JOURNAL FOR DERADICALIZATION

Spring 2022
No. 30
ISSN: 2363-9849

Evaluating an Interactive Film on the Prevention of Political Radicalization

Axel Ebers[a1], Stephan L. Thomsen[b]

[a]Research Assistant, Institute of Economic Policy, Leibniz Universität Hannover, [b]Professor for Economics and Director of Institute of Economic Policy, Leibniz Universität Hannover

Abstract

The functionality of social media permits (and maybe fosters) an increase in political radicalization, which causes immense social harm. In response, authorities have started using social media for prevention but empirical evidence on the effectiveness is scarce The present study evaluates the effects of an interactive film distributed in social media that aims to reduce the individual level of radicalism in attitudes and radicalization intentions. During the film, viewers have to express their opinion on increasingly radical statements by clicking popup buttons. Depending on their opinions, the plot of the film takes a different route. For identification of causal effects, the evaluation uses a randomized controlled trial (RCT) with a two-week follow-up. The empirical results show that the film immediately reduces the level of radicalism in attitudes by 12% and radicalization intentions by 15% of a standard deviation. After two weeks, these effects are still persistent but fade out a little in the general population. There are stronger and more persistent effects among the subgroups of 18-24 year-olds, women, and people on the left of the political spectrum. Because these subgroups resemble the characteristics of the protagonists, we speculate that social identification enhanced treatment effects. Cognitive dissonance, on the other hand, may explain why people on the right of the spectrum did not react to the film. The findings demonstrate the importance of target-group oriented design and early prevention.

Article History

Received Feb 7, 2022

Accepted Mar 21, 2022

Published Mar 25, 2022

Keywords: Preventing Political Radicalization, Interactive Film, Gamification, Randomized Controlled Trial (RCT)

JEL classification: C93, D91

1. Introduction

Political radicalization and terrorism cause significant social harm through the erosion of democratic institutions, human pain and suffering, the costs of medical treatments and lost productivity, destruction of physical capital, and inefficient resource allocation (Bardwell & Iqbal, 2021). To avoid this harm, policy makers initially focused on counterterrorism

[1] Corresponding Author Contact: Axel Ebers, Email: ebers@wipol.uni-hannover.de, Institute of Economic Policy, Leibniz Universität Hannover, Königsworther Platz 1, 30165 Hannover, Phone: +49 (0)511 762 14 628

approaches grounded in the criminal justice system, especially in reaction to the September 11 attacks. Counterterrorism, however, has partially failed to prevent violence and, in extreme cases, even encouraged membership in radical groups (Bhui et al., 2012). Counterterrorism measures by the British government, for example, stigmatized and alienated Muslim communities in the UK by treating them as suspects rather than allies. This isolated a whole religious group and thus damaged social cohesion. Social cohesion, however, contributes to better public health, more equal and just societies, and less crime (McDonald & Mir, 2011). Given the moderate success of counterterrorism approaches (Schmid, 2013), prevention strategies became more and more important (Borum, 2011).

Developing and implementing effective prevention programs requires a profound knowledge of how people radicalize and why they engage in violent actions. Seminal research described the *individual pathway* towards political violence as a process of sequentially moving through particular stages (Moghaddam, 2005; Silber et al., 2007; Wiktorowicz, 2004). These so-called *stage-models* received criticism for assuming that individuals had to pass through each stage of the pathway before reaching the end, and because they did not provide a sufficient explanation for the critical step from radical attitudes to violent actions (Hafez & Mullins, 2015). Taking up this criticism, further seminal research emphasized the importance of *group dynamics* including community support for violent action (Horgan, 2004; Kruglanski et al., 2014), the spiral of violence between terrorist attacks and government responses (Fenstermacher et al., 2010; Pyszczynski et al., 2009), or competition and conflict between fractions of the same movement (Della Porta, 2013). Other research emphasized the differences between the individual *profiles* of mere radicals and violent terrorists (Bartlett et al., 2010). Sageman (2011) even argued that some individuals would develop radical ideas only after they have joined radical groups through relatives or friends. McCauley & Moskalenko (2011) identified the interplay of individual-, group-, and mass-level mechanisms as the root cause of violent radicalization. Mass-level mechanisms would lead to a radicalization of public opinion. Embedded in this context, individual- and group-level mechanisms would lead to a radicalization of actions. With their *two-pyramids model,* McCauley & Moskalenko (2017) went so far as to model the radicalization of attitudes (or

opinions) and the radicalization of actions as two completely different, albeit interrelated, psychological phenomena.

Social media platforms hold great potential for practical prevention work and academic research in the area of radicalization. Due to the ubiquity of smartphones, social media are available as communication channels anytime and anywhere (Silver et al., 2019). Security authorities can use these channels to design crime prevention programs that specifically target vulnerable groups (i.e., *microtargeting*; Winter et al., 2021). The marginal cost of reaching people via social media is relatively low, which allows a quick upscaling of such programs (Castronovo & Huang, 2012). Online interventions, which can be disseminated on social media, allow for customization based on the needs and preferences of the target groups (Lustria et al., 2009). Social media enable dialogue-oriented and interactive communication between authorities and their targets (Tsimonis & Dimitriadis, 2014), potentially strengthening the effectiveness of prevention work. Finally, analytic tools of large platforms offer new possibilities for collecting data (Zhang et al., 2022), which could potentially form the basis for determining the effectiveness and economic efficiency of prevention measures.

With this in mind, it is noteworthy that although we have seen a surge in prevention programs in recent years, the majority lacks a rigorous evaluation based on sound empirical methods. For example, while the *German Federal Criminal Police Office* (BKA) found more than 2,000 projects implemented in Germany alone (Gruber et al., 2017), a recent meta-analysis finds only 9 evaluations that meet rigorous eligibility criteria (Jugl et al., 2020). To fill this research gap, we evaluate an online intervention that aims to prevent a radicalization in attitudes and unfolding of radicalization intentions. The intervention is part of a primary prevention program aiming to prevent the radicalization of public opinion. With our study, we make a significant contribution to the research and practice of preventing political radicalization, since rigorous evaluation allows the replication of successful programs and efficient allocation of scarce public funds.

The online intervention consisted of an interactive film that employs game principles and game design elements. The plot tells the story of Lea and Chris. Lea has a Jewish background. Chris gets increasingly lost in the maelstrom of conspiracy myths and

antisemitism, which puts a strain on their friendship. As the plot unfolds, Chris, his friends and the media contributions shown make increasingly radical statements. The film asks the viewer to take a position on these statements by clicking one of three popup buttons. Depending on these choices, the film takes a good or bad end, which means that an arson attack on Lea's house either will take place or will be prevented.

To evaluate the film's causal impact on the level of *radicalism in attitudes* and *radicalization intentions*, we conducted a *randomized controlled trial* (RCT). For this purpose, we drew a representative sample of the German working population and randomly allocated participants to either the treatment or the control group. While we exposed the treatment group to the film, the control group received no treatment. Subsequently, both groups participated in a survey assessing the level of radicalism in attitudes, radicalization intentions, and a set of secondary outcomes and covariates. Randomization ensured that both groups did not differ, on average, except in exposure to the interactive film. Exposure (i.e. the treatment) therefore causally determined any differences in the outcomes. We conducted a follow-up survey two weeks after treatment to account for potential diminishing of effects over time.

The rest of the paper organizes as follows: Section 2 reviews the process of radicalization using McCauley & Moskalenko's (2017) *two-pyramids model*, aspects of the online context, and the psychology of behavior change using Fishbein & Ajzen's (2011) *reasoned action approach* (RAA). The following section describes the policy intervention in detail and derives our research hypothesis. Section 4 describes our research design, process of data collection and sample characteristics. Section 5 provides our main results and heterogeneity analysis. In section 6, we discuss the empirical results before we give some conclusions, in the final section.

2. Theoretical Considerations

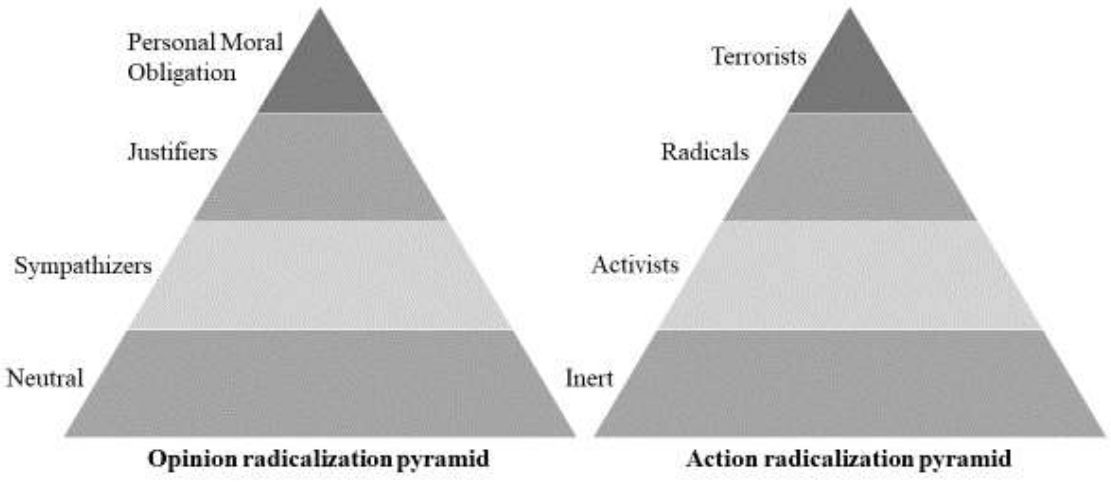
2.1 The Process of Radicalization

Radicalization marks a process of developing beliefs and attitudes that may ultimately culminate in the conduct of radical behaviors, including violence and terrorism at the

extremes (Wolfowicz et al., 2020). An alternative definition summarizes mechanisms that move individual, group or mass opinion to support or participate in political violence under the term radicalization (C. R. McCauley & Moskalenko, 2017). Of course, while almost all of those who engage in radical behaviors hold radical attitudes, most of those who hold radical attitudes will never engage in radial behaviors. In other words, radicalization of *attitudes* (or opinions) is psychologically a different phenomenon from radicalization of *action*. For this reason, the *two-pyramids model* (Figure 1) separately analyzes the two phenomena within the opinion pyramid and the action pyramid (C. R. McCauley & Moskalenko, 2017; Neumann, 2013).

On the one hand, individuals who do not care about any political cause (*neutral*) make up the base of the *opinion pyramid*. Directly above are those who believe in a cause but object to violence (*sympathizers*). Yet higher are those who justify violence in defense of a cause (*justifiers*). At the apex are those who feel a *personal moral obligation* to engage in violent acts to defend the cause. Two characteristic features of the model distinct it from stage models: First, individuals can move up and down the pyramid, and second, they can skip steps within this process.

Figure 1. The Two-Pyramids Model



Source: Own representation based on McCauley & Moskalenko (2017).

On the other hand, at the base of the *action pyramid* are individuals who do nothing for a political group or cause (*inert*). Directly above are those who are engaged in legal political action for the cause (*activists*). Yet higher are those engaged in illegal action for the cause (*radicals*). At the apex of the pyramid are those engaged in illegal action that targets civilians (*terrorists*). Analogously to the opinion pyramid, individuals can skip steps when moving up and down the action pyramid.

To operationalize the empirical content of the two pyramids model, three relevant outcomes can be derived: (1) the level of *radicalism in attitudes*, (2) *radicalization intentions*, and (3) the engagement in *radical actions* such as terrorism. In an experimental setting, however, ethical and practical reasons forbid measuring radical actions. We thus followed the common practice of social psychology and used behavioral intentions as an approximation of actions (Ajzen, 1985; Azjen, 1980; Fishbein & Ajzen, 1977; Sheeran et al., 1999). Consequently, the level of *radicalism in attitudes* and *radicalization intentions* formed the primary outcome variables in our analysis.

Certain risk and protective factors influence the formation of radical attitudes and radicalization intentions as well as the execution of radical behavior. A recent meta-analysis identified the risk and protective factors with the strongest influence and radical attitudes, radicalization intentions, and radical behavior (Wolfowicz et al., 2020). We incorporated the factors with the largest effect sizes as secondary outcomes in our analysis. With respect to radical attitudes, the protective factor with the largest effect size was *law abidance*, while the risk factors with the largest effect sizes were an *authoritarian personality* and ties to *similar peers*. With respect to radicalization intentions, the strongest protective factor was *age*, while the strongest risk factors were *radical attitudes*. Finally, regarding radical behavior, the strongest protective factors include *school bonding*, *age*, *law legitimacy*, and law abidance. The strongest risk factors include *thrill seeking* or *risk-taking*, *radical peers*, *authoritarian personality*, *criminal history*, *low self-control* and *radical attitudes*.

Different types of prevention programs aim to address the described outcomes as well as risk and protective factors. We can distinguish three fundamental types: primary, secondary, and tertiary prevention programs. While *primary* programs target the whole population to prevent a radicalization of public opinion, *secondary* and *tertiary* approaches

focus on disengagement and de-radicalization tactics (Kober, 2017; Mastroe & Szmania, 2016). *Disengagement* programs aim at behavioral change leading individuals to cease engagement in radical action (Doosje et al., 2016; Mastroe & Szmania, 2016). *De-radicalization* programs aim at a rejection of radical attitudes (Berger, 2016), and often take the form of exit programs tailored to the individual needs of the target (Bjørgo & Carlsson, 2005; Mastroe & Szmania, 2016).

2.2 Radicalization in the Online Context

The emergence of social media has enabled any individual with internet access to spread radical messages among the broad public. Previously, traditional media acted as gatekeepers, which prevented the worst excesses. This change in the media landscape poses a major challenge for security authorities since militant individuals or groups can target users who are most receptive to their radical messages on social media (Fink, 2018; Malmasi & Zampieri, 2017; Mathew et al., 2019). Through targeted dissemination of their ideologies, these groups can exert influence, gain sympathizers and supporters, or even recruit new members (Chatfield et al., 2015; Gates & Podder, 2015; Thompson, 2011). The *lone wolf theory* plays an increasingly important role in this context. Accordingly, individuals radicalize themselves on social media and carry out radical actions eventually (Weimann, 2012). In extreme cases, such as the far right terrorist attack in Christchurch, New Zealand, the perpetrators even broadcasted self-filmed video footage of their acts in real time (Rauf, 2021). Such extreme events can attract imitators, as the attack in Halle, Germany has shown. There, the perpetrator also live-streamed footage of his attack on social media (Kessling et al., 2020).

