

EVALUATING DIGITAL POLICY INTERVENTIONS:
STUDIES IN VIOLENCE PREVENTION, DERADICALIZATION,
AND INVESTOR PROTECTION

Von der Wirtschaftswissenschaftlichen Fakultät der
Gottfried Wilhelm Leibniz Universität Hannover
zur Erlangung des akademischen Grades

Doktor der Wirtschaftswissenschaften
– Doctor rerum politicarum (Dr. rer. pol.) –
genehmigte Dissertation

von

M.Sc. Axel Ebers,
geboren am 27. Februar 1983 in Hannover

2022

Referent: Prof. Dr. Stephan L. Thomsen

Korreferentin: Prof. Dr. Marina Schröder

Tag der Promotion: 30.06.2022

Abstract

Abstract Social media platforms allow individuals without great technical knowledge or financial capital to spread violent content and radical messages among a large audience, bypassing traditional media. As violence and radical actions cause significant economic damage and undermine democratic institutions, security agencies have responded by using the same platforms to target at-risk individuals with their prevention programs. The literature has shown that such online programs can be effective in reducing victimization and perpetration rates. However, despite the large number of programs implemented, there is a significant lack of rigorous impact evaluations. This thesis contributes to the literature by developing a conceptual framework for evaluating the impact and economic efficiency of online programs for crime prevention in general. Furthermore, it contributes by evaluating the impact of two online programs in the areas of violence prevention and deradicalization. To reassess the impact of such online programs in a different prevention context, the thesis also evaluates a measure for investor protection in the context of the Bitcoin cryptocurrency.

Keywords: Social Media; Online Interventions; Prevention; Impact Evaluation

JEL Classification: C93; D83; G40

Kurzzusammenfassung Die Social Media Plattformen erlauben es Individuen ohne hohes technisches Wissen oder Finanzkapital, unter Umgehung der klassischen Medien gezielt gewalttätige Inhalte oder radikale Botschaften unter einem großen Publikum zu verbreiten. Da Gewaltkriminalität und radikale Aktionen aber erhebliche wirtschaftliche Schäden verursachen und die demokratischen Institutionen aushöhlen, haben die Sicherheitsbehörden dadurch reagiert, dass sie dieselben Plattformen nutzen, um gefährdete Personen gezielt mit ihren Präventionsprogrammen anzusprechen. Die Literatur hat gezeigt, dass derartige Online-Programme effektiv Viktimisierungs- und Täterschaftsraten senken können. Allerdings herrscht trotz der Vielzahl umgesetzter Programme ein augenfälliger Mangel an Wirkungsevaluationen, die höchsten wissenschaftlichen Standards genügen. Um einen Beitrag zur Literatur zu leisten, entwickelt diese Dissertation einen konzeptionellen Rahmen zur Evaluierung der Wirksamkeit und ökonomischen Effizienz von Online-Programmen zur Kriminalprävention. Des Weiteren leistet sie einen Beitrag, indem sie die Wirksamkeit zweier Online-Programme aus den Bereichen Gewaltprävention und Deradikalisierung evaluiert. Um die Wirksamkeit solcher Programme in einem anderen Präventionskontext zu überprüfen, evaluiert sie außerdem eine Maßnahme zum Investorenschutz im Zusammenhang mit der Kryptowährung Bitcoin.

Schlagwörter: Soziale Medien; Online-Interventionen; Prävention; Wirkungsevaluation

Acknowledgements

Completing a doctoral thesis is both an intellectual and personal challenge. During the writing process, I have benefited from a number of persons. In the following lines, some of them are gratefully acknowledged. It is without a doubt that there are more and these words cannot express the respect and gratitude I feel for every single one of them.

First and foremost, I would like to express my most sincere gratitude and appreciation to my supervisor Stephan L. Thomsen for his guidance, advice and encouragement. The time working at his institute opened up invaluable opportunities and experiences for me. My research papers greatly benefited from his co-authorship and astute comments during feedback cycles. He has motivated and enabled me to participate in numerous conferences and workshops. He also created a very fair and trusting work environment, which enabled me to reconcile work and family life in a good way. His professionalism, understanding, and steadfast optimism greatly inspired my professional and personal development.

I am grateful to have been part of the interdisciplinary research network “Präventive digitale Sicherheitskommunikation (PräDiSiKo) - ein innovativer Ansatz für Kriminalprävention in sozialen Online-Medien [Security Communication via Online Social Networks – An Innovative Approach to Crime Prevention]” Sincere thanks go to the project partners for the fruitful discussions and invaluable inputs from their respective disciplines. I gratefully acknowledge the project funding by the Federal Ministry of Education and Research (BMBF).

Many thanks go to my colleagues at the Leibniz University Hannover. The team has always provided a rich source of advice and proved a recipe for good collaboration. I benefited from and enjoyed the discussions in countless seminars and coffee breaks.

Without my friends, life would have been harder during my doctoral studies. I am deeply grateful for inspiring discussions, welcome distraction and encouragement during challenging periods.

Finally, I would like to express my most sincere gratitude to thank my family for all their love, understanding and support. I am infinitely grateful to my parents who always encouraged me to aim high and live up to my potential. Words cannot express the appreciation I feel for my wife who blessed me with two wonderful children and always had my back.

Table of Contents

Abstract	III
Acknowledgements.....	V
Table of Contents	VI
List of Abbreviations.....	VII
1 Introduction	1
2 Benefit-Cost Analysis of Social Media Facilitated Bystander Programs	6
3 Evaluating a Gamified Bystander Program	7
Abstract	8
3.1 Introduction	9
3.2 Theoretical Considerations	11
3.2.1 The Psychology of Bystander Behavior	11
3.2.2 The Psychology of Behavior Change	11
3.2.3 The Gamification Approach	13
3.3 Program Intervention and Research Hypotheses	14
3.3.1 The Bystander Program	14
3.3.2 Research Hypotheses	16
3.4 The Facebook Experiment	18
3.4.1 Data Collection and Research Design	18
3.4.2 Summary Statistics	20
3.4.3 Checking for Balance	20
3.4.4 Main Results	21
3.4.5 Impact Vectors	23
3.4.6 Heterogeneity Analysis	24
3.4.7 Discussion	25
3.5 The Panel Data Experiment	25
3.5.1 Data Collection and Research Design	25
3.5.2 Summary Statistics	27
3.5.3 Checking for Balance	28
3.5.4 Main Results	28
3.5.5 Impact Vectors	31
3.5.6 Heterogeneity Analysis	31

3.6	Conclusion	33
4	Evaluating an Interactive Film on the Prevention of Political Radicalization	35
5	How do Warnings Affect Retail Demand for Bitcoin?	36
	Bibliography	37
	Appendix B – Appendix for Chapter 3.....	43
B.1	Facebook Experiment Appendix	43
B.2	Panel Data Experiment Appendix	51
B.3	Illustration of the Policy Intervention.....	60

List of Abbreviations

AEA	American Economic Association
AIC	Akaike information criterion
AIS	Activist-Intentions-Scale
API	Application programming interface
ARIS	Activist-Radicalism-Intentions-Scale
ATT	Average Treatment Effects on the Treated
BCR	Benefit-Cost-Ratio
BKA	Bundeskriminalamt
BMBF	Bundesministerium für Bildung und Forschung
CBDC	Central bank digital currency
CFA	Confirmatory factor analysis
CO ₂	Carbon dioxide
CPM	Cost per mille
CSR	Corporate social responsibility
DCE	Discrete choice experiment
DCM	Discrete choice model
e.g.	(lat. <i>exempli gratia</i>) for example
ECB	European Central Ba
EFA	Exploratory factor analysis
ESG	Environmental, Social, and Corporate Governance
GDPR	General Data Protection Regulation
HdM	Hochschule der Medien
i.e.	(lat. <i>id est</i>) in other words

ICS	Intervention capability score
LOC	Locus of control
LR	Likelihood-ratio
MRS	Marginal rate of substitution
MueZI	Muenchener Zivilcourage Instrument
OLS	Ordinary least squares
RAA	Reasoned action approach
RCT	Randomized controlled trial
RIS	Radicalism-Intentions-Scale
RTB	Real-time bidding
SyfoR	Sympathy for Violent Radicalization and Terrorism
U.K.	United Kingdom
WTIS	Willingness to intervene score

1 Introduction

The emergence of social media has enabled any individual who has access to the internet to spread violent or radical content among the broad public (Thompson, 2011). Previously, the traditional media had acted as gatekeepers and prevented the inappropriate depiction of violence, in accordance with the journalistic principles (Deutscher Presserat [German Press Council], 2005). The described change in the media landscape has confronted the security authorities with two major challenges. As the first major challenge, the media change has created a virtual environment where users are easily exposed to violent content due to a lack of regulation (King et al., 2007; Patton et al., 2014). Frequent exposure to violent content, however, increases the likelihood that users will engage in violent behavior themselves. Psychological processes such as priming, arousal, and learning cause the strong relation between the frequent perception and exercise of violence (Anderson et al., 2003; Huesmann et al., 2003).¹ Empirical evidence suggests that this relation also holds in the context of social media (Elsaesser et al., 2021; Gallacher et al., 2021). For example, individuals who exercise aggression in social media are also more likely to believe that violence towards peers is socially acceptable (Hinduja & Patchin, 2013). Due to anonymity, communication on social media also reaches a higher level of hostility, which fosters the escalation of violent conflicts (McKenna & Bargh, 2000).

As the second major challenge for the security authorities, radical groups use social media platforms to target those users who are most receptive to their messages (Fink, 2018; Malmasi & Zampieri, 2017; Mathew et al., 2019). In this way, they can exert more influence, gain sympathizers and supporters, or recruit new members (Chatfield et al., 2015; Gates & Podder, 2015; Thompson, 2011). In this context, the *lone wolf theory* plays an important role. The theory postulates that individuals can become radicalized in social media without actually

¹ Exposure to violent content immediately increases the likelihood of violent behavior by priming aggressive cognitions and behavioral scripts, raising physiological arousal, and triggering the automatic tendency to imitate observed behaviors. In the long term, it causes the acquisition of lasting aggressive behavioral scripts, schemas, and violence-promoting beliefs about social behavior (Anderson et al., 2003; Huesmann et al., 2003).

joining a particular radical group. Some of these individuals carry out terrorist attacks as lone perpetrators. (Weimann, 2012). The far-right terrorist attack in Christchurch, New Zealand introduced a particularly perverse form of media use. In this case, the perpetrator streamed self-filmed video footage of his attack in real time via social media (Rauf, 2021). Six months later, a similar assassination took place in Halle, Germany. The perpetrator was inspired by the events in Christchurch and streamed his attack to incite further imitators in turn (Kessling et al., 2020).

The specific mechanics of social media platforms can reinforce the tendency towards radicalization among users. These mechanics expose each individual user to the content most likely to maximize engagement (i.e., viewing, sharing, and commenting), according to his or her profile. As the content mix is designed to match the user's profile, it tends to support his or her worldview. Content that contradicts the worldview tends to be sorted out. This process encloses the user in a virtual *filter bubble*, where his or her beliefs and attitudes are constantly confirmed and reinforced (Cinelli et al., 2021; Garimella et al., 2018). As confirmation and reinforcement take place on both sides of the political spectrum, they contribute to the polarization of political discourse (Bail et al., 2018; Banks et al., 2021). In extreme cases, they also contribute to the justification, support or execution of illegal or violent political actions (Huey, 2015; Thompson, 2011).

On the other hand, social media hold great potential for crime prevention and deradicalization. While traditional prevention programs take place in small groups and cause high staff expenditures, social media allow the dissemination of prevention programs at marginal costs close to zero, i.e., *upscaling* (Castronovo & Huang, 2012). Due to the ubiquity of mobile devices, users can participate in these online programs at any time and place, which increases reach (Silver et al., 2019). Security authorities can use social media platforms to address target groups who are especially vulnerable to violence or radicalization, i.e., *microtargeting*. Due to reduced scattering losses, microtargeting increases economic efficiency (Winter et al., 2021). Online prevention programs allow customization based on the needs and preferences of vulnerable groups (Lustria et al., 2009). Social media enable dialogue-oriented and interactive communication (Tsimonis & Dimitriadis, 2014). Together, these factors can contribute to increased effectiveness and efficiency of prevention work.

Violent crime and radical actions cause significant damage to individuals and society as a whole. Damages arise from pain and human suffering, as well as the cost of medical treatment and lost productivity (Cohen & Bowles, 2010). Frequent exposure to violent content has detrimental effects on psychosocial development, especially in children, adolescents and young

adults (Marcum et al., 2010). Radical political actions undermine democratic institutions. Terrorist attacks may destruct physical capital. Finally, the fear of violent crime or radical actions leads to changed consumption and investment patterns, where private and public resources are diverted from productive uses to protective measures (Bardwell & Iqbal, 2021).

Given the social damage from violence and radicalism on the one hand, and the potentials of social media on the other, the security authorities have expanded their use of these platforms. The literature has shown that online programs can be effective. For example, such programs were able to promote bystander behavior in violent situations, which reduced perpetration and victimization rates (e.g., Jouriles et al., 2020; Salazar et al., 2019). However, in order to keep up with the permanent evolution of social media technology and implement effective programs in the future, the security agencies must learn from successful programs implemented in the past. Yet, despite a large number of programs implemented, there is a lack of rigorous evaluation studies. While the *Federal Criminal Police Office* (BKA) counted over 2,000 deradicalization projects implemented in Germany alone (Gruber et al., 2017), a recent meta-analysis found only nine studies that complied with the scientific state of the art (Jugl et al., 2020).

Addressing this research gap, this thesis develops a conceptual framework for evaluating the impact and economic efficiency of online interventions for crime prevention (Chapter 2). Based on this framework, the following chapter (Chapter 3) presents evaluation results from an online intervention for violence prevention that applied game principles and game design elements, i.e., *gamification*. The chapter after next (Chapter 4) presents results from an online intervention for deradicalization, which also drew on the gamification approach. The final chapter (Chapter 5) evaluates an online intervention in a different prevention context, i.e., investor protection against *Bitcoin* risks. Together, these studies contribute to filling the research gap mentioned above. The first study extends the traditional benefit-cost model with an application for crime prevention via social media. The second and third studies evaluate pilot programs of crime prevention and deradicalization in the form of interactive films with game elements. The studies expand the spectrum of research methods used in the field of crime prevention in Germany by the economic toolkit. The study results, as well as the evaluated programs, set a benchmark for German prevention work. They contribute to evidence-based crime prevention, which allows the efficient allocation of scarce public resources. The following paragraphs provide a brief summary of the subsequent chapters.

The next chapter (Chapter 2) develops a conceptual framework for benefit-cost analysis of social media facilitated bystander programs, which forms the basis of the empirical studies. Drawing on the potential outcomes of a violent situation that involves at least one bystander, the chapter develops an extended benefit-cost model. The model treats publicly funded programs as investment projects and calculates the benefit-cost ratio as an indicator for assessing funding decisions. Within this framework, social benefit arises from the damages avoided by preventing violent crime. The chapter provides systematic instructions for estimating the benefit side of the model. This includes using social media analytics to estimate program reach, applying machine-learning technology for targeting individuals at risk, and conducting randomized field experiments for evaluating program impact. It also includes using discrete choice experiments for estimating the monetary value of damage per violent crime. Finally, the chapter introduces an alternative approach that draws on the *bid landscaping* methodology² to estimate the monetary value of public attention generated by a prevention program.

Chapter 3 describes the first of the three empirical studies included in this thesis. The study evaluates an online intervention aiming to promote bystander behavior in cases of violence in the public space. The intervention consisted of two components. First, an interactive film simulated a potentially violent situation in the virtual space. Second, a series of online games complemented the content of the film. For impact evaluation, two randomized field experiments were conducted. In the first experiment, a random sample of German *Facebook* users was drawn and randomly assigned to four treatment arms, including three different configurations of the treatment and a control group. In the second experiment, a representative random sample of the German working population was drawn and observed over three survey waves. The results of the first experiment show that the film motivated *Facebook* users to intervene in violent situations and that the game elements reinforced this effect. The results of the second experiment show that the effect also applied to the wider population and persisted over a longer period. They further show that the film not only increased motivation, but improved self-perceived intervention skills.

