. Second, participants rated in the end of the online experiment to which extent the shown treatment contained positive or negative information. Specifically, participants did indicate on a scale ranging from 1 (very positive) to 11 (very negative) the level of tone displayed in the expressions related to refugees.

***Design and Procedure***: The experiment used a 4x2 (communication source: search engine vs. politician x tone: control content vs. neutral vs. positive vs. negative speech about refugees) between-subjects design. Individuals were told that they take part in a scientific study about actual topics in Germany and were asked to provide consent to participate. After answering questions about their age, citizenship, education and profession, they answered questions about their political orientation and populist attitudes. Respondents were then randomly assigned to one of the six experimental groups with content related to a minority group (i.e. refugees):

(1) Respondents seeing a hypothetical search engine presenting hate speech-related content, (2) a hypothetical political candidate presenting hate speech-related content, (3) a hypothetical search engine presenting positive content about refugees, (4) a hypothetical political candidate presenting positive content about refugees, (5) a hypothetical search engine presenting neutral content about refugees, (6) a hypothetical political candidate presenting neutral content about refugees, (7) a hypothetical search engine presenting positive content that is not refugee-related, (8) a hypothetical political candidate presenting neutral content that is not refugee-related (2 x 4-between-subjects design: tone of the content (hate speech vs. positive vs. neutral vs. neutral not-refugee-related) x communication channel (search engine vs. political candidate)).

Participants answered to the questions of interest concerning the trust in the information, attitudes towards immigration and perceived societal divide. Then, they gave additional information about their general online, search engine behavior and the perception of their treatment stimuli. Finally, all respondents were debriefed after the survey experiment, with describing the aim of the study and stressing out again that all scenarios were hypothetical, that the statements such as “refugees are criminal” were fabricated and not true. Participants were also provided with contact details for further questions.

# Results and Discussion

Tests approved that the manipulation worked effectively, as the positive manipulations were rated as significant more positive (p<0.05) and the negative ones as more negative (p<0.001) than neutral experimental conditions.

ATTITUDES TOWARDS IMMIGRATION AND ASYLUM INKLUSIVE RIGHT/LEFT EXTREME:

First, I used multiple linear regressions to analyze whether there are any direct effects of hate speech on attitudes towards immigration and asylum policy. It becomes apparent that – other than expected - hate speech and positive speech has no direct effect on attitudes towards immigration and asylum policy (see M1-M9 Table 1). However, it is crucial that individuals who have a right political identity, have more significant restrictive attitudes towards immigration and asylum policy in all models. Importantly, only extreme right (split at the upper 0.75-quartile) become more hostile in terms of immigration and asylum policy when they are confronted with hate speech about politicians. The same tendency, however nonsignificant, is observable for individuals with populist and extreme populist attitudes (split at the upper 0.75-quartile). (M3, M5; M8, M10)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 1. Predicted scores for attitudes towards immigration and asylum policy for individuals being exposed to hate speech, positive speech and neutral speech about refugees. | | | | | | | | | | |
|  | *Dependent variable:* | | | | | | | | | |
|  | Immigration Policy | | | | | Asylum Policy | | | | |
|  | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Negative | 0.049 | -0.053 | -0.069 | -0.053 | -0.084 | 0.008 | -0.256 | -0.112 | -0.256 | -0.049 |
|  | (0.087) | (0.147) | (0.124) | (0.147) | (0.154) | (0.114) | (0.205) | (0.165) | (0.203) | (0.210) |
| Neutral | 0.083 | 0.127 | 0.036 | 0.127 | -0.063 | 0.023 | 0.053 | 0.018 | 0.053 | -0.011 |
|  | (0.088) | (0.144) | (0.126) | (0.144) | (0.158) | (0.116) | (0.201) | (0.167) | (0.199) | (0.215) |
| Positive | 0.080 | 0.131 | 0.105 | 0.131 | 0.039 | 0.051 | 0.119 | -0.029 | 0.119 | -0.041 |
|  | (0.088) | (0.147) | (0.128) | (0.147) | (0.159) | (0.116) | (0.205) | (0.170) | (0.203) | (0.217) |
| Right |  | 0.645\*\*\* |  |  |  |  | 0.582\*\* |  |  |  |
|  |  | (0.131) |  |  |  |  | (0.183) |  |  |  |
| Negative x Right |  | 0.144 |  |  |  |  | 0.386 |  |  |  |
|  |  | (0.194) |  |  |  |  | (0.271) |  |  |  |
| Neutral x Right |  | -0.080 |  |  |  |  | -0.060 |  |  |  |
|  |  | (0.190) |  |  |  |  | (0.266) |  |  |  |
| Positive x Right |  | -0.115 |  |  |  |  | -0.100 |  |  |  |
|  |  | (0.193) |  |  |  |  | (0.270) |  |  |  |
| Populist |  |  | 0.112 |  |  |  |  | 0.152 |  |  |
|  |  |  | (0.125) |  |  |  |  | (0.166) |  |  |
| Negative x Populist |  |  | 0.255 |  |  |  |  | 0.271 |  |  |
|  |  |  | (0.176) |  |  |  |  | (0.234) |  |  |
| Neutral x Populist |  |  | 0.099 |  |  |  |  | 0.032 |  |  |
|  |  |  | (0.178) |  |  |  |  | (0.237) |  |  |
| Positive x Populist |  |  | -0.066 |  |  |  |  | 0.108 |  |  |
|  |  |  | (0.178) |  |  |  |  | (0.236) |  |  |
| Extreme Right |  |  |  | 0.703\*\* |  |  |  |  | 0.487 |  |
|  |  |  |  | (0.229) |  |  |  |  | (0.317) |  |
| Negative x Extreme Right |  |  |  | 0.522† |  |  |  |  | 1.235\*\* |  |
|  |  |  |  | (0.303) |  |  |  |  | (0.419) |  |
| Neutral x Extreme Right |  |  |  | 0.678† |  |  |  |  | 0.949 |  |
|  |  |  |  | (0.401) |  |  |  |  | (0.554) |  |
| Positive x Extreme Right |  |  |  | 0.054 |  |  |  |  | 0.119 |  |
|  |  |  |  | (0.320) |  |  |  |  | (0.442) |  |
| Extreme Populist |  |  |  |  | 0.346 |  |  |  |  | 0.338 |
|  |  |  |  |  | (0.322) |  |  |  |  | (0.440) |
| Negative x Extreme Populist |  |  |  |  | 0.227 |  |  |  |  | -0.010 |
|  |  |  |  |  | (0.520) |  |  |  |  | (0.710) |
| Neutral x Extreme Populist |  |  |  |  | -0.044 |  |  |  |  | -0.853 |
|  |  |  |  |  | (0.570) |  |  |  |  | (0.777) |
| Positive x Extreme Populist |  |  |  |  | 0.104 |  |  |  |  | 0.591 |
|  |  |  |  |  | (0.570) |  |  |  |  | (0.778) |
| Constant | 2.219\*\*\* | 1.843\*\*\* | 2.173\*\*\* | 1.843\*\*\* | 2.261\*\*\* | -0.020 | -0.394\*\* | -0.090 | -0.394\*\* | -0.026 |
|  | (0.062) | (0.094) | (0.092) | (0.094) | (0.118) | (0.081) | (0.132) | (0.122) | (0.130) | (0.161) |
| Observations | 602 | 422 | 584 | 218 | 240 | 602 | 422 | 584 | 218 | 240 |
| R2 | 0.002 | 0.176 | 0.023 | 0.277 | 0.023 | 0.0004 | 0.103 | 0.019 | 0.215 | 0.017 |
| Adjusted R2 | -0.003 | 0.162 | 0.011 | 0.253 | -0.006 | -0.005 | 0.088 | 0.007 | 0.189 | -0.013 |
|  | | | | | | | | | | |
| *Note:* | †p <0.1; \*p<0.05; \*\*p<0.01; \*\*\*p<0.001 | | | | | | | | | |