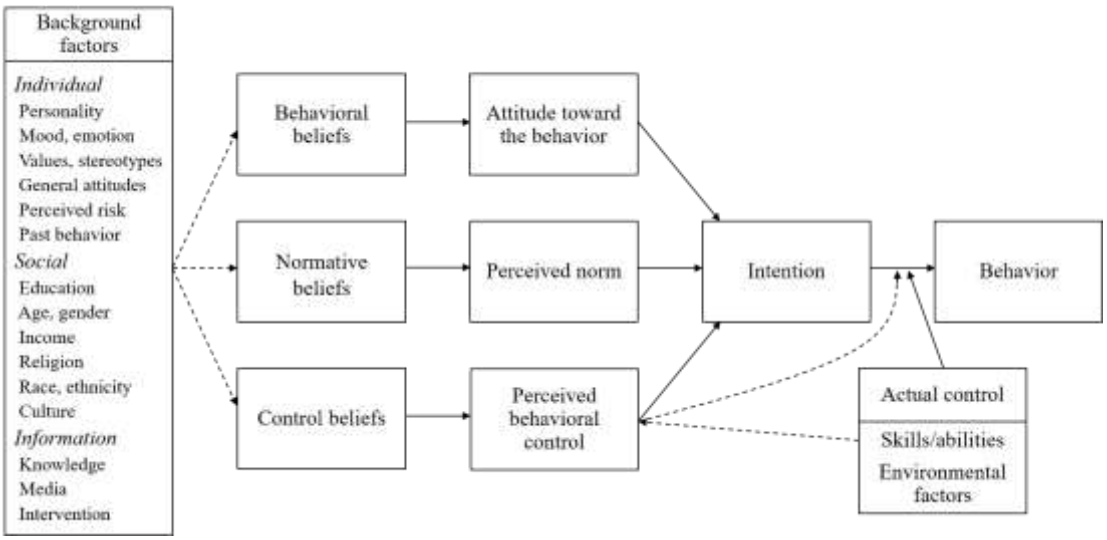
The specific mechanics of social media reinforce the tendency toward political radicalization among some users. The platforms' algorithms analyze user behavior and use the data to show content based on users' preferences. This leads to a higher likelihood of the user viewing, sharing or commenting the content. However, these posts usually correspond to the user's beliefs, attitudes and preferences and therefore tend to support his or her existing worldview. Posts that critically question the existing worldview, on the other hand, tend to be sorted out. The user moves into a virtual *filter bubble* that continuously reinforces his or her own beliefs and attitudes (M. Cinelli et al., 2021; Garimella et al., 2018). The continuous

reinforcement in turn contributes to a shift in voter potential from the center to the fringes of the political spectrum (Bail et al., 2018; Banks et al., 2021). In extreme cases, it even contributes to the justification, support or actual execution of illegal or violent – i.e., radical – political actions (Huey, 2015; Thompson, 2011).

2.3 The Psychology of Behavior Change

Following the *reasoned action approach* (RAA), radical behavior may be modified by changing the underlying beliefs through communication measures (Fishbein & Ajzen, 2011). In general, beliefs represent the individual state of information regarding a particular behavior. New information changes the current state. This immediately (and often involuntarily) leads to changes in attitudes, perceived social pressure, and perceived control over the behavior. Taken together, these three factors determine behavioral intentions, which reflect the individual level of motivation and constitute the single best predictor of actual performance (Fishbein & Ajzen, 1977). For example, if a person has strong radical intentions, he or she will probably engage in radical actions – at least if no personal or environmental factors prevent him or her from doing so (see Figure 2).

Figure 2. The Reasoned Action Approach (RAA)



Notes: This figure shows a schematic representation of the *reasoned action approach*. Own representation based on Fishbein & Ajzen (2011).

The RAA distinguishes between behavioral, normative, and control beliefs. Most important to our research purpose are *behavioral beliefs*. They represent the individual level of information regarding the positive and negative outcomes of a particular behavior and determine a person's attitudes towards that behavior. Besides, *normative beliefs* represent the level of information regarding social norms. Normative beliefs can refer to injunctive or descriptive norms. *Injunctive norms* describe the degree of approval or disapproval of a certain behavior by the relevant peer group. *Descriptive norms* describe whether members of the relevant peer group would commit the behavior themselves. They determine perceived social pressure. *Control beliefs* refer to personal or environmental factors that promote or impede the behavior and determine perceived control, i.e., *self-efficacy*. Attitudes, perceived norms, and perceived control together determine behavioral intention, as mentioned above. The relative weight of the different beliefs depends on the behavior and situation at hand.

3. Policy Intervention and Research Hypotheses

3.1 Policy Intervention

This study evaluates an online policy intervention designed to prevent political radicalization or, more specifically, to lower the level of *radicalism in attitudes* and *radicalization intentions*. German police authorities have developed a growing interest in such interventions because of the increasing importance of social media for security related issues and the public outrage following the terrorist attacks in Halle (October 9, 2019) and Hanau (February 20, 2020). In both cases, the perpetrators had apparently radicalized themselves in the pertinent social media. The intervention to be evaluated consists of an interactive film that tells the story of teenage friends Lea and Chris. Chris (the male protagonist) is increasingly drawn into a maelstrom of conspiracy myths and antisemitism through the influence of false friends and pertinent social media. The conspiracy myths in the film claim that climate change is a hoax used by a Jewish financial elite to bring climate refugees to Europe and displace the European population. Lea (the female protagonist) comes from a Jewish family and tries to

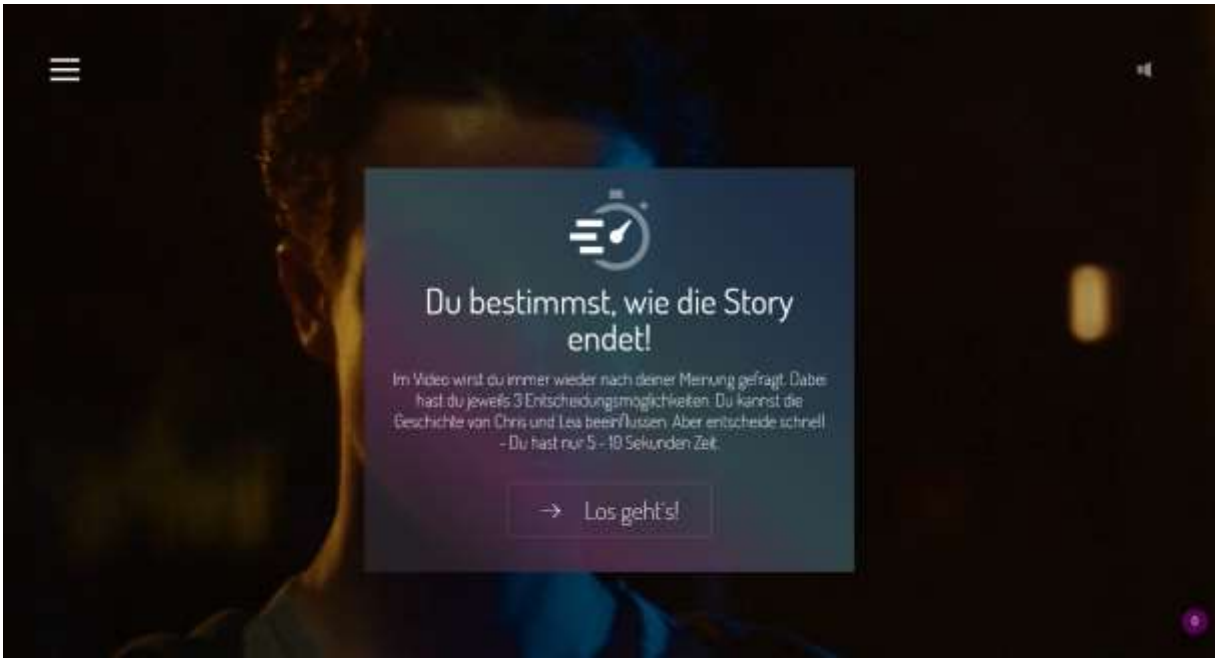
bring him back to the center of society by means of *counterspeech*.[2] Due to Chris's progressive radicalization, the relationship between the two increasingly deteriorates. The situation finally culminates into a friend of Chris planning an arson attack on the car of Lea's parents.

The film was implemented on a proprietary website to allow the use of game principles and game design elements.[3] Figure 3 shows some examples of the game design elements. Before the film starts, a pop up window informs the viewer that he or she will be asked for his or her opinion during the film, which will determine the ongoing of the plot (a). More specifically, the viewer must take a position on 12 increasingly radical statements made by the characters.[4] By clicking one of three buttons, he or she can take a *negative*, *neutral* or *positive* position to be exact (b). Depending on his or her positioning, the viewer will gain scores, which will in turn determine the end of the film. If the viewer agreed with the radical statements made, the film will take a bad end in the form of the arson attack. The same happens if the viewer fails to click on the buttons in time. The buttons are only shown for five to ten seconds. If he or she disagrees with the radical statements, the film will take a good end as the arson attack is prevented. The users had only a limited time for each decision, because the next one follows already a short time later. Following the respective end of the film, the viewer receives additional feedback (besides the film end) in graphic and text form. As a graphical feature, a traffic light shows where the viewer stands on the spectrum from *non-radical* (green) o *latent radical* (yellow) to *manifest radical* (red) according to his or her score (c). A short text verbalizes this result and, if necessary, invites the viewer to try again, to share the campaign content on social media, or to browse in-depth information on the campaign website. In addition, the viewer receives a detailed feedback to each of his or her clicks in text format (d).

[2] *Counterspeech* broadly describes citizens' responses to hate speech (or misinformation) in order to stop it, reduce its harmful consequences, and discourage it (Rieger et al., 2018). These responses usually consist of showing empathy and introducing alternative narratives instead of censorship or hate speech in the opposite direction (Kohn, 2018).
[3] The film was produced and published by the joint organization on crime prevention of the federal police and the polices of the federal states (Polizeiliche Kriminalprävention der Länder und des Bundes (ProPK)). The address of the website is https://www.zivile-helden.de/.
[4] Table A.1 in the Appendix shows the items and possible reactions (buttons).

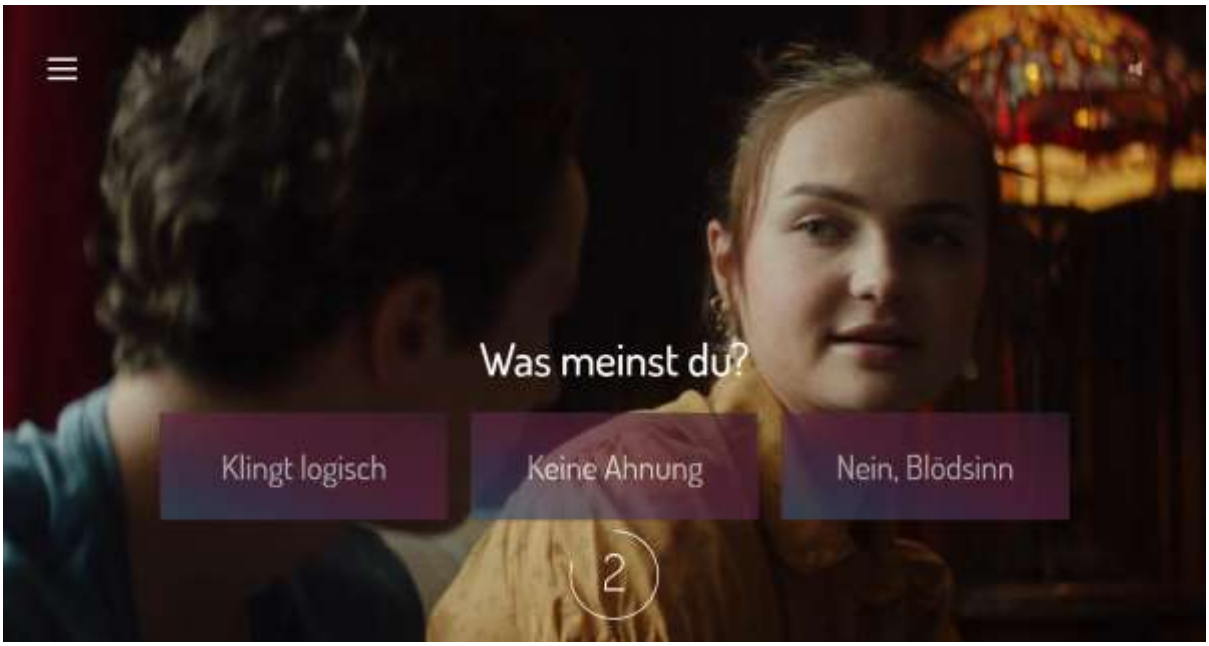
Figure 3. Screenshots from the Interactive Film

a) Start Screen



Notes: This screenshot shows the start screen of the interactive film. The text says, "*You decide how the story ends! In the video you will be asked for your opinion again and again. You have 3 options to make a decision. You can influence the story of Chris and Lea. But decide quickly - you only have 5-10 seconds.*"

b) Example of a Decision Situation

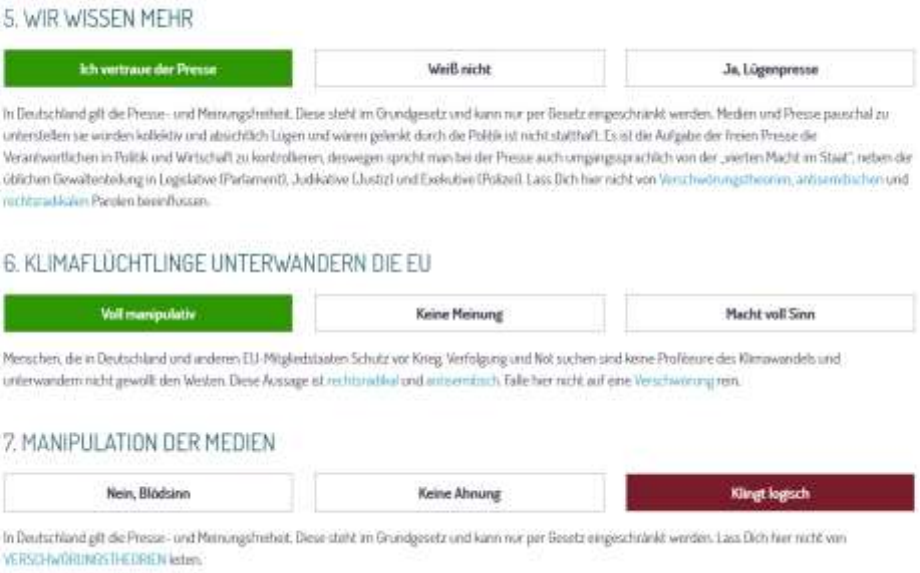


Notes: This screenshot shows an exemplary decision situation. It refers to item 7, which states that a small group benefits from climate change, the whole thing is for scaremongering, and European society is being infiltrated. The question says, "*What do you think?"* The buttons say, "*Sounds logical*", "*No idea*", and "*No, nonsense*".