Chapter 4 describes the second empirical study included in this thesis, which evaluates an online intervention for deradicalization. The intervention consisted of an interactive film that portrayed the radicalization process of a teenage student. The film asked viewers to position themselves in relation to the increasingly radical statements of the student. For impact

² Bid landscaping is a method to determine the monetary value of online advertisements (AdWords API, 2016).

evaluation, a randomized field experiment was conducted. For this purpose, a representative sample of the German working population was drawn, randomly assigned to one treatment and one control group, and observed over two survey waves. The results of this experiment show that the interactive film reduced both the radicalism of political attitudes and the radicalization intentions among participants. After two weeks, these effects were still significant, but less pronounced. The effects among the 18-24 age group, women, and people on the left of the political spectrum were stronger and more persistent. Among people on the right of the spectrum, on the other hand, the film had no effect at all.

The final chapter (Chapter 5) describes the third study included in this thesis. The study was conducted to reassess the impact of online interventions in a different prevention context, i.e., investor protection against risks emanating from the *Bitcoin* cryptocurrency. The study was motivated by the fact, that the emergence of social media had significant implications for financial markets, especially in the context of cryptocurrencies. While social media platforms can be used for market manipulation³ and fraud⁴, they hold great potential for investor protection. For example, regulators can use these platforms to warn investors who are most vulnerable to fraud and financial risk.⁵ To the best of our knowledge, there has been no empirical evidence on the impact of such regulatory measures on investor behavior. Therefore, the final study presented in this thesis used information experiments to investigate how warnings affect demand for *Bitcoin*. For this purpose, a representative sample was drawn from each of the four largest European economies. Participants were randomly assigned to five treatment arms, namely four different information treatments and a control group. The results of the study show that information about Bitcoin's privacy issues had the strongest negative effect on demand. Information on the lack of deposit insurance or carbon dioxide (CO₂) emissions due to Bitcoin mining only had an effect on certain subgroups. Positive information about the broad public acceptance increased demand.

³ For example, when *Tesla* founder Elon Musk added the hashtag *#bitcoin* to his Twitter bio, the price of the largest cryptocurrency increased by 17% (Nadeem, 2021).

⁴ A prominent case revolves around "Crypto Queen" Dr. Ruja Ignatova and her alleged cryptocurrency One Coin. Ignatova hosted a promotional event at London's Wembley Arena in 2016, where spectators could buy educational brochures and supposedly exchange them for One Coin later. Ignatova reportedly raised a total of four to five billion dollars. On October 25, 2017, she disappeared without a trace.

⁵ For example, the *European Central Bank* (ECB) used twitter to warn about the specific risks of cryptocurrencies (European Central Bank [@ecb], 2021), or to promote *digital central bank currencies* (CBDC) (CBDC; European Central Bank [@ecb], 2020).

2 Benefit-Cost Analysis of Social Media Facilitated Bystander Programs*

with:
Stephan L. Thomsen

This paper has been published as:

Ebers, A., & Thomsen, S. L. (2021). Benefit–Cost Analysis of Social Media Facilitated Bystander Programs. *Journal of Benefit-Cost Analysis*, 12(2), 367–393.

DOI: <https://doi.org/10.1017/bca.2020.34>

* We like to thank the editor, Amanda Ross, and two anonymous referees for their valuable comments and suggestions. We also thank our project partners at the Hochschule der Medien (HdM), Police Crime Prevention of the States and the Federal Government (ProPK), German Police University (DHPol), MOSAIQ, the Federal Criminal Police Office (BKA), and the State Office of Criminal Investigations (LKA) Lower Saxony for their helpful comments. Financial support by the German Federal Ministry of Education and Research (BMBF) via the project “Security Communication via Online Social Networks – An Innovative Approach to Crime Prevention” (PräDiSiKo) is gratefully acknowledged.

3 Evaluating a Gamified Bystander Program

Evidence from Two Randomized Online Field Experiments*

with:

Stephan L. Thomsen

* We thank our project partners at the University of Media (HdM), Police Crime Prevention of the States and the Federal Government (ProPK), German Police University (DHPol), MOSAIQ, the Federal Criminal Police Office (BKA), and the State Office of Criminal Investigations (LKA) Lower Saxony for their input. We further acknowledge helpful comments from participants at the 2018 Seminar on Crime Prevention at the German Police University (Münster), the 23rd German Prevention Congress (DPT) in Dresden, the 2018 Workshop on Police Crime Prevention in the Digital Age at the German Police University (Münster), the 2019 Momentum Congress (Hallstatt), the 2020 Workshop on Microeconomics at the Leuphana University of Lüneburg, the 25th German Prevention Congress (DPT), and the 2021 Virtual Congress of the International Association for Research in Economic Psychology (IAREP). Financial support by the German Federal Ministry of Education and Research (BMBF) via the project “Security Communication via Online Social Networks – An Innovative Approach to Crime Prevention” (PräDiSiKo) is gratefully acknowledged. The experiment is registered at the American Economic Association (AEA) registry: “Gaming for Good: Evaluating an Interactive Online Bystander Intervention” (AEARCTR-0005779).

Abstract

Objective: Previous social-psychological research has demonstrated the positive effects of online bystander programs on various crime-related outcomes, while information systems research has demonstrated the ability of gamification to improve motivation, engagement, and learning. This study bridges the gap between social psychology and information systems research by evaluating a bystander program that combines the simulation of a dangerous situation in a virtual environment with the application of game principles and game design elements. **Method:** We developed three research hypotheses and tested them using two randomized online field experiments. During the first experiment, we collected data from 4,188 users on Facebook and randomly assigned them to four treatment arms, including three different configurations of the treatment and one control group. During the second experiment, we collected data from a representative sample of the population and observed them across three waves. **Results:** The results from the first experiment support the hypotheses that the bystander program motivates people to intervene in violent situations and that gamification enhances the motivational effect. The results from the second experiment support the hypothesis that the program makes people feel more capable of intervening. They also show that the treatment effects persist over a long period of time and hold for the overall population. **Conclusions:** We conclude that the gamification approach offers great potential for bystander education and that social media are well suited for the dissemination and upscaling of bystander programs. Policymakers can use these findings to improve the effectiveness and efficiency of future bystander programs or similar prevention measures.

Keywords: Bystander Intervention; Gamification; Program Evaluation; Field Experiments; Social Media; Facebook

JEL Classification: C93; D91; K42

3.1 Introduction

Violent crime causes tangible damage, such as the cost of medical treatment, law enforcement, or lost productivity, as well as intangible damage, such as pain, suffering, or lost quality of life (Cohen & Bowles, 2010). As incarceration appears less efficient for fighting crime, immense social damage leads to a need for effective prevention strategies (Welsh et al., 2015). Bystander programs aim to contribute to prevention by motivating people to intervene when they observe warning signs or incidents of violence and teaching the skills necessary for safe and effective intervention (Banyard et al., 2007). Evidence suggests the strategy is successful. Various bystander programs manage to improve crime-related outcomes, including violent victimization, perpetration, acceptance of violence, or bystander behavior (e.g., Gidycz et al., 2015; Miller et al., 2014; Potter et al., 2008; Shaw & Janulis, 2016). Traditional bystander programs consist of face-to-face training in small groups (e.g., Banyard et al., 2007) and therefore require many staff members to train many people. Since training takes place at a fixed time and location, some prospects may not be willing or able to participate. Despite their successes in crime prevention, traditional programs are thus difficult to scale up.

The described disadvantages sparked the development of bystander programs that provide training via the internet. So-called online bystander programs allow for training many people without many staff members (Cugelman et al., 2011). Indeed, the reproduction of digital content at a marginal cost close to zero enables the rapid upscaling of such programs. Prospects can participate in the training on their computer at their preferred time and location (White et al., 2010). Digital technologies, therefore, allow expanding program reach without significantly increasing costs. Recent studies found that online bystander programs can have the same positive effects on real-life behaviors as traditional programs, with *TakeCare* (Jouriles et al., 2020) and *RealConsent* (Salazar et al., 2019) being the most discussed examples. Similar programs successfully targeted deviant online behaviors such as *cyberbullying* or social media harassment (Wang, 2020; Wong et al., 2021).

As training potential bystanders is most effective under realistic conditions (Baumert et al., 2013), the simulation of dangerous situations in virtual environments promises great potential for bystander education (Röderer et al., 2019). In this regard, the *gamification approach* may offer another lever to improve effectiveness. *Gamification* describes the use of game principles and game-design elements in nongame contexts (Deterding et al., 2011). Games evoke more interest, engagement, and motivation than traditional educational materials,

which lack implicit rules, objectives, and pursuits (Deater-Deckard et al., 2013). Research shows that *gamification experiences* could influence psychological and behavioral outcomes from different areas, including education, health, and prevention (Hamari et al., 2014). For example, *serious games* could not only raise awareness of *cyberbullying* but also induce effective coping behaviors (Calvo-Morata et al., 2020; DeSmet et al., 2018). In this context, social media offer the ideal platform to reach large and relevant target groups with a gamified bystander intervention (Ebers & Thomsen, 2021).

To evaluate the impact of an online bystander program that combines the simulation of a dangerous situation with game principles and game-design elements, we develop three research hypotheses and test them using two randomized field experiments (i.e. *randomized controlled trials*, RCTs). The bystander program consists of an interactive film, which simulates a violent situation, and a series of online games. During the film, the user has to make choices that determine how the storyline developed. The subsequent online games test knowledge about the film to train the desired behaviors in such a situation. Based on the relevant theory, we hypothesize that, first, the program motivates people to intervene in violent situations; second, the use of game principles and game-design elements enhances this effect; and third, participants feel more capable of intervening due to program treatment.

We conducted the first experiment on *Facebook* to consider the environment where people usually first encounter new digital content. The results confirmed that the program motivates people to intervene, and that the application of the *gamification approach* enhances this effect. To test whether the motivational effect is long-term and holds for the overall population, we conducted a second field experiment with panel data. The results confirmed the motivational effect shown by the *Facebook* experiment. They further confirmed that the program makes people feel more capable of intervening and that both effects are long-lasting and hold for the overall population. The program achieves these effects by reducing all the psychological barriers to intervention and changing the beliefs regarding intervention behavior.

The paper proceeds as follows. Section 2 lays out the theoretical foundations for our empirical work. They include the psychological processes underlying bystander behavior, our behavior change model, and the mechanisms of the gamification approach. Section 3 describes the bystander intervention in detail and states the research hypotheses. The research design, data collection, and key findings of the two experiments are presented in sections 4 and 5. The final section provides our conclusions.

3.2 Theoretical Considerations

3.2.1 The Psychology of Bystander Behavior

Bystander behavior includes responses to the observation of physical violence, including ignorance and intervention. The various types of intervention behavior can be categorized into four Ds: direct, distract, delegate, and delay (Banyard et al., 2005; Berkowitz, 2002). *Direct* tactics involve direct intervention aiming to prevent or stop the violence. *Distraction* tactics distract the attention of the offender to rescue the victim. *Delegation* tactics involve at least one other person and a plan for cooperation. For example, one bystander could distract the perpetrator, while the other bystander called the police. *Delay* tactics apply after the violent situation has taken place. The bystander may give first aid or consolation. The appropriate tactic depends on the characteristics of the situation at hand. More specifically, bystanders should save direct tactics for dangerous emergencies in which no other options remain.

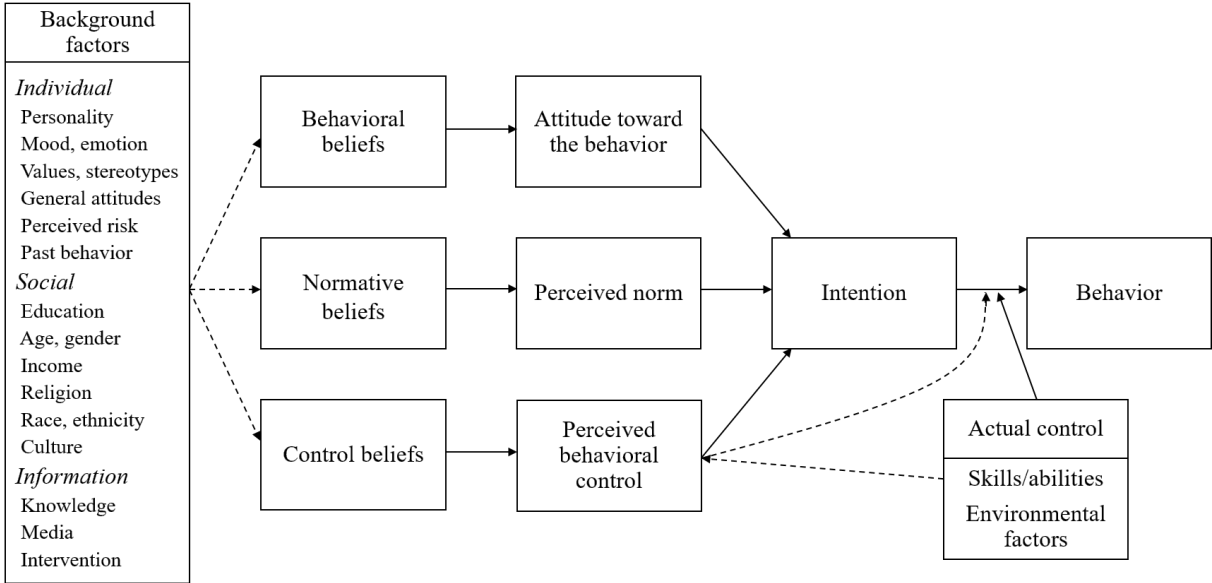
When observing warning signs or incidents of physical violence, bystanders have to overcome a series of psychological barriers before they intervene (Latané & Darley, 1970). They have to notice the event in the first place (detection). Assuming they do, they have to understand that the event marks a case of violence (interpretation), feel responsible for helping (assumption of responsibility), and know they have the skills necessary for intervention (perceived control). Finally, they have to think that the benefits of helping clearly outweigh the costs (cost-benefit analysis). Only if they overcome all the psychological barriers will they take action (see Figure 2.1). Different factors (e.g., empathy, the acceptance of negative social consequences, anticipated guilt, indignation, and audience inhibition; Halmburger et al., 2017) determine whether bystanders manage to overcome a particular barrier. Interdependencies and feedback loops connect the different barriers. A core objective of any bystander program is teaching participants how to overcome the psychological barriers to motivate intervention.

3.2.2 The Psychology of Behavior Change

Following the *reasoned action approach* (RAA), we can motivate intervention behavior 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 attitude, perceived social pressure, and perceived control over the behavior. Taken

together, these three factors determine behavioral intention, which captures the individual level of motivation to perform a behavior. Within the model framework, intentions are the best predictor of the actual performance of the behavior (Fishbein & Ajzen, 1977). For example, if a person has a strong intention to intervene in a violent situation, she probably will – at least if no personal or environmental factors prevent her from doing so (see Figure 3.1).

Figure 3.1. The Reasoned Action Approach



Notes: This figure shows a schematic representation of the Reasoned Action Approach. Source: Own representation based on Fishbein & Ajzen (2011).

The *reasoned action approach* distinguishes between behavioral, normative, and control beliefs. Behavioral beliefs represent the individual level of information regarding the behavior’s positive and negative outcomes. They determine the attitude toward that behavior. Normative beliefs represent the level of information about injunctive and descriptive norms. Injunctive norms refer to the approval or disapproval of a behavior by the relevant reference group. Descriptive norms refer to the behavior of the reference group itself. Together, injunctive and descriptive norms determine perceived social pressure. Control beliefs refer to personal or environmental factors that promote or impede behavior. They determine perceived control. 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. In conclusion, any bystander program would have to change the underlying beliefs of trainees to motivate intervention behavior.

3.2.3 The Gamification Approach

Applying the *gamification approach* could further enhance the motivational effect of a bystander program. Gamification contributes to behavior change by leveraging two key motivational drivers of human behavior: *reinforcement* and *emotion* (Robson et al., 2015). Game designers use rewards or punishments to reinforce particular behavior (Sailer et al., 2014). In this process, called *operant conditioning*, the reinforcements induce behavior change by evoking affective responses or emotions. Positive reinforcements likely lead to repetition, and negative reinforcements likely lead to avoidance of a behavior (Skinner, 2019). Thus, if players should repeat or sustain a particular behavior, the intervention should use reinforcements that lead to rewarding outcomes (Rothschild & Gaidis, 1981). Game designers can leverage the motivational drivers of reinforcement and emotion to turn desired behaviors into habits or automatic behavioral processes (Duhigg, 2012). More specifically, they can manipulate the formation of habits by repeatedly setting cues that elicit a behavior, then rewarding execution. Through the repeated *behavioral loop* consisting of cues, behavior, and rewards, execution requires fewer and fewer cognitive resources.

