Scientific Report

Scientific Evaluation of the
Impact of the Online Campaign

PRECOBIAS

#thinkaboutit

Prevention of Youth RadicalisationThrough Self-Awareness of Cognitive Biases

Brigitte Naderer & Diana Rieger
Department of Media and
Communication; LMU Munich

Content

1. Introduction

1.1 Purpose of the Campaign

PRECOBIAS (Prevention of Youth Radicalisation Through Self-Awareness on Cognitive Biases) is a user-centered counter-narrative campaign project that focuses on the role of mental processes and cognitive biases when adolescents, especially the ones either vulnerable to radicalization or already radicalized, are faced with radical and violent content on social media.

PRECOBIAS aims to counter radicalization in the long term by enhancing adolescents' digital resilience and critical thinking. PRECOBIAS wants to help adolescents to get to know themselves better by revealing the mental processes and cognitive biases that underlie their everyday (often unconscious/automatic) thinking patterns.

To reach these objectives, PRECOBIAS targets **vulnerable and radicalized youngsters** directly through:

- A social media campaign, where the content is presented on a **website**
- Ten topical **videos**
- A quiz that is designed to increase self-awareness
- An Instagram **contest**

The campaign aims:

- To enhance **digital resilience** and **critical thinking** of the adolescents by focusing on their structures of thinking, instead of stigmatizing extremist narratives, which can be counterproductive.
- To help adolescents to **increase self-awareness** by revealing the mental processes and cognitive biases that underlie their everyday (often unconscious/ automatic) thinking patterns.
- To **counter radicalization processes** in the long term.

1.2 Content of the Campaign

The PRECOBIAS campaign is based on ten cognitive biases that research deems important in the context of radicalization and social media.

1. **Authority Bias:** This bias refers to our tendency to regard the opinions and instructions of an authority figure as highly influential, which is why we are more inclined to follow these instructions. This is why TV commercials use doctors to

appeal to the persuasive potential of an authority figure. The authority figure is a relevant actor in radical narratives which often claims the necessity of a strong, decisive leader (Bouko et al., 2021a, 2021b). Claiming this authority might lead to persuasive effects within those subjected to the content shared by radical groups.

2. **Bandwagon Effect:** This phenomenon describes the effect that the rate of adopting beliefs, opinions, and ideas increases the more they have been adopted by others. In other words, if we come to the belief that a certain opinion is very popular, we tend to join in on this opinion so as to be part of the "winning team". This phenomenon can, for instance, be helpful to political parties or candidates in an election race (Barnfield, 2020) or might also be useful when trying to engage new recruits with a radical group by claiming the majority view.

3. **Confirmation Bias:** This bias explains the tendency to search for, favor, and interpret information in a way that affirms our existing beliefs and opinions. People display this bias when they gather or remember information selectively or when they interpret it in a biased way. The effect tends to be stronger when we already have the desired outcome in mind or for emotionally charged issues and beliefs. Thus, people tend to confirm themselves in their beliefs, which might contribute to radicalization dynamics (MacDonald & Whittaker, 2019).

4. **Hostile Media Effect:** This effect refers to the tendency of individuals with a strong pre-existing attitude on an issue to perceive media coverage as biased against their own views and in favor of their antagonists' point of view. For instance, both republicans and democrats tend to describe mainstream media biased against their opinions (Lee et al., 2018). Hostile media perceptions are linked to distrust in mainstream media, and more openness to populist views (Schulz et al., 2020), which might foster radicalization.

5. **Humor Effect:** This effect causes people to remember information better when they perceive the information as humorous. For example, a teacher could use the humor effect to help students learn a certain concept by illustrating this concept using a funny story. Radical groups use this bias, for instance, by posting memes, thus employing humor as a gateway for radicalization processes (Frischlich, 2021) and making viewers aware and open to their messages.

6. **Ingroup-Outgroup-Bias:** This bias describes a pattern of favoring members of one's ingroup over outgroup members. This can be expressed in the evaluation of others, e.g., by assuming that an outgroup is a homogenous entity and generally is assessed less positively. The ingroup-outgroup bias can be commonly found

in the content shared by radical groups as they often make use of an "us versus them" narrative (Bouko et al., 2021a).

7. **Negativity Bias:** This effect describes the notion that, even when of equal intensity, information of a more negative nature has a greater effect on our psychological wellbeing and memory than neutral or positive information. For instance, if we receive twenty compliments and one harsh critique about an assignment, the critical remark will stick more to our memory and affect our mood and actions to a greater extent than the compliments. The salience of negative information can also play into the hands of radical groups when it comes to what information sticks and what mobilizes people (Bar-Tal et al., 2007).

8. **Picture Superiority Effect:** Pictures and images are often more likely to be remembered than words and can help make a piece of information memorable. The effect is explained by human memory being extremely sensitive to the symbolic modality of presentation. Radical groups might use the picture superiority effect to their advantage by curating highly aesthetic Instagram feeds (Frischlich, 2021) to appeal to a bigger group of people and thus find new recruits.

9. **Rosy Retrospection:** This phenomenon refers to our tendency to disproportionately judge the past more positively than the present. Rosy retrospection is therefore very closely related to the concept of nostalgia. Right-wing radicaliats, for instance, tend to make claims about how things have been so much better in the past (Menke & Wulf, 2021). We are inclined to believe these statements as we are all prone to distort the past in a positive way.

10. **Sleeper Effect:** This effect describes the tendency to forget where, when, or how previously learned information has been acquired while retaining the factual knowledge. For instance, message information sticks to our memory, but we forget where we retrieved this information. This way, disinformation, exaggerated numbers, etc., might stick in our memory, and we "forget" to be critical about them, as we have forgotten about the reliability of the source (Reuter et al., 2020). Being guided by disinformation in the decision-making process is potentially harmful for the democratic process and might fuel radical views.

1.3 Campaign Instruments

The campaign includes four main instruments, which are all interconnected.

1. **Brain.fail Website:** The brain.fail website is the landing page for all people made aware of the campaign. Here the videos and the quiz are linked. In addition, detailed information about the biases is provided here. The text explains how biases might affect people but also how they could relate to radicalization processes.
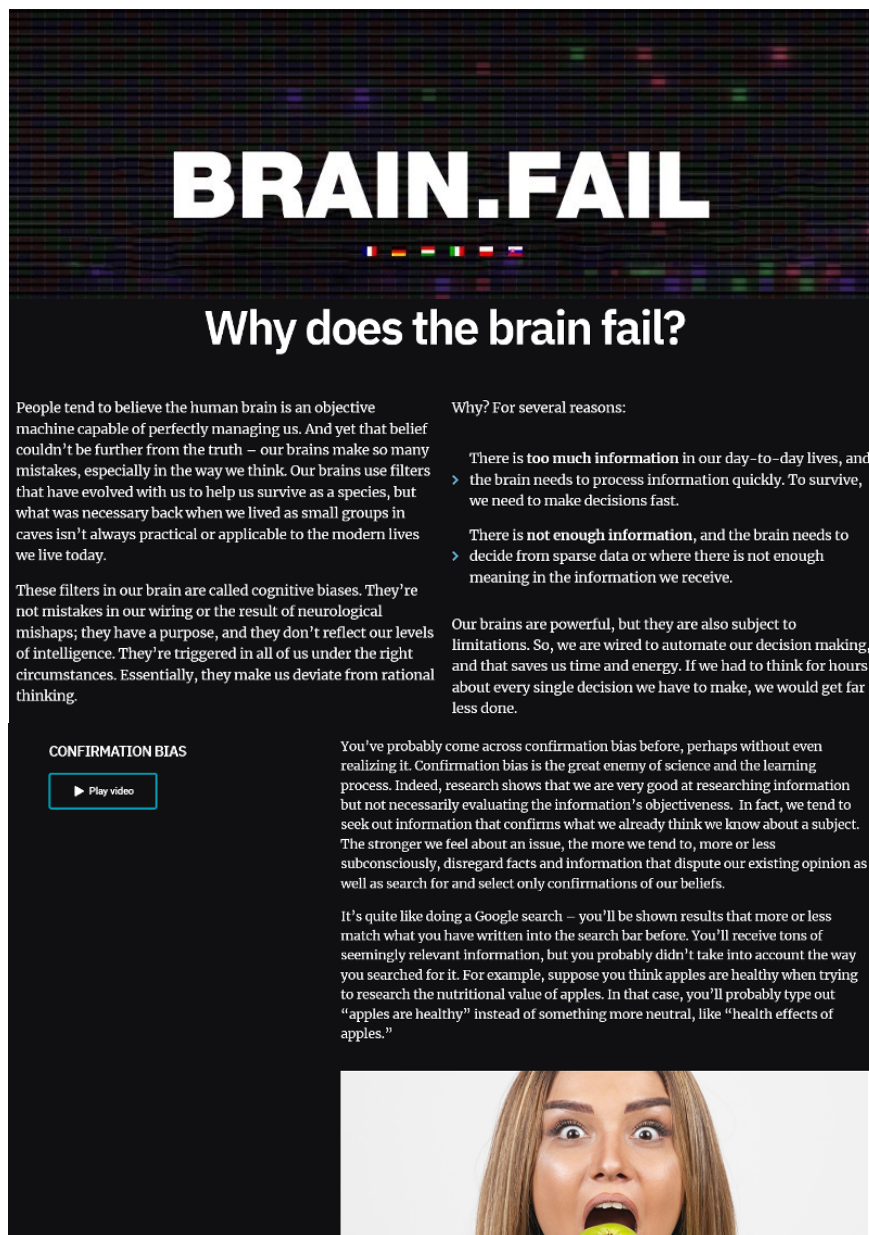
2. **Ten YouTube Videos**: The YouTube videos provide short (31-60 sec.) explanations for all biases. They vary stylistically, with some videos being more colorful in comic style, in a more simplistic black and white optic, or some that provide mostly textual explanations about the biases. The videos showcase examples, speak about day-to-day occurrences of biases but do not explicitly mention radicalization to show the relevance of biases in everyday life and not to deter affected audiences from the campaign content.