TRUST INTO THE CONTENT, COMMUNICATION SOURCE: Next, I used multiple linear regressions to analyze whether the participants trust more into the content provided by a search engine and compare it to content provided by a politician, which is a popular communication source (see Table 2, left side of Figure 1). Overall, it revealed that participants trust generally most into the content provided by search engines and politicians when it contains neutral statements about refugees, as well as neutral content about a general political topic. They trust the content significantly much less if it contains positive statements about refugees (p<0.01), and they trust significantly least in negative content about refugees (p<0.001). Remarkably, as expected, the general trust in the communication source is higher for the search engine than for politicians (see Table 2, right side of Figure 1). The largest difference between search engines and politicians in trust is found in the control group and the neutral group, with by far more trust being placed in search engines than in politicians (see Table 2, general trust M4). The general trust in politicians remains almost unchanged, regardless of whether the participants have previously been presented with positive, neutral or negative content by a politician. Meanwhile, the general trust in search engines is significant less when participants are confronted with both positive content and negative content about refugees in search engines (p<0.001). The general trust in search engines, however, always remains higher than the general trust, both when they provide hate speech and positive content about refugees. Overall, it appears that individuals seem to trust search engines to a large extent, but once there is a positive or negative bias, they perceive them as politicized and general trust in search engines moves to a similar level as general trust in politicians. Interestingly, this reflects the results of a recent study and new study that combines tracked online behavior with fake news survey results (((kann ich in dem anderen Paper als Ausgangspukt nehmen,dass es auch ein Massmedium ist))// Interestingly, this finding on trust in search engines reflects the results of a recent study (Ogyanova) on exposure to misinformation and trust in mainstream media in general, which found that trust in mainstream media declines over time when online users were exposed to tracked online misinformation.

results/analyses_final/plot_trust.pdfresults/analyses_final/plot_generaltrust.pdfFigure 1. Predicted scores for trust in the content and general trust in the communication source for the experimental treatments search engine (vs. politician) and hate speech (vs. neutral vs. positive expressions) and their interactions.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | | | |
| Table 2. Predicted scores for trust into the content and general trust into the communication source for individuals being exposed to hate speech, positive speech and neutral speech about refugees. | | | | |
|  | *Dependent variable:* | | | |
|  |  | | | |
|  | Trust into the content | | General trust into the source | |
|  | (1) | (2) | (3) | (4) |
|  | | | | |
| Neutral | 0.271 | 0.344 | 0.084 | -0.086 |
|  | (0.203) | (0.288) | (0.189) | (0.263) |
|  |  |  |  |  |
| Positive | -0.611\*\* | -0.409 | -0.374\* | 0.007 |
|  | (0.203) | (0.289) | (0.189) | (0.264) |
|  |  |  |  |  |
| Negative | -1.661\*\*\* | -1.658\*\*\* | -0.383\* | -0.134 |
|  | (0.202) | (0.289) | (0.188) | (0.264) |
|  |  |  |  |  |
| Search engine |  | 0.147 |  | 0.921\*\*\* |
|  |  | (0.287) |  | (0.262) |
|  |  |  |  |  |
| Neutral x Search engine |  | -0.143 |  | 0.372 |
|  |  | (0.407) |  | (0.372) |
|  |  |  |  |  |
| Positive x Search engine |  | -0.402 |  | -0.753\* |
|  |  | (0.407) |  | (0.372) |
|  |  |  |  |  |
| Negative x Search engine |  | -0.008 |  | -0.497 |
|  |  | (0.405) |  | (0.370) |
|  |  |  |  |  |
| Constant | 5.638\*\*\* | 5.564\*\*\* | 5.492\*\*\* | 5.027\*\*\* |
|  | (0.143) | (0.204) | (0.133) | (0.187) |
|  |  |  |  |  |
|  | | | | |
| Observations | 1,200 | 1,200 | 1,200 | 1,200 |
| R2 | 0.082 | 0.083 | 0.008 | 0.040 |
| Adjusted R2 | 0.080 | 0.078 | 0.006 | 0.034 |
|  | | | | |
| *Note:* | \*p<0.05; \*\*p<0.01; \*\*\*p<0.001 | | | |