To create a *behavioral loop* and reinforce a desired behavior, game designers can shape three basic characteristics of a game: mechanics, dynamics, and emotions (Robson et al., 2015). *Mechanics* include the setup, rules, and progression of the game. Setup mechanics determine the setting and necessary objects (Elverdam & Aarseth, 2007). Rule mechanics determine the goal of the game as well as permitted actions. Progression mechanics are especially important for gamification, as they determine reinforcement. For example, players could earn points for desired behaviors as they progress through the game. These *achievement rewards* are especially effective when they indicate social standing within a peer group community. *Dynamics* describe how players utilize the mechanics of the game (Camerer, 2011). They strongly depend on the players' structure and the presence of observers. A multiplayer structure promotes cooperation, while a single-player structure, or the presence of observers, promotes competition.¹ *Emotions* result from mechanics and player dynamics. Creating positive emotions is the most important goal for player engagement (Sweetser & Wyeth, 2005). Because of the multiple interactions between mechanics, dynamics, and emotions, game designers must carefully fine-tune these factors to achieve the desired behavioral change (Robson et al., 2015).

¹ In this context, the prevailing competitive structures have particular effects on engagement and learning (Santhanam et al., 2016).

3.3 Program Intervention and Research Hypotheses

3.3.1 The Bystander Program

We evaluate a unique bystander program that combines the simulation of a dangerous situation in a virtual environment using game principles and game design elements. The program uses heroism as a guiding principle, which is reflected in the program title, logo, and individual messages transported.² It targets young people who have received a higher education and show a strong affinity for technology. The program's main objective is motivating people to intervene in violent situations and teaching skills for safe and effective intervention. To achieve these objectives, the program employs an interactive film and six online games located on a proprietary website³. The interactive film puts the player in the position of a bystander to a potentially violent situation. During the film, the player has to make choices that determine the progression of the storyline. The subsequent online games test the player's knowledge about the film to practice the desired intervention behaviors.

The interactive film strongly relies on game principles and game design elements. Its *setup mechanics* closely reflect the real-life conditions of a violent situation. The scene takes place in an underground car park, where a group of teenagers attacks a young couple. The film has a single-player structure and occurs in real-time. The *rule mechanics* provide three *choice points* where the player has five seconds to choose between two courses of action (Figure 3.2). If the player makes the desirable choice, the storyline takes a positive turn, while the opposite happens if she makes the undesirable choice or misses the five-second deadline. For example, choosing to get help from other bystanders prevents the situation from escalating. Choosing to stay passive leads to escalation and fatal injury of the victim. The 5-second deadline puts additional pressure on the player. Depending on the choices made, the *progression mechanics* credit the player points in real-time (see Appendix B for further screenshots of the bystander program).

² The program title is “*Zivile Helden*” (for civilian heroes). The logo is designed reminiscent of superheroes from comic books. For example, the individual messages emphasize that one does not have to put oneself in danger to be a hero.

³ The web address is www.zivile-helden.de.

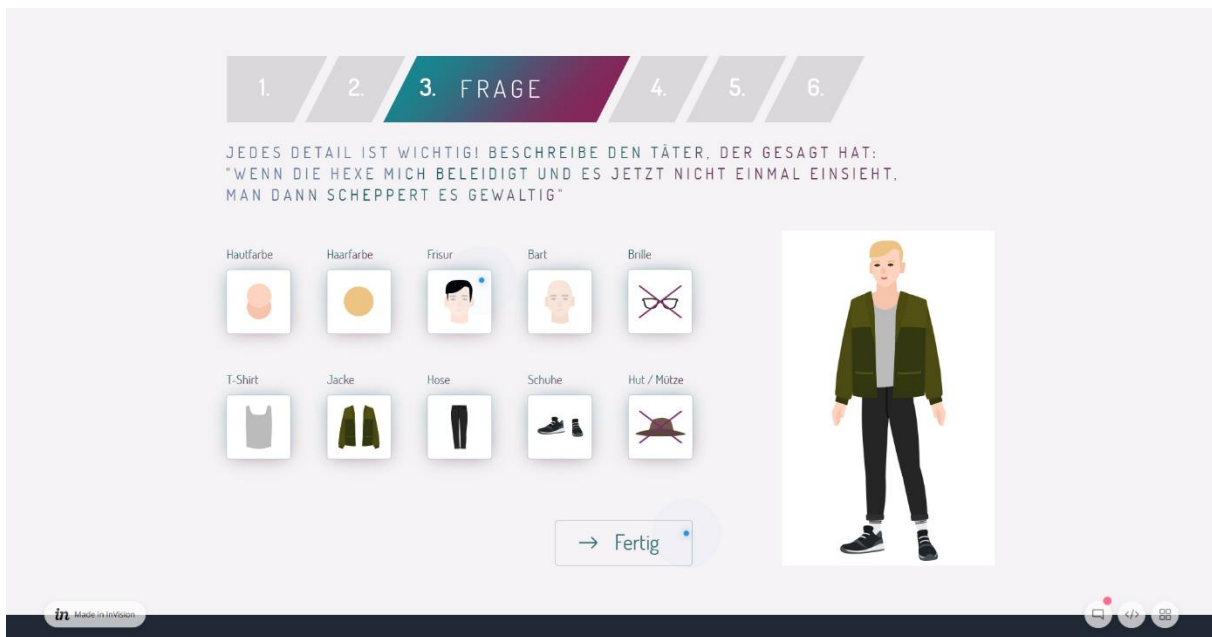
Figure 3.2. Exemplary Decision Situation within the Interactive Film



Notes: The figure shows an exemplary choice point from the interactive film. The header says, “What would you do now?” The first choice is, “I call the police.” The second choice is, “I wait and see what happens.” The user has five seconds to make a choice. If she misses the deadline, the system continues with the default, which is the undesirable choice. At the decision point, the film pauses, the music quiets, and the image in the background darkens. *Source:* <https://www.zivile-helden.de/>.

The *setup mechanics* of the online games allow the player to practice the desired intervention behaviors in a fun way without time pressure. The user interface animates the player by relying mainly on graphical elements. The *rule mechanics* make the player solve tasks testing her level of knowledge about the film. For example, she must design an avatar to make its appearance reflect the perpetrator's characteristics from the film (Figure 3.3). In another game, she must complete a puzzle to reconstruct the progression of events. After each game, the system displays the sample solution and short feedback. The *progression mechanics* of the online games reward the player with points for solving tasks. Depending on the total score from the interactive film and the online games, the player reaches the status of beginner, advanced or professional. Finally, the system asks her to share her score and status on social media to invite her friends to participate in the game.

Figure 3.3. Online Game – The Avatar



Notes: The figure shows the task of the third online game. It asks the user to make the avatar look like the main perpetrator from the interactive film. Source: <https://www.zivile-helden.de/>.

3.3.2 Research Hypotheses

The described program uses several lever to motivating people to intervene. By engaging with the interactive film, the player learns to recognize the subtle indications of violence. This increases the likelihood that she will be able to notice a violent situation in the future and interpret it as such. Moreover, if the player remains passive during the film, the victim receives no support, and the situation escalates. If, in contrast, the player takes the first step and becomes active, she receives support from other bystanders, and the situation is resolved. This experience illustrates the central importance of taking responsibility. As the storyline takes different paths based on her choices, the player learns that she can actually influence the situation. Consequently, her perception of control improves. Finally, the film dramatically illustrates that the benefits of intervening outweigh the costs. Saving the victim's life represents an immense benefit to society, while the cost of getting help from other bystanders is relatively manageable. The player learns that she can prevent the worst without endangering herself in the process. In summary, we hypothesize that the bystander program significantly reduces the psychological barriers to intervention postulated by the theoretical model (Hypothesis 1.1).

New information on the favorable benefit-cost ratio of intervening constitutes a significant change in behavioral beliefs. This instantaneously improves the player's attitudes toward intervention behavior. The guiding principle of heroism gives the impression that

society expects her to intervene. Knowing that friends on social media achieved a high score gives the impression that they would intervene themselves. Changed beliefs about injunctive and descriptive norms increase the social pressure perceived by the player. Practicing the desired intervention behaviors during online games increases perceived control. The experience of being able to influence the situation in the film reinforces this effect. We thus hypothesize that the bystander program significantly shifts attitudes, perceived norms, and perceived control in a favorable direction with respect to intervention behavior (Hypothesis 1.2). Overall, the removal of psychological barriers increases the likelihood that a person will intervene in a violent situation. Improved attitudes combined with high social pressure and greater perception of control will automatically result in the behavioral intention to engage in the desired intervention behaviors. Together, this leads to our first testable research hypothesis:

Hypothesis 1: The bystander program significantly increases the behavioral intention to perform the desired intervention behaviors in a violent situation.

The program leverages the motivational drivers of reinforcement and emotions to motivate the desired intervention behaviors. The interactive film rewards active intervention with a sense of joy and satisfaction, or *warm glow*, from helping others (Andreoni, 1990). Together with the score and status achieved, this leads to a positive reinforcement of the desired intervention behavior. Sharing achievement rewards on social media indicates standing within the community, which enhances the reinforcing effect. Sharing also creates a dynamic of repeated play, which leads to the formation of a *behavioral loop*. Intervention behavior finally becomes an automatic behavioral process or habit. This leads to our second research hypothesis:

Hypothesis 2: The use of game principles and game-design elements enhances the motivational effect of the bystander program.

Learning the desired behaviors in a violent situation equips the player with a versatile set of appropriate tactics she can adapt to the situation at hand. Practicing the behaviors during the online games makes her feel more confident in performing these behaviors, which leads to our third research hypothesis:

Hypothesis 3: The bystander program significantly increases the perceived capability to perform the desired intervention behaviors in a violent situation.

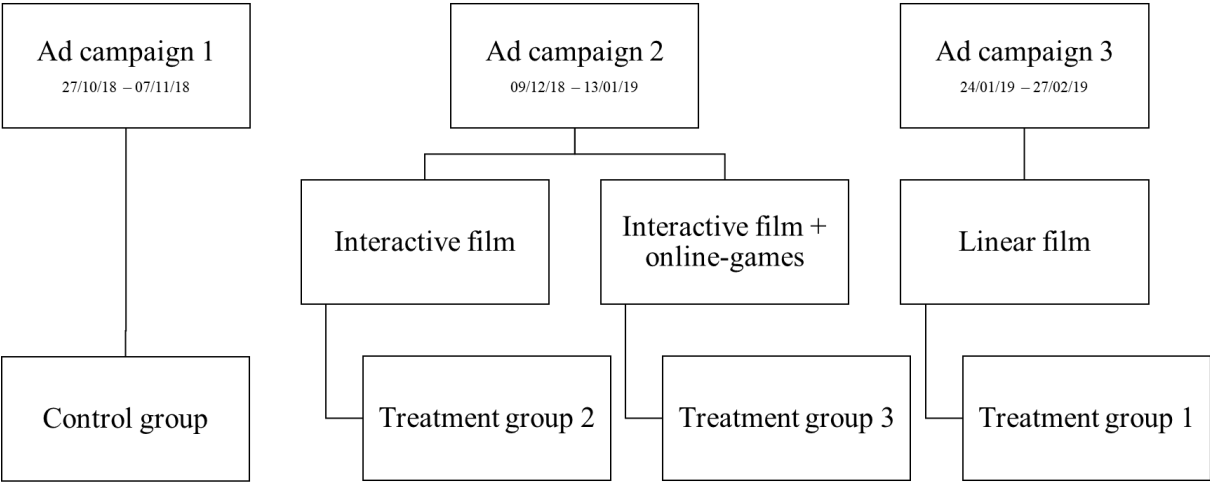
3.4 The Facebook Experiment

To test the three research hypotheses, we conducted two randomized online field experiments. The first experiment took place on *Facebook*. Since potential participants first encountered the bystander program in their *Facebook* newsfeed, the experiment mirrors the real-world conditions of an online prevention campaign. During the experiment, we compared three configurations of the program treatment to test whether gamification increased effectiveness. The configurations included (1) a linear, noninteractive version of the film, (2) the interactive film, and (3) the interactive film in combination with the online games. The *Facebook* experiment was a one-shot game. To ensure that the control group members had no access to the bystander program, we collected their data just before the program went live. In contrast, we collected data from the treatment group after the program had gone live. The timing of data collection thus provided the main randomization mechanism for assigning participants to the treatment and control groups. Moreover, since we recruited the experiment participants on *Facebook*, privacy regulations prevented us from collecting their contact information to conduct a follow-up survey.

3.4.1 Data Collection and Research Design

Data collection for the *Facebook* experiment took place between October 27, 2018, and February 27, 2019. We collected data using three *Facebook* advertising campaigns (Figure 3.4). The first advertising campaign took place before the bystander program went live on November 7, 2018. As part of the campaign, we ran ads that included a link to our online survey. The participants in this first survey formed our control group. After going live, we launched our second advertising campaign in December 2018. In this campaign, we placed ads with a link to the website containing the interactive film and the online games. A random number generator implemented on the website assigned people to one of the two treatment groups. While one treatment group engaged with the interactive film only, the other played the subsequent online games in addition. After participants had finished their respective treatment, a popup incentivized participation in the second online survey.

Figure 3.4. Research Design of the Facebook Experiment



Notes: The figure shows a schematic representation of the Facebook experiment’s research design. The treatment groups are numbered according to treatment intensity.

Finally, we launched a third Facebook advertising campaign in January 2019. In this campaign, the ads included a link to a hidden website that contained a linear, noninteractive version of the film. The linear version corresponded to the mid-case scenario of the interactive film, meaning the situation escalated into violence, but the police arrived before the victim was fatally injured. Afterward, the system directed participants to the online survey. They formed a third treatment group. We numbered the treatment groups according to treatment intensity. That is, treatment group 1 watched the linear movie, treatment group 2 the interactive movie, and treatment group 3 the interactive movie in combination with the online games.

We used the same questionnaire for all of the treatment and control groups.⁴ The questionnaire consisted of four main parts. Part 1 surveyed our primary outcome, part 2 surveyed the parameters of the *reasoned action approach*, part 3 surveyed the typical determinants of bystander behavior, and part 4 surveyed a set of socioeconomic characteristics as covariates. The primary outcome of our Facebook experiment was the behavioral intention to intervene in a violent situation or *willingness to intervene*. We measured the *willingness to intervene* with the violence subscale of the *Munich civil courage instrument* (de. *Muenchener Zivilcourage Instrument*, MueZI; Kastenmüller et al., 2007) and operationalized our primary outcome using the *MueZI score*, which is defined as the sum of the answers given to the subquestions.⁵

⁴ The questionnaire is available upon request.

⁵ The scale describes four situations of violence and asks people to express their agreement with two statements each. The first statement maps the anticipated negative social consequences and reads as follows: "If I take any

To measure attitudes, perceived social pressure, and perceived control in relation to bystander behavior, we developed scales based on the work of Fishbein & Ajzen (2011).⁶ To measure the typical determinants of bystander behavior, we used the relevant scales from the literature. They include *self-efficacy* (Beierlein et al., 2012), *justice sensitivity* (Baumert et al., 2014), *empathy* (Leibetseder et al., 2001), *responsibility denial* (Schwartz, 1977), and *propensity to violence* (Ulbrich-Herrmann, 2014). To analyze potential effect heterogeneity, we assessed a set of sociodemographic characteristics. We checked the survey data carefully to ensure the validity of our results. That is, we identified straight liners, checked for outliers or implausible answers, and dropped incomplete interviews or duplicates. Straight-liners give the same answer to every single question. After we cleaned the data, the total sample contained 4,118 observations.

3.4.2 Summary Statistics

The summary statistics of our sample confirm that we have reached the intended target group with our advertising campaigns. The sample consists of people who are younger and better educated than the overall population. At two-thirds, the proportion of 18- to 34-year-olds in the sample is exactly twice as high as that in the overall population. The proportion of academics is also more than one and a half times greater in the sample than in the overall population. Furthermore, we observe a slight preponderance of women. The share of employed persons does not deviate much from the population average. The share of singles and parents, on the other hand, is somewhat lower, and the number of children is somewhat higher. Urban residents are severely underrepresented, with a share that is approximately one-third as high as in the overall population (see Table B.1.1 in Appendix B for details).