3. **Quiz**: The quiz was designed to help participants confront their own biases. Here, participants answer a series of questions associated with certain biased views and behaviors, such as a hypothetical scenario of a past vacation in which good and bad things happened, and what aspect participants are likely to focus on when talking about the vacation later on (referring only to positive aspects would suggest rosy retrospection)

4. **Contest:** Finally, the campaign includes an online contest that invites participants to submit their own videos or posts about biases. These posts could unmask biased behavior, highlight the consequences of biases, or show possible tricks to avoid biases. The contest was shared on social media, and participants were entered into a drawing for prizes for their submissions.
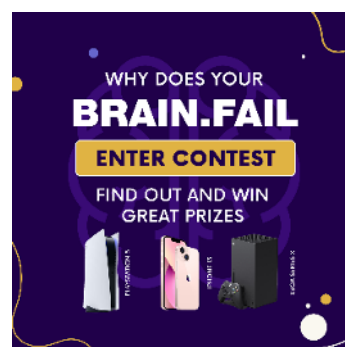
1.4 Objectives of the Campaign

Through the campaign, the adolescents should:

- Become acquainted with the topic of the campaign and find out more about their possible personal biases.
- Sharpen their critical thinking by being able to correlate forms of radicalization to their potential impacts on their thinking and feeling patterns.
- Shift from passive receivers to critically engaged participants in online discourses, thanks to a user-centered contest.

2. Campaign evaluation

In order to learn about the effectiveness of the online campaign, it is scientifically evaluated. The evaluation tackles all four instruments of the campaign through three different studies.

1. We conducted a **video and web-campaign evaluation study** with adolescents from eight European countries of the website, the videos, and the quiz. Here we examined whether the campaign increases topical awareness for radicalization and cognitive biases, the cognitive and evaluative assessment of the material, and whether the campaign material affects the perceived relevance of the topic.
2. We conducted an **effect study** with adolescents of the videos and the quiz. Here we examined whether the campaign was able to significantly increase literacy about biases, the adolescents' confidence in being able to recognize biases, and a pre-post measurement of radical attitudes.
3. Finally, we conducted a **qualitative workshop evaluation** of the contest in a school. Here we examined wether the campaign was able to engage adolescents as active recivers and participants. Furthermore we investigated how the contest is evaluated and aimed to generate qualitative assessments of how the contest was perceived.

3. Video and web campaign evaluation

3.1 Sample

The evaluation study was conducted online via the survey institute respondi in **eight European** countries: Austria (n = 73; 7%), Germany (n = 172; 16.5%), France (n

= 127; 12.2%), Hungary (n = 125; 12.0%), Italy (n = 155; 14.9%), Poland (n = 139; 13.3%), Slovakia (n = 139; 13.3%), and the UK (n = 113; 10.8%). The survey was conducted from September 24 to October 4, 2021. A total of 1,043 adolescents (16-22 years old; M = 19.73; SD = 1.61) were interviewed (48.1% male; 50.0% female; 1.9% other). With regard to education, 26.2% (n = 273) had no school-leaving qualification, completed lower-secondary education, or vocational school education, 52.0% (n = 542) completed secondary education, 19.8% (n = 206) completed university education, and 2.1% (n = 22) did not want to answer this question.

We started by asking **how familiar our respondents** were with **content that incites violence and radical views**. Overall, **familiarity** with problematic online content was rather high, with a mean of 60.7%. Adolescents were most familiar with content that is racist or xenophobic (n = 756; 72.5%); followed by posts that hostile people based on gender and/or sexual identity (n = 735; 70.5%); and by posts that explicitly incite violence (n = 719; 68.9%). Online content promoting violent jihad was least familiar to our respondents (n = 320; 30.7%; see Figure 1).
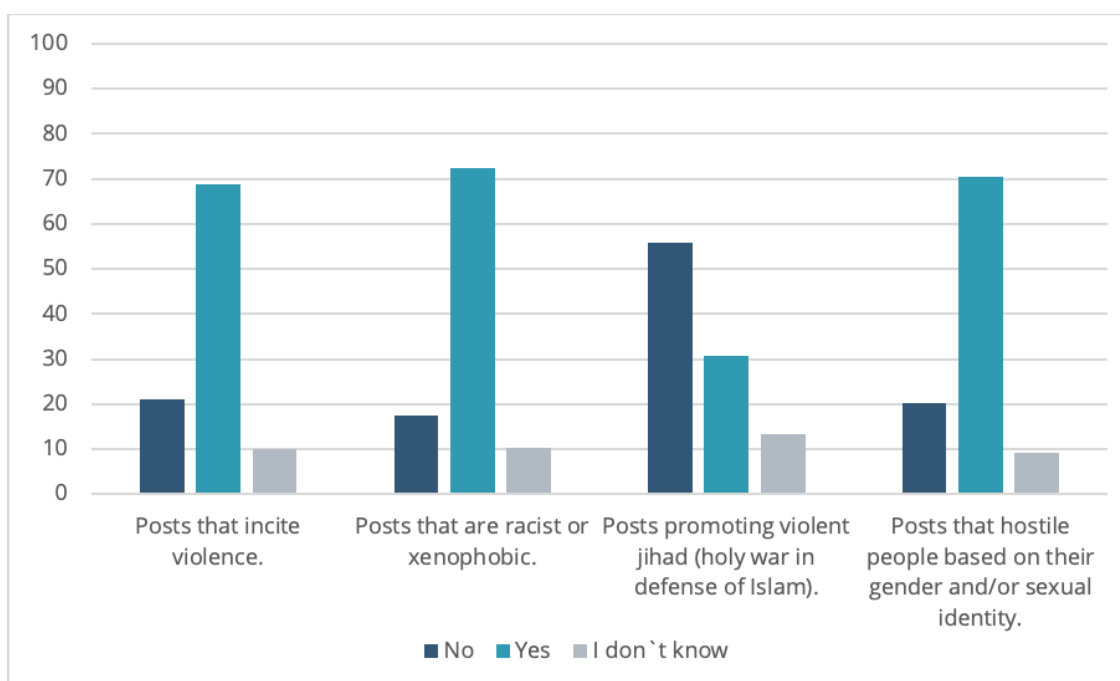