c) Final Screen with Traffic Light

Notes: This screenshot shows the final screen of the interactive film. The text says, "*Just gone well again! Your attitudes are not completely in the radical spectrum, but maybe you look again more closely.*" The call to action below the traffic light says, "*Are your friends also on `Zivile Helden´ (civilian heroes)? Ask them on Facebook, Instagram or Twitter!*" The last low says, "*Tips for dealing with anti-Semitism and conspiracy theories*", and contains a link to more in-depth information.

d) Feedback to Individual Decisions



Notes: This screenshot shows examples of the in-depth feedback given after the film to the individual choices made. For the example, point 5 refers to trust in the press. It says, "*In Germany, freedom of the press and freedom of opinion apply. This is laid down in the Basic Law and can only be restricted by law. It is not permissible to make sweeping accusations that the media and the press are collectively and deliberately lying and are controlled by politics. It is the task of the free press to control those responsible in politics and business. That is why the press is colloquially referred to as the "fourth power in the state," in addition to the usual division of powers into the legislative, judicial and executive branches. Don't let yourself be influenced by conspiracy theories, anti-Semitic and radical right-wing slogans.*" *Source*: https://www.zivile-helden.de/verschwoerungsmythen/.

The elaboration of the plot closely mirrors the *two-pyramids model*. On the one hand, the character of Chris illustrates the *opinion pyramid* by evolving from *neutral* to *sympathizer*. While initially uninterested in the political cause, he begins to believe in it more and more. Whether he also takes the next step from *sympathizer* to *justifier* depends on viewer's choices during the film. If these choices indicate agreement with the radical statements, Chris will justify violence and let his friend carry out the arson attack. In reverse, if they indicate disagreement with the radical statements, Chris will object to violence and prevent the arson attack by calling the police. On the other hand, the character of Chris' friend

(the bully in class) illustrates the *action pyramid* by evolving from *inert* to (potential) *terrorist*.

3.2 Research Hypotheses

As the basis for our evaluation, we derived two testable research hypotheses from the theoretical considerations above. Following the RAA, we can change a given behavior by changing the underlying beliefs, attitudes, and perceptions of social pressure and self-efficacy. The policy intervention evaluated by this study aims to reduce the level of radicalism in viewers' *attitudes* by changing their underlying *behavioral beliefs*, i.e., their level of information regarding the costs and benefits of radical behavior. Indeed, the interactive film shows vividly that the costs of radical behavior exceed its benefits by far. In the course of their radicalization process, Chris and his friends bear increasing costs such as the deterioration of their reputation, punishment at school, and eventually even arrest. Additionally, Chris forfeits social capital by losing his friendship with Lea. He also bears psychological costs in the form of a guilty conscience in the aftermath of the attack. Compared to this, the benefit from radicalizing is vanishingly small. It includes the usual factors such as feelings of connectedness within the inside group and superiority over outsiders, as well as a sense of purpose and meaning. As the viewer learns the costs of radicalization exceed the benefits, his or her attitudes will shift in a favorable direction leading to our first research hypothesis:

Hypothesis 1: The interactive film has a negative causal effect on the level of radicalism in individual attitudes.

Moreover, according to the RAA, the favorable shift in attitudes caused by the policy intervention will directly translate into a favorable shift in behavioral intention – unless there are opposing effects on perceived social pressure or self-efficacy. It is plausible to assume that the interactive film will shift perceived social pressure and self-efficacy in a favorable direction. Alternatively, if we take a conservative view, we could assume that both factors remain constant. Perceived social pressure is a product of normative beliefs, i.e., the individual level of information regarding the expectations (inductive norms) or behavior

(descriptive norms) of the relevant peer group. The interactive film conveys the general message of society disapproving of radical behavior. The perceived social pressure to abstain from radical behavior will thus increase, or remain constant under a conservative view.

Finally, self-efficacy is a product of control beliefs, i.e., the individual level of information regarding the factors that promote or impede a given behavior. The interactive film describes the factors that promote political radicalization in detail such as socializing with radical peers or consuming radical media. Knowing these factors will increase the viewer's perceived self-efficacy regarding political radicalization. Conservatively, self-efficacy will remain constant. Thus, if the user has already adopted a less radical attitude because of the interactive film, this will translate directly into a lower radicalization intention due to increased (or constant) self-efficacy. Taken together, this leads to our second research hypothesis:

Hypothesis 2: The interactive film has a negative causal effect on individual radicalization intentions.

4. Research Design

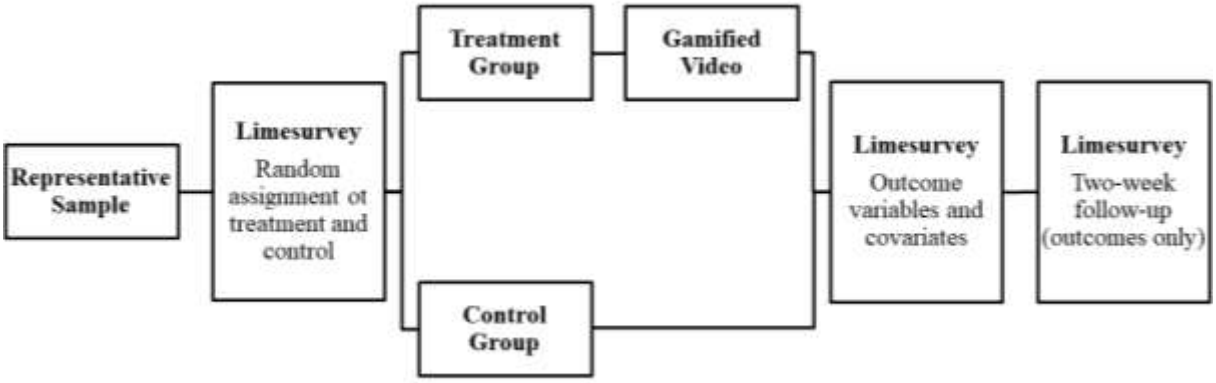
4.1 Data Collection

To test our research hypotheses, we conducted a *randomized controlled trial* (RCT).[5] The trial comprised the random allocation of individuals in the analysis sample into a treatment and a control group by equal chance. To measure potential dynamics in treatment effects over two weeks after treatment, the RCT was set up as a panel survey with two waves (Figure 4), which could be defined as a post-test only control group design. Based on the assumption of randomization, this design ensures that participants of the treatment and control groups do not differ in observable or unobservable characteristics except the treatment. The design thus yields unbiased estimation results of the treatment effect. We collected the two waves of data between October 4 and November 9, 2021. The first wave took place between October 4 and 18, 2021. The analysis sample is representative in terms of age and gender *(cross-quoted*) of the German working population and was provided by a professional market

[5] As mentioned above, the experiment is registered at the *American Economic Association* (AEA) registry: *"Evaluating a gamified online intervention to prevent political radicalization."* (AEARCTR-0008329).

research company. The gross sample size was 4,122. For the second wave, each individual participant was invited exactly two weeks after he or she had participated in the first wave. Here, gross sample size was 3,447.

Figure 4. Research Design



Notes: Own representation.

We used *LimeSurvey* for data collection, where we integrated a random number generator that assigned participants to the treatment group and the control group. The treatment group was redirected to the website with the interactive film. After completing the film, members of the treatment group were redirected to *LimeSurvey*. The control group received no treatment and was directed through the survey directly. In both groups, we measured our primary outcomes, secondary outcomes, and covariates. Covariates contained age (in groups), gender, educational attainment (by highest degree), occupational status (categories), relationship status (cohabitation), parenthood (children), urban/non-urban resident, immigrant background, federal state, and two personality measures (self-esteem and locus of control). With these covariates, we can separately analyze treatment effects for different socio-demographic groups.[6]

The chosen personality measures are closely related to political radicalization. Locus of control describes the individual tendency to attribute achievements or failures in life to either one's own abilities (interal locus of control) or to external factors (external locus of

[6] Table A.2 in the Appendix compares means of selected covariates for treatment and control group.

control; Rotter, 1966). Therefore, locus of control influences whether an individual attributes perceived grievances to his or her own failings or to a particular image of the enemy (e.g. the West, unwanted foreigners etc.), the latter often triggering a process of radicalization (Vergani et al., 2020). Self-esteem is a concept of self-image and describes the subjective evaluation of one's own worth (Rosenberg, 2015). Threats to self-esteem may increase group identification, which may also trigger radicalization (Moskalenko & McCauley, 2009). We employed 5-point Likert scales for both measures, for locus of control from external to internal, and for self-esteem from low to high.

We checked the data carefully to ensure the validity of our results. For this purpose, we used a *screening* question and identified *speeders* and *straight liners*. We assumed speeding if a candidate's interview time was below one-third of the median interview time. Straight-liners were candidates who gave exactly the same answer to every single choice question. If a candidate failed at least two of the three quality criteria (i.e., screening, speeding, or straight lining), the observation was excluded from the analysis. We also checked for outliers and implausible answers. Furthermore, the market research company assigned a unique user id to each individual participant, in order to prevent *ballot box stuffing*, and to ensure that candidates could participate only once. In wave 1, the average completion time was approximately 8 minutes and 7 seconds, and the median completion time was 6 minutes and 32 seconds. In wave 2, the average was approximately three and a half minutes, and the median was 2 minutes and 25 seconds. The relatively short interview times in the second wave came about because, we only measured the primary and secondary outcomes and left out the covariates to avoid redundancy.

After we cleaned the data, the total sample sizes were 3,991 in wave 1 and 3,237 in wave 2. Panel attrition with about 16% was quite low. In the control group, there were 2,006 observations in the first wave and 1,685 in the second wave, which is equivalent to an attrition rate of about 16%. In the treatment group, there were 1,985 and 1,552 observations in the first and second wave, respectively. This is equivalent to an attrition rate of about 22%. Although panel attrition was within a normal range, we tested for non-random attrition (*cf.* Williams, 2021) to avoid biased estimates. The results from linear probability models (see Appendix A.3 for results) show that some factors increase the attrition probability. For example, participants

who were assigned to the treatment group, who belonged to the youngest age group, or who had no educational attainment had a lower probability of participating in the second survey wave. Nevertheless, as shown below, we checked for potential bias by re-estimating the treatment effects for the balanced panel.

4.2 Operationalization of the Outcome Variables

As explained above, the interactive film aimed to lower the individual level of *radicalism in attitudes* and *radicalization intentions*. To operationalize these two outcomes, we employed the *Sympathy for Violent Radicalization and Terrorism* (SyfoR) scale by Bhui et al. (2014) and the radicalism subscale of the *Activist-Radicalism-Intentions-Scale* (ARIS) by Moskalenko & McCauley (2009). Validity, objectivity and reliability of these two scales are well established. Both are widely used in academic research, which allows comparing our results with those from other studies.

To measure the level of *radicalism in attitudes*, we used an adjusted and translated version of the SyfoR scale. In social psychology, an attitude is an individual evaluation of an attitude object (i.e. a person, thing, or event), which can range from extremely positive to extremely negative (Fishbein & Ajzen, 2011). In line with this, the SyfoR scale surveys the individual evaluations of 12 political actions on a 5-point Likert-Scale ranging from 1 for "*fully condemn*" to 5 for "*fully support*". We adjusted the original SyfoR scale by shortening it to eight items and translating it into German.

To measure the level of respondents' *radicalization intentions*, we used the radicalism subscale of the ARIS. Behavioral intention is the individual likelihood of performing a particular action (Fishbein & Ajzen, 2011). Political actions, in particular, always refer to a specified social group or political cause. Accordingly, the ARIS first surveys the social group or political cause that is most important to the respondent. Subsequently, it surveys the individual likelihood of performing eight political actions to support (or defend) the respective group (or cause) using a 5-point Likert-scale from 1 for "*very unlikely*" to 5 for "*very likely*". The ARIS dissects into two subscales: The *Activist-Intentions-Scale* (AIS) measures the likelihood of performing legal and non-violent political actions. The *Radicalism-Intentions-Subscale* (RIS), on the other hand, refers to illegal and violent political

actions. Remember we chose radicalization intentions as a primary outcome because we cannot observe illegal or violent actions in an experimental setting and behavioral intentions represent the single best predictor of actual behavior (Fishbein & Ajzen, 2011). Surveying both subscales ensured comparability with other studies and allowed investigating whether the interactive film was able to specifically address radicalism without suppressing the quite desirable activism.

For each of the four scales described above (i.e. SyfoR, RIS, AIS, and ARIS), we generated a standardized variable (*z-score*) by summing over the scale items, subtracting the mean of the control group, and dividing by the standard deviation of the control group (*cf.* Kling et al., 2007). The *z-scores* made up our primary outcome variables including the (1) *SyfoR Score*, (2) *RIS Score*, (3) *AIS Score*; and (4) *ARIS Score*. Due to the way they are calculated, the scores provided *standardized effects*, which can be interpreted as the differences between treatment and control group in percentages of a standard deviation. Standardized effects allow us, for example, to factor out any *level effects* when comparing different subgroups of the population. However, for each scale, we also looked at the simple sums over the items to get a sense of the level effects.