3.4.3 Checking for Balance

Random assignment worked well. We checked for balance by regressing assignment to the respective treatment arm jointly on all covariates and again separately on each covariate alone. The relatively large share of insignificant covariates, together with the insignificant F -test on joint significance, and the low adjusted coefficient of determination (adjusted R -

action, I'm threatened with negative consequences from the perpetrators." The second statement maps the *willingness to intervene* and reads as follows: "I am prepared to take action against it." People can express their agreement on a 4-point scale from 0 ("Not agree") to 3 ("Fully agree").

⁶ These scales ask people for their agreement with a series of statements including "I am expected to intervene when the situation calls for it", "It is advantageous for me to intervene when the situation requires it", and "It is difficult for me to act in a civil manner when the situation calls for it."

squared) indicate that the treatment and control groups have the same characteristics on average. Thus, no systematic selection into the groups occurred that could have biased our main results (i.e., no *selection bias*). Table B.1.2 in Appendix B shows the combined results of the different balancing checks.

3.4.4 Main Results

Table 3.1 shows the main results of our *Facebook* experiment. They come from three separate regressions of the *MueZI score* on assignment to the respective treatment and a set of covariates. The covariates include age group, gender, academic degree, employment and relationship status (single versus relationship), parenthood, and residence (rural versus urban). We included these covariates to improve the statistical precision of the estimated treatment effects. We estimated a series of different models for the final specification and chose the one with the highest adjusted coefficient of determination for our analysis. Since we obtained data from a randomized experiment, the estimated treatment effects are quite robust across model specifications.

Table 3.1. Main Results of the Facebook Experiment

Dependent variable: MueZI score				
	(1)	(2)	(3)	(4)
	Control group mean	Estimated coefficients		
Variable		Linear film	Interactive film	Interactive film + online- games
Treatment	10.12	0.30*** (0.07)	0.66*** (0.08)	0.45*** (0.08)
Age group				
35-44		-0.24** (0.11)	-0.19* (0.11)	-0.29** (0.12)
45-64		-0.35*** (0.11)	-0.31** (0.13)	-0.43*** (0.13)
Female		0.23*** (0.07)	0.24*** (0.08)	0.25*** (0.08)
Academic		-0.17** (0.08)	-0.20** (0.08)	-0.22*** (0.08)
Employed		0.16 (0.12)	0.16 (0.13)	0.05 (0.12)
Single		-0.05 (0.07)	-0.16** (0.08)	-0.21*** (0.08)
Parent		0.24** (0.10)	0.32*** (0.11)	0.30*** (0.11)
Urbanite		0.16** (0.08)	0.14* (0.09)	0.10 (0.08)
Constant		9.82*** (0.17)	9.69*** (0.20)	9.79*** (0.19)
F-value		2.80	5.23	4.36
df		24	24	24
Prob > <i>F</i>		0.00	0.00	0.00
R ²		0.02	0.05	0.04
Adjusted R ²		0.02	0.04	0.03
Observations	1,696	2,854	2,224	2,432

Notes: This table shows the point estimates from three separate linear regressions (*ordinary least squares*, OLS) of the *Munich civil courage instrument score* (MueZI score) on the respective treatment indicator (dummy) and a set of covariates. The covariates include age group (3 categories), female (dummy), academic (dummy), employed (dummy), single (dummy), parent (dummy), urbanite (dummy), and fixed effects at the federal state-level. The last row gives the number of observations in the control group in column (1), and the combined observations in the control and the respective treatment group in the columns (2) to (4). Robust standard errors in parentheses. *p<0.10, **p<0.05, ***p<0.01.

The treatment effects are positive and highly significant in all three cases. However, the interactive film increases the *MueZI score* more than twice as much as the linear film. Given a control group mean of 10.12 and a point estimate of 0.66, the magnitude of the effect reaches 6.5%. This means that 6-7 out of 100 people are willing to intervene in a violent situation just because they engaged with the interactive film. In combination with online games, the effect still exceeds the linear film by half. The results clearly demonstrate that the program has achieved its main goal of motivating people to intervene in a violent situation. Notably, they also show that the application of game principles and game design elements reinforces the treatment effect.

3.4.5 Impact Vectors

The bystander program unfolds its motivational effect through all of the channels predicted by the *reasoned action approach*. Table 3.2 shows the results from separately regressing the parameters of the *reasoned action approach* on the treatment indicator and the same set of covariates as above. Contrary to our expectations, the attitude toward intervention behavior worsens after people see the linear film. In contrast, engagement with the interactive film improves attitudes. Engagement also convinces people that their relevant peer group would expect them to intervene and that they have control over their intervention behavior. People who additionally played the online games believed that members of their relevant peer group would intervene themselves.

Table 3.2. Impact Vectors as Predicted by the Reasoned Action Approach

Dependent variable: Row-wise				
Variables	(1)	(2)	(3)	(4)
	Control group mean	Estimated coefficients (treatment indicators)		
		Linear film	Interactive film	Interactive film + online-games
Attitude	1.58	-0.07** (0.03)	0.09** (0.04)	0.12*** (0.04)
Injunctive norm	2.23	0.04 (0.03)	0.07** (0.03)	0.12*** (0.03)
Descriptive norm	1.70	0.00 (0.03)	0.06 (0.04)	0.10*** (0.03)
Control	1.68	0.00 (0.03)	0.08** (0.04)	0.11*** (0.04)

Notes: This table shows the control group means and treatment effects on the parameters of the *reasoned actions approach*. The point estimates come from separately regressing each of the parameters on the respective treatment indicator (dummy) and a set of covariates. The covariates include age group (3 categories), female (dummy), academic (dummy), employed (dummy), single (dummy), parent (dummy), and urbanite (dummy). Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The bystander program also works through some of the typical determinants of bystander behavior. More specifically, it mainly strengthens the sense of responsibility. All three treatments decrease *responsibility denial*. The interactive film also reduces the *propensity to violence* – both on its own and in combination with online games. All three treatments also affected the other determinants of bystander behavior. However, these effects are rather scattered and thus less robust (see Table B.1.3 in Appendix B).

3.4.6 Heterogeneity Analysis

Effect heterogeneity seems to play only a minor role in explaining the results. The interactive film motivates intervention across age groups, genders, educational attainments, and employment status. In combination with the online games, the effect persists in all subgroups

except the high age group. The linear film has no impact in the subgroups of high- and middle-aged, and unemployed individuals.⁷

3.4.7 Discussion

The results of the Facebook experiment support the hypotheses that the bystander program is able to motivate intervention in violent situations and that the use of game principles and game-design elements further enhances the motivational effect. The sample, however, consisted of typical *Facebook* users who tended to be younger, better educated, and more technophilic than average. The question, therefore, remains whether the positive treatment effect holds for the overall population as well. Moreover, data regulation did not allow conducting a follow-up survey to evaluate the long-term effect. Finally, the outcome measure we used is rather broad, making it difficult to distinguish between mere priming and actual learning effects. To compensate for these limitations, we conducted a second experiment.

3.5 The Panel Data Experiment

In the second experiment, we tested whether the positive treatment effect (1) persists over a longer period and (2) holds for the overall population. Pursuing these two major objectives, we drew a random sample that was representative of the population and randomly assigned participants into a treatment and a control group. Because the interactive film had the strongest effect in the *Facebook* experiment, we exposed participants to this treatment only and left out the online games. A screening question ensured that members of the treatment group had never seen the interactive film before. As in the *Facebook* experiment, the control group received no treatment. In both groups, we conducted follow-up surveys after four and eight weeks. To differentiate mere priming from actual learning, we developed two precise outcome measures.

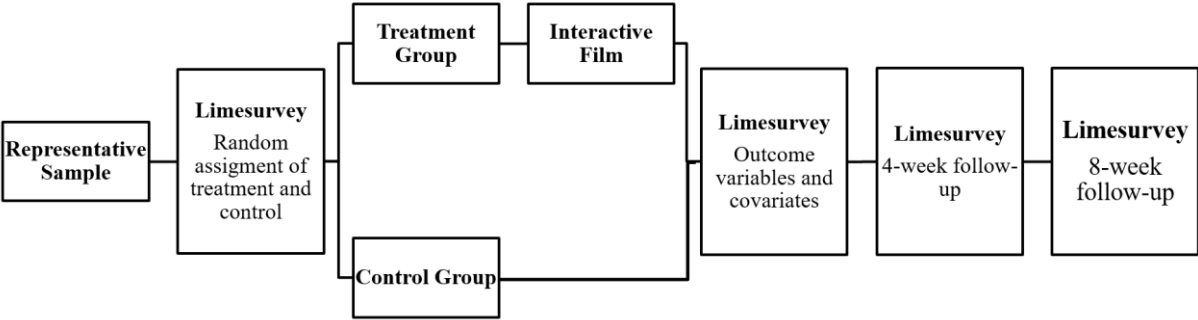
3.5.1 Data Collection and Research Design

We collected survey data over three waves between May 18 and August 13, 2020. A market research firm carried out the sampling. In the first wave, they drew a representative sample of the working population in terms of age and gender (cross-quoted). For the second wave, participants received an e-mail invitation exactly 4 weeks after they had answered the questionnaire in the first wave. Likewise, they received an invitation for the third wave exactly

⁷ Table B.1.4 in Appendix B shows the estimated treatment effects in the subgroups in detail.

4 weeks after answering the second questionnaire (Figure 3.5). We ensured the representativeness of the sample using two screening questions on age and gender. After potential participants had answered these two questions, a random number generator assigned the eligible candidates to a treatment group or control group. The system redirected participants from the treatment group to the interactive film, which started automatically. After the film, the system automatically redirected them to the online survey. Users in the control group went directly to the survey.

Figure 3.5. Research Design of the Panel Data Experiment



Notes: The figure shows a schematic representation of the panel data experiment’s research design.

We used the *LimeSurvey* app as the technical infrastructure and conducted the survey in German. Both the treatment and control groups had to answer the same questionnaire with three main parts.⁸ Part 1 measured the primary outcomes, part 2 measured the secondary outcomes, and part 3 measured a set of socioeconomic characteristics as covariates. Our two primary outcomes include the *willingness to intervene* and the *perceived intervention capability*, which capture the two major objectives of any bystander intervention as explained above. To measure the primary outcomes, we developed two scales based on existing instruments (see Banyard et al., 2007; Levine & Crowther, 2008). The *willingness to intervene scale* asked participants about their willingness to perform each of eight different intervention behaviors on a scale from 1 (“*highly unlikely*”) to 7 (“*highly likely*”). We selected the eight intervention behaviors from the existing instruments according to the recommendations of expert reviewers from police crime prevention. The *intervention capability scale* asked participants for their ability to perform each of these eight intervention behaviors on a scale from 0 (“*I cannot do that.*”) to 10 (“*I am absolutely sure that I can do that.*”). We asked the questions after we had shown participants a short vignette describing a violent situation. To distinguish between priming and

⁸ The questionnaire is available upon request.

learning effects, we randomly varied the storyline of the vignette. While one vignette accurately depicted the situation from the interactive film, the other described a typical violent situation in which a man threatens to beat a woman on the street. Assuming the outcomes would systematically differ between the two vignettes, this would suggest a priming effect.

We constructed two scores as primary outcome measures. The *willingness to intervene score* (WTIS) is the sum of the six desired intervention behaviors minus the sum of the two undesired intervention behaviors from the *willingness to intervene scale*. The *intervention capability score* (ICS) is calculated analogously but uses the answers from the *intervention capability scale*. In the later analysis, we calculated these scores once separately for the different vignettes and once pooled for the entire treatment and control groups. Finally, we conducted *exploratory factor analysis* (EFA) and *confirmatory factor analysis* (CFA) to prove that the two scales have a well-defined factor structure and high validity. The results from these analyses are available upon request. Subsequently, we measured the secondary outcomes, including the parameters of the *barriers to intervention model* (Burn, 2009) and the *reasoned action approach* (Fishbein & Ajzen, 2011). They help to explain the impact vector between our program treatment and the primary outcomes. To analyze potential effect heterogeneity in subgroups of the population, we included questions about sociodemographic characteristics.

We checked the data carefully to ensure the validity of the results. In addition to implementing an additional screening question in the survey questionnaire, we identified speeders and straight liners. We assumed speeding if a candidate's interview time was below one-third of the median interview time. If a candidate failed at least two of the three quality criteria (i.e., screening, speeding, or straight-lining), we excluded the observation from the analysis. We also checked for outliers and implausible answers and dropped incomplete interviews and duplicates. Finally, we excluded participants who had already received the treatment (i.e., engaged with the interactive film) before the panel data experiment. After we cleaned the data, the total sample size was 1,587 in the first survey wave, 1,388 in the second survey wave, and 1,253 in the third survey wave.

3.5.2 Summary Statistics

Summary statistics of selected covariates confirm that the sample is nationally representative with respect to age and gender. However, there are some deviations with respect to other characteristics. The proportion of academics is twice as high as in the total population, and the employment rate is slightly lower. Singles are slightly underrepresented, while parents

are overrepresented. The average number of children per household is below the population average of 1.6. The urbanization rate is also 10 percentage points lower than in the population as a whole (see Table B.2.1 in Appendix B for details).

3.5.3 Checking for Balance

Randomization worked well in the panel data experiment. We checked for balance by regressing assignment to treatment jointly on all covariates and separately on each covariate alone. As in the *Facebook* experiment, the large share of insignificant covariates, the insignificant *F*-test on joint significance, and the close to zero adjusted coefficient of determination (adjusted R-squared) indicate that the treatment and control groups have the same characteristics on average. Therefore, no systematic selection into the groups occurred. Table B.2.2 in Appendix B shows the combined results of the balancing checks.

3.5.4 Main Results

Our main results confirm that engagement with the interactive film motivates people to intervene in violent situations, as demonstrated by the *Facebook* experiment above. Notably, they further show that the effect persists over time. Table 3.3 contains the interactive film's treatment effects on the *willingness to intervene score* across the three survey waves. The effects are point estimates from separately regressing the two outcome measures on the treatment indicator, fixed effects at the federal state level, and the set of covariates. In this specification, we used the pooled scores. The results for the scores that we differentiated by vignettes are available upon request.

Table 3.3. Regression Table – Willingness to Intervene

Dependent variable: Willingness to intervene score				
	(1)	(2)	(3)	(4)
	Control group mean (wave 1)	Estimated coefficients		
Variable		Post- treatment (wave 1)	4 weeks follow-up (wave 2)	8 weeks follow-up (wave 3)
Treatment	27.65	1.56*** (0.34)	1.36*** (0.35)	1.05*** (0.37)
Constant		25.16*** (0.83)	25.10*** (0.82)	24.83*** (0.93)
F statistic		5.15	4.28	4.90
p-value		0.00	0.00	0.00
Degrees of freedom		24	24	24
Adjusted R ²		0.06	0.05	0.07
No. of observations	796	1,587	1,378	1,247

Notes: This table shows the control group means and treatment effects across the three survey waves. The point estimates come from separately regressing the *willingness to intervene score* (WTIS) on the treatment indicator (dummy) and a set of covariates. The covariates include age group (3 categories), female (dummy), academic (dummy), employed (dummy), single (dummy), parent (dummy), urbanite (dummy), and fixed effects at the federal state-level. The table also shows the regression statistics including the F-statistics from an F-test of joint significance, the corresponding p-values as well as the degrees of freedom, and adjusted R-squares. The last row gives the number of observations in the control group in columns (1) and (5), and the combined observations in the control and the treatment groups in the columns (2) to (4) and (6) to (8). Robust standard errors in parentheses. *p<0.10, **p<0.05, ***p<0.01.

Compared with a control group mean of 27.65 points, the *willingness to intervene score* increases by 1.56 points or 5.64% immediately after participants have engaged with the interactive film. In other words, five to six people of one hundred participants will intervene in a violent situation because they engaged with the film when they otherwise would not have done so. Four weeks after treatment, the effect still had a magnitude of 1.36 or 4.85%. Eight weeks later, it was still 1.05 or 3.74%.⁹ In other words, even after eight weeks, there were still three to four program participants intervening because of engagement with the interactive film.

⁹ According to the t-test performed, the control group means of the *willingness to intervene score* are relatively stable across the three waves, with values of 28.05 and 28.06 in waves 2 and 3, respectively.