Figure 1: Familiarity with Violent and Radical Online Content

We then inquired about **how adolescents would deal with these postings**. Participants' willingness to report postings was much higher than their willingness to actively participate in counter speech. Their willingness to conduct counter speech against explicitly violent postings was particularly low, with only 13.8% willing to comment in defense of the person being attacked in any case. Willingness to report the content was much higher: between 38% and 45.4% being willing to report the author or the post in any case (see Figure 2 for more information).
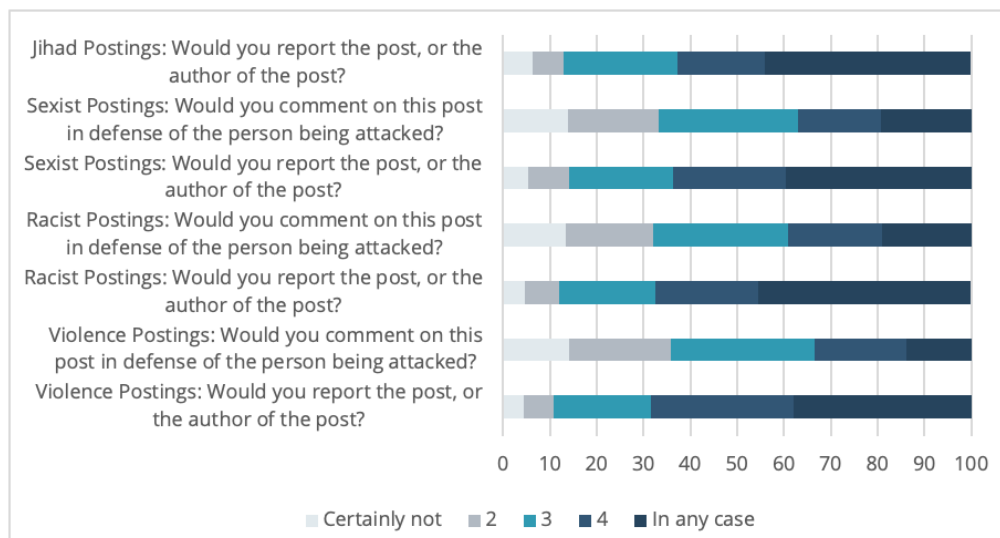

Figure 2: Familiarity with Violent and Radical Online Content

We furthermore asked about the **participants' predisposition to radicalization** (mean index based on 3 items assessed on a 5-point scale; e.g., "I can understand when people resort to violence to achieve political goals, even in Germany/Austria.", Reinemann et al., 2019; M = 2.31; SD = 1.13; α= .76). The **predisposition was generally rather low**. Over 30% showed a low predisposition to radicalization, while only 3.6% were very highly predisposed to radicalization (see Figure 3).
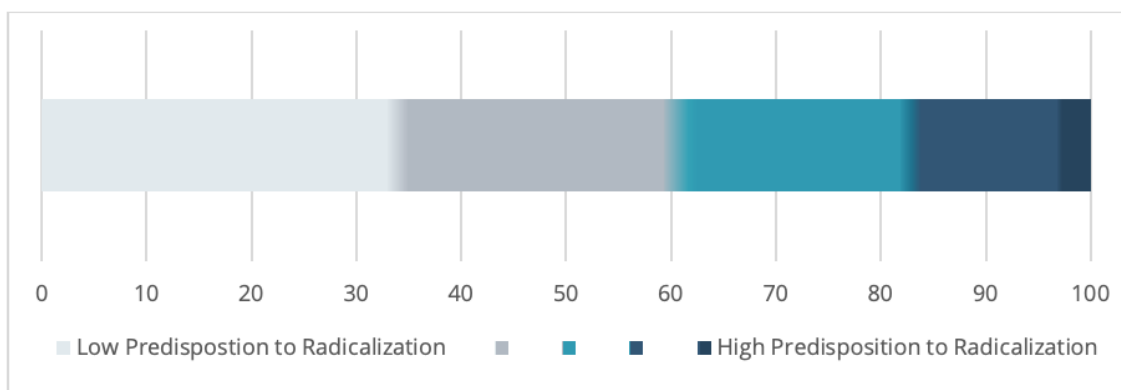
Figure 3: Predisposition to Radicalization

3.2 Procedure

After participants had answered these questions about sociodemographic data, their familiarity with problematic online content and how they would deal with this content, as well as their predisposition to radicalization, we either showed them the campaign website with a randomized selection of three cognitive biases descriptions (n = 354), a random selection of three videos[1] (n = 369;), or we asked them to fill out the quiz (n = 320).

We then examined whether the campaign elements succeeded in **creating awareness** for the campaign topics (radicalization and cognitive biases) by asking about which topics the campaign, in their opinion, was able to address. A list of seven topics was provided: politics, sports, nutrition, cognitive biases, radicalization, sustainability, psychology. Indications (1 = Yes, this was thematized in the campaign; 2 = No, this was not thematized in the campaign) about radicalization and cognitive biases were considered for the evaluation study.

In the next step, we examined how the **campaign was evaluated on three different dimensions:**

1. Intention to share the campaign material; based on 4 items for each instrument assessed on a 5-point scale (1 = disagree; 5 = agree); e.g., "I would recommend these videos/this website/this quiz to others.", (Boerman et al., 2017; M = 3.39;

[1] **Note:** It was not possible to expose participants to all campaign videos because we had to adhere to time constraints of a maximum of 15 minutes.

SD = 1.06; α = .84).

2. Evaluation of professionalism and information level of the content; based on 3 items for each instrument assessed on a 5-point scale (1 = disagree; 5 = agree); e.g., "The videos/this website/this quiz were/was informative.", (Frischlich et al., 2017; M = 3.70; SD = 0.96; α = .79).

3. Evaluation of the content being interesting and exciting; based on 4 items for each instrument assessed on a 5-point scale; e.g., "The videos/this website/this quiz were/was 1 = boring; 5 = exciting.", (Matthes & Naderer, 2016; M = 3.77; SD = 0.89; α = .82).

Finally, we assessed whether the campaign affected their **perceived relevance to learn about biases and radicalization** (measured before and after being confronted with the campaign; single item; "I think learning about cognitive biases/radicalization is relevant."; based on Tunney et al., 2021; perceived relevance to learn about biases before M = 3.95; SD = 1.05 and after the campaign M = 4.00; SD = 0.99; perceived relevance to learn about radicalization before: M = 3.85; SD = 1.11 and after the campaign M = 3.89; SD = 1.06).

3.3 Results of the Campaign Evaluation

With regard to the **awareness about the campaign topics**, the tested campaign instruments significantly differed in how successful they communicated the topic of cognitive biases ($\chi2$ (4, N = 1,043) = 21.34, p < .001). The quiz was significantly less successful in communicating the campaign topic of cognitive biases (75.3%), compared to the videos (88.3%; ASR ± 3.8).

Awareness for the topic of radicalization was generally lower, with a maximum of 49.2%. Again, the three tested campaign materials significantly differed in how successful they communicated the topic ($\chi2$ (4, N = 1,043) = 16.829, p =.002). The quiz was significantly less successful in communicating the campaign topic of radicalization (34.1%), compared to the website (49.4%; ASR ± 3.1). See Figure 4 for more details.
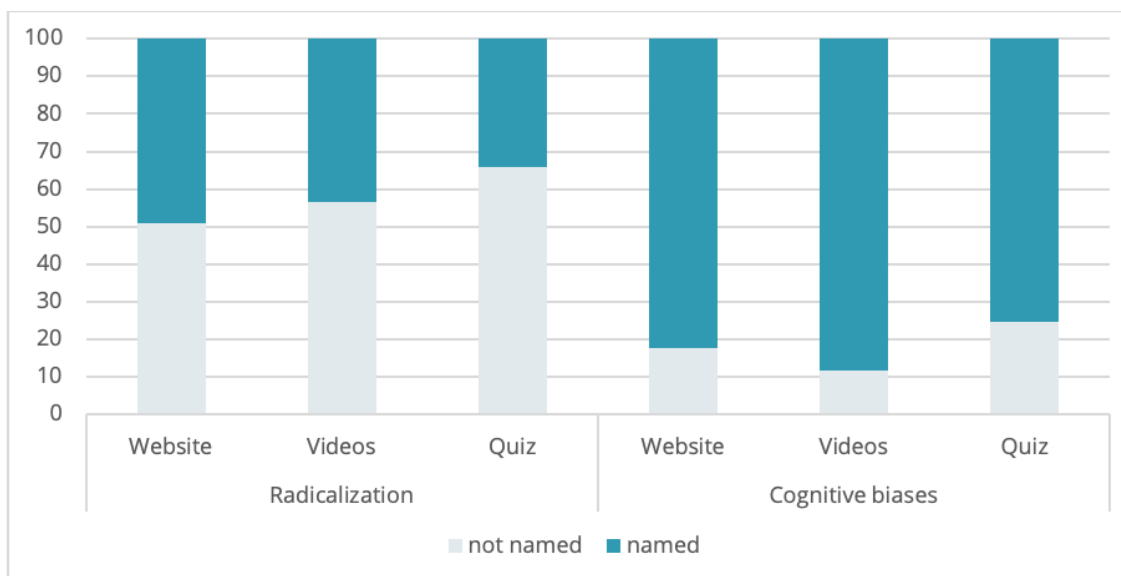
Figure 4: Topic Awareness about the Campaign Topics