IMI ASYLUM/TRUST

As we have seen trust into the search engine plays a significant role and recent research suggests that political trust is moderated by individual’s political orientation by showing that only conservative US citizens increased their trust into the US government while it was decreased for strong liberal ones (Ogyanova). The political orientation may also play a role for trust into search engine content. Therefore, I take a deeper look into the trust in the content to examine whether it varies with the participants’ political orientation. It becomes apparent that hate speech is trusted least but the trust is much higher into hate speech when individuals have a right political identity (see Model 1, Table 3). Populist individuals (see Model 2) do also trust more into hate speech than less populist individuals, but the difference is not statistically significant. Populist individuals trust significant less into positive speech than less populist individuals. Knowing that the political identity and populist attitudes affect trust into the content about refugees, I analyzed whether trust into the content affects how hate speech about refugees affects individuals. Those who trust into the content become more restrictive towards immigration (Model 3) and asylum policy (Model 3-4), see also Figure XYZ), however, the effect is insignificant for asylum policy in Model 5 . Those who trust into positive speech about refugees become significant less restrictive towards immigration and asylum policy when being exposed to positive search suggestions about refugees than individuals that don’t trust the content (Model 3-4). Figure XYZ visualizes these findings by looking only at individuals who trust the search content (trust above 5).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | | | | |
| Table 3. Predicted scores for attitudes towards immigration and asylum policy for individuals who trust into the content | | | | | |
|  | Dependent variable: | | | | |
|  |  | | | | |
|  | trust | | Immigration Policy | Asylum policy |  |
|  | (1) | (2) | (3) | (4) |  |
| Negative | -2.432\*\*\* | -1.881\*\*\* | -0.298 | -0.627\* |  |
|  | (0.515) | (0.393) | (0.207) | (0.272) |  |
|  |  |  |  |  |  |
| Neutral | 0.463 | 0.450 | 0.472 | 0.759\* |  |
|  | (0.504) | (0.397) | (0.251) | (0.330) |  |
|  |  |  |  |  |  |
| Positive | -0.774 | 0.078 | 0.663\*\* | 0.567 |  |
|  | (0.515) | (0.405) | (0.220) | (0.289) |  |
|  |  |  |  |  |  |
| Right | 0.175 |  |  |  |  |
|  | (0.458) |  |  |  |  |
|  |  |  |  |  |  |
| Negative x Right | 1.392\* |  |  |  |  |
|  | (0.681) |  |  |  |  |
|  |  |  |  |  |  |
| Neutral x Right | -0.220 |  |  |  |  |
|  | (0.667) |  |  |  |  |
|  |  |  |  |  |  |
| Positive x Right | -0.276 |  |  |  |  |
|  | (0.676) |  |  |  |  |
|  |  |  |  |  |  |
| Populist |  | 0.184 |  |  |  |
|  |  | (0.396) |  |  |  |
|  |  |  |  |  |  |
| Negative x Populist |  | 0.529 |  |  |  |
|  |  | (0.557) |  |  |  |
|  |  |  |  |  |  |
| Neutral x Populist |  | -0.576 |  |  |  |
|  |  | (0.565) |  |  |  |
|  |  |  |  |  |  |
| Positive x Populist |  | -1.513\*\* |  |  |  |
|  |  | (0.564) |  |  |  |
|  |  |  |  |  |  |
| Trust |  |  | 0.022 | -0.007 |  |
|  |  |  | (0.030) | (0.039) |  |
|  |  |  |  |  |  |
| Negative x Trust |  |  | 0.095\*\* | 0.154\*\* |  |
|  |  |  | (0.036) | (0.048) |  |
|  |  |  |  |  |  |
| Neutral x Trust |  |  | -0.067 | -0.124\* |  |
|  |  |  | (0.041) | (0.053) |  |
|  |  |  |  |  |  |
| Positive x Trust |  |  | -0.115\*\* | -0.107\* |  |
|  |  |  | (0.038) | (0.049) |  |
|  |  |  |  |  |  |
| Constant | 5.537\*\*\* | 5.588\*\*\* | 2.092\*\*\* | 0.022 |  |
|  | (0.331) | (0.291) | (0.179) | (0.235) |  |
|  | | | | | |
| Observations | 422 | 584 | 602 | 602 |  |
| R2 | 0.107 | 0.107 | 0.079 | 0.085 |  |
| Adjusted R2 | 0.092 | 0.096 | 0.068 | 0.074 |  |
|  | | | | | |
| Note: | †p<0.1; \*p<0.05; \*\*p<0.01; \*\*\*p<0.001 | | | | |