Engagement with the interactive film also made people feel more capable of intervening. We repeated the regressions described above using the *intervention capability score* as the dependent variable (Table 3.4). Compared with a control group mean of 40.68 points, the *intervention capability score* increased by 2.64 points or 6.49% immediately after treatment. Thus, people who engaged with the film rated their intervention capability nearly 7% higher than people who did not. . The effect also persists over time, even though it declines slightly more in this case. After 4 and 8 weeks, they still rated their capability 5.27% and 3.38% higher, respectively.¹⁰

Table 3.4. Regression Table – Intervention Capability

Dependent variable: Intervention capability score				
	(1)	(2)	(3)	(4)
	Control group mean (wave 1)	Estimated coefficients		
Variable		Post- treatment (wave 1)	4 weeks follow-up (wave 2)	8 weeks follow-up (wave 3)
Treatment	40.68	2.64*** (0.54)	2.16*** (0.56)	1.40** (0.61)
Constant		35.10*** (1.31)	35.97*** (1.35)	34.67*** (1.50)
F statistic		7.06	5.95	5.60
p-value		0.00	0.00	0.00
Degrees of freedom		24	24	24
Adjusted R ²		0.08	0.07	0.08
No. of observations	796	1,587	1,378	1,247

Notes: This table shows the control group means and treatment effects across the three survey waves. The point estimates come from separately regressing the *Intervention Capability Score* (ICS) on the treatment indicator (dummy) and a set of covariates. The covariates include age group (3 categories), female (dummy), academic (dummy), employed (dummy), single (dummy), parent (dummy), urbanite (dummy), and fixed effects at the federal state-level. The table also shows the regression statistics including the F-statistics from an F-test of joint significance, the corresponding p-values as well as the degrees of freedom, and adjusted R-squares. The last row gives the number of observations in the control group in columns (1) and (5), and the combined observations in the control and the treatment groups

¹⁰ According to the t-test performed, the control group means of the *intervention capability score* are relatively stable across the three waves, with values of 41.00 and 41.37 in waves 2 and 3, respectively.

in the columns (2) to (4) and (6) to (8). Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3.5.5 Impact Vectors

Having shown that interactive film exerts a significant effect that persists over time, we now analyze the channels through which the interactive film unfolds its effect. For this purpose, we ran several regressions. First, we separately regressed each parameter of the *barriers to intervention model* on the treatment indicator and the set of covariates. The results imply that engagement with the interactive film leads to people being more able to detect a violent situation and interpret it as such (see Table B.2.3 in Appendix B). These effects persisted even after 4 and 8 weeks, although they faded slightly at the end. The interactive film also makes people more likely to take responsibility. Members of the treatment group also perceive that they have more control over their intervention behavior. The interactive film was thus able to reduce all of the psychological barriers to intervention predicted by the model.

Different factors determine whether an individual can overcome a particular barrier to intervention. Using separate regressions, we evaluated the interactive film's effect on these factors. According to the regression results, the film seems to have the strongest and most persistent effect on empathy (see Table B.2.4 in Appendix B). Acceptance of negative social consequences increases immediately after treatment but does not persist over time. In contrast, the effect on anticipated guilt appears only after 4 weeks and fades out again after 8 weeks. The same is true for the effect on indignation and audience inhibition.

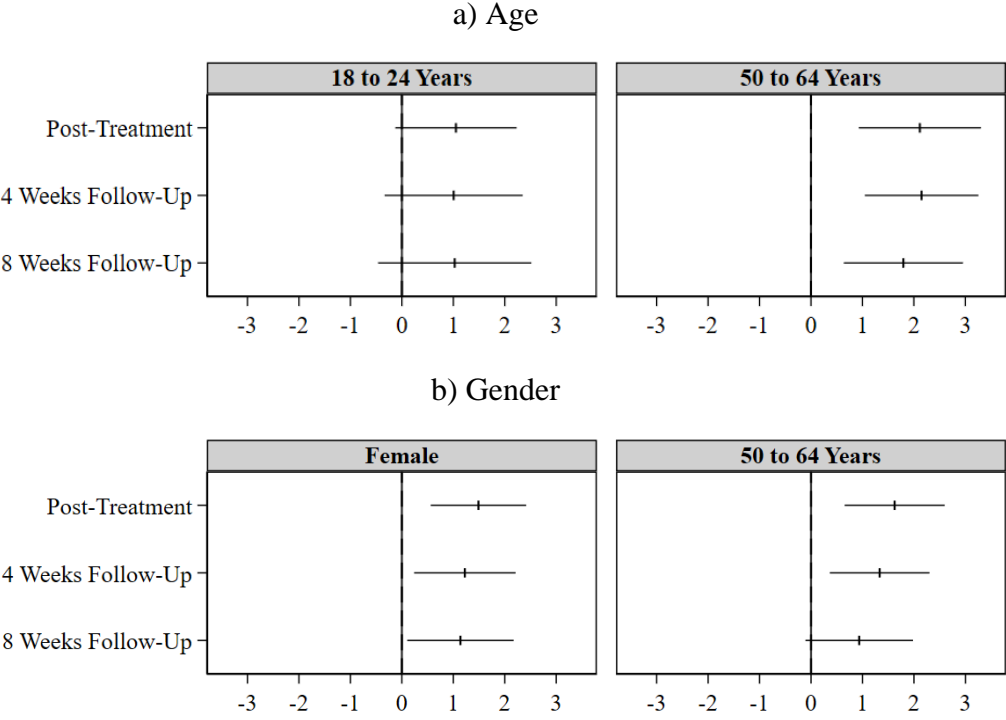
Engagement with the interactive film also influences all factors of the *reasoned action approach* (see Table B.2.5 in Appendix B). Immediately after the interactive film, people have a more positive attitude toward calling the police in a violent situation. The change in attitudes lasts even after 4 or 8 weeks. On the other hand, perceived social pressure only increases immediately after the film and fades out again after only 4 weeks. Perceived control increases only 4 weeks after people have seen the interactive film and then fades out again. However, we must note that the base level for these questions was already relatively high.

3.5.6 Heterogeneity Analysis

The panel data experiment highlights that effect heterogeneity plays a larger role than revealed by the *Facebook* experiment. Figure 3.6 shows the treatment effects on the *willingness*

to intervene score in major demographic subgroups.¹¹ Age seems to be a decisive factor here. For the older subgroup, we see a highly significant, positive effect on the willingness to intervene that persists over time and exceeds the effect in the younger subgroup by approximately twice. In the younger group, the effect was only weakly significant and disappeared after only four weeks. Gender seems to be less decisive. For both genders, the effect ranges roughly in the same order of magnitude and remains relatively stable over time. The factor of education again plays a greater role. The interactive film exerts a highly significant and persistent effect on nonacademics. The effect on academics turns out to be weaker and disappears after four weeks. An even clearer picture emerges for employment status. While the film has a highly significant impact on the employed that persists over the entire period, we see no impact at all on the unemployed.

Figure 3.6. Treatment Effects by Demographic Subgroups



¹¹ For full results in all subgroups, see Table B.2.6 in Appendix B.



Notes: This figure shows the treatment effects and confidence intervals for selected sociodemographic subgroups across the three survey waves. The point estimates come from regressing the *willingness to intervene score* (WTIS) on the treatment indicator (dummy) and a set of covariates. The covariates include age group (3 categories), female (dummy), academic (dummy), employed (dummy), single (dummy), children (dummy), urbanite (dummy), and fixed effects at the federal state-level. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3.6 Conclusion

Gamification offers great potential for bystander education. The findings of the *Facebook* experiment not only support the hypothesis that the interactive film motivates people to intervene but also that the use of game principles and game design elements enhances the motivational effect. The film achieves this effect at both the rational and emotional levels. On the rational level, it changes the beliefs underlying intervention behavior in the desired way. On the emotional level, it leverages the appropriate reinforcements to elicit the desired affective responses. Finding out that the film has a stronger impact without the online games could indicate that the emotional scenes in the film represent stronger reinforcements than the points awarded in the online games.

Social media are well suited for the dissemination and upscaling of online bystander programs. The *Facebook* experiment shows that bystander education works within the social media environment, where people usually first encounter new digital content. As the *Facebook* experiment had minor caveats in terms of the persistence and external validity of the results, we conducted a second experiment with panel data. The panel data experiment supports the

hypothesis that the film equips people with the skills necessary for safe and effective intervention. It also shows that the treatment effects persist over the long-term period and apply to the population as a whole.

Research suggests that prevention provides a more effective crime-fighting measure than incarceration. In this context, the education of potential bystanders is a particularly effective prevention strategy. However, traditional programs have inherent disadvantages in terms of scaling, which can be solved with online programs distributed via social media. The use of game principles and game-design elements can add to the already great effectiveness of such programs. Our key contributions to social psychology and information systems research amount to showing that gamification does indeed increase effectiveness and that gamified bystander programs do work in the social media environment. Policymakers can take advantage of these findings to make future prevention programs more effective and highly scalable.

4 Evaluating an Interactive Film on the Prevention of Political Radicalization*

with:
Stephan L. Thomsen

This paper has been published as:

Ebers, A., & Thomsen, S. L. (2022). Evaluating an Interactive Film on the Prevention of Political Radicalization. *Journal for Deradicalization*, (30), 169-222.

Link: <https://journals.sfu.ca/jd/index.php/jd/article/view/579>

* We like to thank Harald Schmidt, Sophie von Bissingen, Frank Buchheidt, Constantin Gerber, Jan-Gerrit Keil, Alexander Kuchling, Andreas Ostwaldt, Wolfram Pemp, and Ralf Richter for valuable inputs and fruitful discussions. The paper also profited from discussions with participants at the Best Practice Conference (BPC) of the European Crime Prevention Network (ECPN). The experiment is registered at the American Economic Association (AEA) registry: “*Evaluating a gamified online intervention to prevent political radicalization.*” (AEARCTR-0008329).

5 How do Warnings Affect Retail Demand for Bitcoin?

Evidence from an international Survey Experiment*

with:

Prof. Dr. Stephan L. Thomsen

This paper has been published as:

Ebers, A., & Thomsen, S. L. (2021). *How do warnings affect retail demand for Bitcoin? Evidence from an international survey experiment*. *Journal of Behavioral and Experimental Finance*, 32.

DOI: <https://doi.org/10.1016/j.jbef.2021.100567>

* We thank Marina Schröder and Lars Brausewetter for their helpful comments. The paper also profited from discussions with participants of the 2021 Virtual European Meeting of the Econometric Society (ESEM). Financial support by the *Deutsche Bundesbank* is gratefully acknowledged. The experiment is registered at the American Economic Association (AEA) registry: "*The effect of warnings on consumer attitudes towards cryptocurrency and central bank digital currency: A survey experiment*" (AEARCTR-0006573).

Bibliography

- AdWords API. (2016). *Bid Landscapes*.
<https://developers.google.com/adwords/api/docs/guides/bid-landscapes>
- Anderson, C. A., Berkowitz, L., Donnerstein, E., Huesmann, L. R., Johnson, J. D., Linz, D., Malamuth, N. M., & Wartella, E. (2003). The influence of media violence on youth. *Psychological Science in the Public Interest*, 4(3), 81–110.
- Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *The Economic Journal*, 100(401), 464–477.
- 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.
- Banyard, V. L., Moynihan, M. M., & Plante, E. G. (2007). Sexual violence prevention through bystander education: An experimental evaluation. *Journal of Community Psychology*, 35(4), 463–481.
- Banyard, V. L., Plante, E. G., & Moynihan, M. M. (2005). Rape prevention through bystander education: Bringing a broader community perspective to sexual violence prevention. *US Department of Justice*, 1–206.
- 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.
- Baumert, A., Beierlein, C., Schmitt, M., Kemper, C. J., Kovaleva, A., Liebig, S., & Rammstedt, B. (2014). Measuring four perspectives of justice sensitivity with two items each. *Journal of Personality Assessment*, 96(3), 380–390.
- Baumert, A., Halmburger, A., & Schmitt, M. (2013). Interventions against norm violations: Dispositional determinants of self-reported and real moral courage. *Personality and Social Psychology Bulletin*, 39(8), 1053–1068.
- Beierlein, C., Kovaleva, A., Kemper, C. J., & Rammstedt, B. (2012). *Ein Messinstrument zur Erfassung subjektiver Kompetenzerwartungen: Allgemeine Selbstwirksamkeit Kurzskala (ASKU)*.
- Berkowitz, A. D. (2002). Fostering men's responsibility for preventing sexual assault.

- Preventing Violence in Relationships: Interventions across the Life Span*, 163–196.
- Burn, S. M. (2009). A situational model of sexual assault prevention through bystander intervention. *Sex Roles*, 60(11–12), 779–792.
- Calvo-Morata, A., Alonso-Fernández, C., Freire, M., Martínez-Ortiz, I., & Fernández-Manjón, B. (2020). Serious games to prevent and detect bullying and cyberbullying: A systematic serious games and literature review. *Computers & Education*, 157, 103958.
- Camerer, C. F. (2011). *Behavioral game theory: Experiments in strategic interaction*. Princeton university press.
- Castronovo, C., & Huang, L. (2012). Social media in an alternative marketing communication model. *Journal of Marketing Development and Competitiveness*, 6(1), 117–134.
- 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.
- 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).
- Cohen, M. A., & Bowles, R. (2010). Estimating costs of crime. In *Handbook of quantitative criminology* (pp. 143–162). Springer.
- Cugelman, B., Thelwall, M., & Dawes, P. (2011). Online interventions for social marketing health behavior change campaigns: a meta-analysis of psychological architectures and adherence factors. *Journal of Medical Internet Research*, 13(1).
- Deater-Deckard, K., Chang, M., & Evans, M. E. (2013). Engagement states and learning from educational games. *New Directions for Child and Adolescent Development*, 2013(139), 21–30.
- DeSmet, A., Bastiaensens, S., Van Cleemput, K., Poels, K., Vandebosch, H., Deboutte, G., Herrewijn, L., Malliet, S., Pabian, S., & Van Broeckhoven, F. (2018). The efficacy of the Friendly Attac serious digital game to promote prosocial bystander behavior in cyberbullying among young adolescents: a cluster-randomized controlled trial. *Computers in Human Behavior*, 78, 336–347.
- Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: defining gamification. *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, 9–15.
- Deutscher Presserat [German Press Council] (2005). Publizistische Grundsätze [Journalistic Principles]. *Bonn: Deutscher Presserat [German Press Council]. i i*.
- Duhigg, C. (2012). *The power of habit: Why we do what we do in life and business*. Random House.
- Ebers, A., & Thomsen, S. L. (2021). Benefit–Cost Analysis of Social Media Facilitated Bystander Programs. *Journal of Benefit-Cost Analysis*, 12(2), 367–393.
- Elsaesser, C., Patton, D. U., Weinstein, E., Santiago, J., Clarke, A., & Eschmann, R. (2021). Small becomes big, fast: Adolescent perceptions of how social media features escalate online conflict to offline violence. *Children and Youth Services Review*, 122, 105898.

- Elverdam, C., & Aarseth, E. (2007). Game classification and game design: Construction through critical analysis. *Games and Culture*, 2(1), 3–22.
- European Central Bank [@ecb]. (2020). *Lagarde on cryptocurrencies: We are looking into the feasibility and merits of a central bank digital currency that would ensure that people can continue to use central bank money even if the use of physical cash eventually declines. We want to play an ac.* Twitter.
<https://twitter.com/ecb/status/1214926506570452993>
- European Central Bank [@ecb]. (2021). *Crypto-assets are volatile. They lack any intrinsic value and there is no reliable institution backing them. In contrast, a digital euro would offer Europeans the same level of confidence as with cash, since it would be backed by the ECB #AskECB [Tweet].* Twitter; Twitter.
<https://twitter.com/ecb/status/1371852009989668877>
- 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.
- Gallacher, J. D., Heerdink, M. W., & Hewstone, M. (2021). Online engagement between opposing political protest groups via social media is linked to physical violence of offline encounters. *Social Media+ Society*, 7(1), 2056305120984445.
- 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.
- Gidycz, C. A., Orchowski, L. M., Probst, D. R., Edwards, K. M., Murphy, M., & Tansill, E. (2015). Concurrent administration of sexual assault prevention and risk reduction programming: Outcomes for women. *Violence against Women*, 21(6), 780–800.
- Gruber, F., Lützing, S., & Kemmesies, U. E. (2017). Extremismusprävention in Deutschland—Erhebung und Darstellung der Präventionslandschaft [Prevention of Extremism in Germany - Survey and Presentation of the Prevention Landscape]. *Modulabschlussbericht. Wiesbaden: Bundeskriminalamt.*
- Halmburger, A., Baumert, A., & Schmitt, M. (2017). Everyday heroes: Determinants of moral courage. In S. T. Allison, G. R. Goethals, & R. M. Kramer (Eds.), *Handbook of heroism and heroic leadership* (pp. 165–184). Routledge/Taylor & Francis Group.
- Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does Gamification Work?-A Literature Review of Empirical Studies on Gamification. *HICSS*, 14(2014), 3025–3034.
- Hinduja, S., & Patchin, J. W. (2013). Social influences on cyberbullying behaviors among middle and high school students. *Journal of Youth and Adolescence*, 42(5), 711–722.
- Huesmann, L. R., Moise-Titus, J., Podolski, C.-L., & Eron, L. D. (2003). Longitudinal relations between children's exposure to TV violence and their aggressive and violent behavior in young adulthood: 1977-1992. *Developmental Psychology*, 39(2), 201.