The differences in awareness could be explained by how detailed and explicit the information was in the three tools studied. While the website provided detailed and explicit information about how biases work and how they might relate to radicalization, this information was less extensive in the videos without explicit reference to radicalization. The information in the quiz was even less extensive. Here it was always assumed that the quiz should be coupled with either the website or the videos.

To assess differences between the three campaign instruments in our three dimensions of evaluations, we conducted ANOVAs. For **participants' intention to share the campaign material**, we found that it was above average concerning the employed 5-point scale (M = 3.39; SD = 1.06). Furthermore, the intention to share the campaign material differed significantly between the campaign instruments (F(1040, 2) = 14.70; p < .001). It was significantly higher for the website compared to both the video (p < .001) and the quiz (p < .001). The video and the quiz did not differ significantly (p = .998) from each other (See Figure 5).

Interestingly, when examining a possible interaction effect of the predisposition to radicalization, we found that the intention to share the website, which according to the awareness measure most explicitly mentioned radicalization, was significantly lower (M = 3.48; SD = 1.00; p = .002) for people who scored above 2.33 on predisposition to radicalization measure than people below this threshold (M = 3.81; SD = 0.99).

Adolescents' **evaluation of professionalism and information level of the content** was also above average concerning the employed 5-point scale (M = 3.70; SD = 0.96). Furthermore, the evaluation differed significantly between the campaign instruments (F(1040, 2) = 18.04; p < .001). The evaluation of professionalism and information level was significantly higher for the video compared to the quiz (p < .001) and for the website compared to the quiz (p < .001). The video and the website did not differ significantly (p = .279) from each other (See Figure 6).
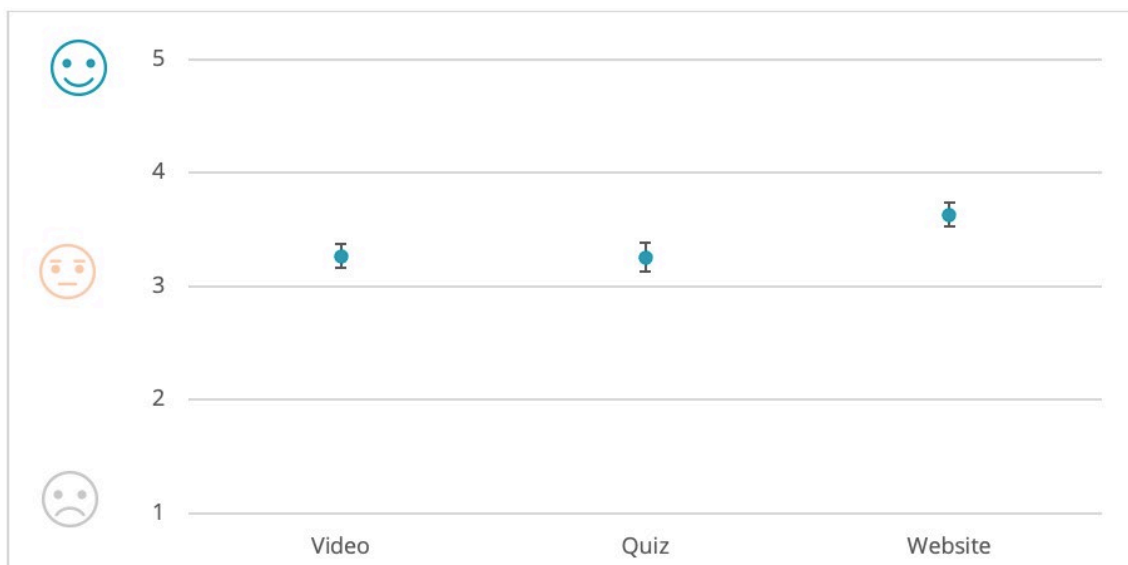

Figure 5: Intention to Share the Campaign Material

When examining a possible interaction effect of the predisposition to radicalization, we found that the evaluation of professionalism and information level was overall significantly lower (M = 3.69; SD = 0.87; p < .001) for people who scored above 2.33 on the predisposition to radicalization measure, and thus had a higher predisposition to radicalization than people below this threshold (M = 4.08; SD = 0.79).
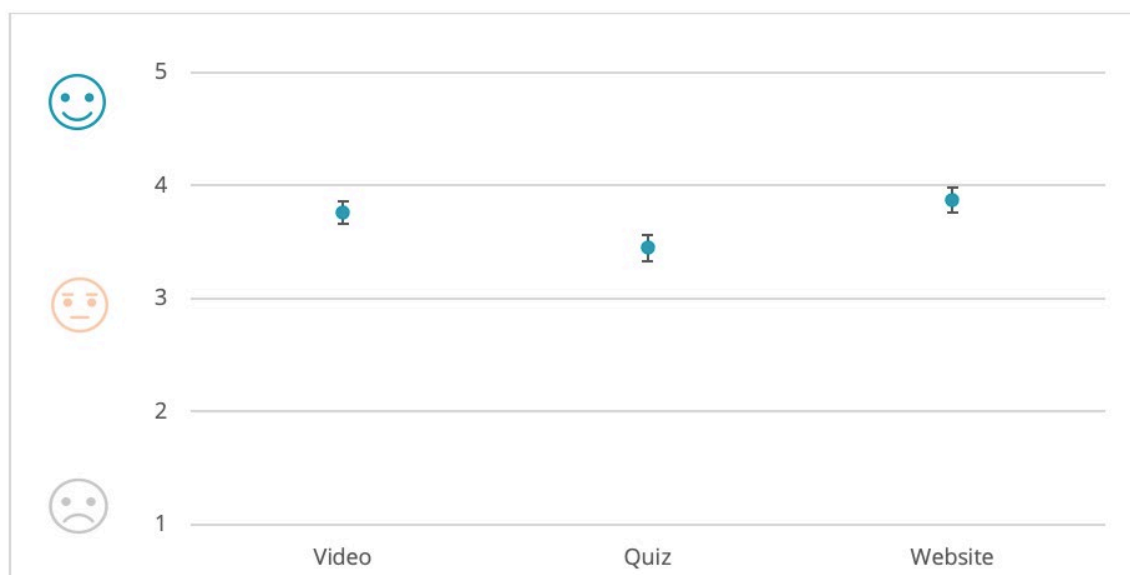
Figure 6: Evaluation of Professionalism and Information Level

Adolescents' **evaluation of the content being interesting and exciting** was also above average with regard to the employed 5-point scale (M = 3.77; SD = 0.89). The evaluation however did not differ significantly between the campaign instruments (F(1040, 2) = 0.91; p = .403). Hence, the evaluation of the campaign content being interesting and exciting was equal for the website, the videos, and the quiz (see Figure 7). We, furthermore, found no interaction effect with the predisposition to radicalization.