Similarly, we surveyed our secondary outcomes using the relevant psychological scales and calculated a *z-score* for each one. As mentioned above, we incorporated some of the risk and protective factors against radical attitudes, intentions, and behavior that had the largest effect sizes according to meta-analytic evidence (Wolfowicz et al., 2020). In particular, we incorporated *law abidance* (Bergmann & Baier, 2015), *propensity to authority* (Hübner et al., 2014), ties to *similar peers* (Wojcieszak, 2010), *risk taking* (Falk et al., 2016, 2018), and *self-control* (Seipel, 2014). In contrast, we left out *school bonding*, *law legitimacy*, *thrill seeking*, *radical peers*, and *criminal history*, because they were either difficult to measure in the context of an online survey, redundant with other outcomes, or represented sensitive information whose collection would have led to an increased dropout rate.

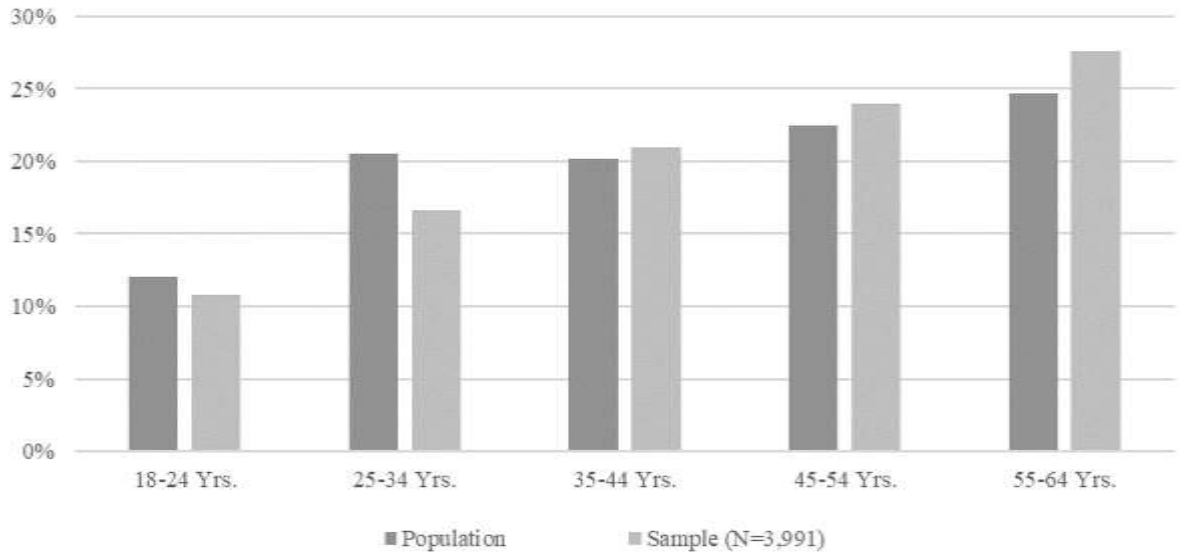
4.3 Sample Description

To ensure the external validity of our results, we aimed to draw a representative sample of the German working population. Figure 5 compares selected characteristics of our

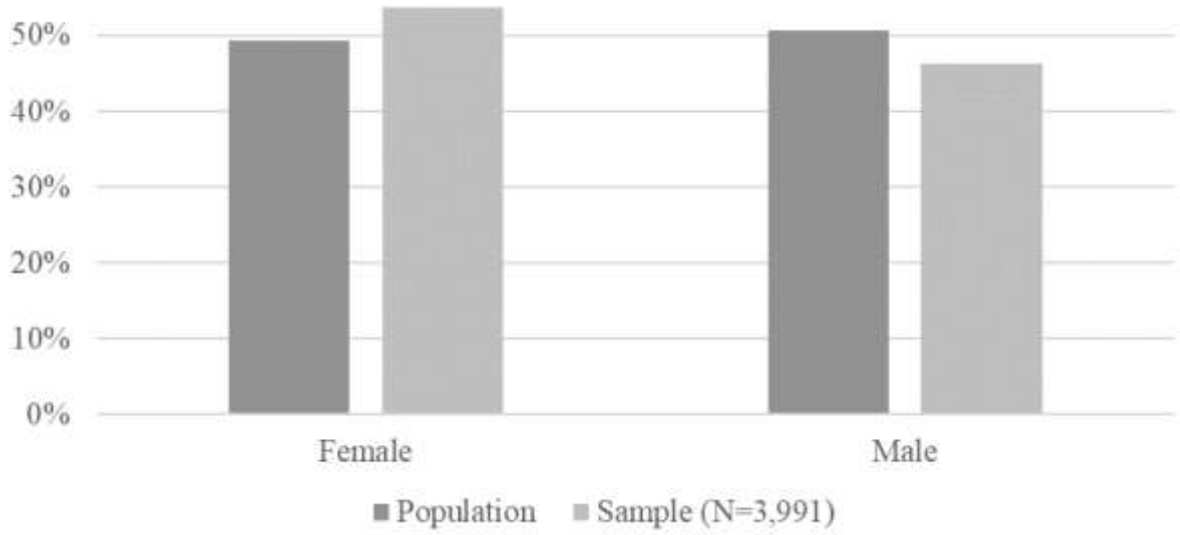
sample with the population. We selected age (a) and gender (b) as the sample should be representative along these two dimensions. Although the oldest age group is slightly overrepresented, the sample reflects the age structure of the working population quite well. The same applies to the gender distribution, although women are slightly overrepresented, here. Overall, it seems reasonable to say that the sample is representative of the German working population along the dimensions of age and gender.

Figure 5. Comparison of Age and Gender Distributions

a) Age Distribution



b) Gender Distribution



Notes: The figure compares the age distribution and the gender distribution in the sample and in the German working population See Table A.3 in the appendix for detailed numbers. DESTATIS (2020) & own data.

Checking for Balance

The internal validity of our results depends heavily on whether we were able to randomly assign participants into a treatment and a control group. Only in this case, both groups have the same characteristics on average and differ only by exposure to treatment – i.e., the groups are *balanced*. Random assignment worked well in our experiment. We checked for balance by regressing assignment to treatment jointly on all of the covariates. The large share of insignificant covariates and the close to zero adjusted coefficient of determination (adj. R^2) indicate that the groups have the same characteristics on average and only differ by treatment (Table 1). We can thus assign any differences in outcomes to this single factor. In other words, the experiment provides unbiased results and has high internal validity. Table A.4 in the appendix gives a detailed comparison of both groups for variables considered in the estimation (see below).

Table 1. Balancing Checks (Wave 1)

Dependent Variable: Assignment to Treatment	
Share of insignificant covariates	20/22
F-value	2.627
Prob > F	0.049
R-Squared	0.012
Adjusted R-Squared	0.008
No. of Observations	3,991

Notes: This table shows the statistics from regressing the treatment indicator on all of the selected covariates. The covariates comprise sociodemographic characteristics including age group (5 categories), female (dummy), educational attainment (5 categories), and employed status (5 categories). They further comprise indicators of social integration including relationship status (dummy), parenthood status (dummy), urban residency (dummy), and migration status (dummy). Finally, the comprise preferences including locus of control (5-point Likert-scale from 1-5) and self-esteem (5-point Likert-scale from 1-5).

5. Empirical Findings

5.1 Main Results

The goal of the experiment was to test our hypotheses that the interactive film reduces the level of *radicalism in attitudes* (H1) and *radicalization intentions* (H2). Table 2 shows the

main results of the experiment from the first wave of the survey. To improve statistical precision, we estimated linear regression models (*ordinary least squares*, OLS). For this purpose, we modeled each of the four primary outcome variables as a function of assignment to the treatment group and different sets of covariates. As we used standardized outcome variables (see section 4.2 above), the estimated coefficients represent *average standardized effects*, which can be interpreted as treatment effects in percentages of one standard deviation.

Columns (1)-(4) contain the results for different model specifications. For the baseline specification, we used assignment to the treatment group as the only explanatory variable. The results of this specification are thus equivalent to a simple mean comparison and represent the core experimental estimates. For the second specification, we additionally controlled for demographic characteristics such as age group, gender, educational attainment (academic versus non-academic), and occupational status. For the third specification, we added indicators of social integration, including relationship status, parenthood, urban residence, and immigrant background. In the final specification, we additionally controlled for locus of control and self-esteem. The row blocks of three lines each refer to one of the four different outcome variables considered.

JOURNAL FOR DERADICALIZATION

Spring 2022
No. 30
ISSN: 2363-9849

Table 2. Main Results (Wave 1)

	(1)	(2)	(3)	(4)
	Specification			
Outcome	Baseline	Demographics	Integration	Preferences
Radicalism in Attitudes (SyfoR)	-0.13***	-0.14***	-0.14***	-0.12***
	(0.03)	(0.03)	(0.03)	(0.03)
Adjusted R^2	0.00	0.10	0.10	0.14
Radicalization Intentions (RIS)	-0.17***	-0.18***	-0.18***	-0.15***
	(0.03)	(0.03)	(0.03)	(0.03)
Adjusted R^2	0.01	0.08	0.08	0.13
Activist Intentions (AIS)	-0.08***	-0.11***	-0.11***	-0.10***
	(0.03)	(0.03)	(0.03)	(0.03)
Adjusted R^2	0.00	0.05	0.06	0.07
Activist-Radicalization Intentions (ARIS)	-0.14***	-0.16***	-0.16***	-0.15***
	(0.03)	(0.03)	(0.03)	(0.03)
Adjusted R^2	0.01	0.07	0.08	0.11
Number of Observations	3,991			

Notes: This table shows standardized treatment effects on our four primary outcome variables in the first survey wave. The effects are estimated coefficients from linear regressions (*Ordinary Least Squares*, OLS). For the regression models, we specified the standardized outcome variables as functions of assignment to the treatment group and different sets of covariates. The columns (1)-(4) show the coefficients from different model specifications. For the baseline specification, we controlled for assignment to the treatment group only. For the second specification, we additionally controlled for demographic characteristics such as age group, gender, educational attainment (5 categories), and occupational status (5 categories). For the third specification, we added indicators of social integration, including relationship status, parenthood, urban residence, and immigrant background. In the final specification, we additionally controlled for locus of control and self-esteem. The full results from specification 4 including the estimated coefficients of the covariates can be found in Table A.5 in the Appendix. Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

The results from the first survey wave show that the interactive film had the expected immediate effects on all of our primary outcomes. These results are quite robust as the effects were highly significant without exception and relatively constant across the different model specifications. Immediately after exposure, the film decreased the average level of *radicalism in attitudes* by 12% to 14% of a standard deviation. Notably, it achieved a meaningful

discrimination between activism and radicalism. While *radicalization intentions* decreased by 15% to 18% of a standard deviation, *activist intentions* only decreased by 8% to 11%. As expected, the values for the *ARIS score* as a comprehensive measure lay roughly in between.

The table also shows that the adjusted R^2 increases as we add more covariates. However, this merely means that certain covariates influence the levels of the outcomes – or the variation in the levels of the outcomes, to be precise. For example, older individuals, women, and individuals with an internal locus of control tend to be less radical (see Table A.5 in the Appendix for detailed regression results). Because these and other covariates appear to affect the levels of outcomes, we also performed a heterogeneity analysis (see section 5.2). The key point, however, is that the estimated treatment effects remain unchanged despite the addition of further covariates. This shows that even the baseline specification yielded unbiased estimation results, as would be expected in an RCT.

Because we observed higher sample attrition in the treatment group than in the control group, we repeated the estimations above using only the observations that participated in both waves of the survey. The results of these re-estimations do not systematically differ from those shown above, indicating that sample attrition did not bias the treatment effects (see Table A.7. in the appendix).

Table 3. Main Results (Wave 2)

	(1)	(2)	(3)	(4)
	Specification			
Outcome	Baseline	Demographics	Integration	Preferences
Radicalism in Attitudes (SyfoR)	-0.07**	-0.06**	-0.07**	-0.05
	(0.03)	(0.03)	(0.03)	(0.03)
Adjusted R^2	0.00	0.08	0.08	0.11
Radicalization-Intentions (RIS)	-0.07*	-0.06*	-0.06**	-0.05
	(0.03)	(0.03)	(0.03)	(0.03)
Adjusted R^2	0.00	0.09	0.10	0.13
Activist-Intentions (AIS)	0.00	-0.02	-0.02	-0.02
	(0.03)	(0.03)	(0.03)	(0.03)
Adjusted R^2	0	0.05	0.06	0.07
Activist-Radicalization-Intentions (ARIS)	-0.04	-0.04	-0.05	-0.04
	(0.03)	(0.03)	(0.03)	(0.03)
Adjusted R^2	0.00	0.08	0.09	0.11
Number of Observations	3,237			

Notes: This table shows standardized treatment effects on our four primary outcome variables in the second survey wave. The effects are estimated coefficients from linear regressions (*Ordinary Least Squares*, OLS). For the regression models, we specified the standardized outcome variables as functions of assignment to the treatment group and different sets of covariates. The columns (1)-(4) show the coefficients from different model specifications. For the baseline specification, we controlled for assignment to the treatment group only. For the second specification, we additionally controlled for demographic characteristics such as age group, gender, educational attainment (5 categories), and occupational status (5 categories). For the third specification, we added indicators of social integration, including relationship status, parenthood, urban residence, and immigrant background. In the final specification, we additionally controlled for locus of control and self-esteem. The full results from specification 4 including the estimated coefficients of the covariates can be found in Table A.6 in the Appendix. Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

The film still had the expected negative effects on the two most important outcomes, even though the pattern of results was no longer that clear two weeks later (Table 3). Among the participants we had exposed to the film, the level of *radicalism in attitudes* was still between 6% and 7% of a standard deviation lower than among those who received no treatment. Except for the final model specification, these effects were still statistically

significant. The effects on *radicalization intentions* lay in same range, even though they were only weakly significant to significant (or even insignificant for the final specification). The film still achieved a meaningful discrimination between radicalism and activism, as the effect on *activist intentions* became insignificant for all specifications. The same applies to the *ARIS score*. The patterns of covariates driving the outcome levels are robust over the two weeks of observation (see Table A.6 in the Appendix for full estimation results).