- 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).
- Jouriles, E. N., Sargent, K. S., Salis, K. L., Caiozzo, C., Rosenfield, D., Cascardi, M., Grych, J. H., O'Leary, K. D., & McDonald, R. (2020). TakeCARE, a video to promote bystander behavior on college campuses: Replication and extension. *Journal of Interpersonal Violence*, 35(23–24), 5652–5675.
- Kastenmüller, A., Greitemeyer, T., Fischer, P., & Frey, D. (2007). Das Münchner Zivilcourage-Instrument (MüZI) [The Munich Civil Courage Instrument (MueZI)]. *Diagnostica*, 53(4), 205–217.
- 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.
- King, J. E., Walpole, C. E., & Lamon, K. (2007). Surf and turf wars online—growing implications of internet gang violence. *Journal of Adolescent Health*, 41(6), S66–S68.
- Latané, B., & Darley, J. M. (1970). *The unresponsive bystander: Why doesn't he help?* Appleton-Century-Crofts.
- Leibetseder, M., Laireiter, A. R., Riepler, A., & Köller, T. (2001). E-Skala: Fragebogen zur Erfassung von Empathie—Beschreibung und psychometrische Eigenschaften [E-scale: Empathy Questionnaire - Description and Psychometric Properties]. *Zeitschrift Für Differentielle Und Diagnostische Psychologie*, 22(1), 70–85.
- Levine, M., & Crowther, S. (2008). The responsive bystander: How social group membership and group size can encourage as well as inhibit bystander intervention. *Journal of Personality and Social Psychology*, 95(6), 1429.
- 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*.
- Marcum, C. D., Higgins, G. E., & Ricketts, M. L. (2010). Potential factors of online victimization of youth: An examination of adolescent online behaviors utilizing routine activity theory. *Deviant Behavior*, 31(5), 381–410.
- 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.
- McKenna, K. Y. A., & Bargh, J. A. (2000). Plan 9 from cyberspace: The implications of the Internet for personality and social psychology. *Personality and Social Psychology Review*, 4(1), 57–75.
- Miller, E., Das, M., Tancredi, D. J., McCauley, H. L., Virata, M. C. D., Nettiksimmons, J., O'Connor, B., Ghosh, S., & Verma, R. (2014). Evaluation of a gender-based violence prevention program for student athletes in Mumbai, India. *Journal of Interpersonal Violence*, 29(4), 758–778.
- Nadeem, D. (2021). *How tweets by Tesla's Elon Musk have moved markets*. Reuters. <https://www.reuters.com/business/finance/how-tweets-by-teslas-elon-musk-have-moved-markets-2021-11-08/>

- Patton, D. U., Hong, J. S., Ranney, M., Patel, S., Kelley, C., Eschmann, R., & Washington, T. (2014). Social media as a vector for youth violence: A review of the literature. *Computers in Human Behavior, 35*, 548–553.
- Potter, S. J., Stapleton, J. G., & Moynihan, M. M. (2008). Designing, implementing, and evaluating a media campaign illustrating the bystander role. *Journal of Prevention & Intervention in the Community, 36*(1–2), 39–55.
- 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.
- Robson, K., Plangger, K., Kietzmann, J. H., McCarthy, I., & Pitt, L. (2015). Is it all a game? Understanding the principles of gamification. *Business Horizons, 58*(4), 411–420.
- Röderer, K., Schwarz, S., & Tscheligi, M. (2019). Engaging Bystanders Using Persuasive Technology: A Meta-analysis of Influencing Factors on Moral Courage. *Persuasive Technology: Development of Persuasive and Behavior Change Support Systems: 14th International Conference, PERSUASIVE 2019, Limassol, Cyprus, April 9-11, 2019, Proceedings, 11433*, 202.
- Rothschild, M. L., & Gaidis, W. C. (1981). Behavioral learning theory: Its relevance to marketing and promotions. *Journal of Marketing, 45*(2), 70–78.
- Sailer, M., Hense, J., Mandl, J., & Klevers, M. (2014). Psychological perspectives on motivation through gamification. *Interaction Design and Architecture Journal, 19*, 28–37.
- Salazar, L. F., Vivolo-Kantor, A., & Schipani-McLaughlin, A. M. (2019). Theoretical mediators of RealConsent: a web-based sexual violence prevention and bystander education program. *Health Education & Behavior, 46*(1), 79–88.
- Santhanam, R., Liu, D., & Shen, W.-C. M. (2016). Research Note—Gamification of technology-mediated training: Not all competitions are the same. *Information Systems Research, 27*(2), 453–465.
- Schwartz, S. H. (1977). Normative influences on altruism. In *Advances in experimental social psychology* (Vol. 10, pp. 221–279). Elsevier.
- Shaw, J., & Janulis, P. (2016). Re-evaluating sexual violence prevention through bystander education: A latent growth curve approach. *Journal of Interpersonal Violence, 31*(16), 2729–2750.
- Silver, L., Huang, C., & Taylor, K. (2019). In emerging economies, smartphone and social media users have broader social networks. *Pew Research Center*.
- Skinner, B. F. (2019). *The behavior of organisms: An experimental analysis*. BF Skinner Foundation.
- Statista. (2021). *Statista*. <https://de.statista.com/>
- Sweetser, P., & Wyeth, P. (2005). GameFlow: a model for evaluating player enjoyment in games. *Computers in Entertainment (CIE), 3*(3), 3.
- 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.

- Ulbrich-Herrmann, M. (2014). Gewaltbefürwortende Einstellungen [Attitudes in favor of violence]. *Zusammenstellung Sozialwissenschaftlicher Items Und Skalen*.
- Wang, S. (2020). Standing up or standing by: Bystander intervention in cyberbullying on social media. *New Media & Society*, 1461444820902541.
- Weimann, G. (2012). Lone wolves in cyberspace. *Journal of Terrorism Research*.
- Welsh, B. C., Farrington, D. P., & Gowar, B. R. (2015). Benefit-Cost Analysis of Crime Prevention Programs. *Crime and Justice*, 44(1), 447–516.
- White, A., Kavanagh, D., Stallman, H., Klein, B., Kay-Lambkin, F., Proudfoot, J., Drennan, J., Connor, J., Baker, A., & Hines, E. (2010). Online alcohol interventions: a systematic review. *Journal of Medical Internet Research*, 12(5).
- Winter, S., Maslowska, E., & Vos, A. L. (2021). The effects of trait-based personalization in social media advertising. *Computers in Human Behavior*, 114, 106525.
- Wong, R. Y. M., Cheung, C. M. K., Xiao, B., & Thatcher, J. B. (2021). Standing Up or Standing By: Understanding Bystanders' Proactive Reporting Responses to Social Media Harassment. *Information Systems Research*.

Appendix B – Appendix for Chapter 3

B.1 Facebook Experiment Appendix

TABLE B.1.1. SUMMARY STATISTICS – FACEBOOK EXPERIMENT

Variable	(1)	(2)	(3)
	Population Mean	Sample Mean Standard deviation	
Age Group			
18-34	0.33	0.66	0.47
35-44	0.20	0.16	0.37
45-64	0.47	0.18	0.38
Female	0.51	0.59	0.49
Academic	0.19	0.31	0.46
Employed	0.93	0.90	0.3
Single	0.50	0.40	0.49
Parent	0.34	0.27	0.44
Urbanite	0.77	0.26	0.44
No. of observations		4,118	

Notes: This table shows population means, sample means, and sample standard deviations of the covariates from the *Facebook* experiment. Age group is a categorical variable, meaning that it provides the share of observations in the respective class. Female, academic, employed, single, parent, and urbanite are dummy variables. They each take on a value of one if the observation is female, has an academic degree, is employed, single, has at least one child, or lives in the urban area. Otherwise, they each take on a value of zero. *Source:* The population means come from Statista (2021).

TABLE B.1.2. BALANCING TABLE – FACEBOOK EXPERIMENT

Dependent variable: Assignment to treatment				
	(1)	(2)	(3)	(4)
	Control group mean		Estimated coefficients	
Variable		Non-interactive film	Interactive film	Interactive film + online-games
<i>Results from separate regressions</i>				
Age group				
18-34	0.65	0.06*** (0.02)	0.00 (0.02)	-0.01 (0.02)
35-44	0.14	0.00 (0.03)	0.08*** (0.03)	0.11*** (0.03)
45-64	0.21	-0.08*** (0.02)	-0.07*** (0.02)	-0.09*** (0.02)
Female	0.65	-0.13*** (0.02)	-0.08*** (0.02)	-0.04** (0.02)
Academic	0.32	-0.05** (0.02)	0.00 (0.02)	-0.02 (0.02)
Employed	0.89	0.04 (0.03)	0.03 (0.03)	0.00 (0.03)
Single	0.41	-0.03 (0.02)	-0.03 (0.02)	-0.01 (0.02)
Parent	0.26	0.00 (0.02)	0.01 (0.02)	0.03 (0.02)
Urbanite	0.28	-0.03 (0.02)	-0.01 (0.02)	-0.04** (0.02)
<i>Statistics from joint regressions</i>				
F-value		6.02	3.51	4.70
Prob > F		0.00	0.00	0.00
R ²		0.051	0.043	0.047
Adjusted R ²		0.042	0.031	0.036
No. of Observations	1,696	2,854	2,224	2,432

Notes: This table shows the combined results from the different balancing checks. The first panel of the table shows the results from separately regressing the treatment indicators on each of the covariates. The second panel shows the statistics from regressing the treatment indicators on all of the selected covariates. The treatments include the non-interactive film, the interactive film, and the interactive film in combination with the online-games. The covariates comprise fixed effects on the federal state-level and sociodemographic characteristics including

age group (3 categories), female (dummy), academic (dummy), employed (dummy), single (dummy), parent (dummy), and urbanite (dummy). The last row gives the number of observations in the control group in column (1), and the combined observations in the control and the respective treatment group in the columns (2) to (4). Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE B.1.3. IMPACT VECTORS – FACEBOOK EXPERIMENT

Dependent variable: Row-wise				
	(1)	(2)	(3)	(4)
	Control group mean	Estimated coefficients (treatment indicator)		
Independent variables		Film	Interactive film	Interactive film + online-games
Responsibility denial 1	1.18	-0.01 (0.03)	-0.05 (0.04)	-0.08 ** (0.04)
Responsibility denial 2	0.67	-0.08*** (0.03)	-0.06 (0.04)	-0.06** (0.03)
Responsibility denial 3	0.86	-0.08*** (0.03)	-0.14*** (0.04)	-0.06 (0.04)
Responsibility denial 4	0.29	-0.04** (0.02)	-0.07*** (0.02)	-0.06*** (0.02)
Propensity to violence	1.15	0.03 (0.03)	-0.10** (0.04)	-0.12*** (0.04)
Justice sensitivity 1	1.65	0.03 (0.03)	-0.01 (0.04)	-0.05 (0.04)
Justice sensitivity 2	2.08	0.06** (0.03)	0.04 (0.04)	0.08*** (0.03)
Self-efficacy 1	2.18	0.07*** (0.02)	0.02 (0.03)	0.05** (0.02)
Self-efficacy 2	2.23	0.03 (0.02)	-0.01 (0.03)	0.04 (0.03)
Self-efficacy 3	2.12	-0.01 (0.02)	-0.01 (0.03)	0.00 (0.03)
Empathy 1	1.89	0.05* (0.03)	0.03 (0.04)	0.00 (0.04)
Empathy 2	2.04	0.01 (0.03)	0.09** (0.04)	0 (0.03)
Empathy 3	1.52	-0.05* (0.03)	-0.01 (0.04)	-0.05 (0.04)
Self-esteem	1.96	-0.04 (0.03)	-0.06 (0.04)	-0.06** (0.03)

Notes: This table shows the control group means and treatment effects on the typical determinants of bystander behavior. The point estimates come from separately regressing each of the determinants on the respective treatment indicator (dummy) and a set of covariates. The covariates include age group (3 categories), female (dummy), academic (dummy), employed (dummy), single (dummy), parent (dummy), and urbanite (dummy). Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE B.1.4. REGRESSION WITH DEMOGRAPHIC SUBGROUPS – FACEBOOK EXPERIMENT

Dependent variable: MueZI score				
Variable	(1) CG mean	(2) Film	(3) Interactive film	(4) Interactive film + online- games
<i>Pooled sample</i>				
Treatment indicator	10.12	0.30*** (0.07)	0.66*** (0.08)	0.45*** (0.08)
No. of observations	1,696	2,854	2,224	2,432
<i>Low-age sample</i>				
Treatment indicator	10.15	0.30*** (0.08)	0.57*** (0.09)	0.52*** (0.09)
No. of observations	1,098	1,909	1,442	1,570
<i>Mid-age sample</i>				
Treatment indicator	10.07	0.25 (0.18)	0.94*** (0.19)	0.36* (0.19)
No. of observations	235	395	339	390
<i>High-age sample</i>				
Treatment indicator	10.05	0.26 (0.18)	0.91*** (0.21)	0.27 (0.21)
No. of observations	363	550	443	472
<i>Male sample</i>				
Treatment indicator	9.93	0.35*** (0.11)	0.75*** (0.12)	0.51*** (0.13)
No. of observations	592	1,147	828	881
<i>Female sample</i>				
Treatment indicator	10.22	0.26*** (0.09)	0.62*** (0.10)	0.41*** (0.09)
No. of observations	1,104	1,707	1,396	1,551
<i>Non-academic sample</i>				
Treatment indicator	10.17	0.27*** (0.08)	0.68*** (0.09)	0.48*** (0.09)
No. of observations	1,150	1,982	1,511	1,665
<i>Academic sample</i>				

Treatment indicator	10.01	0.28** (0.13)	0.61*** (0.14)	0.36** (0.14)
No. of observations	546	872	713	767
<i>Unemployed sample</i>				
Treatment indicator	9.91	0.43* (0.24)	1.15*** (0.24)	1.09*** (0.24)
No. of observations	183	292	232	262
<i>Employed sample</i>				
Treatment indicator	10.15	0.27*** (0.07)	0.62*** (0.08)	0.37*** (0.08)
No. of observations	1,513	2,562	1,992	2,170
<i>Non-single sample</i>				
Treatment indicator	10.22	0.14 (0.09)	0.60*** (0.10)	0.42*** (0.09)
No. of observations	996	1,709	1,324	1,439
<i>Single sample</i>				
Treatment indicator	9.97	0.54*** (0.11)	0.78*** (0.13)	0.50*** (0.12)
No. of observations	700	1,145	900	993
<i>Non-parent sample</i>				
Treatment indicator	10.08	0.35*** (0.08)	0.61*** (0.09)	0.51*** (0.09)
No. of observations	1,250	2,105	1,631	1,770
<i>Parent sample</i>				
Treatment indicator	10.24	0.19 (0.14)	0.85*** (0.15)	0.34** (0.15)
No. of observations	446	749	593	662
<i>Rural sample</i>				
Treatment indicator	10.08	0.32*** (0.08)	0.69*** (0.09)	0.52*** (0.09)
No. of observations	1,221	2,086	1,607	1,777
<i>Urban sample</i>				
Treatment indicator	10.23	0.22* (0.13)	0.58*** (0.14)	0.23 (0.15)
No. of observations	475	768	617	655

Notes: This table shows the control group means and treatment effects for selected sociodemographic subgroups and the pooled sample as a reference point. The point estimates come from separately regressing the *Munich civil courage instrument score* (MueZI score) on the respective treatment indicator (dummy) and a set of covariates. The covariates include age group (3 categories), female (dummy), academic (dummy) employed (dummy), single (dummy), children (dummy), urbanite (dummy), and fixed effects at the federal state-level. Robust standard errors in parentheses. *p<0.10, **p<0.05, ***p<0.01.