Overall, the evaluation of the campaign instruments, hence, was above average on three different dimensions. We, however, **observed slight preferences for the website** and the videos, while particularly evaluation of the website was dependent on the predisposition to radicalization. This might be explained by the more explicit links to the topic of radicalization on the website.
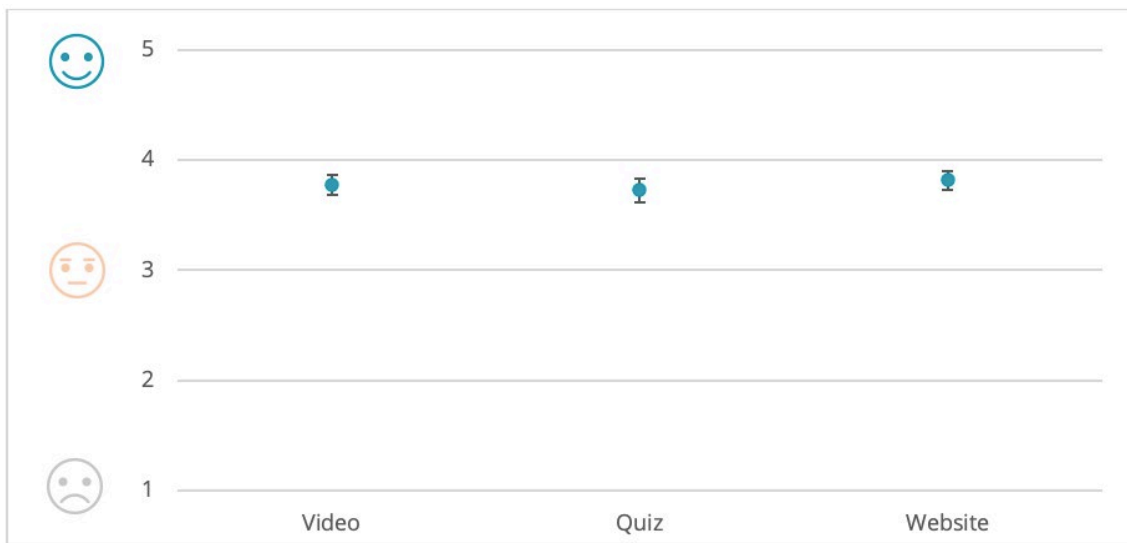
Figure 7: Evaluation of Campaign Content Being Interesting and Exciting

Finally, we examined **adolescents' perceived relevance of learning about biases and radicalization.** The perceived relevance of learning about radicalization was above average with regard to the employed 5-point scale before (M = 3.85; SD = 1.11) and after the campaign (M = 3.89; SD = 1.06). The campaign did not significantly affect the within-subject perception of topic relevance (F(849, 2) = 0.95; p = .331), nor did we observe a between-subject difference of the different campaign instruments (F(849, 2) = 0.35; p = .707). Interestingly, predisposition to radicalization had a main effect on learning about radicalization, as those with a low predisposition (below the threshold of 2.33) considered learning about radicalization as significantly more relevant (F(836, 1) = 20.53; p < .001). In addition, we observed an interaction effect over time (F(836, 1) = 4.74; p = .030), as those predisposed to radicalization even decreased in their perceived relevance of learning about radicalization, while those not predisposed to radicalization increased their perceived relevance after compared to before the campaign.

The perceived relevance of learning about biases was also above average before (M = 3.95; SD = 1.05) and after the campaign (M = 4.00; SD = 0.99). The campaign did not significantly affect the within-subject perception of topic relevance (F(803, 2) = 2.14; p = .144), nor did we observe a between-subject difference of the different campaign instruments (F(803, 2) = 0.01; p = .986). Interestingly, predisposition to radicalization had a main effect on learning about cognitive biases, as those with a low predisposition (below the threshold of 2.33) considered learning about biases as significantly more relevant (F(792, 1) = 27.96; p < .001). Yet, no interaction effect

with the campaign content or the pre-post measure occurred.

For both of these measures, we only considered adolescents who already knew what the terms radicalization (n = 190; 18.2% did indicate to not understand this term before seeing the campaign) and cognitive biases (n = 236; 22.6% did indicate to not understand this term before seeing the campaign) meant and thus, we might have observed a ceiling effect for the perceived relevance of both concepts as it was already rather high at the beginning.

4. Campaign effects

4.1. Sample

For the campaign effect study, we recruited **adolescents** (16-22 years old; M = 20.05, SD = 1.66) from **Austria and Germany** [2] via the survey institute respondi. The survey was conducted from September 24 to October 4, 2021. In sum, N = 223 adolescents fully completed the effect study (38.1% male, 60.5% female, 1.3% other; 0.9% no school-leaving qualification, 26.5 % complete lower-secondary education or vocational school education, 55.2% complete secondary education, 17.0% complete university education, 0.4% did not want to answer this question).

4.2. Procedure

We employed a 1x3 experimental design manipulating **whether or how adolescents were confronted with the campaign material**. It was not possible to expose participants to all campaign materials because we had to adhere to the time constraints of the survey institute of a maximum of 15 minutes. Thus, we chose **four biases** that were rather easily distinguishable and recognizable in social media content. We chose these four biases: rosy retrospection, the ingroup-outgroup bias, the confirmation bias, and the hostile media effect.

Based on previous literacy experiments, we created three conditions (Fernandes et al., 2020; Naderer & Opree, 2021). In the **control condition** (n = 70, 31.3%), we provided short information about cognitive biases, but **did not show any campaign material**. In the **video condition** (n = 72, 32.3%), we also showed the short information about the cognitive biases, and four videos developed for the campaign. In the **quiz +video** condition (n = 81, 36.3%), we asked participants to **take the quiz** where we included only the questions that related to the four chosen biases and afterwards to **watch the campaign videos** (see Figure 8).

[2] **Note:** Due to the developed bias literacy measure, only German-speaking adolescents were considered for this study. The internal consistency of this measurement instrument had to be ensured.
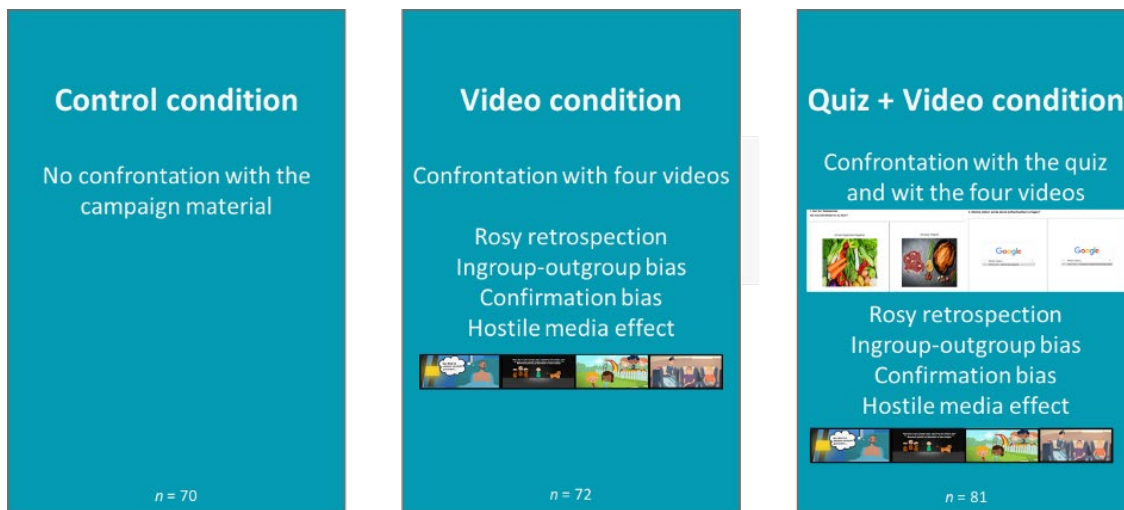
Figure 8: Study Design

Participants were randomly assigned to one of these three conditions. Before they saw the stimulus, we asked about their age, gender, educational background, and we also inquired about their **radical attitudes** (pre-stimulus measure; mean index based on 3 items assessed on a 5-point scale; e.g., "I can understand when people resort to violence to achieve political goals, even in Germany/Austria.", Reinemann et al., 2019; M = 1.91; SD = 1.01; α = .80). After the stimulus, we conducted a literacy test, which was designed to measure participants ability to use the acquired knowledge from the campaign to recognize cognitive biases. For this **literacy test** (see Figure 9), our participants had to classify whether they recognized a bias in a posting and if so, which one. In the survey, we informed participants how many of the posts they had correctly classified. Participants literacy score could range from 0-4 (M = 1.98; SD = 1.42; Hostile media effect post: 50.7% correct; rosy retrospection: 50.2% correct; confirmation bias: 40.4% correct; ingroup-outgroup bias: 57% correct). We, furthermore, enquired about their **confidence to recognize biases** (mean index based on three items assessed on 5-point scale; e.g., "I am convinced that I can recognize cognitive biases." based on Naderer & Opree, 2021; M = 3.23; SD = 1.04; α = .91), as well as their level of **radical attitudes as a post-measurement** (same items as for the pre-measurement; M = 1.96; SD = 1.05; α = .87).