To test whether participants responded honestly to the questions during the interactive film, that is, whether they took the policy intervention seriously, we estimated another set of models, where we regressed each of the outcomes on the score achieved in the film and the same covariates as in the models above (see Table A.8 in the Appendix). The results of these estimates show that there is a significant, statistical relation between the outcomes and the score. Participants who were more likely to agree with the radical statements in the film also scored higher on radical attitudes and radicalization intentions in the subsequent survey, but not on activism intentions. This suggests that they responded honestly and took the policy intervention seriously.

Overall, the presented results support our first research hypothesis that the interactive film exerted a negative causal effect on the individual level of *radicalism in attitudes*. They also support our second hypothesis that the film had a negative causal effect on individual *radicalization intentions*.

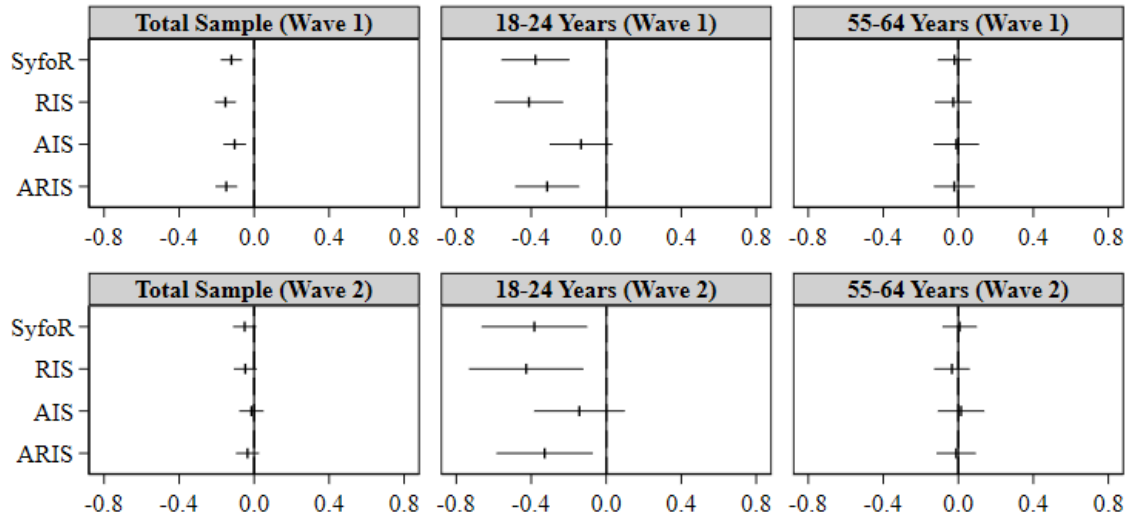
5.2 Effect Heterogeneity

Because some covariates appear to influence the levels of outcomes, we explored potential effect heterogeneity. More specifically, we analyzed whether treatment effects differed for particular subgroups. In fact, differences in age, gender, or cognitive ability to understand new information could explain differences in reactions to the film. We thus started by comparing the treatment effects between different demographic subgroups.

Figure 6 compares the youngest with the oldest age group and the total sample as a reference point. In the youngest age group, most notably, the interactive film had pronounced negative effects on the two most important outcomes that persisted even after two weeks. Immediately after exposure to the film, *radicalism in attitudes* decreased by 38%,

radicalization intentions by 41% of a standard deviation. Two weeks later, these effects remained relatively constant at 38% and 43%. All of these estimated treatment effects were highly significant and more pronounced than in the total sample. The impact on *activist intentions* was far less pronounced and insignificant from the beginning. Thus, the film not only had a greater impact on the young but also was also more accurate in distinguishing between radicalism and activism (the effect on the *ARIS score* was again in between that on the *AIS score* and the *RIS score*). In stark contrast to the youngest, the film had no significant impact on the oldest age group even immediately after exposure.

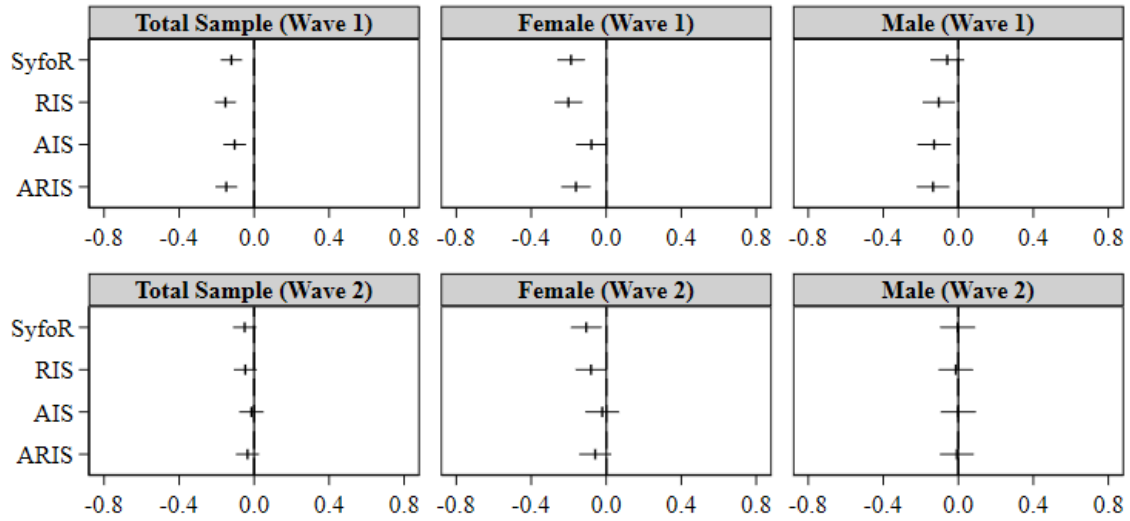
Figure 6. Treatment Effects by Age Group



Notes: This figure shows the estimated coefficients and confidence intervals for selected age groups with the total sample as a reference point across survey waves. The point estimates come from regressing the four primary outcome variables on the treatment indicator (dummy) and a set of covariates. Because the outcome variables are z-scores, the coefficients are average standardized effects. The covariates include age group (5 categories), gender (dummy), educational attainment (5 categories), occupational status (5 categories), relationship status (dummy), parenthood (dummy), urban residency (dummy), migrant background (dummy), locus of control (5-point Likert-Scale), self-esteem (5-point Likert-Scale), and fixed effects at the federal state-level. The scale of the abscissa goes from -0.8 to 0.8, i.e. -80% to 80% of a standard deviation. In wave 1, the total sample includes 3,991, the 18 to 24 age group includes 432, and the 55 to 64 age group includes 1,103 observations. In wave 2, the groups include 3,237, 177 and 993 observations, respectively.

We also identified crucial gender differences (Figure 7). Immediately after exposure, the film had highly significant negative effects on *radicalism in attitudes* and *radicalization intentions* but no significant effect on *activist intentions* among women. Therefore, it had the expected impact in this group. Among men, in contrast, the film had significant negative effects only on *radicalization intentions* and, contrary to our expectations, *activist intentions*. The effect on the *ARIS score* was highly significant in both gender groups. Two weeks later, the gender differences became fully apparent. While the film still had significant, or at least weakly significant, effects on *radicalism in attitudes* and *radicalization intentions* among women, the effects completely disappeared for men.

Figure 7. Treatment Effects by Gender

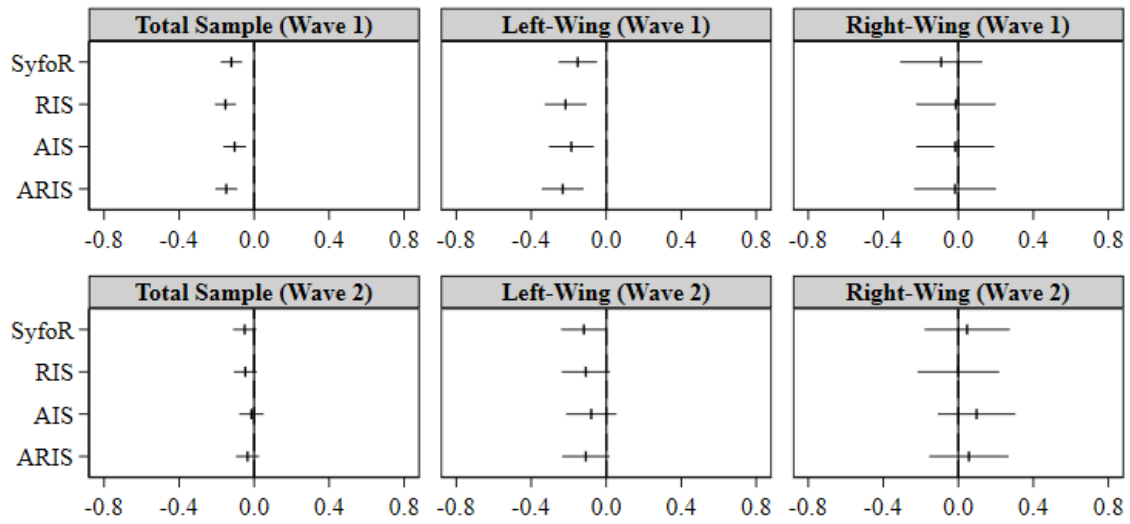


Notes: This figure shows the estimated coefficients and confidence intervals by gender with the pooled sample as a reference point across survey waves. The point estimates come from regressing the four primary outcome variables on the treatment indicator (dummy) and a set of covariates. Because the outcome variables are z-scores, the coefficients are average standardized effects. The covariates include age group (5 categories), gender (dummy), educational attainment (5 categories), occupational status (5 categories), relationship status (dummy), parenthood (dummy), urban residency (dummy), migrant background (dummy), locus of control (5-point Likert-Scale), self-esteem (5-point Likert-Scale), and fixed effects at the federal state-level. The scale of the abscissa goes from -0.8 to 0.8, i.e. -80% to 80% of a standard deviation. In wave 1, the total sample includes 3,991, the female group includes 2,142, and the male group includes 1,849 observations. In wave 2, the groups include 3,237, 1,690 and 1,547 observations, respectively.

The differences between people at the two edges of the political spectrum became fully apparent right away (Figure 8). Among people on the left edge, the film immediately had the expected negative effects on all of the outcomes. In contrast, it had almost no immediate effect among people on the right edge of the spectrum. Two weeks later, we could still observe negative effects on *radicalism in attitudes* and *radicalization intentions* among the leftists, even though they were only weakly significant. In the right-wing political camp, the effects even turned positive at that time, but were invariably insignificant.

We also analyzed differences in treatment effects by educational attainment, relationship status, parenthood, rural vs. urban and past East vs. West Germany residence, as well as migration background, but found no considerable differences between these demographic groups. Employment status mattered, but only immediately after treatment. At that time, all primary outcomes declined for the employed, while this was not the case for the unemployed. Two weeks later, however, these differences disappeared completely. The differences between groups with varying psychological traits were ambiguous. The treatment effects persisted for a longer period among individuals with high self-esteem, while they quickly faded out among those with low self-esteem. Locus of control made no difference neither immediately after treatment, nor after two weeks.

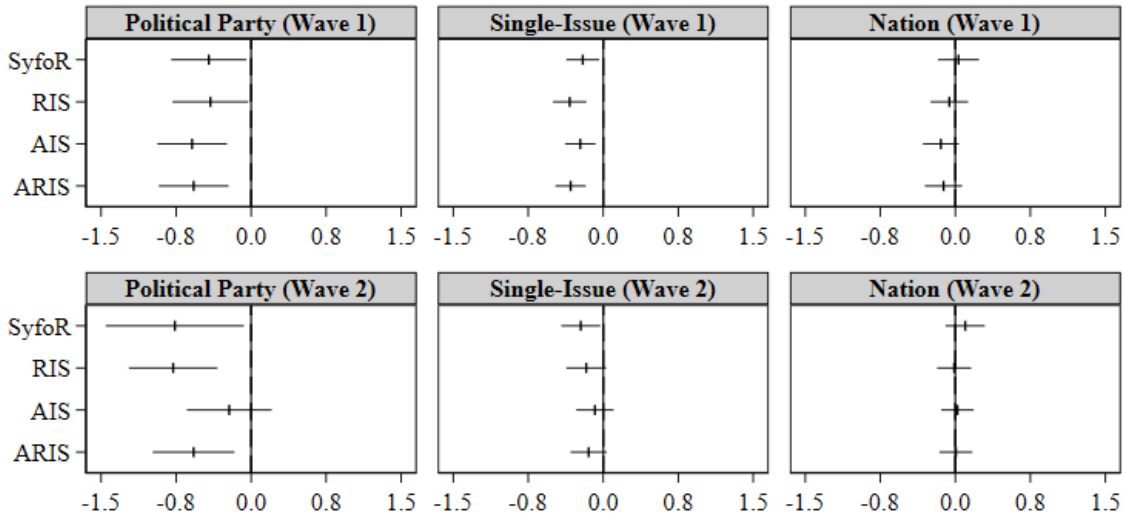
Figure 8. Treatment Effects by Political Positioning



Notes: This figure shows the estimated coefficients and confidence intervals by political positioning (self-reported) with the pooled sample as a reference point across survey waves. The point estimates come from regressing the four primary outcome variables on the treatment indicator (dummy) and a set of covariates. Because the outcome variables are z-scores, the coefficients are average standardized effects. The covariates include age group (5 categories), gender (dummy), educational attainment (5 categories), occupational status (5 categories), relationship status (dummy), parenthood (dummy), urban residency (dummy), migrant background (dummy), locus of control (5-point Likert-Scale), self-esteem (5-point Likert-Scale), and fixed effects at the federal state-level. The scale of the abscissa goes from -0.8 to 0.8, i.e. -80% to 80% of a standard deviation. In wave 1, the total sample includes 3,991, the left-wing group includes 981, and the right-wing group includes 462 observations. In wave 2, the groups include 3,237, 767 and 373 observations, respectively.