B.2 Panel Data Experiment Appendix

TABLE B.2.1. SUMMARY STATISTICS – PANEL DATA EXPERIMENT

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Population		Sample				
		<i>Post-treatment</i>		<i>4 weeks follow-up</i>		<i>8 weeks follow-up</i>	
Variable	Mean	Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation
<i>Age group</i>							
18-34	0.33	0.33	0.47	0.31	0.46	0.30	0.46
35-49	0.30	0.31	0.46	0.32	0.46	0.32	0.46
50-64	0.37	0.36	0.48	0.37	0.48	0.39	0.49
Female	0.51	0.50	0.50	0.49	0.50	0.49	0.50
Academic	0.19	0.37	0.48	0.36	0.48	0.36	0.48
Employed	0.93	0.84	0.37	0.83	0.37	0.84	0.37
Single	0.50	0.32	0.47	0.32	0.47	0.33	0.47
Parent	0.34	0.48	0.50	0.49	0.50	0.49	0.50
Urbanite	0.77	0.67	0.47	0.66	0.47	0.66	0.47
Number of observations		1,587		1,388		1,253	

Notes: This table shows population means, sample means, and sample standard deviations of the covariates from the panel data experiment. Age group is a categorical variable meaning it provides the share of observations in the respective class. Female, academic, employed, single, parent, and urbanite are dummy variables. They each take on a value of one if the observation is female, has an academic degree, is employed, single, has at least one child, or lives in the urban area. Otherwise, they each take on a value of zero.

TABLE B.2.2. BALANCING TABLE – PANEL DATA EXPERIMENT

Dependent variable: Assignment to treatment						
	(1)	(2)	(3)	(4)	(5)	(6)
	Post-treatment		4 weeks follow-up		8 weeks follow-up	
Variable	Control group mean	Estimated coefficient	Control group mean	Estimated coefficient	Control group mean	Estimated coefficient
<i>Results from separate regressions</i>						
Age group						
18-34	0.32	0.02 (0.03)	0.31	0.01 (0.03)	0.30	0.00 (0.03)
35-44	0.30	0.02 (0.03)	0.30	0.04 (0.03)	0.29	0.07** (0.03)
45-64	0.38	-0.04 (0.03)	0.40	-0.05* (0.03)	0.41	-0.06** (0.03)
Female	0.51	-0.02 (0.03)	0.50	-0.02 (0.03)	0.50	-0.02 (0.03)
Academic	0.37	0 (0.03)	0.35	0.01 (0.03)	0.35	0.02 (0.03)
Employed	0.85	-0.03 (0.03)	0.85	-0.04 (0.04)	0.85	-0.03 (0.04)
Single	0.32	0 (0.03)	0.33	-0.01 (0.03)	0.34	-0.03 (0.03)
Parent	0.49	0 (0.03)	0.48	0 (0.03)	0.48	0.03 (0.03)
Urbanite	0.68	-0.03 (0.03)	0.67	-0.02 (0.03)	0.66	-0.01 (0.03)
<i>Statistics from joint regressions</i>						
F-value		1.20		1.26		1.03
Prob > F		0.75		0.76		0.70
R ²		0.017		0.020		0.018
Adjusted R ²		0.002		0.002		-0.001
No. of observations	796	1,587	679	1,378	605	1,247

Notes: This table shows the combined results from the different balancing checks across the three survey waves. The first panel of the table shows the results from separately regressing the treatment indicator on each of the covariates. The second panel shows the statistics from regressing the treatment indicator jointly on all of the selected covariates. The covariates comprise fixed effects on the federal state-level and sociodemographic characteristics including age

group (3 categories), female (dummy), academic (dummy), employed (dummy), single (dummy), parent (dummy), and urbanite (dummy). The last row gives the number of observations in the control group in column (1), and the combined observations in the control and the respective treatment group in the columns (2) to (4). Robust standard errors in parentheses.*p<0.10, **p<0.05, ***p<0.01.

TABLE B.2.3. IMPACT VECTORS I – PANEL DATA EXPERIMENT

Dependent variable: Row-wise						
	(1)	(2)	(3)	(4)	(5)	(6)
	Post-treatment		4 weeks later		8 weeks later	
Variables	Control group mean	Estimated coefficient	Control group mean	Estimated coefficient	Control group mean	Estimated coefficient
Detection (positive)	5.12	0.19*** (0.07)	5.07	0.17** (0.08)	5.10	0.15* (0.08)
Detection (negative)	2.59	-0.16** (0.07)	2.78	-0.20** (0.08)	2.76	-0.11 (0.08)
Interpretation (positive)	4.51	0.30*** (0.07)	4.53	0.16** (0.07)	4.58	0.11 (0.08)
Interpretation (negative)	3.79	-0.15* (0.08)	3.84	-0.23*** (0.08)	3.84	-0.18** (0.09)
Responsibility (negative)	3.25	-0.18** (0.08)	3.31	-0.16* (0.09)	3.31	-0.19** (0.09)
Responsibility (positive)	5.17	0.13 (0.08)	5.12	0.25*** (0.08)	5.19	0.05 (0.09)
Control (negative) 1	3.56	-0.13 (0.09)	3.68	-0.33*** (0.09)	3.66	-0.30*** (0.10)
Control (negative) 2	3.61	-0.19** (0.09)	3.59	-0.25*** (0.09)	3.57	-0.17* (0.10)
No. of observations	796	1,587	679	1,378	605	1,247

Notes: This table shows the control group means and treatment effects on the parameters of the *Barriers to Intervention Model* across the three survey waves. The estimates come from separately regressing the parameters on the treatment indicator (dummy) and a set of covariates. The covariates include age group (3 categories), female (dummy), academic (dummy), employed (dummy), single (dummy), parent (dummy), and urbanite (dummy). The last row gives the number of observations in the control group in columns with uneven numbers, and the combined observations in the control and the treatment groups in the columns with even numbers. Robust standard errors in parentheses. *p<0.10, **p<0.05, ***p<0.01.

TABLE B.2.4. IMPACT VECTORS II – PANEL DATA EXPERIMENT

Dependent variable: Row-wise						
	(1)	(2)	(3)	(4)	(5)	(6)
	Post-treatment		4 weeks later		8 weeks later	
Variable	Control group mean	Estimated coefficient	Control group mean	Estimated coefficient	Control group mean	Estimated coefficient
Empathy	4.80	0.30*** (0.09)	4.80	0.25*** (0.09)	4.86	0.15* (0.09)
Acceptance	4.98	0.17** (0.07)	5.01	0.07 (0.08)	5.03	0.10 (0.08)
Guilt	5.46	0.12 (0.08)	5.41	0.17** (0.08)	5.37	0.15* (0.09)
Indignation	5.30	0.09 (0.07)	5.32	0.19*** (0.07)	5.42	0.08 (0.08)
Audience 1	3.95	-0.1 (0.09)	4.04	-0.23** (0.09)	3.96	-0.15 (0.10)
Audience 2	3.01	-0.15* (0.09)	3.09	-0.22** (0.09)	3.11	-0.16* (0.09)
No. of observations	796	1,587	679	1,378	605	1,247

Notes: This table shows the control group means and treatment effects on the typical determinants of bystander behavior across the three survey waves. The estimates come from separately regressing the parameters on the treatment indicator (dummy) and a set of covariates. The covariates include age group (3 categories), female (dummy), academic (dummy), employed (dummy), single (dummy), parent (dummy), and urbanite (dummy). The last row gives the number of observations in the control group in columns with uneven numbers, and the combined observations in the control and the treatment groups in the columns with even numbers. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE B.2.5. IMPACT VECTORS III – PANEL DATA EXPERIMENT

Dependent variable: Row-wise						
	(1)	(2)	(3)	(4)	(5)	(6)
	Post-treatment		4 weeks later		8 weeks later	
Variable	Control group mean	Est. coeff.	Control group mean	Est. coeff.	Control group mean	Est. coeff.
Attitudes	5.62	0.11*	5.54	0.16**	5.54	0.14**
		(0.06)		(0.07)		(0.07)
Norms	5.84	0.16***	5.91	0.05	5.86	0.1
		(0.06)		(0.07)		(0.07)
Control	5.98	0.06	6.05	0.22**	6.15	0.01
		(0.10)		(0.10)		(0.10)
No. of observations	796	1,587	679	1,378	605	1,247

Notes: This table shows the control group means and treatment effects on the parameters of the *reasoned action approach* across the three survey waves. The estimates come from separately regressing the parameters on the treatment indicator (dummy) and a set of covariates. The covariates include age group (3 categories), female (dummy), academic (dummy), employed (dummy), single (dummy), parent (dummy), and urbanite (dummy). The last row gives the number of observations in the control group in columns with uneven numbers, and the combined observations in the control and the treatment groups in the columns with even numbers. Robust standard errors in parentheses. *p<0.10, **p<0.05, ***p<0.01.

TABLE B.2.6. REGRESSION WITH DEMOGRAPHIC SUBGROUPS – PANEL DATA EXPERIMENT

Independent variable: Willingness to intervene score (WTIS)						
	(1)	(2)	(3)	(4)	(5)	(6)
	Post-treatment		4 weeks later		8 weeks later	
Variable	CG mean	Coefficient	CG mean	Coefficient	CG mean	Coefficient
<i>Pooled sample</i>						
Treatment indicator	27.66	1.56*** (0.34)	28.05	1.36*** (0.35)	28.06	1.05*** (0.37)
No. of observations	796	1,587	679	1,378	605	1,247
<i>Low-age sample</i>						
Treatment indicator	26.50	1.05* (0.60)	26.49	1.01 (0.68)	26.12	1.03 (0.76)
No. of observations	255	524	208	428	181	372
<i>Mid-age sample</i>						
Treatment indicator	27.63	1.47** (0.64)	28.59	0.62 (0.63)	28.79	-0.06 (0.66)
No. of observations	238	491	201	435	173	393
<i>High-age sample</i>						
Treatment indicator	28.65	2.12*** (0.60)	28.84	2.15*** (0.56)	28.97	1.8*** (0.59)
No. of observations	303	572	270	515	251	482
<i>Male sample</i>						
Treatment indicator	26.97	1.62*** (0.50)	27.34	1.33*** (0.49)	27.42	0.94* (0.53)
No. of observations	391	797	337	701	302	638
<i>Female sample</i>						
Treatment indicator	28.32	1.49*** (0.47)	28.74	1.22** (0.50)	28.70	1.14** (0.53)
No. of observations	405	790	342	677	303	609
<i>Non-academic sample</i>						
Treatment indicator	27.86	1.88*** (0.43)	28.31	1.45*** (0.43)	28.40	0.89* (0.47)
No. of observations	503	1,001	440	886	393	796
<i>Academic sample</i>						

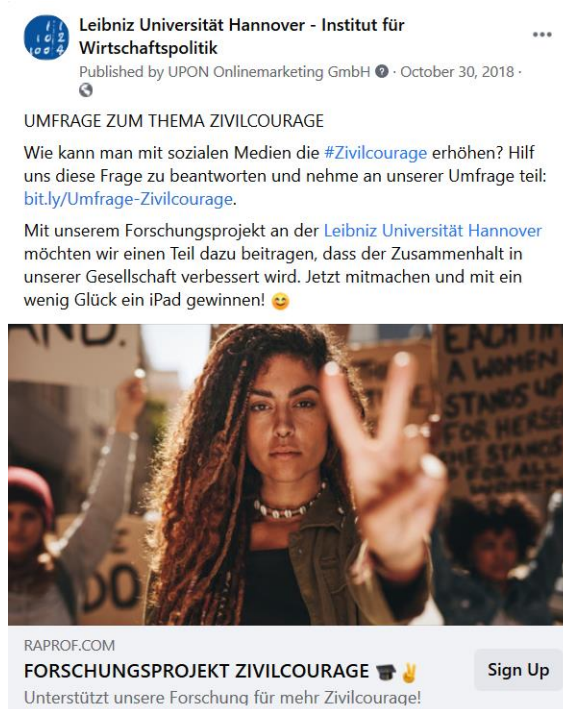
Appendix B – Appendix for Chapter 3

Treatment indicator	27.31	0.94 (0.57)	27.55	1.01 (0.62)	27.44	1.08* (0.62)
No. of observations	293	586	239	492	212	451
<i>Unemployed sample</i>						
Treatment indicator	28.19	-0.14 (0.96)	28.75	-0.61 (1.04)	28.32	-0.87 (1.18)
No. of observations	122	258	105	230	92	200
<i>Employed sample</i>						
Treatment indicator	27.56	1.8*** (0.37)	27.92	1.7*** (0.37)	28.02	1.32*** (0.39)
No. of observations	674	1,329	574	1,148	513	1,047
<i>Non-single sample</i>						
Treatment indicator	28.11	1.32*** (0.41)	28.40	1.46*** (0.40)	28.56	1.15*** (0.41)
No. of observations	541	1,076	458	934	398	836
<i>Single sample</i>						
Treatment indicator	26.70	2.05*** (0.62)	27.32	1.27* (0.70)	27.11	1 (0.75)
No. of observations	255	511	221	444	207	411
<i>Non-parent sample</i>						
Treatment indicator	26.65	1.48*** (0.50)	27.36	0.76 (0.53)	26.82	1.09* (0.56)
No. of observations	409	819	351	709	317	635
<i>Parent sample</i>						
Treatment indicator	28.72	1.67*** (0.47)	28.78	1.94*** (0.46)	29.43	1.04** (0.47)
No. of observations	387	768	328	669	288	612
<i>Rural/suburban sample</i>						
Treatment indicator	28.06	2.18*** (0.59)	28.18	2.06*** (0.60)	28.08	1.99*** (0.62)
No. of observations	251	519	225	468	203	424
<i>Urban sample</i>						
Treatment indicator	27.47	1.25*** (0.44)	27.98	1.03** (0.44)	28.05	0.61 (0.47)
No. of observations	545	1,068	454	910	402	823

Notes: This table shows the control group means and treatment effects for selected sociodemographic subgroups and the pooled sample as a reference point. The point estimates come from separately regressing the *willingness to intervene score* (WTIS) on the respective treatment indicator (dummy) and a set of covariates. The covariates include age group (3 categories), female (dummy), academic (dummy), employed (dummy), single (dummy), children (dummy), urbanite (dummy), and fixed effects at the federal state-level. Robust standard errors in parentheses. *p<0.10, **p<0.05, ***p<0.01.

B.3 Illustration of the Policy Intervention

FIGURE B.3.1. EXEMPLARY FACEBOOK AD



The image shows a Facebook post from 'Leibniz Universität Hannover - Institut für Wirtschaftspolitik'. The post is dated October 30, 2018, and was published by UPON Onlinemarketing GmbH. The text of the post asks how social media can increase #Zivilcourage and invites users to participate in a survey at bit.ly/Umfrage-Zivilcourage. It also mentions a research project at Leibniz Universität Hannover aimed at improving social cohesion and offers a chance to win an iPad. Below the text is a photograph of a woman with long dark hair making a peace sign. At the bottom of the post, there is a 'Sign Up' button and the text 'FORSCHUNGSPROJEKT ZIVILCOURAGE' with a small icon of a hand holding a flame.

Leibniz Universität Hannover - Institut für Wirtschaftspolitik
Published by UPON Onlinemarketing GmbH · October 30, 2018 ·

UMFRAGE ZUM THEMA ZIVILCOURAGE

Wie kann man mit sozialen Medien die #Zivilcourage erhöhen? Hilf uns diese Frage zu beantworten und nehme an unserer Umfrage teil: bit.ly/Umfrage-Zivilcourage.