Figure 9: Literacy Test

4.3. Results

We calculated an ANOVA to assess how the campaign material affected the literacy score. Results indicated a significant difference between conditions ($F_{(2, 220)} = 3.041$, $p = .050$). For the literacy score, we found that both the video on its own and the quiz+video did improve the participants abilities to recognize biases (control group: $M = 1.67$, $SD = 1.36$; quiz+video condition: $M = 2.01$, $SD = 1.42$; video condition: $M = 2.25$, $SD = 1.43$), yet only the video condition significantly increased the literacy score compared to the control group ($p = 0.045$). The quiz+video condition did not significantly increase literacy compared to the control group ($p = 0.415$). The two campaign conditions also did not differ significantly from each other ($p = 0.892$; see Figure 10). Interestingly, predisposition to radicalization had a main effect on

learning about biases, as those with a low predisposition (below the threshold of 2.33) had a significantly higher literacy score in all three conditions (M = 2.14; SD = 1.38) compared to adolescents with a predisposition to radicalization (M = 1.65; SD = 1.44; F(216, 1) = 6.27; p = .013).
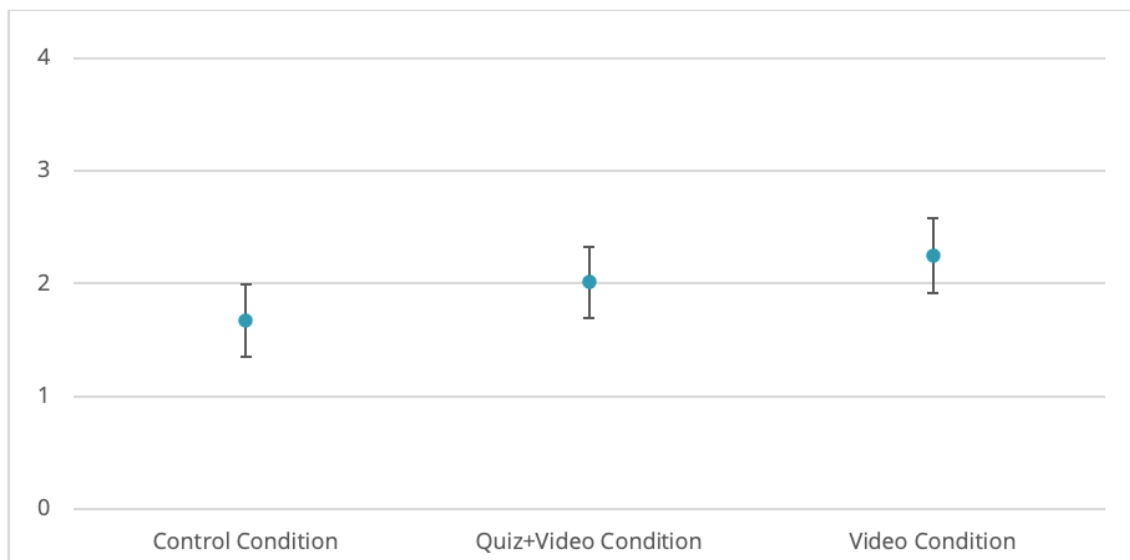

Figure 10: Literacy Score

In a next step, we examined whether the campaign material builds confidence in recognizing biased content in the future by conducting a mediation analysis. While we did not observe a main effect of the campaign material condition compared to the control group (quiz+video condition: b = 0.06; SE = 0.14; p = .674; video condition: b = -0.15; SE = 0.15; p = .332), the literacy score increased the effect on self-confidence (b = 0.40; SE = 0.04; p < .001). Hence, via the increased literacy score the video condition compared to the control condition significantly increased adolescents' self-confidence (b = 0.23; SE = 0.10; LLCI = 0.05; ULCI = 0.42)

We furthermore examined the effect of the campaign material on participants' level of radicalization. For that, we conducted a repeated measures ANOVA. Results indicate a marginally significant interaction effect of the pre-post measure (F(218, 2) = 2.33; p = .099). While the radical attitudes stayed largely unchanged for the control condition and the video condition, the reported radical attitudes slightly increased for the quiz+video condition in the post measure compared to before being confronted with the campaign material (see Figure) 11. This indicates that confrontation with the quiz can potentially lead to unintended reactance effects. If adolescents are confronted with their own biases, this might lead to them manifesting their pre-existing radical views.
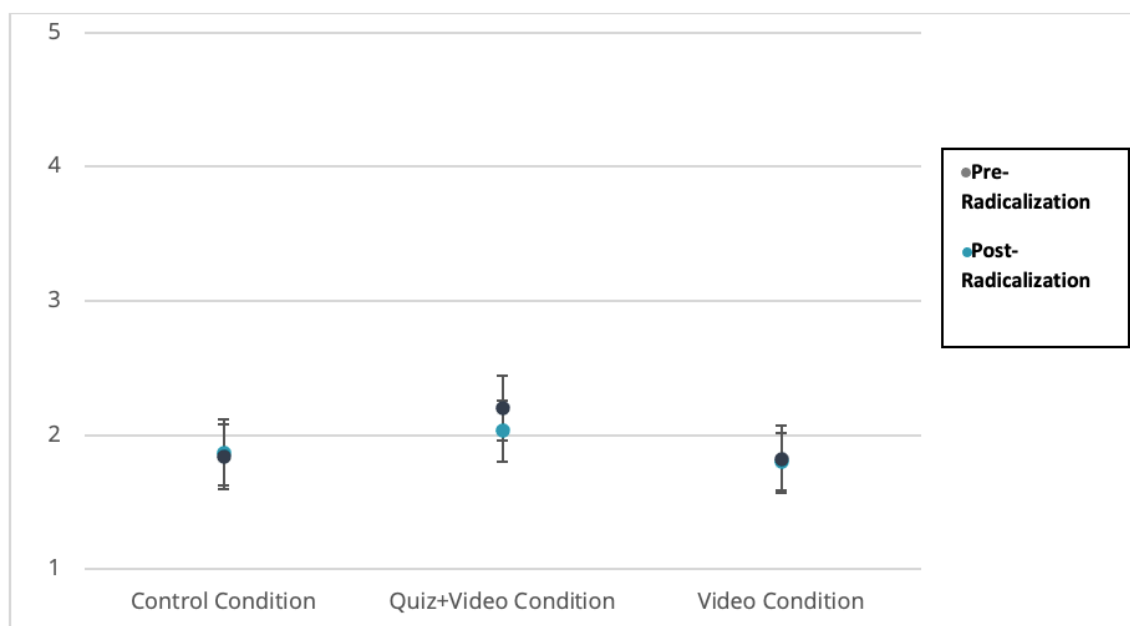
*Figure 11: Development of Level of Radicalization Before and After
Seeing the Campaign Material*

5. Qualitive contest workshop

5.1 Sample

We conducted a school workshop in Vienna with 34 adolescents between 16-18 years in the first week of November 2021. The workshop consisted of 2 lessons. We conducted the workshop in two separate classes.