To explore more closely the extent to which *group identification* determined treatment effects, we compared the outcomes among people who identified with different social groups. Figure 9 compares the outcomes among people who identified with either a political party, a single-issue movement, or their nation. Politics seem to have mattered most here. On the one hand, we can observe the strongest and most persistent effects for people who identified with a political party or a single-issue movement. The effects among party supporters were among the strongest in the whole study, comparable only to those among the youngest. Contrary to our expectations, the film also had a conspicuously strong immediate effect on *activism intentions* among people who identified with a political party. However, this effect became insignificant two weeks later. On other hand, the film had no impact on people who identify the most with their respective nation. Overall, our empirical findings confirm the importance of group identification, which had already been established by Moskalenko & McCauley (2009). Again, we also looked at other subgroups, including people who identified with their family, religion, or an association. People who identified the most with their family made up the majority of the sample and thus resembled the main results. In the other two subgroups, we did not observe any significant effects.

Figure 9. Treatment Effects by Peer Group

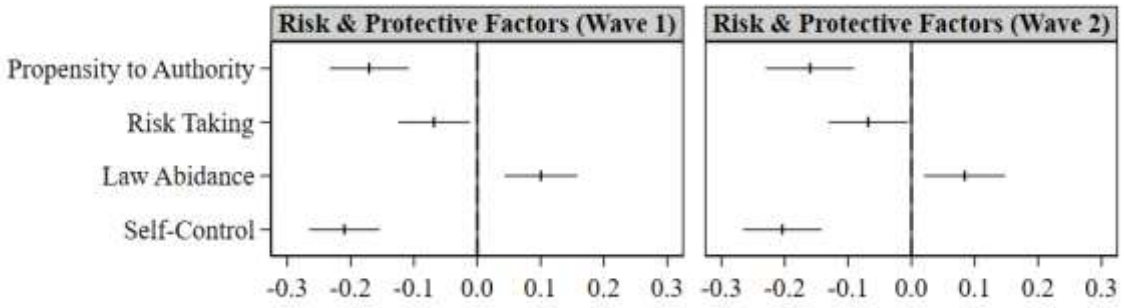


Notes: This figure shows the estimated coefficients and confidence intervals by stated peer group across survey waves. The point estimates come from regressing the four primary outcome variables on the treatment indicator (dummy) and a set of covariates. Because the outcome variables are z-scores, the coefficients are average standardized effects. The covariates include age group (5 categories), gender (dummy), educational attainment (5 categories), occupational status (5 categories), relationship status (dummy), parenthood (dummy), urban residency (dummy), migrant background (dummy), locus of control (5-point Likert-Scale), self-esteem (5-point Likert-Scale), and fixed effects at the federal state-level. The scale of the abscissa goes from -1.5 to 1.5, i.e. -150% to 150% of a standard deviation. In wave 1, the "political party" group includes 153, the "single-issue movement" group includes 559, and the "family" group includes 2,379 observations. In wave 2, the groups include 85, 331 and 2,002 observations, respectively.

5.3 Channel Analysis/Secondary Results

As mentioned above, particular risk and protective factors influence radical attitudes, intentions, and behavior. In this study, we used *propensity to authority*, *risk-taking*, *law-abidance* and *self-contro*l as secondary outcome variables because these are the risk and protective factors with the strongest effects according to meta-analytic evidence (Wolfowicz et al., 2020). To analyze the extent to which the interactive film influenced them, we repeated the regression analyses described above, with the only difference that the mentioned risk and protective factors now served as dependent variables. Figure 10 shows the estimated coefficients and confidence intervals obtained by these regressions.

Figure 10. Treatment Effects on Risk and Protective Factors



Notes: This figure shows the estimated coefficients and confidence intervals for the treatment effects on the risk and protective factors. The point estimates come from separately regressing the risk and protective factors on the treatment indicator (dummy) and a set of covariates. Because the outcome variables are z-scores, the coefficients are average standardized effects. The covariates include age group (5 categories), gender (dummy), educational attainment (5 categories), occupational status (5 categories), relationship status (dummy), parenthood (dummy), urban residency (dummy), migrant background (dummy), locus of control (5-point Likert-Scale), self-esteem (5-point Likert-Scale), and fixed effects at the federal state-level. The scale of the abscissa goes from -0.3 to 0.3, i.e. -30% to 30% of a standard deviation. N=3,237.

Immediately noticeable, all effects are persistent over the study period considered. We can further observe that the risk factors of *propensity to authority* and *risk taking* decreased due to the interactive film as we had expected. At the same time, we can observe that the film significantly strengthened the protective factor of *law abidance*. Contrary to our expectations, the film decreased the protective factor of *self-control*.

6. Discussion

Strikingly, the groups that respond most strongly to the film share the characteristics of the protagonists or general values transported by the film. First, all the main characters are students and thus closest in age to 18 to 24 year olds. Second, the person who endorses the values of the liberal, enlightened majority society and tries to have a positive influence on her social environment is a young woman. Third, the film generally represents values that tend to be located in the left-liberal milieu, such as climate protection and refugee relief. Thus, from

our point of view, it is plausible to assume that *identification* with the protagonists or transported values reinforces the treatment effects of the film.

From a theoretical perspective, when people belief they share the values, interests, and characteristics with another person, they are more likely to adopt his or her beliefs, attitudes, and behaviors (Cialdini, 2007; Kelman, 2006). *Identification* can thereby encompass actual or perceived similarities (Hoffner & Buchanan, 2005). In advertising research, it is long established that identification reinforces advertising effectiveness (e.g. Basil, 1996; Schouten et al., 2020). More recent studies on the impact of *social media influencers* support these findings (e.g. Chapple & Cownie, 2017; Djafarova & Rushworth, 2017). Based on this theoretical reasoning and empirical evidence, it seems plausible that the film had the greatest impact on the youngest age group, women and people on the (self-reported) left of the political spectrum.

The observation that radicalism among on the (self-reported) right of the political spectrum tended to even increase in the long term (albeit statistically insignificant) maybe related to the psychological phenomenon of *cognitive dissonance*. As people seek consistency between their expectations and experienced reality, new information that contradicts their beliefs causes psychological distress (Festinger, 1962). To restore consistency and relieve some distress, people will undertake great effort to justify maintaining their beliefs. This can involve misperception, rejection, or refutation of the contradicting information and thus even reinforce existing beliefs (Dillard & Harmon-Jones, 2002). The interactive film tends to contradict the existing beliefs of people on the far right of the spectrum. These will tend to justify maintaining their beliefs by misperceiving, rejecting or refuting the presented information. In extreme cases, these processes might lead to further radicalization as expressed by the (insignificant) positive effects among the rightists in the second survey wave. In fact, relieve tactics can include seeking moral support from people who share the same beliefs (Dillard & Harmon-Jones, 2002), which could explain why the effects turned positive only after two weeks. Due to the missing statistical significance of the estimates, however, this is only a tentative interpretation and should be taken with caution. Moreover, one should not overemphasize the aspect of cognitive dissonance at this point, since the

intervention has a lasting impact on the youngest age group, for example – regardless of their political views.

7. Conclusion

With this study, we evaluated an online intervention that consisted of an interactive film distributed via social media. Based on Fishbein & Ajzen's (2011) *reasoned action approach*, we had developed the hypothesis that the interactive film would lower the radicalism in individual attitudes by showing that the costs of radical behavior outweigh its benefits by far. Assuming the film would leave perceived social pressure and self-efficacy unchanged, if not shift them in a favorable direction, we further hypothesized that improved (i.e., less radical) attitudes would directly translate into a decline in radicalization intentions. The results of our experiment strongly support these two hypotheses. Among the participants we had exposed to the interactive film, we could observe a significant decline in both the level of radicalism in attitudes and radicalization intentions. As behavioral intentions are the single best predictor of actual behavior, we may conclude that the interactive film was able to prevent radical actions.

Among the general population, the described effects of the film persisted two weeks after exposure but tended to fade out a little. Among the subgroups of the 18 to 24 year olds, women, and people on the left of the political spectrum, the effects of exposure were stronger and more persistent. The observation that the interactive film has a particularly lasting impact in the youngest age group further supports and strengthens our main results, because, in this group, radical attitudes and radicalization intentions tend to be on a higher level anyway.

The described differences in treatment effects might be explained by social identification as certain subgroups closely resembled the characteristics of the film's protagonists. Thus, they were more likely to adapt the protagonists' beliefs, attitudes, and behaviors. Among people on the right, in contrast, the effects might have failed to materialize because the message of the film contradicts their core beliefs, creating cognitive dissonance. Because it causes stress, people tend to counteract the cognitive dissonance by misperceiving, rejecting or refuting opposing opinions.

Furthermore, the film was able to strengthen participants' law-abidingness as one of the most important protective factors against radical attitudes. At the same time, it lowered risk taking and propensity for authority, counteracting two of the most important risk factors of radical behavior. Suppression effects may explain the unexpected negative impact on self-control as risk-taking reflects parts of the self-control construct. However, the explanation remains an open question to be answered by subsequent research.

Our results hold important implications for researchers, practitioners, and policy-makers in the field of deradicalization. First, online interventions in general and interactive films in particular represent effective tools for preventing violent radicalism. Second, design matters. In order to and avoid the type of cognitive dissonance described above, the intervention should aim to achieve the highest possible social identification among the target group with the film (or any other online intervention). For this purpose, the setting, plot, characters, and other design elements should be closely adapted to the realities of life of those most vulnerable to radicalization. Instead of condemning their beliefs and attitudes across the board, the intervention should take their perspective and address their fears and needs. To achieve this, one could involve people in the development of the intervention who had radicalized themselves in the past and successfully made the exit from the scene. In this way, one could prevent the intervention from overly reflecting the beliefs, attitudes, preferences, and thus perhaps biases of the filmmakers. Third, timing matters. Effect heterogeneity suggests that primary prevention programs should address the target group at the earliest possible stage of the radicalization process. For people at the later stages, it might be sensible to resort to secondary or tertiary programs. Moreover, the slight fade out of effects among the general population suggests two tactical options. On the one hand, short online interventions could be disseminated at regular intervals in order to achieve a sustainable effect. On the other hand, such interventions could form the prelude to more in-depth measures, which could also take place in person.

Political radicalization will continue to pose a threat to the liberal and democratic majority society in the future. While classic media had a gatekeeper function against the worst excesses, the functionality of social media allows anyone to use their smartphone for spreading hate speech among a large audience – at any time, anywhere, very easily and

practically free of charge. It is therefore imperative that the security authorities use the same advantages of this technology to counteract the threat. As technology and the tactics of political agitators continue to evolve, authorities need to keep up to date. In order to keep learning, they must draw on reliable, scientific knowledge about programs that have already been successfully implemented and have shown to be effective in preventing radicalization. Since there has been a lack of rigorous evaluations so far, our study and results may contribute to the ongoing development of a modern prevention strategy.

References

Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. In *Action control* (pp. 11–39). Springer.

Azjen, I. (1980). Understanding attitudes and predicting social behavior. *Englewood Cliffs*.

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. B. F., Lee, J., Mann, M., Merhout, F., & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, *115*(37), 9216–9221.

Banks, A., Calvo, E., Karol, D., & Telhami, S. (2021). # polarizedfeeds: Three experiments on polarization, framing, and social media. *The International Journal of Press/Politics*, *26*(3), 609–634.

Bardwell, H., & Iqbal, M. (2021). The Economic Impact of Terrorism from 2000 to 2018. *Peace Economics, Peace Science and Public Policy*, *27*(2), 227–261.

Bartlett, J., Birdwell, J., & King, M. (2010). The edge of violence: A radical approach to extremism. *Demos*, 5–75.

Basil, M. D. (1996). Identification as a mediator of celebrity effects. *Journal of Broadcasting & Electronic Media*, *40*(4), 478–495.

Berger, J. M. (2016). *Promoting disengagement from violent extremism* (No. 5; ICCT Policy Brief). JSTOR.

Bergmann, M. C., & Baier, D. (2015). *Wir hier-Zukunft in Aachen: Ergebnisse einer Befragung von Aachener Kindern und Jugendlichen*. Kriminologisches Forschungsinst. Niedersachen.

Bhui, K., Hicks, M. H., Lashley, M., & Jones, E. (2012). A public health approach to understanding and preventing violent radicalization. *BMC Medicine*, *10*(1), 1–8.

Bhui, K., Warfa, N., & Jones, E. (2014). Is violent radicalisation associated with poverty, migration, poor self-reported health and common mental disorders? *PloS One*, *9*(3), e90718.

Bjørgo, T., & Carlsson, Y. (2005). *Early intervention with violent and racist youth groups* (No. 677; NUPI Paper). NUPI.

Borum, R. (2011). Radicalization into violent extremism I: A review of social science theories. *Journal of Strategic Security*, *4*(4), 7–36.

Castronovo, C., & Huang, L. (2012). Social media in an alternative marketing communication model. *Journal of Marketing Development and Competitiveness*, *6*(1), 117–134.

Chapple, C., & Cownie, F. (2017). An investigation into viewers' trust in and response towards disclosed paid-for-endorsements by YouTube lifestyle vloggers. *Journal of Promotional Communications*, *5*(2).

Chatfield, A. T., Reddick, C. G., & Brajawidagda, U. (2015). Tweeting propaganda, radicalization and recruitment: Islamic state supporters multi-sided twitter networks. *Proceedings of the 16th Annual International Conference on Digital Government Research*, 239–249.

Cialdini, R. B. (2007). *Influence: The psychology of persuasion* (Vol. 55). Collins New York.

Cinelli, C., Forney, A., & Pearl, J. (2020). *A crash course in good and bad controls* (No. 3689437; Social Science Research Network Papers, Vol. 3689437).

Cinelli, M., Morales, G. D. F., Galeazzi, A., Quattrociocchi, W., & Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, *118*(9).

Della Porta, D. (2013). *Clandestine political violence*. Cambridge University Press.

DESTATIS. (2020). *Bevölkerung: Altersjahre, Geschlecht*. https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Bevoelkerungsstand/_inhalt.html

Dillard, J., & Harmon-Jones, C. (2002). A cognitive dissonance theory perspective on persuasion. *The Persuasion Handbook: Developments in Theory and Practice*, 99.

Djafarova, E., & Rushworth, C. (2017). Exploring the credibility of online celebrities' Instagram profiles in influencing the purchase decisions of young female users. *Computers in Human Behavior*, *68*, 1–7.