Finally, I conducted a single logistic regression to predict clicks on negative search suggestions about refugees compared to neutral and positive suggestions (see Table 4, Figure 2). The results indicate that the political identity predicts significantly the likelihood to click on a negative search suggestion and a positive search suggestion. Having a right political identity versus having a left political orientation  changes the log odds of clicking on a negative search suggestion significantly by 0.671 (p<0.001). In contrast, having a right political identity versus having a left political orientation decreases the log odds of clicking on positive search suggestions significantly by -0.763 (p<0.001). Having strong populist attitudes versus having a low populist attitudes increases the log odds of clicking on negative search suggestions significantly by 0.348 (p<0.05). However, it decreases the log odds of clicking on positive search suggestions significantly by -0.367 (p<0.01).

Figure 7 shows the probabilities for clicking on negative and positive search suggestions by the political identity of individuals. Here it is shown that the probability to click on negative search suggestions is about twice as high for the users with a right identity than for those with a left one. Leftists are almost twice as likely to click on positive search engine suggestions.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | | | |
| Table 4. Logistic regression results for clicks on hate speech suggestions and positive search suggestions. | | | | |
|  | Dependent variable: | | | |
|  |  | | | |
|  | Negative Clicks | | Positive Clicks | |
|  | (1) | (2) | (3) | (4) |
|  | | | | |
| Right | 1.340\*\*\* |  | -1.065\*\*\* |  |
|  | (0.185) |  | (0.161) |  |
|  |  |  |  |  |
| Populist |  | 0.348\* |  | -0.367\*\* |
|  |  | (0.139) |  | (0.132) |
|  |  |  |  |  |
| Constant | -1.852\*\*\* | -1.331\*\*\* | -0.388\*\*\* | -0.772\*\*\* |
|  | (0.159) | (0.102) | (0.111) | (0.089) |
|  |  |  |  |  |
|  | | | | |
| Observations | 814 | 1,152 | 814 | 1,152 |
| Log Likelihood | -448.785 | -632.280 | -459.300 | -678.682 |
| Akaike Inf. Crit. | 901.569 | 1,268.561 | 922.599 | 1,361.365 |
|  | | | | |
| Note: | \*p\*\*p\*\*\*p<0.001 | | | |

results/clicks_hatespeech.pdfresults/clicks_positivespeech.pdf

results/clicks_hatespeech_populist.pdfresults/clicks_positivespeech_lesspop.pdf

Figure 2. By political identities: Probabilities for clicking on hate speech and positive speech in search engines.

# Discussion and Conclusion

es zeigt sich, dass es potential hat gezielt Personen auf die Riskien von Suchmaschienn aufmerksam zu Machen, da hezigt wurde, dass besoners die gefährdet sind, die stark vertrauen [main results] The results show that the information source has a large effect on how hate speech, positive speech and neutral speech are processed and what effects they have on individuals. Importantly, trust in search engines is very high and is generally higher compared to politician, which is also a popular communication source about refugees. The trust in the search engine, however, declines to a level comparable to trust in politicians as the content becomes politicized by having positively biased content or negatively biased content. Individuals trust most into the content about refugees when it is neutral content, followed by trust in positive content about refugees and the lowest trust in negative content about refugees. Trust into hate speech is associated with higher hostility towards refugees, that is, being more restrictive in terms of the immigration and asylum policy. In general, individuals with participants with a right political identity were significantly more restrictive in terms of immigration and asylum policy. Moreover, those who have an extreme right political identity reinforced their hostile attitude towards refugees and supportive actions after being exposed to hate speech in search engines. They are already more hostile towards refugees and supportive actions, and this attitude is being reinforced. (/They already have a high level of restrictive attitudes and these are being reinforced.) It could indicate that negative content about refugees triggers for these individuals certain negative opinions (e.g. criminal; asylum fraud) and this is then generally reflected in reinforced restrictive attitudes towards asylum policy.