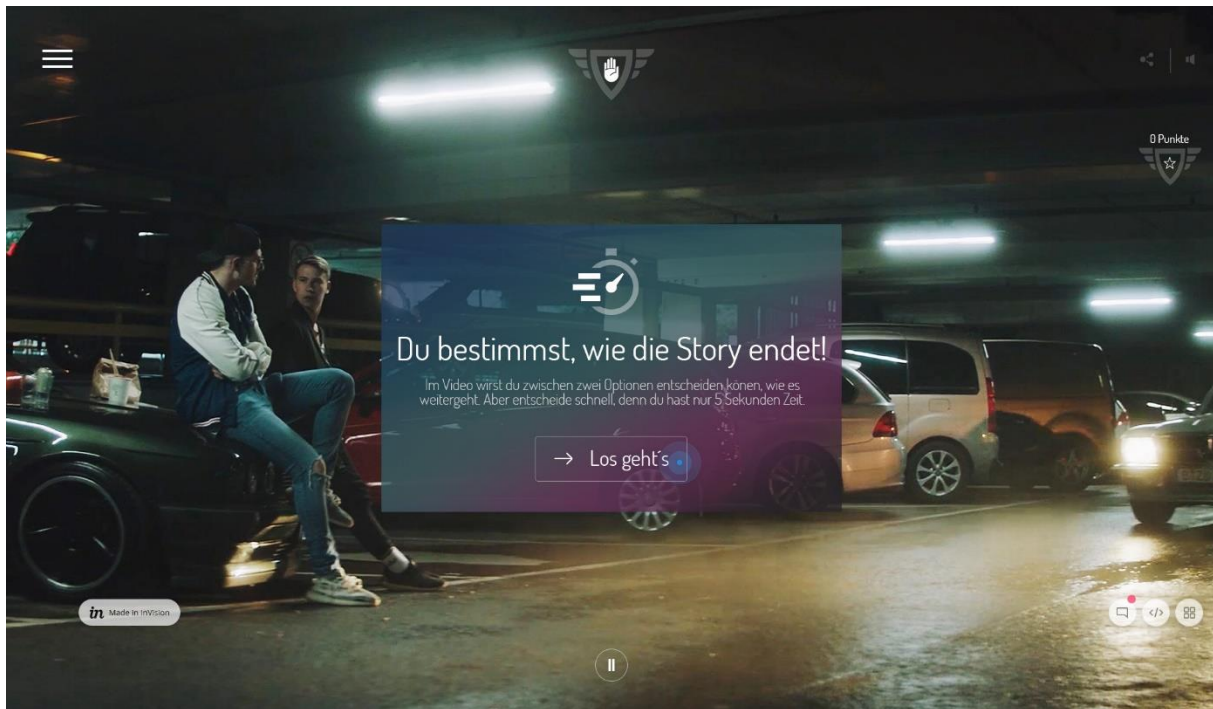
Mit unserem Forschungsprojekt an der Leibniz Universität Hannover möchten wir einen Teil dazu beitragen, dass der Zusammenhalt in unserer Gesellschaft verbessert wird. Jetzt mitmachen und mit ein wenig Glück ein iPad gewinnen! 😊

RAPROF.COM
FORSCHUNGSPROJEKT ZIVILCOURAGE 🙌
Unterstützt unsere Forschung für mehr Zivilcourage!

Sign Up

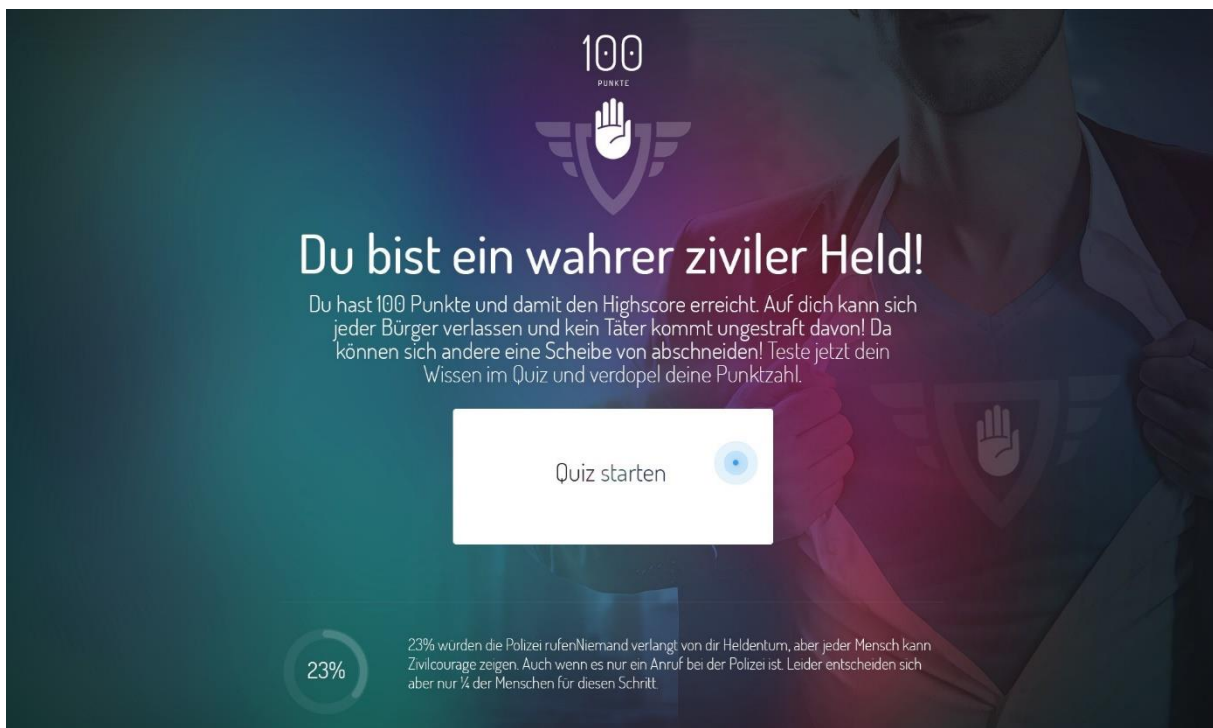
Notes: The figure shows an example of the Facebook ads used during the first experiment. *Source:* <https://www.facebook.com/>.

FIGURE B.3.2. START SCREEN OF THE INTERACTIVE FILM



Notes: The figure shows the start screen of the interactive film. The header in the pop-up windows says, “You decide how the story ends!” The text says, “In the video you will be able to decide between 2 options how to proceed. But decide quickly, because you have only 5 seconds.” The button says, “Let’s go!” and leads the user directly to the interactive film. *Source:* <https://www.zivile-helden.de/>.

FIGURE B.3.3. FINAL SCREEN OF THE INTERACTIVE FILM

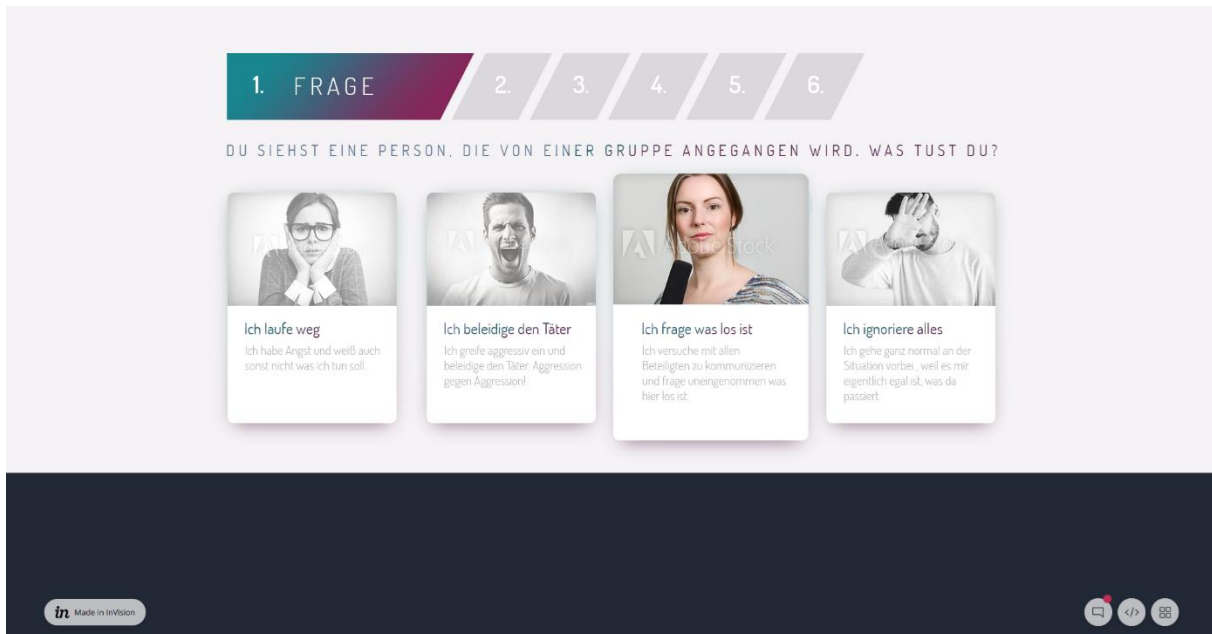


Notes: The figure shows the final screen of the interactive film. The header says, “You are a true civilian hero!” The text tells the user that she has achieved the maximum score of 100 and gives her positive feedback. It also invites her to test her knowledge in the online games and thus double her score. At the very bottom, the user is

informed that 23% of the population would call the police, but that it does not take much to be a hero. The button leads the user directly to the online games. Source: <https://www.zivile-helden.de/>.

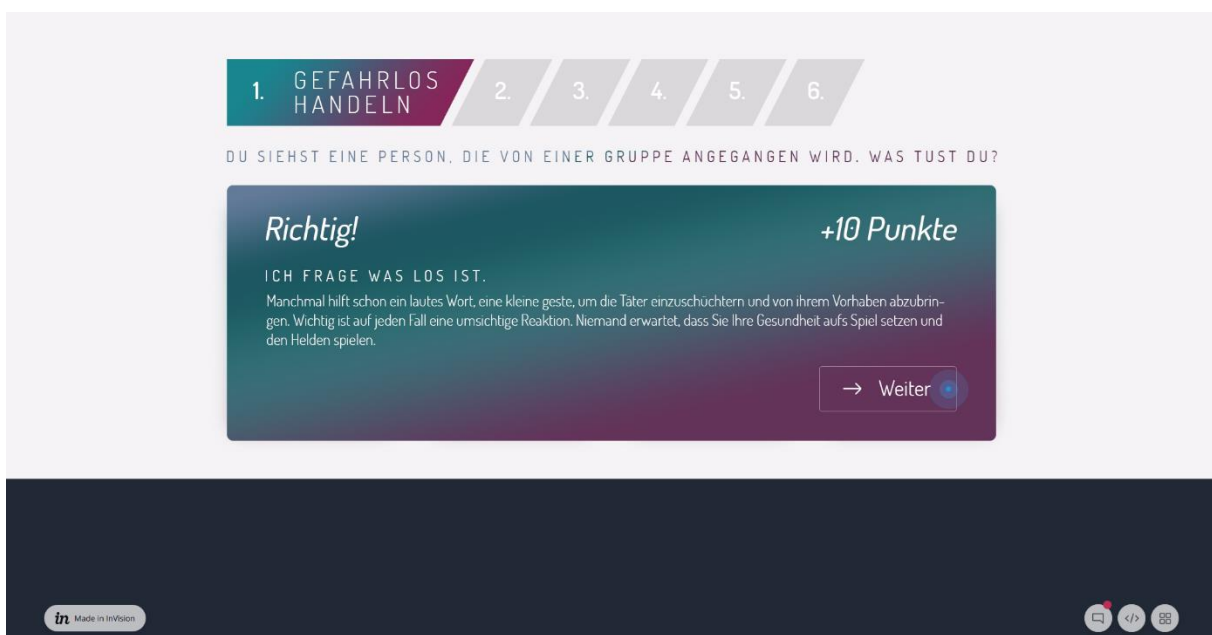
FIGURE B.3.4. ONLINE GAME 1

(a) Task



Notes: The figure shows the task of first the online game. The task asks the user, “You see a person being attacked by a group. What do you do?” The options for action are, “I run away”, “I insult the perpetrator”, “I ask what is going on”, and “I ignore everything.” Source: <https://www.zivile-helden.de/>.

(b) Feedback



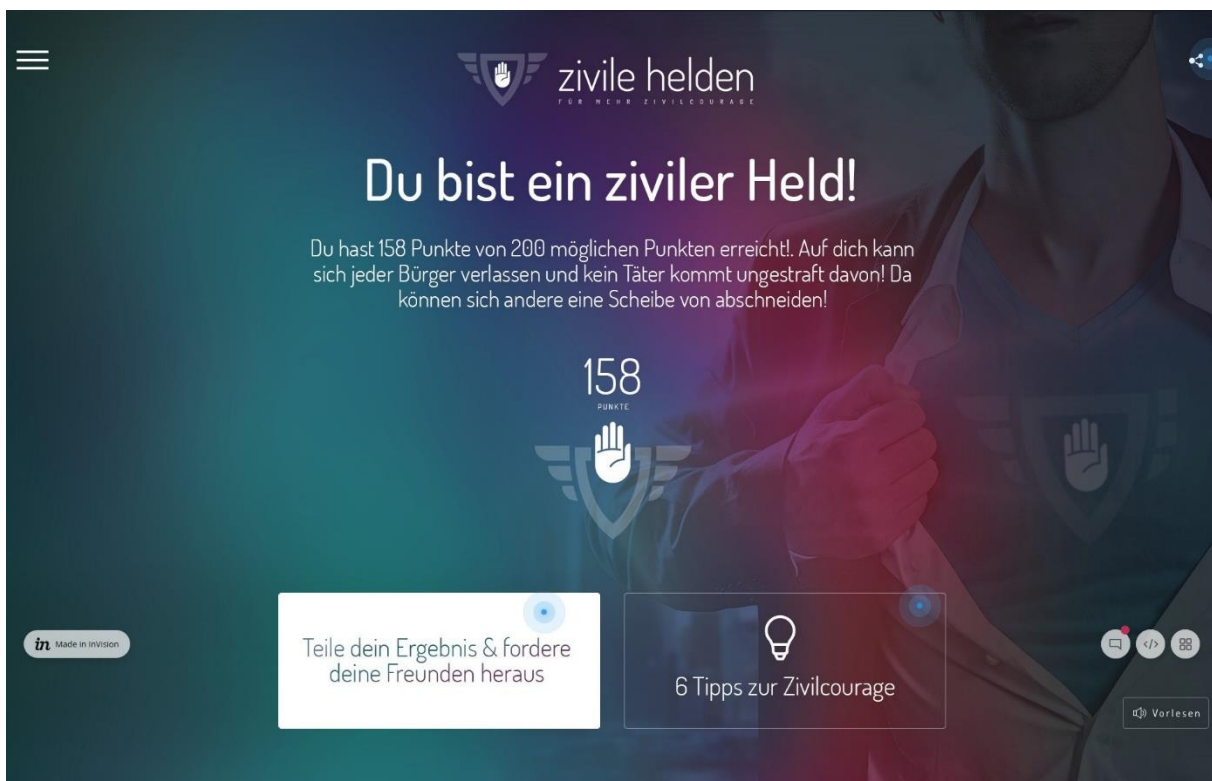
Notes: The figure shows the feedback to the task. The game tells the user that she acted correctly and rewards her with 10 points. It also briefly explains why the decision was right. Source: <https://www.zivile-helden.de/>.

FIGURE B.3.5. ONLINE GAME 3 - FEEDBACK



Notes: The figure shows the feedback to the third task. It tells the user that she got most of it right, rewards her with 27 points, and compares her solution to the sample solution. It also explains why it is important to memorize perpetrator characteristics in a violent situation. Source: <https://www.zivile-helden.de/>.

FIGURE B.3.6. FINAL SCREEN – ONLINE GAMES



Notes: The figure shows the final screen of the online games. The header says, “You are a civilian hero!” The text tells the user that she has achieved a score of 158 and gives her positive feedback. The left button lets the user share her result on social media. The right button takes her to a detailed explanation of the 6 desired behaviors in a violent situation. Source: <https://www.zivile-helden.de/>.

FIGURE B.3.7. ILLUSTRATION OF THE INCENTIVE



Notes: The figure shows the incentive for participating in the Facebook experiment. The header says, “Win an iPad!” The button leads the user directly to the online survey. *Source:* <https://www.zivile-helden.de/>.