Doosje, B., Moghaddam, F. M., Kruglanski, A. W., De Wolf, A., Mann, L., & Feddes, A. R. (2016). Terrorism, radicalization and de-radicalization. *Current Opinion in Psychology*, *11*, 79–84.

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., & Sunde, U. (2018). Global evidence on economic preferences. *The Quarterly Journal of Economics*, *133*(4), 1645–1692.

Falk, A., Becker, A., Dohmen, T., Huffman, D., & Sunde, U. (2016). *The preference survey module: A validated instrument for measuring risk, time, and social preferences* (No. 9674; IZA Discussion Paper).

Fenstermacher, L., NSI, L. K., Rieger, T., & Speckhard, A. (2010). Protecting the homeland from international and domestic terrorism threats. *White Paper: Counter Terrorism, 178*.

Festinger, L. (1962). *A theory of cognitive dissonance* (Vol. 2). Stanford university press.

Fink, C. (2018). Dangerous speech, anti-Muslim violence, and Facebook in Myanmar. *Journal of International Affairs*, *71*(1.5), 43–52.

Fishbein, M., & Ajzen, I. (1977). *Belief, attitude, intention, and behavior: An introduction to theory and research*.

Fishbein, M., & Ajzen, I. (2011). *Predicting and changing behavior: The reasoned action approach*. Taylor & Francis.

Garimella, K., De Francisci Morales, G., Gionis, A., & Mathioudakis, M. (2018). Political discourse on social media: Echo chambers, gatekeepers, and the price of bipartisanship. *Proceedings of the 2018 World Wide Web Conference*, 913–922.

Gates, S., & Podder, S. (2015). Social media, recruitment, allegiance and the Islamic State. *Perspectives on Terrorism*, *9*(4), 107–116.

Gruber, F., Lützinger, S., & Kemmesies, U. E. (2017). Extremismusprävention in Deutschland–Erhebung und Darstellung der Präventionslandschaft. *Modulabschlussbericht. Wiesbaden: Bundeskriminalamt*.

Hafez, M., & Mullins, C. (2015). The radicalization puzzle: A theoretical synthesis of empirical approaches to homegrown extremism. *Studies in Conflict & Terrorism*, *38*(11), 958–975.

Hoffner, C., & Buchanan, M. (2005). Young adults' wishful identification with television characters: The role of perceived similarity and character attributes. *Media Psychology*, *7*(4), 325–351.

Horgan, J. (2004). *The psychology of terrorism*. Routledge.

Hübner, M., Schmidt, P., Schürhoff, R., & Schwarzer, S. (2014). Allgemeine Autoritarismus-Kurzform. *Zusammenstellung Sozialwissenschaftlicher Items Und Skalen. Mannheim: GESIS*.

Huey, L. (2015). This is Not Your Mother's Terrorism: Social Media, Online Radicalization and the Practice of Political Jamming. *Journal of Terrorism Research*, *6*(2).

Jugl, I., Lösel, F., Bender, D., & King, S. (2020). Psychosocial prevention programs against radicalization and extremism: a meta-analysis of outcome evaluations. *European Journal*

of Psychology Applied to Legal Context, *13*(1), 37–46.

Kelman, H. C. (2006). Interests, relationships, identities: Three central issues for individuals and groups in negotiating their social environment. *Annu. Rev. Psychol.*, *57*, 1–26.

Kessling, P., Kiessling, B., Burkhardt, S., & Stöcker, C. (2020). Dynamic Properties of Information Diffusion Networks during the 2019 Halle Terror Attack on Twitter. *International Conference on Human-Computer Interaction*, 568–582.

Kling, J. R., Liebman, J. B., & Katz, L. F. (2007). Experimental analysis of neighborhood effects. *Econometrica*, *75*(1), 83–119.

Kober, M. (2017). Zur Evaluation von Maßnahmen der Prävention von religiöser Radikalisierung in Deutschland (On the evaluation of measures to prevent religious radicalization in Germany). *Journal for Deradicalization*, *11*, 219–257.

Kohn, S. (2018). *The opposite of hate: A field guide to repairing our humanity*. Algonquin Books.

Kruglanski, A. W., Gelfand, M. J., Bélanger, J. J., Sheveland, A., Hetiarachchi, M., & Gunaratna, R. (2014). The psychology of radicalization and deradicalization: How significance quest impacts violent extremism. *Political Psychology*, *35*, 69–93.

Lustria, M. L. A., Cortese, J., Noar, S. M., & Glueckauf, R. L. (2009). Computer-tailored health interventions delivered over the Web: review and analysis of key components. *Patient Education and Counseling*, *74*(2), 156–173.

Malmasi, S., & Zampieri, M. (2017). Detecting hate speech in social media. *ArXiv Preprint ArXiv:1712.06427*.

Mastroe, C., & Szmania, S. (2016). Surveying CVE metrics in prevention, disengagement and deradicalization programs. *Report to the Office of University Programs, Science and Technology Directorate, Department of Homeland Security*.

Mathew, B., Dutt, R., Goyal, P., & Mukherjee, A. (2019). Spread of hate speech in online social media. *Proceedings of the 10th ACM Conference on Web Science*, 173–182.

McCauley, C., & Moskalenko, S. (2011). *Friction: How radicalization happens to them and us*. oxford university Press.

McCauley, C. R., & Moskalenko, S. (2017). Understanding political radicalization: The two-pyramids model. *American Psychologist*, *72*(3), 205.

McDonald, B., & Mir, Y. (2011). Al-Qaida-influenced violent extremism, UK government prevention policy and community engagement. *Journal of Aggression, Conflict and*

Peace Research.

Moghaddam, F. M. (2005). The staircase to terrorism: A psychological exploration. *American Psychologist*, *60*(2), 161.

Moskalenko, S., & McCauley, C. R. (2009). Measuring political mobilization: The distinction between activism and radicalism. *Terrorism and Political Violence*, *21*(2), 239–260.

Neumann, P. (2013). The trouble with radicalization. *International Affairs*, *89*(4), 873–893.

Pyszczynski, T., Motyl, M., & Abdollahi, A. (2009). Righteous violence: killing for God, country, freedom and justice. *Behavioral Sciences of Terrorism and Political Aggression*, *1*(1), 12–39.

Rauf, A. A. (2021). New moralities for new media? Assessing the role of social media in acts of terror and providing points of deliberation for business ethics. *Journal of Business Ethics*, *170*(2), 229–251.

Rieger, D., Schmitt, J. B., & Frischlich, L. (2018). Hate and counter-voices in the Internet: Introduction to the special issue. *SCM Studies in Communication and Media*, *7*(4), 459–472.

Rosenberg, M. (2015). *Society and the adolescent self-image*. Princeton university press.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs: General and Applied*, *80*(1), 1.

Sageman, M. (2011). *Understanding terror networks*. University of Pennsylvania press.

Schmid, A. P. (2013). Radicalisation, de-radicalisation, counter-radicalisation: A conceptual discussion and literature review. *ICCT Research Paper*, *97*(1), 22.

Schouten, A. P., Janssen, L., & Verspaget, M. (2020). Celebrity vs. Influencer endorsements in advertising: the role of identification, credibility, and Product-Endorser fit. *International Journal of Advertising*, *39*(2), 258–281.

Seipel, C. (2014). Deutsche Version der Self-Control Skala. *Zusammenstellung Sozialwissenschaftlicher Items Und Skalen*.

Sheeran, P., Orbell, S., & Trafimow, D. (1999). Does the temporal stability of behavioral intentions moderate intention-behavior and past behavior-future behavior relations? *Personality and Social Psychology Bulletin*, *25*(6), 724–734.

Silber, M. D., Bhatt, A., & Analysts, S. I. (2007). *Radicalization in the West: The homegrown threat*. Police Department New York.

Silver, L., Huang, C., & Taylor, K. (2019). In emerging economies, smartphone and social media users have broader social networks. *Pew Research Center*.

Thompson, R. (2011). Radicalization and the use of social media. *Journal of Strategic Security*, *4*(4), 167–190.

Tsimonis, G., & Dimitriadis, S. (2014). Brand strategies in social media. *Marketing Intelligence & Planning*.

Vergani, M., Iqbal, M., Ilbahar, E., & Barton, G. (2020). The three Ps of radicalization: Push, pull and personal. A systematic scoping review of the scientific evidence about radicalization into violent extremism. *Studies in Conflict & Terrorism*, *43*(10), 854.

Weimann, G. (2012). Lone wolves in cyberspace. *Journal of Terrorism Research*.

Wiktorowicz, Q. (2004). Joining the cause: Al-Muhajiroun and radical Islam. *The Roots of Radical Islam*.

Williams, M. (2021). Attrition happens (and what to do about it). *Williams, MJ (2021). Attrition Happens (and What to Do about It). Journal for Deradicalization*, *26*, 217–226.

Winter, S., Maslowska, E., & Vos, A. L. (2021). The effects of trait-based personalization in social media advertising. *Computers in Human Behavior*, *114*, 106525.

Wolfowicz, M., Litmanovitz, Y., Weisburd, D., & Hasisi, B. (2020). A field-wide systematic review and meta-analysis of putative risk and protective factors for radicalization outcomes. *Journal of Quantitative Criminology*, *36*(3), 407–447.

Zhang, H., Zang, Z., Zhu, H., Uddin, M. I., & Amin, M. A. (2022). Big data-assisted social media analytics for business model for business decision making system competitive analysis. *Information Processing & Management*, *59*(1), 102762.

JOURNAL FOR DERADICALIZATION

Spring 2022
No. 30
ISSN: 2363-9849

Appendix

Table A.1 – Items of the Decision Situations in the Interactive Film

Statements (OWTTE)		**Possible Reactions (Buttons)**		
		Disagreement	**Neutral**	**Agreement**
1.	Why are you hanging out with the eco chick?	Outrageous!	Could be	Haha, eco chick!
2.	The climate has been changing for thousands of years - even without human intervention.	Wrong	No idea	True
3.	Climate change is a political tool for scaremongering.	The threat is real	No idea	Yes, scaremongering
4.	Climate refugees are infiltrating the European continent.	Refugees welcome	I don't care	Refugees out
5.	We are not part of the broad identity-less masses that can be manipulated by mainstream media. We know more.	I trust the press	Don't know	Yes lying press
6.	This all makes so much sense.	Completely manipulative	No opinion	Completely makes sense
7.	A small group benefits from climate change. The whole thing is for scaremongering. We are being infiltrated.	No, nonsense	No idea	Sounds logical
8.	Of all people, you want to play the victims of society. Really?	Lea is right	I don't care	Yes, victims of society
9.	Man has the need to give meaning to inexplicable things.	Some things are inexplicable	No idea	There is meaning everywhere
10.	Climate change is a lie of the Jews.	Conspiracy myth	No idea	Yes, lie of the Jews
Actions		**Possible Reactions (Buttons)**		
		Disagreement	**Neutral**	**Agreement**
11.	The bully sprays a Star of David on Lea's locker.	Stop it		Yeah, show her!
12.	The bully threatens Lea with a gesture.	You have gone too far, stop!	I stay out of it	She'll get that back!

Notes: This table contains the items of the 12 decision situations in the interactive video.

Table A.2 – Summary Statistics (Wave 1)

	(1)	(2)	(3)
	Population	**Sample**	
Variable	Mean	Mean	Standard Deviation
Age Group			
18-24	0.12	0.11	0.31
25-34	0.21	0.17	0.37
35-44	0.20	0.21	0.41
45-54	0.23	0.24	0.43
55-64	0.25	0.28	0.45
Female	0.51	0.54	0.50
Number of Observations		3,991	

Notes: This table shows the population means, sample means and sample standard deviations of selected covariates. Age group is a categorical variable meaning it provides the shares of observations in the respective classes. Female is a dummy variables meaning it takes on a value of one if the observation is female or zero otherwise.

Table A.3 – Testing for Non-Random Panel Attrition

Dependent variable: Observation remains in sample over both waves (dummy)				
	(1) Baseline	(2) + Outcome	(3) + Demographics	(4) + Demographics & Outcome
Treatment indicator	-0.22***	-0.23***	-0.22***	-0.22***
	(0.05)	(0.05)	(0.05)	(0.05)
Radicalization intentions		-0.08***		0.01
		(0.02)		(0.03)
Age group (5 categories)				
25-34			0.89***	0.89***
			(0.09)	(0.09)
35-44			1.19***	1.20***
			(0.10)	(0.10)
45-54			1.42***	1.43***
			(0.10)	(0.10)
55-64			1.44***	1.44***
			(0.10)	(0.10)
Female (dummy)			-0.09*	-0.09*
			(0.05)	(0.05)
Educational attainment (5 categories)				
Secondary			0.52	0.52
			(0.37)	(0.37)
A-levels			0.68*	0.68*
			(0.37)	(0.37)
Vocational degree			0.64*	0.65*
			(0.37)	(0.37)
Academic degree			0.68*	0.68*
			(0.37)	(0.37)
Occupational status (5 categories)				
Unemployed			-0.10	-0.10
			(0.12)	(0.12)
In training/study			-0.20	-0.20
			(0.14)	(0.14)
Part-time			-0.04	-0.04
			(0.11)	(0.11)
Full-time			-0.06	-0.06
			(0.10)	(0.10)
In committed relationship (dummy)			-0.05	-0.05
			(0.06)	(0.06)
Has children (dummy)			0.01	0.01
			(0.06)	(0.06)

Urban resident (dummy)			0.06	0.06
			(0.06)	(0.06)
Migrant background (dummy)			-0.10	-0.10
			(0.07)	(0.07)
Locus of control (5 pt. Likert-scale)			-0.01	-0.01
			(0.01)	(0.01)
Self-esteem (5 pt. Likert-scale)			-0.03	-0.03
			(0.03)	(0.03)
Constant	0.99***	1.00***	-0.45	-0.45
	(0.03)	(0.03)	(0.40)	(0.40)
Number of observations		3,991		

Notes: This table shows the estimated coefficients from probabilistic regression models. The dependent variable is a binary variable that takes a value of one if the observation participated in both survey waves and zero otherwise. The estimated coefficients thus give the change in the probability of participating in both waves. The columns (1)-(4) show the coefficients from different model specifications. For the baseline specification, we controlled for assignment to the treatment group only. For the second specification, we additionally controlled for outcome variable radicalization intentions. For the third specification, we controlled for assignment to the treatment group, demographic characteristics, indicators of social integration, and preferences. The demographic characteristics include age group, gender, educational attainment (5 categories), and occupational status (5 categories). The indicators of social integration include relationship status, parenthood, urban residence, and immigrant background. The preferences include self-esteem and locus of control. In the final specification, we additionally controlled for radicalization intentions. Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table A.4 – Comparison of Treatment and Control Groups (Wave 1)

	(1)	(2)
Variable	**Control group mean**	**Mean difference**
Age group		
18-24	0.11	-0.01
25-34	0.15	0.07***
35-44	0.21	0.01
45-54	0.26	-0.05***
55-64	0.28	0.00
Female (dummy)	0.54	-0.01
Educational attainment (5 categories)		
None	0.00	0.07
Secondary	0.15	-0.05**
A-levels	0.16	0.03
Vocational	0.45	-0.03
Academic	0.24	0.05***
Occupational status (5 categories)		
Retired/pension	0.12	-0.01
Unemployed	0.10	-0.02
In training/study	0.06	0.08**
Part-time	0.19	-0.05**
Full-time	0.53	0.02
In committed relationship (dummy)	0.67	-0.01
Has children (dummy)	0.51	-0.01
Urban resident (dummy)	0.66	0.02
Migrant background (dummy)	0.16	0.01
Locus of control (5 pt. Likert-scale)	1.94	0.01***
Self-esteem (5 pt. Likert-scale)	3.35	0.01
Number of observations	2,006	1,985

Notes: This table shows the control group means and estimated coefficients from separately regressing the treatment indicator on the respective covariate. Due to the specification of the regression models, the estimated coefficients are equivalent to mean differences between treatment and control group. In case of the categorical variables, we generated a dummy variable for each category and ran a separate regression using that variable. * p<0.10, ** p<0.05, *** p<0.01.