5.2 Procedure

The **workshop** started with an introduction of the presenter. Then, without introducing the topic of the workshop further, the workshop began with three **playful exercises**. Here the adolescents were confronted with their own biases. Examples from the teacher toolkits were employed. Specifically, the exercises showed our predispositions to stereotypes of certain professional groups, as we assume a shy, introvert, accurate human who likes to flee reality is much more likely to be a librarian and not a farmer. Furthermore, the exercises highlighted the ingroup-outgroup bias. For this, the class was split into coffee and tea drinkers, and students were then instructed to make assumptions about their own and the other group. The discussion quickly led to saying mostly positive things about

the own group (e.g., "we have good taste", "we are more health conscious") and mostly negative things about the other group (e.g., "coffee makes people addicted", "drinking tea is for softies"). Finally, students were presented with a number linefor which they had to figure out the underlying rule by asking which numbers belonged to the line. The series was 2, 4, 6, and students named many even numbers before asking for odd or negative numbers. Odd numbers were also included because the series is actually based on the principle that full, ascending numbers are included in the series. However, students only asked for numbers that confirmed their pre-existing hypothesis rather than attempting to falsify their initial assumptions. This illustrates the confirmation bias.

Then the **theory behind cognitive biases**, in general, was explained. Five biases were selected, and defined further: the ingroup-outgroup bias, the negativity bias, the hostile media effect, the confirmation bias, and the bandwagon effect. Due to the time constraints, more biases could not be discussed. After the explanation of each bias, students were called upon for **active discussion to gather examples for the biases.**

After this introduction, students were given 20 minutes to work in small teams of 3-4 students to **create a post or video that explained biases,** illustrated the consequences of biases, or showed what could be done to counter biases. After the creation time, the remaining time was used to present and discuss the created content.

5.3 Workshop Outputs

In total, students created 10 pieces of content that included videos and posts on the negativity bias, the bandwagon effect, and the ingroup-outgroup bias. Most of the content featured videos that were prepared in **TikTok format**. The students most frequently addressed the **bandwagon effect** in their content. According to their own statements, they could identify most strongly with this bias because they had already followed group majorities in their decision-making processes. There was a wide variety of interpretations for the bandwagon effect, ranging from the switch of a well-known soccer player to a new team and the accompanying increased popularity of the team on social media; fashion phenomena; the popularity of the series Squid Game; and group dynamics they had observed in their class. Regarding the other biases, there were also short explanatory videos, or posts that addressed the cognitive biases in a soulful text. The students showed a lot of **self-reflection** and understanding of how biases work in their self-created content, and largely came up with their **own scenarios** that were not already brought up in the examples during the introductory unit.

5.4 Workshop Evaluation

Finally, there was a short Q&A session about the PRECOBIAS project and scientific studies. In addition, the students were asked to evaluate the workshop.

They evaluated whether the workshop was informative, professional, has given them food for thought on a scale from 1 (I don't agree) - 5 (I fully agree). This **cognitive evaluation** of the workshop was rather positive (mean index based on three items; Frischlich et al., 2017; M = 3.94; SD = 1.21). In addition, the students were asked about the attitudinal evaluation (mean index based on three items negative-positive; uninteresting-interesting; boring-exciting; Matthes & Naderer, 2016), and the evaluation was very positive (M = 4.59; SD = 0.56). We also asked about their **self-confidence in recognizing cognitive biases**. The self-confidence was above average concerning the employed 5-point scale (mean index based on three items assessed on 5-point scale; e.g., "I am convinced that I can recognize cognitive biases." based on Naderer & Opree, 2021; M = 3.96; SD = 0.88). Lastly, we asked whether they **consider learning about biases as relevant**. Again, the relevance learning was above average concerning the employed 5-point scale (M = 3.98; SD = 1.02).

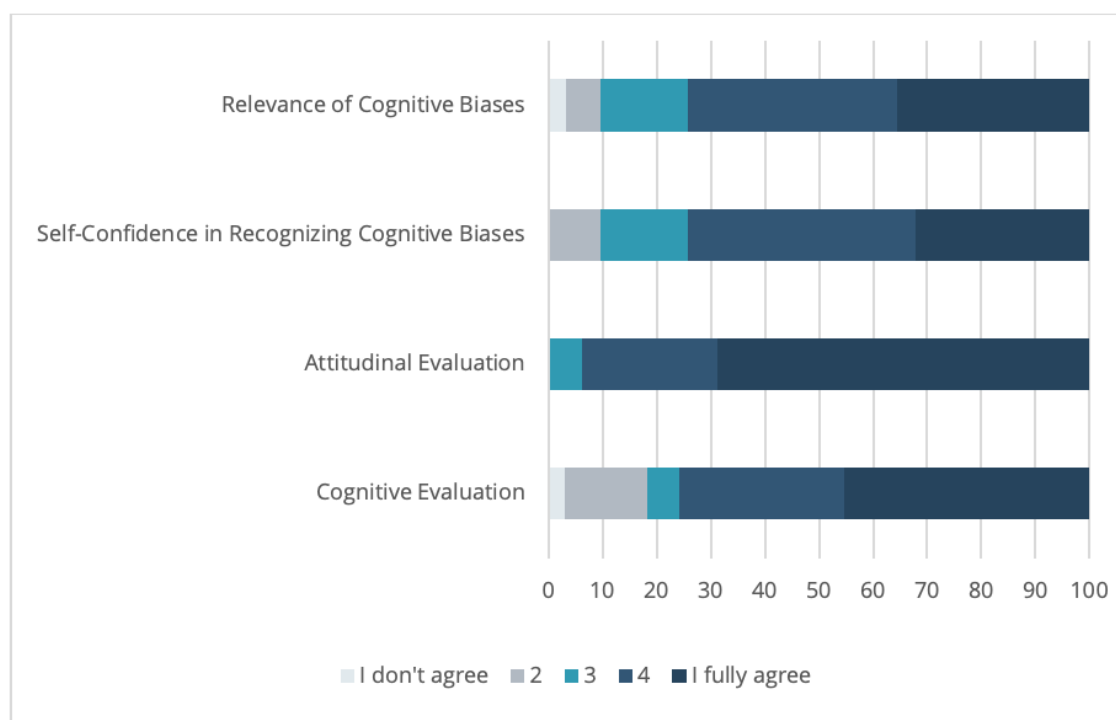*Figure 12: Evaluation of the Workshop*

6. Discussion

Our campaign evaluation study, based on a sample of European adolescents, highlighted **how relevant the issue of radicalization** is and **how often young people are actually confronted with content inciting violence and spreading prejudiced and hostile content about certain groups online**. Thus, the campaign tackles a very relevant issue in the lives of European adolescents. While in our sample, we overall found a rather low predisposition to radicalization of adoelscents, we were able to include the propensity to radical attitudes in our analysis to examine the vulnerability of adolescents in our results.

6.1 Fulfillment of the Campaign Objectives

Overall, the campaign material was rated very positively. In terms of **raising awareness of the campaign issues**, we found that the campaign materials worked quite well for raising awareness of cognitive biases. In particular, the website and videos, which provided more detailed information about biases, performed well in this regard. In the campaign, the quiz was always intended to be used in combination with the videos or the website rather than alone. So, it was not the perfect tool for raising awareness on its own, which is also reflected in our results. On the topic of radicalization, the campaign did not explicitly convey much information. This was also a conscious decision so to not confront at-risk youth too directly with this topic.

In terms of building **relevance to learning about radicalization and cognitive biases**, we conducted a pre-post measurement. We did not find an increase in perceived relevance to learning about these topics. However, since we only considered those who had heard of the terms prior to the campaign, this could also indicate a ceiling effect, as the relevance of learning about these topics was already relatively high before seeing the campaign material. Still, it was interesting to observe that those prone to radicalization rated the relevance of learning about radicalization and prejudice significantly lower. While those who were not prone to radicalization rated the relevance of learning about radicalization higher after the campaign than before the campaign. This suggests that youth who are predisposed to radicalization reject the idea of others being exposed to information about the topic and also do not consider it relevant to themselves.

In the effect study, we found that the campaign was able to increase **adolescents' knowledge about biases**. Thus, by being confronted with the video material, young people were able to apply their learned knowledge about cognitive biases to other social media content. Interestingly, the combination of quiz and video was not able

to achieve this effect. A possible reason behind this might be that the quiz provided more information than the videos about how the participants themselves are affected by biases. This may have counteracted the objective information provided by the videos alone. The effects on literacy were small, but this could be due to the fact that we also provided the control condition with some information about cognitive biases. Otherwise, it would not have been possible to measure the detection ability in the control group and thus compare it with the campaign instruments. The small effects we observed migh therefore be due to methodological decisions.