Political identities, i.e. whether people are more likely to identify themselves as left or right-wing, and populist attitudes of individuals also play a significant role in how individuals interact with hatred and positive expressions about refugees. Right-wing participants were more than twice likely to click on hate speech than left-wing participants and participants with populist attitudes were more likely to click on it compared to those with weaker populist attitudes. In contrast, left-wing and individuals with rather weak populist attitudes were more likely to click on positive suggestions instead of hate suggestions about refugees in order to see the related search results.

Positive speech is like hate speech part of everyday (media) communication, but it is not yet very well researched. It should be noted that positive speech in search engines can also lead to a backfire effect by causing more restrictive attitudes towards asylum policy. (This is especially the case for populist and left-identity) The importance of positive speech is also evident in the further exploration, where it is shown that citizens perceive society most divisively when they are confronted with positive expressions about refugees in search engines.

The study gives a further indication of how polarized society in Germany is in general when it comes to politics, especially when it comes to refugees. This should also be examined more closely in future research to what extent current political issues divide society and this can lead potentially to radicalization and conflict. This has been demonstrated not least by the recent developments on the Covid pandemic, where there were also divided opinions and a shift from health experts and politicians towards conspiracy theories (BBC News 2020; Brennen et al. 2020; McConald-Gibson 2020). [implications] In the digital age, it is not only important for the society to investigate biases in search engines. It is also worthwhile for search engine providers like Google to pay attention to biases in search engine content. The results have shown the general confidence into search engines is the highest for neutral content. It could also be useful to deliberate whether search suggestions should appear at all with regard to social groups. This is particularly important because of the recent increase in xenophobia (Decker and Brähler 2018).A good sign is that participants seem to perceive both positive and negative formulations and their confidence in the information is adjusted.

Future research should focus on whether positive, negative and neutral speech has an impact on different social groups and the extent to which contagious effects occur between groups (i.e. spill over). This means that if there is hatred against refugees, attitudes and behavior towards other groups can be influenced, as research suggests (Ziegele et al. 2018).

[abschlusssätze, 3 main take-awyas: max 3 Sätze: 1) Trust höher bei S2) Hate speech: Verschlecjterung 3) Poistive kann backfiren]

To conclude, hate speech can lead to more restrictive attitudes towards asylum policy, but positive speech can backfire by also leading to more restrictive attitudes. That is why particular caution is needed when providing information on a political issue. In particular, technologies that are perceived as neutral play a major role here, as it has been shown that they are by far more trusted and can lead to reinforced political attitudes (while participants among politicians tended to distance themselves). The nature of the information source and search engines are still an under-researched topic, although they are part of our everyday life and are the most used platform to search for information (e.g. Google as the world's most visited website (Pradel 2020)). Critically, individuals having a right political identity and those with stronger populist attitudes, were also more likely to click on hate speech search and less likely to click on positive ones than those with a left political identity and weaker populist attitudes. This is particularly appalling, since for these groups even hate speech has led to hostile attitudes towards refugees. Once people take the path towards hate speech, this could be the start of them getting caught in this filter bubble and confronting themselves mainly with selected negative information.

Different information sources and the interaction with different social identities should be considered in future research on information processing and opinion formation. This is even more important because, as shown here, it could lead to a polarization of society and to extremism. In view of the recent and alarming right-wing terrorist attacks like the shootings in Hanau (Hoffman et al. 2020; Institute for Economics & Peace - **2019**), the need to counteract information structures that promote xenophobia is once again becoming apparent. It is evident that extreme right-wing citizens and individuals become more hostile towards refugees, when they are confronted with negative speech in search engines, and users may turn to fake news and conspiracy theories. This suggests that we need to rethink and develop a strategy for sharing content about minorities, who are already vulnerable and disadvantaged.

**Literature**

1. The pre-registration is available on <http://egap.org/registration/6647>. [↑](#footnote-ref-2)