Table A.5 – Estimation Results (Specification 4, Wave 1)

Dependent variables: separately in columns				
	(1) Radicalism in attitudes (SyfoR)	(2) Radicalization intentions (RIS)	(3) Activist intentions (AIS)	(4) ARIS score
Treatment Indicator	-0.12***	-0.15***	-0.10***	-0.15***
	(0.03)	(0.03)	(0.03)	(0.03)
Age group (in years, reference cat.: 18-24 years))				
25-34	-0.28***	-0.25***	-0.26***	-0.29***
	(0.07)	(0.06)	(0.06)	(0.06)
35-44	-0.51***	-0.34***	-0.31***	-0.38***
	(0.07)	(0.07)	(0.06)	(0.06)
45-54	-0.75***	-0.56***	-0.46***	-0.59***
	(0.07)	(0.07)	(0.06)	(0.06)
55-64	-0.90***	-0.66***	-0.53***	-0.68***
	(0.07)	(0.07)	(0.06)	(0.06)
Female (dummy)	-0.25***	-0.13***	-0.06*	-0.11***
	(0.03)	(0.03)	(0.03)	(0.03)
Educational attainment (reference cat.: no degree)				
Secondary	-0.00	-0.20	-0.12	-0.18
	(0.32)	(0.25)	(0.26)	(0.27)
A-levels	-0.08	-0.27	0.09	-0.10
	(0.32)	(0.24)	(0.25)	(0.27)
Vocational degree	-0.11	-0.40*	-0.13	-0.31
	(0.32)	(0.24)	(0.25)	(0.27)
Academic degree	-0.06	-0.35	0.13	-0.13
	(0.32)	(0.24)	(0.25)	(0.27)
Occupational status (reference cat.: retired/pension)				
Unemployed	-0.04	0.04	-0.02	0.01
	(0.06)	(0.06)	(0.07)	(0.06)
In training/study	-0.28***	-0.07	0.13	0.03
	(0.08)	(0.08)	(0.09)	(0.08)
Part-time	-0.04	0.15***	0.03	0.10*
	(0.05)	(0.05)	(0.06)	(0.06)
Full-time	0.03	0.25***	0.12**	0.21***
	(0.05)	(0.05)	(0.06)	(0.05)
In committed relationship (dummy)	-0.00	0.00	-0.01	-0.00
	(0.03)	(0.03)	(0.04)	(0.03)
Has children (dummy)	0.11***	0.14***	0.17***	0.18***
	(0.03)	(0.03)	(0.03)	(0.03)
Urban resident (dummy)	0.02	0.07**	0.03	0.06*
	(0.03)	(0.03)	(0.03)	(0.03)
Migration background (dummy)	0.08*	0.08*	0.02	0.06
	(0.04)	(0.04)	(0.04)	(0.04)
Locus of control (5 pt. Likert-scale)	-0.07***	-0.08***	-0.02***	-0.06***
	(0.01)	(0.01)	(0.01)	(0.01)
Self-esteem (5 pt. Likert-scale)	0.06***	0.12***	0.11***	0.13***
	(0.02)	(0.02)	(0.02)	(0.02)
Constant	0.73**	0.40	-0.08	0.19
	(0.32)	(0.26)	(0.27)	(0.28)
Adjusted R^2	0.14	0.13	0.07	0.11
Number of observations	3,991			

Notes: This table shows coefficients estimates from linear regressions (*Ordinary Least Squares*, OLS) on our four primary outcome variables in the first survey wave. Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table A.6 – Estimation Results (Specification 4, Wave 2)

Dependent variables: separately in columns				
	(1) Radicalism in attitudes (SyfoR)	(2) Radicalization intentions (RIS)	(3) Activist intentions (AIS)	(4) ARIS score
Treatment Indicator	-0.05	-0.05	-0.01	-0.04
	(0.03)	(0.03)	(0.03)	(0.03)
Age group (in years, reference cat.: 18-24 years))				
25-34	-0.30***	-0.36***	-0.19**	-0.31***
	(0.09)	(0.09)	(0.08)	(0.08)
35-44	-0.50***	-0.49***	-0.27***	-0.44***
	(0.09)	(0.09)	(0.08)	(0.08)
45-54	-0.73***	-0.74***	-0.35***	-0.63***
	(0.09)	(0.09)	(0.08)	(0.08)
55-64	-0.88***	-0.88***	-0.40***	-0.74***
	(0.08)	(0.09)	(0.08)	(0.08)
Female (dummy)	-0.17***	-0.15***	-0.08**	-0.13***
	(0.03)	(0.03)	(0.04)	(0.03)
Educational attainment (reference cat.: no degree)				
Secondary	-0.32	-0.54*	-0.07	-0.35
	(0.39)	(0.28)	(0.32)	(0.33)
A-levels	-0.34	-0.63**	0.13	-0.29
	(0.39)	(0.28)	(0.32)	(0.33)
Vocational degree	-0.38	-0.75***	-0.16	-0.53
	(0.39)	(0.28)	(0.32)	(0.32)
Academic degree	-0.43	-0.74***	0.07	-0.39
	(0.39)	(0.28)	(0.32)	(0.32)
Occupational status (reference cat.: retired/pension)				
Unemployed	-0.00	-0.05	-0.09	-0.08
	(0.06)	(0.06)	(0.07)	(0.07)
In training/study	-0.24**	-0.22**	0.11	-0.07
	(0.10)	(0.10)	(0.10)	(0.09)
Part-time	0.02	0.05	0.12*	0.10*
	(0.05)	(0.06)	(0.07)	(0.06)
Full-time	0.04	0.16***	0.22***	0.22***
	(0.05)	(0.05)	(0.06)	(0.05)
In committed relationship (dummy)	-0.04	0.04	0.08**	0.07**
	(0.04)	(0.03)	(0.04)	(0.04)
Has children (dummy)	0.05	0.09***	0.12***	0.12***
	(0.03)	(0.03)	(0.04)	(0.04)
Urban resident (dummy)	0.06*	0.10***	0.05	0.09**
	(0.04)	(0.03)	(0.04)	(0.03)
Migrant background (dummy)	0.08	0.01	0.00	0.01
	(0.05)	(0.04)	(0.05)	(0.04)
Locus of control (5 pt. Likert-scale)	-0.06***	-0.07***	-0.02**	-0.05***
	(0.01)	(0.01)	(0.01)	(0.01)
Self-esteem (5 pt. Likert-scale)	0.06***	0.08***	0.10***	0.10***
	(0.02)	(0.02)	(0.02)	(0.02)
Constant	1.02**	1.06***	-0.16	0.52
	(0.41)	(0.31)	(0.34)	(0.35)
Adjusted R^2	0.11	0.13	0.07	0.11
Number of observations	3,237			

Notes: This table shows coefficients estimates from linear regressions (*Ordinary Least Squares*, OLS) on our four primary outcome variables in the second survey wave. Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table A.7 – Regression Table for Retention Sample – Wave 1

	(1)	(2)	(3)	(4)
	Specification			
Outcome	Baseline	Demographics	Integration	Preferences
Radicalism in Attitudes (SyfoR)	-0.11***	-0.11***	-0.11***	-0.09***
	(0.03)	(0.03)	(0.03)	(0.03)
Adjusted R^2	0.00	0.08	0.09	0.12
Radicalization Intentions (RIS)	-0.16***	-0.15***	-0.16***	-0.14***
	(0.03)	(0.03)	(0.03)	(0.03)
Adjusted R^2	0.01	0.08	0.08	0.13
Activist Intentions (AIS)	-0.10***	-0.12***	-0.12***	-0.12***
	(0.04)	(0.03)	(0.03)	(0.03)
Adjusted R^2	0.00	0.05	0.05	0.06
Activist-Radicalization Intentions (ARIS)	-0.15***	-0.15***	-0.16***	-0.15***
	(0.03)	(0.03)	(0.03)	(0.03)
Adjusted R^2	0.01	0.07	0.08	0.10
Number of Observations		3,237		

Notes: This table shows standardized treatment effects on our four primary outcome variables in the first survey wave for the observations that participated in both survey waves. The effects are estimated coefficients from linear regressions (*Ordinary Least Squares*, OLS). For the regression models, we specified the standardized outcome variables as functions of assignment to the treatment group and different sets of covariates. The columns (1)-(4) show the coefficients from different model specifications. For the baseline specification, we controlled for assignment to the treatment group only. For the second specification, we additionally controlled for demographic characteristics such as age group, gender, educational attainment (5 categories), and occupational status (5 categories). For the third specification, we added indicators of social integration, including relationship status, parenthood, urban residence, and immigrant background. In the final specification, we additionally controlled for locus of control and self-esteem. Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table A.8 – Regression of Main Outcomes on Game Score (Treatment Group)

	(1)	(2)	(3)	(4)
			Wave 1	
Outcome	Baseline	Demographics	Integration	Preferences
Radicalism in Attitudes (SyfoR)	0.02***	0.02***	0.02***	0.02***
	(0.00)	(0.00)	(0.00)	(0.00)
Radicalization Intentions (RIS)	0.02***	0.02***	0.02***	0.02***
	(0.00)	(0.00)	(0.00)	(0.00)
Activist Intentions (AIS)	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Activist-Radicalization Intentions (ARIS)	0.01***	0.01***	0.01***	0.01***
	(0.00)	(0.00)	(0.00)	(0.00)
Number of Observations		1,985		
			Wave 2	
Radicalism in Attitudes (SyfoR)	0.01***	0.01***	0.01***	0.01***
	(0.00)	(0.00)	(0.00)	(0.00)
Radicalization Intentions (RIS)	0.01***	0.01***	0.01***	0.01***
	(0.00)	(0.00)	(0.00)	(0.00)
Activist Intentions (AIS)	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Activist-Radicalization Intentions (ARIS)	0.01***	0.01***	0.01***	0.01***
	(0.00)	(0.00)	(0.00)	(0.00)
Number of Observations		1,552		

Notes: This table shows standardized effects of the score on our four primary outcome variables for treated individuals. Score values empirically range between 4 and 106, with mean 59.50 (std. dev. 15.18). The effects are estimated coefficients from linear regressions (*Ordinary Least Squares*, OLS). For the regression models, we specified the standardized outcome variables as functions of the game score and different sets of covariates. The columns (1)-(4) show the coefficients from different model specifications. For the baseline specification, we controlled for for game score only. For the second specification, we additionally controlled for demographic characteristics such as age group, gender, educational attainment (5 categories), and occupational status (5 categories). For the third specification, we added indicators of social integration, including relationship status, parenthood, urban residence, and immigrant background. In the final specification, we additionally controlled for locus of control and self-esteem. Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

JOURNAL FOR DERADICALIZATION

Spring 2022
No. 30
ISSN: 2363-9849

About the JD Journal for Deradicalization

The JD Journal for Deradicalization is the world's only peer reviewed periodical for the theory and practice of deradicalization with a wide international audience. Named an "essential journal of our times" (Cheryl LaGuardia, Harvard University) the JD's editorial board of expert advisors includes some of the most renowned scholars in the field of deradicalization studies, such as Prof. Dr. John G. Horgan (Georgia State University); Prof. Dr. Tore Bjørgo (Norwegian Police University College); Prof. Dr. Mark Dechesne (Leiden University); Prof. Dr. Cynthia Miller-Idriss (American University Washington D.C.); Prof. Dr. Julie Chernov Hwang (Goucher College); Prof. Dr. Marco Lombardi, (Università Cattolica del Sacro Cuore Milano); Dr. Paul Jackson (University of Northampton); Professor Michael Freeden, (University of Nottingham); Professor Hamed El-Sa'id (Manchester Metropolitan University); Prof. Sadeq Rahimi (University of Saskatchewan, Harvard Medical School), Dr. Omar Ashour (University of Exeter), Prof. Neil Ferguson (Liverpool Hope University), Prof. Sarah Marsden (Lancaster University), Prof. Maura Conway (Dublin City University), Dr. Kurt Braddock (American University Washington D.C.), Dr. Michael J. Williams (The Science of P/CVE), Dr. Mary Beth Altier (New York University) and Dr. Aaron Y. Zelin (Washington Institute for Near East Policy), Prof. Dr. Adrian Cherney (University of Queensland).

For more information please see: www.journal-derad.com

Twitter: @JD_JournalDerad
Facebook: www.facebook.com/deradicalisation

The JD Journal for Deradicalization is a proud member of the Directory of Open Access Journals (DOAJ).

ISSN: 2363-9849