The queried **competence regarding biases** in turn led to the young people being more **self-confident in recognizing and dealing with biases in the future**. This is a positive signal, as the campaign succeeded in creating awareness of cognitive biases among adolescents that can also be transferred to other posts in social media they encounter. Both knowledge and self-confidence could be helpful in the long run to protect adolescents against radicalization

Concerning short-term effects, however, parts of the campaign should also be viewed critically, specifically when aiming to achieve **deradicalization**. Overall, the predisposition to radicalization in our studied sample was rather low, so that no significant, immediate decrease could be observed after the campaign. Rather, the results suggest that confronting one's own shortcomings and prejudices through the quiz might even lead to a very slight increase in radical attitudes. As in the campaign evaluation study, in the effect study we found that the quiz has some relevant downsides that need to be considerd. The quiz combined with the videos did not build youths' competence and confidence to identify biases as we observed for the videos alone, and could even lead to slight boomerang effects for the campaign's deradicalization intentions immediately after being confronted with the campaign material.

Finally, the qualitative school workshop conducted to evaluate the campaign contest showed that **adolescents were very open to learning about cognitive biases**. They were able to engage with the topic and bring in examples and experiences from their own lives. When creating content about biases, they particularly referred to biases they had observed in their own lives and connected these to current social developments and discourses, hence the topic motivated them to **actively engage** with the campaign topic.

6.2 Recommendations for Future Counter-Radicalization Campaigns

Our findings suggest that future counter-radicalization campaigns will need to walk the **fine line of informing and educating at-risk youth** without becoming too explicit and thus avoiding a dismissive response. Interestingly, the website that contained the most explicit references to radicalization, and thus raised awareness of this issue the most, was rated least positively by those who were already vulnerable to radical ideas. Thus, the results of our study indicate that those who are vulnerable to radical attitudes are less open to learning about this issue than youth who are at low risk for radicalization. This prevents adolescents for whom the content would be most relevant from engaging with this information themselves or sharing the material they have learned with their relevant networks. Thus, the conscious decision not to mention radicalization too explicitly seems to be a promising strategy to avoid alienating vulnerable groups for whom the content is particularly relevant. However, not making an explicit connection from the topic of the campaign (in this case cognitive biases) to radicaliazion processes has the disadvantage that no explicit awareness is created, leaving it up to the target audience to make the appropriate deductions from the campaign material. Whether this leads to the desired results in the long term would still have to be found out through further studies.

In terms of infomation and awareness-raising, even the website that made the most explicit connections to radicalization, failed to create a high level of awareness of the issue. Therefore, when **planning a campaign**, it is important to decide whether the **main goal** should be to create **explicit awareness** of the issue of radicalization or to reach the **intended target audience** and possibly build skills, confidence, and self-reflection that they can use in their future confrontation with radical content.

Furthermore, the scientific evaluation shows that it can be **risky** in a counter-radicalization campaign to focus too much on **self-reflection** and thus confront at-risk youth with their own shortcomings. The sientific evaluation has shown that although the quiz is perceived as entertaining and interesting, it can lead to reactance and thus a slight increase in radical attitudes. Future campaigns should therefore consider whether such an instrument should be part of the campaign or whether there is another way to confront people with the results of such a quiz. For example, the results could be presented in a way that highlights how many other people are also affected by biases in their processing so that people do not feel attacked and therefore do not see the need to overcompensate.

Finally, choosing a **topic** that was **relevant to young people's everyday lives** and not necessarily associated with radicalization proved to be a successful decision.

The content of the campaign was very well received by the target group overall. In particular, the school workshop showed that the topic of cognitive bias was truly taken from the lives of young people, as they were able to contribute many of their own examples and experiences. Future campaigns could build on this insight to ensure that at-risk youth actively engage with a campaign, ideally with the long-term result of becoming more engaged, critical media users and citizens.

7. Conclusion

In conclusion, the scientific evaluation showed that the campaign did indeed **help adolescents to learn about the campaign topic and find out more about their potential personal biases**. The campaign impact study also showed that the videos, in particular, were successful in promoting critical thinking by **improving** their **knowledge** of biases, which increased their **confidence** in their ability to identify this content in the future. However, the campaign's deradicalization goal was only partially achieved, as the quiz may have lead to reactance.

The scientific evaluation indicates that **deradicalization campaigns tread a very fine line between achieving information and increased reflection or experiencing rejection and even boomerang effects**. This is demonstrated especially among vulnerable adolescents and when topics connected to radicalization are mentioned too explicitly. Furthermore, when youths are confronted with their own shortcomings, this can make them even more convinced in their pre-existing attitudes.

However, the content of the campaign and the issue of cognitive biases was overall very well received by the target group of young people. This was equally evident in the evaluation of all campaign elements.

This high level of personal relevance and comprehensibility gives hope that a change from a passive recipient of media information to a **critically engaged participant** in online discourse could be initiated or supported by the campaign.

8. References

Barnfield, M. (2020). Think twice before jumping on the bandwagon: Clarifying concepts in research on the bandwagon effect. Political Studies Review, 18(4), 553-574.

Bar-Tal, D., Halperin, E., & de Rivera, J. (2007). Collective emotions in conflict situations: Societal implications. Journal of Social Issues, 63, 441-460.

Bouko, C., Naderer, B., Rieger, D., Van Ostaeyen, P., & Voué, P. (2021a). Discourse patterns used by extremist Salafists on Facebook: identifying potential triggers to cognitive biases in radicalized content. Critical Discourse Studies, online-first 1-22.

Bouko, C., Van Ostaeyen, P., & Voué, P. (2021b). How Jihadi Salafists Sometimes breach, but mostly circumvent, Facebook's community standards in crisis, identity and solution frames. Studies in Conflict & Terrorism, online-first 1-26.

Frischlich, L. (2021). # Dark inspiration: Eudaimonic entertainment in extremist Instagram posts. New Media & Society, 23(3), 554-577.

Frischlich, L., Rieger, D., Morten, A., & Bente, G. (2017). Videos gegen Extremismus? Counter-Narrative auf dem Prüfstand (Vol. 51). Bundeskriminalamt.

Lee, T. K., Kim, Y., & Coe, K. (2018). When social media become hostile media: An experimental examination of news sharing, partisanship, and follower count. Mass Communication and Society, 21(4), 450-472.

Macdonald, S., & Whittaker, J. (2019). Online radicalization. In: J. R. Vacca (ed.). Online Terrorist Propaganda, Recruitment, and Radicalization (pp. 33-45), CRC Press: Boca Raton, London, New York.

Matthes, J., & Naderer, B. (2016). Product placement disclosures: Exploring the moderating effect of placement frequency on brand responses via persuasion knowledge. International Journal of Advertising, 35(2), 185-199.

Menke, M., & Wulf, T. (2021). The dark side of inspirational pasts: An investigation of nostalgia in right-wing populist communication. Media and Communication, 9(2), 237–249.

Naderer, B., & Opree, S. J. (2021). Increasing advertising literacy to unveil disinformation in green advertising. Environmental Communication, online.-first.

Reinemann, Carsten, Nienierza, Angela, Fawzi, Nayla, Riesmeyer, Claudia, Neumann, Katharina (2019). Jugend – Medien – Extremismus. Wo Jugendliche mit Extremismus in Kontakt kommen und wie sie ihn erkennen. Wiesbaden: Springer VS.

Reuter, C., Hartwig, K., Kirchner, J., & Schlegel, N. (2019). Fake news perception in Germany: A representative study of people's attitudes and approaches to counteract disinformation. Presented at the 14th International Conference on Wirtschaftsinformatik.

Schulz, A., Wirth, W., & Müller, P. (2020). We are the people and you are fake news: A social identity approach to populist citizens' false consensus and hostile media perceptions. Communication Research, 47(2), 201-226.

Prevention of Youth Radicalisation Through Self-Awareness on Cognitive Biases

(PRECOBIAS)